Computational Model for Human 3D Shape Perception From a Single Specular Image

Takeaki Shimokawa[1]*, Akiko Nishio[2,3], Masa-aki Sato[1], Mitsuo Kawato[1] and Hidehiko Komatsu[3]

[1] Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International (ATR), Seika-cho, Japan, [2] Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Okazaki, Japan, [3] Brain Science Institute, Tamagawa University, Machida, Japan

In natural conditions the human visual system can estimate the 3D shape of specular objects even from a single image. Although previous studies suggested that the orientation field plays a key role for 3D shape perception from specular reflections, its computational plausibility, and possible mechanisms have not been investigated. In this study, to complement the orientation field information, we first add prior knowledge that objects are illuminated from above and utilize the vertical polarity of the intensity gradient. Then we construct an algorithm that incorporates these two image cues to estimate 3D shapes from a single specular image. We evaluated the algorithm with glossy and mirrored surfaces and found that 3D shapes can be recovered with a high correlation coefficient of around 0.8 with true surface shapes. Moreover, under a specific condition, the algorithm's errors resembled those made by human observers. These findings show that the combination of the orientation field and the vertical polarity of the intensity gradient is computationally sufficient and probably reproduces essential representations used in human shape perception from specular reflections.

Keywords: 3D shape perception, specularity, gloss, orientation field, illumination prior

INTRODUCTION

Specular reflections, which are seen in many daily objects, provide information about their material and surface finish (Adelson, 2001; Motoyoshi et al., 2007; Fleming, 2014), enhance the reality of animation and computer graphics, support 3D shape perception (Blake and Bülthoff, 1990; Norman et al., 2004; Khang et al., 2007), and increase the 3D appearance of images (Mooney and Anderson, 2014). A specular reflection component in a single image can be regarded as a marking that is pasted on an object's surface. However, the human visual system solves inverse optics, and we intuitively recognize that an image pattern is generated by a specular reflection (Todd et al., 2004). The regularity of the image patterns of specular reflections is closely related to 3D shape, and the human visual system perceives and evaluates specular reflection through coupled computation with 3D shape perception (Anderson and Kim, 2009; Marlow et al., 2011, 2015).

A previous psychophysical study showed that humans could recover 3D shapes from a single mirrored surface image under unknown natural illumination (Fleming et al., 2004). Furthermore, they hypothesized that the human visual system uses the orientation field for 3D shape perception from specular reflection and texture (Breton and Zucker, 1996; Fleming et al., 2004, 2011).

The orientation field is a collection of dominant orientations at every image location (**Figure 1A**), and this information is represented in the primary visual cortex (V1), which contains cells tuned to specific orientations (Hubel and Wiesel, 1968). In support of their hypothesis, they showed that 3D shape perception is modulated by psychophysical adaptation to specific orientation fields (Fleming et al., 2011). However, how 3D shapes are reconstructed from the orientation field, and whether it is adequate for 3D shape recovery remains unknown. Tappen (2011) proposed a shape recovery algorithm and recovered the 3D shape of simple mirrored surfaces with curvature constraints by an orientation field from a single image under an unknown natural illumination. This suggests a possible mechanism of 3D shape perception from specular reflections. However, since the method is limited to convex shapes, it only explains a small part of human shape perception, which can recover more general shapes including both convex and concave regions (Fleming et al., 2004).

Other algorithms have also been proposed to recover 3D shapes from specular images. They employed either a known calibrated scene (Savarese et al., 2005; Tarini et al., 2005; Liu et al., 2013) or multiple images such as specular flows (Adato et al., 2010), motions of reflection correspondences (Sankaranarayanan et al., 2010), or line tracking (Jacquet et al., 2013). Although they are useful in some situations, they cannot recover 3D shapes from a single specular image in an unknown scene. Li et al. recovered shapes using reflection correspondences extracted by SIFT (Li et al., 2014) just using a single image under an unknown illumination environment like our proposed algorithm. However, their method is limited because it requires the known surface normal values of several surface points to constrain their results.

In this study, we recover general shapes containing both convex and concave surface regions using the orientation field. However, an innate problem prevents the recovery of general shapes from it. Here, we briefly explain the information of 3D shapes contained in the orientation field and its limitation as well as a strategy to overcome that limitation.

Figure 1 shows the relationship between the orientation field and the second order derivatives of the surface depth, which can be decomposed into two orthogonal orientations (left side of **Figure 1A**). These decomposed second derivatives are closely related to the principal curvatures, but these are not strictly the same (see Materials and Methods). The right side of **Figure 1A** represents the orientation field. In specular reflection, the illumination environment is reflected and appears in the image. At that time, the illumination environment is compressed toward a strong surface second derivative orientation and elongated along a weak surface second derivative orientation (Fleming et al., 2004, 2009). As a result, image orientation θ is generated along small surface second derivative orientation θ_s. Moreover, the image anisotropy (the degree of the image orientation's clarity, see Materials and Methods) also approximates the surface anisotropy (the ratio of the large and small surface second derivatives, see Materials and Methods; Fleming et al., 2004, 2009). The proposed algorithm uses this relationship for 3D surface recovery. Here the problem is whether the shape is concave or convex is ambiguous (**Figure 1**). The image orientations are identical among **Figures 1A–D** at the mid-point because the surface orientations are identical for all the images. However, the two signs of the surface second derivatives are different across these images. The orientation field cannot distinguish among these four types.

We overcome the problem of concave/convex ambiguity by imposing a prior that light comes from above (Ramachandran, 1988; Sun and Perona, 1998; Gerardin et al., 2010; Andrews et al., 2013) (hereafter called the "light from above prior"). In utilizing this prior knowledge, we actively use both a diffuse and a specular reflection component. Since most objects that give specular reflection also give diffuse reflection, a natural extension is to combine the features of both reflection components. Note that this prior also works for mirrored surfaces (see the Results section) and the human performance to resolve the concave/convex ambiguity from a mirrored surface increased when the illumination environment was brighter in the upper hemisphere (Faisman and Langer, 2013).
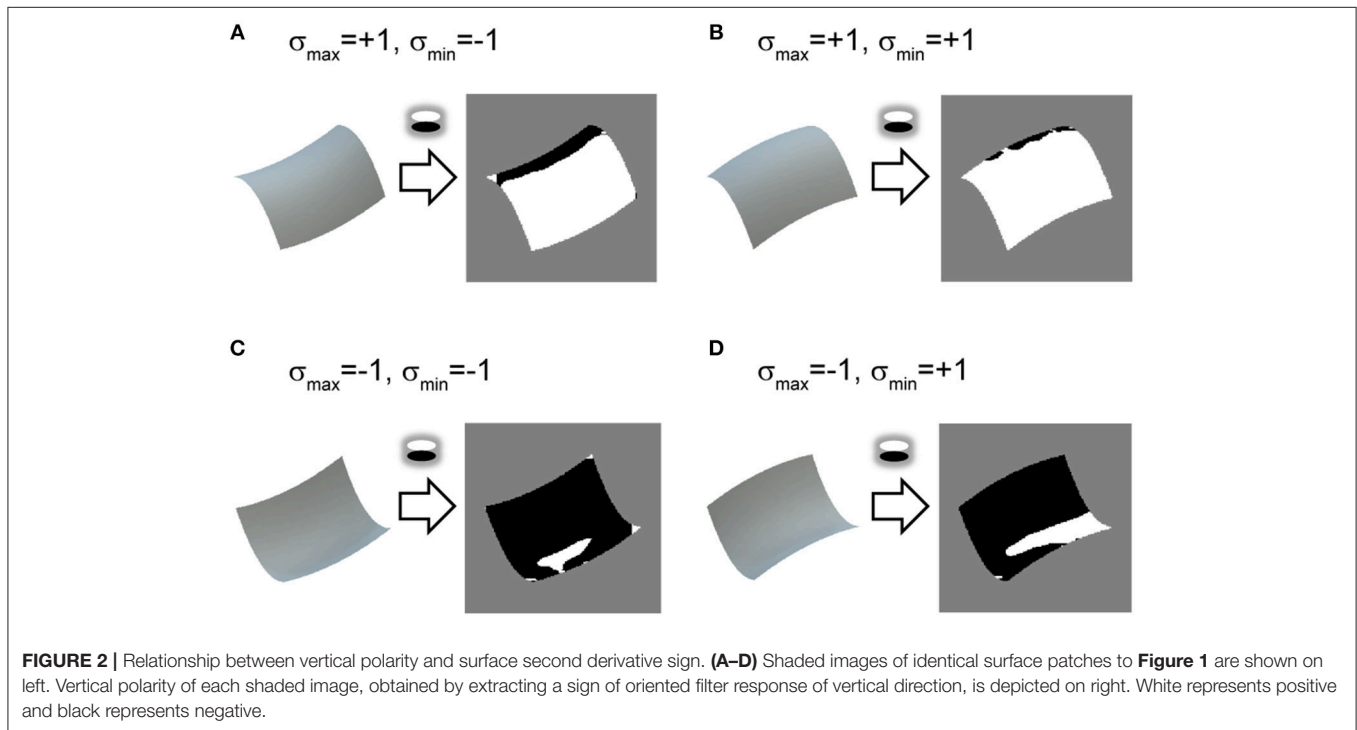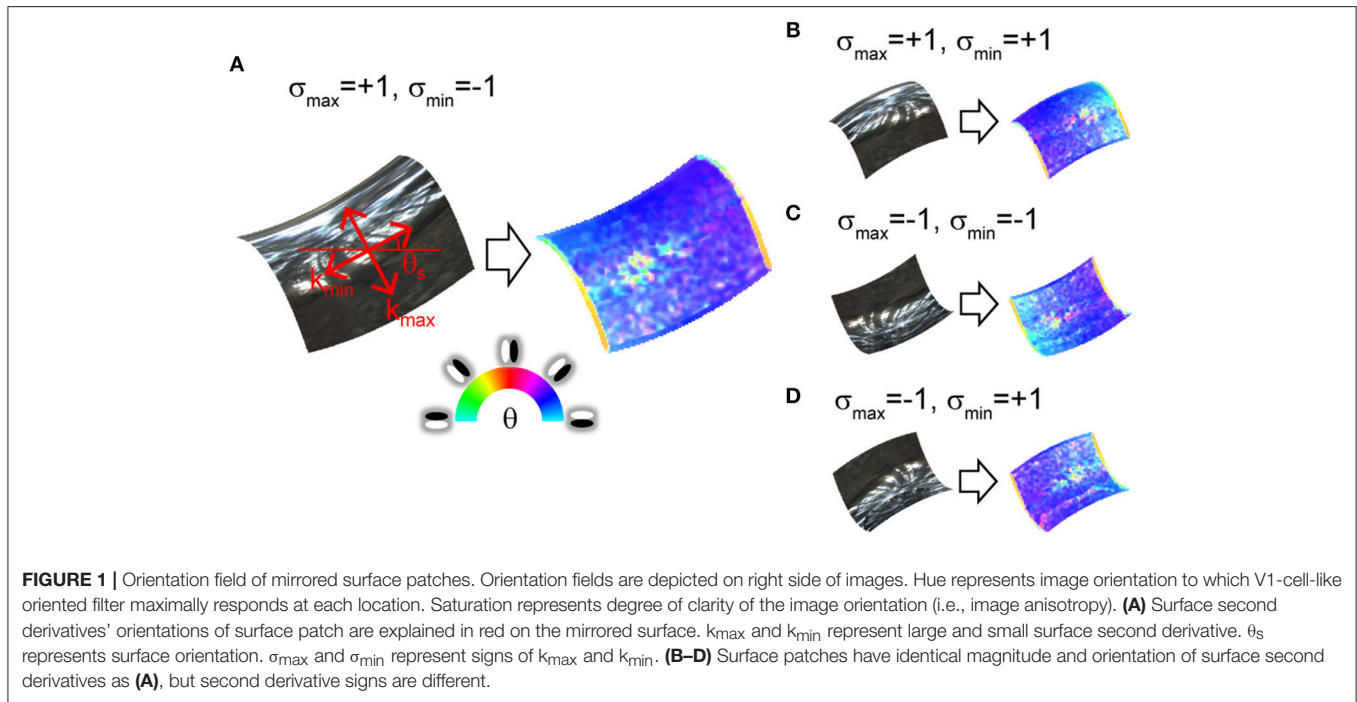
We propose using the vertical polarity of the intensity gradient (hereafter vertical polarity) as an image cue (**Figure 2**; see also **Appendix A**) to realize the prior knowledge. As with the orientation field, the polarity of the intensity gradient can be obtained by a V1-like filter (DeAngelis et al., 1995) and its relation with 3D shape perception was reported (Sawayama and Nishida, 2018). Neurophysiological studies also showed that the activity amplitudes of human early visual areas to the oriented shading gradients differed between the vertical and horizontal directions (Humphrey et al., 1997), and the unidirectional tuning of monkey V4 cells was biased toward vertical directions (Hanazawa and Komatsu, 2001), suggesting the significance of vertical polarity among other directions in the visual system. Assuming lighting from above and Lambert reflectance, vertical polarity corresponds to the surface second derivative sign of vertical orientation (see Materials and Methods). This prior is used only as an initial value for the optimization for 3D shape recovery. Because physically possible shape patterns given the orientation field are restricted (Huffman, 1971; Malik, 1987), it is expected that the remaining ambiguity (i.e., the surface second derivative sign of the horizontal orientation) is implicitly resolved and erroneous initial values are corrected through optimization.

Our proposed algorithm, which can recover general shapes including both convex and concave regions under an unknown natural illumination, is based on the information used by the human visual system. Therefore, it makes a critical contribution to understanding the mechanism of 3D shape perception from specular reflections.

RESULTS

Flowchart of Proposed Algorithm

Figure 3 shows the flowchart of the proposed algorithm that recovers the 3D surface depth from a single specular image. The main procedure is as follows. First, the orientation field is extracted from an image; second, the cost function is formulated based on the orientation field; finally, the 3D shape is recovered by minimizing the cost function. Additionally, we extracted the vertical polarity from the image
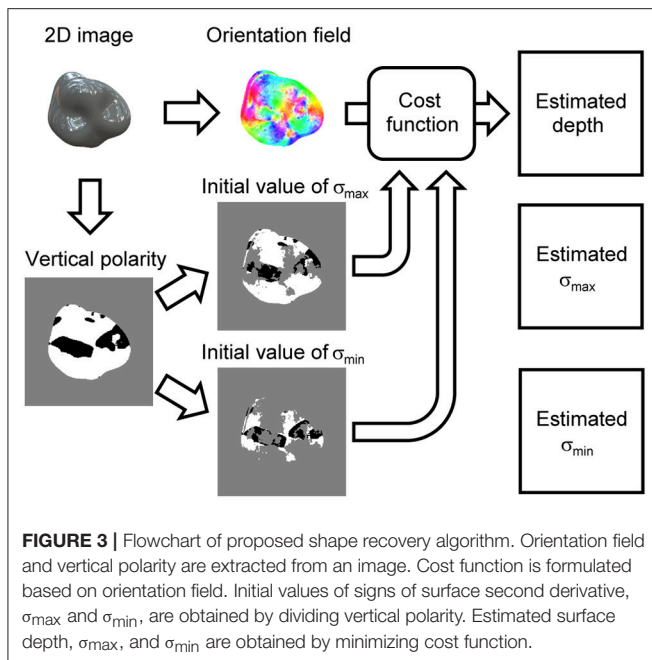
FIGURE 1 | Orientation field of mirrored surface patches. Orientation fields are depicted on right side of images. Hue represents image orientation to which V1-cell-like oriented filter maximally responds at each location. Saturation represents degree of clarity of the image orientation (i.e., image anisotropy). **(A)** Surface second derivatives' orientations of surface patch are explained in red on the mirrored surface. k_{max} and k_{min} represent large and small surface second derivative. θ_s represents surface orientation. σ_{max} and σ_{min} represent signs of k_{max} and k_{min}. **(B–D)** Surface patches have identical magnitude and orientation of surface second derivatives as **(A)**, but second derivative signs are different.

FIGURE 2 | Relationship between vertical polarity and surface second derivative sign. **(A–D)** Shaded images of identical surface patches to **Figure 1** are shown on left. Vertical polarity of each shaded image, obtained by extracting a sign of oriented filter response of vertical direction, is depicted on right. White represents positive and black represents negative.

to resolve the concave/convex ambiguity. The initial values of the surface second derivative signs, σ_{max} and σ_{min}, are calculated based on the vertical polarity and used to minimize the cost function. The proposed algorithm outputs not only the recovered 3D surface depth but also the estimated surface second derivative signs, σ_{max} and σ_{min}, due to minimizing the cost function.

Here we briefly explain the cost function (see section Formulation of cost function for details), which consists of two terms: second derivative constraint C and boundary condition B:

$$E = C + B. \qquad (1)$$

C is defined as the sum of the squared differences of the surface second derivative from the constraint given by the orientation

FIGURE 3 | Flowchart of proposed shape recovery algorithm. Orientation field and vertical polarity are extracted from an image. Cost function is formulated based on orientation field. Initial values of signs of surface second derivative, σ_{max} and σ_{min}, are obtained by dividing vertical polarity. Estimated surface depth, σ_{max}, and σ_{min} are obtained by minimizing cost function.

field. Boundary condition B consists of the following three terms: $B_0 + B_1 + B_c$. B_0 and B_1 were introduced to resolve the solution's ambiguity. B_0 resolves the translation ambiguity along the depth direction by making the mean depth value zero at the boundary region. B_1 resolves the ambiguity of the affine transformation (Belhumeur et al., 1999) by making the solution not slanted in both the x and y directions. B_c incorporates the knowledge that, near the boundary, $\sigma_{max} = +1$ and σ_{min} equals the sign of the apparent curvature of the 2D contour (Koenderink, 1984), assuming that the 3D surface near the boundary is smooth and differentiable. Thus, we calculated the apparent curvature sign of the 2D contour to formulate B_c. Cost function E depends on σ_{max} and σ_{min} because C and B_c depend on them. We optimized them with the initial values obtained by the vertical polarity (see **Appendix C** for details).

Shape Recovery of Glossy Surfaces

We used 12 glossy surfaces to validate our proposed algorithm (**Figure 4**). We generated them by computer graphics assuming both specular and diffuse reflections of the object's surface. The 3D shapes of objects #1-6 were randomly generated with spherical harmonics, whose complexity rose as the number was increased. The 3D shapes of objects #7-12 were human-made and used in our previous electrophysiological studies of gloss perception (Nishio et al., 2012, 2014). The details of the images and shapes are described in Materials and Methods. The recovered shapes from these glossy surfaces with the ground-truth shapes are shown in **Figure 4**. The depths are represented in grayscale; nearer surfaces are lighter and more distant surfaces are darker. Additionally, 15 contour lines are superimposed. The estimated surface second derivative signs, σ_{max} and σ_{min}, with the ground-truth signs are shown in **Supplementary Figure 1**.

We evaluated the image cues (i.e., orientation field and vertical polarity) and the estimation results as follows. The

orientation field error was quantified by the mean absolute errors throughout the object region between the image and surface orientations and between the image and surface anisotropies. We quantified the error of the vertical polarity by the correct ratio between the initial and true values of σ_{max} and σ_{min}, where the initial values exist. The shape recovery performance was quantified with two measures: global depth correlation r_g and local interior depth correlation r_{li}. The global depth correlation is simply the correlation coefficient of the recovered and true depths throughout the object region. The local interior depth correlation is the averaged value of the correlation coefficients of the recovered and true depths calculated in the local regions except near the boundary. The local interior depth correction is more sensitive to the agreement of the concavity and convexity inside the object region than the global depth correlation. Note that both depth correlations are calculated after the affine transformation so that the slant of the true surface depth becomes zero, because there is ambiguity about the recovered shape's affine transformation (Belhumeur et al., 1999). No values were obtained of the local interior depth correlation of objects #9 and #11 because most of the object region is near the boundary. The details of the measures are described in Materials and Methods. The estimation performance of the surface second derivative signs, σ_{max} and σ_{min}, was quantified by the correct ratio with true values throughout the object region.

We evaluated the images cues before the shape recovery. The average values of the mean absolute error of the orientation and anisotropy for the 12 objects were 11.3° and 0.15. The average values of the correct ratio of the initial values of σ_{max} and σ_{min} for the 12 objects were 0.79 and 0.70. The initial values are shown in **Supplementary Figure 1**.

The estimation performances of the 12 objects are evaluated and summarized in **Table 1**. The average values of global depth correlation r_g and local interior depth correlation r_{li} for the 12 objects were 0.85 and 0.76. As an impression of appearance, the shape recovery seems successful if both the global and local interior depth correlations exceed 0.7. The recovered shapes of objects #1, #2, #5, #6, #7, #8, and #9, where both r_g and r_{li} exceed 0.7, resemble the 3D surface impressions received from the corresponding images in **Figure 4**. The global depth correlations of #10 and #11 and the local interior depth correlations of #3, #4, #10, and #12 were below 0.7. The recovered shapes of #3, #4, #11, and #12 were roughly good but lacked accuracy. The shape of object #10 was not well recovered. The average values of the correct ratios of the estimated σ_{max} and σ_{min} for the 12 objects were 0.86 and 0.72. The correct ratios of the estimated σ_{max} and σ_{min} exceeded those of the initial values even though the initial values exist only in half of the object region.

Shape Recovery of Mirrored Surfaces

The proposed algorithm is applicable to mirrored surfaces without shading although we assumed that shading exists to obtain good initial values of the surface second derivative signs by calculating the vertical polarity. **Figure 5** shows the mirrored surfaces used to validate our proposed algorithm.

The average values of the correct ratio of the initial values of σ_{max} and σ_{min} for the 12 objects were 0.64 and 0.62. These
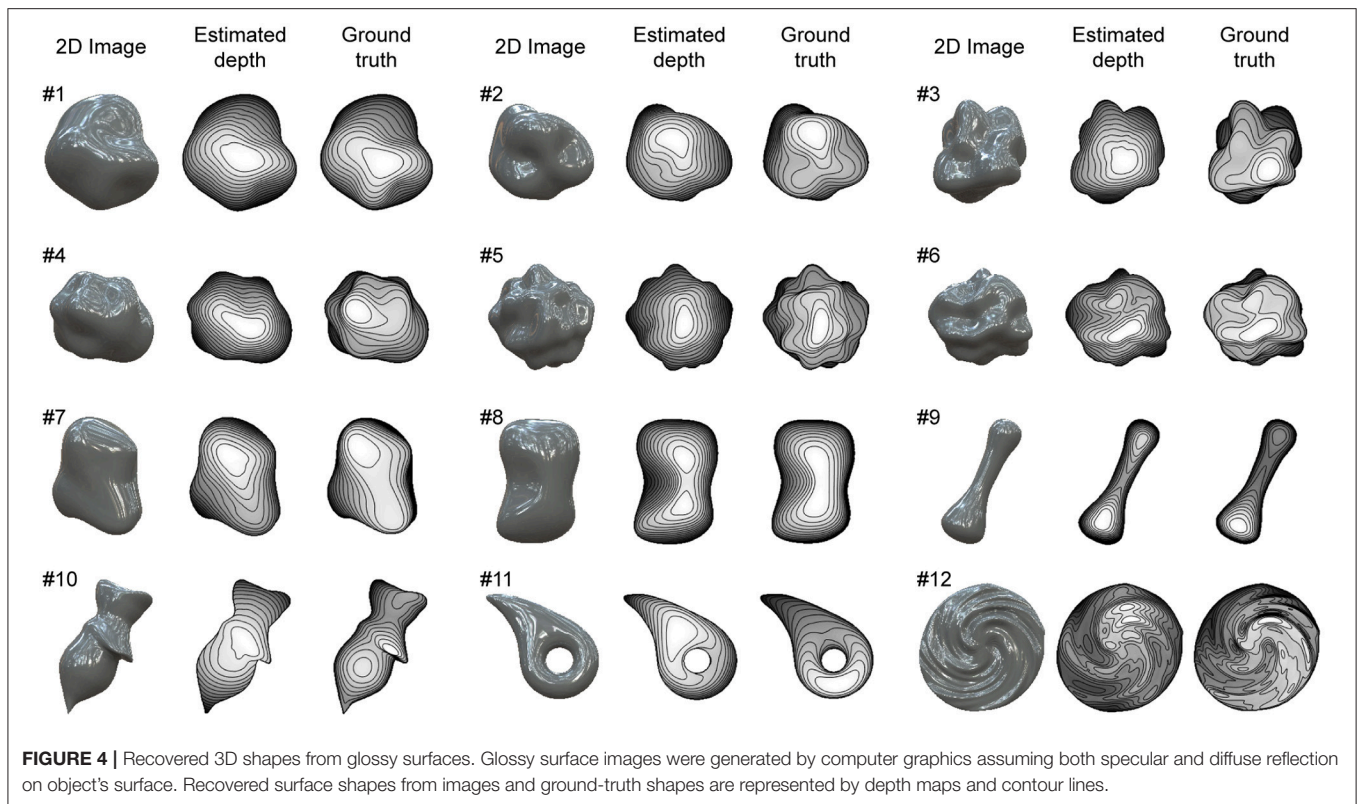
FIGURE 4 | Recovered 3D shapes from glossy surfaces. Glossy surface images were generated by computer graphics assuming both specular and diffuse reflection on object's surface. Recovered surface shapes from images and ground-truth shapes are represented by depth maps and contour lines.

TABLE 1 | Estimation performance of each glossy surface.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	Average
r_g	0.98	0.91	0.87	0.82	0.89	0.88	0.90	0.95	0.89	0.65	0.65	0.80	0.85
r_{li}	0.97	0.71	0.67	0.66	0.91	0.86	0.84	0.95	–	0.45	–	0.60	0.76
σ_{max}	0.90	0.90	0.80	0.81	0.79	0.86	0.87	0.93	0.99	0.86	0.98	0.70	0.86
σ_{min}	0.77	0.74	0.63	0.68	0.64	0.68	0.69	0.82	0.82	0.75	0.78	0.60	0.72

Global and local interior depth correlations of recovered shapes and correct ratios of estimated signs of surface second derivative.

correct ratios were substantially lower than those of the glossy surfaces, but still higher than a chance level of 0.5. The initial values and the correct ratios of all the objects are shown in **Supplementary Figure 2**. The average values of the mean absolute error of the orientation and the anisotropy for the 12 objects were 10.9° and 0.13. These orientation field errors were slightly lower than those of the glossy surfaces, suggesting that the shading component slightly disturbed the relationship between the orientation field and the surface second derivative based on specular reflections.

The recovered shapes from the mirrored surfaces with the ground-truth shapes are shown in **Figure 5**. The estimation performances of the 12 objects are summarized in **Table 2**. The average values of global depth correlation r_g and local interior depth correlation r_{li} for the 12 objects were 0.84 and 0.75. Although the appearances of the recovered shapes from the mirrored surfaces look noisier than those from the glossy surfaces (e.g., #1 and #8), the averaged global and local interior depth correlations differ by only 0.01 and 0.01, indicating that the

proposed shape recovery algorithm is applicable to both mirrored and glossy surfaces. The average values of the correct ratios of the estimated σ_{max} and σ_{min} for the 12 objects were 0.80 and 0.70. The noisier appearance of the recovered shapes of the mirrored surfaces is related to the lower correct ratio of the estimated σ_{max} than that of the glossy surfaces. The estimated surface second derivative signs, σ_{max} and σ_{min}, with the ground-truth signs are shown in **Supplementary Figure 2**.

Estimation Accuracy in Different Conditions

We tested the proposed algorithm in four different conditions. The first and second conditions are the shape recoveries from the glossy and mirrored surfaces shown in **Figures 4**, **5** (denoted as glossy and mirrored conditions). In the third condition, the shapes were recovered from the glossy surfaces shown in **Figure 4**, but the light from above prior was not used (denoted as the noLFAP condition). And in the fourth, the shapes were recovered from the shape orientation fields that were obtained
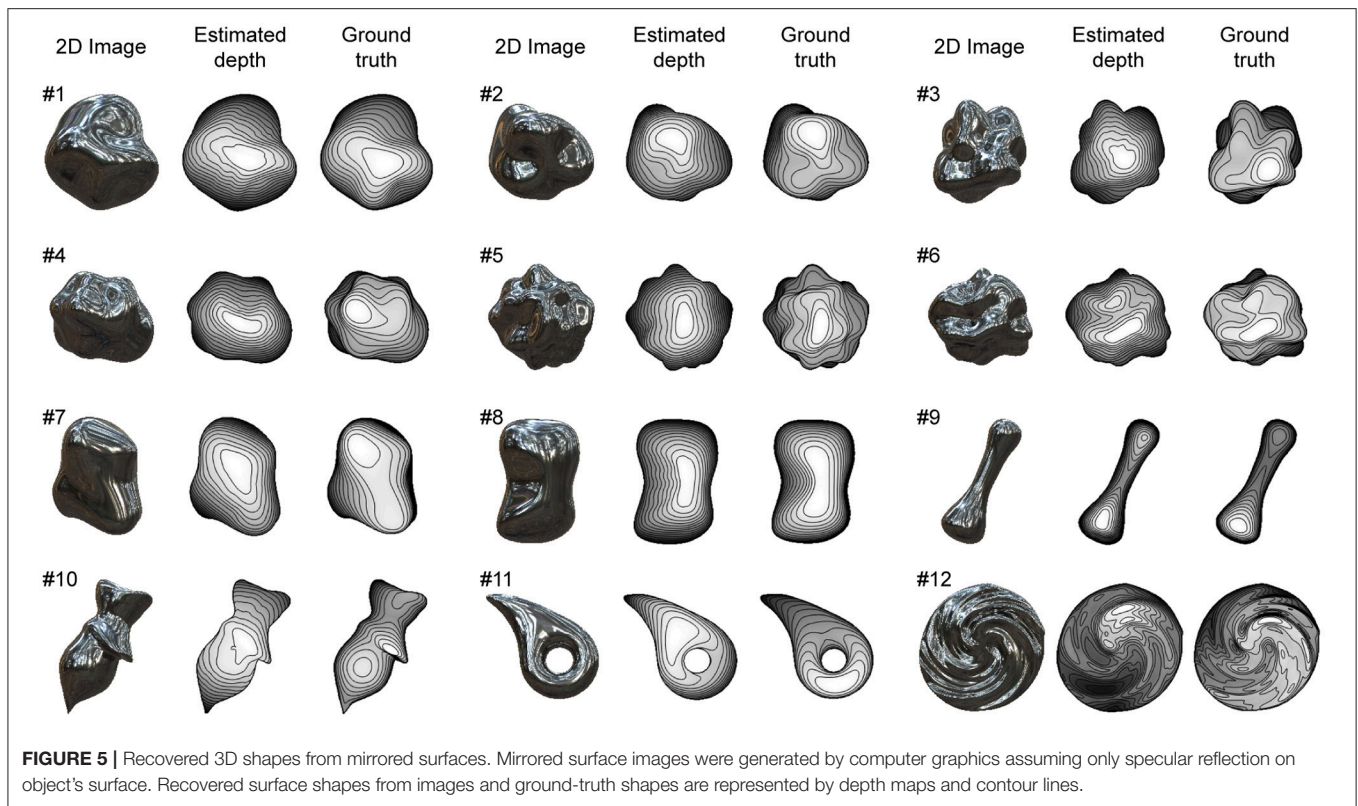
FIGURE 5 | Recovered 3D shapes from mirrored surfaces. Mirrored surface images were generated by computer graphics assuming only specular reflection on object's surface. Recovered surface shapes from images and ground-truth shapes are represented by depth maps and contour lines.

TABLE 2 | Estimation performance of each mirrored surface.

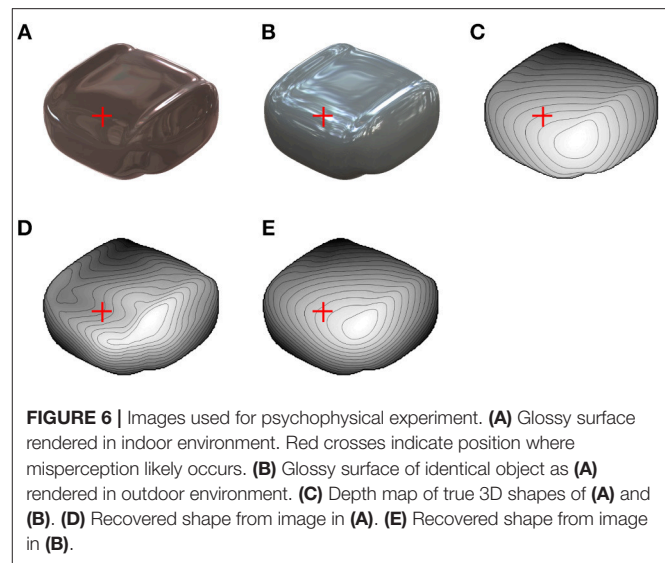|	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	Average
r_g	0.95	0.93	0.78	0.81	0.89	0.89	0.81	0.93	0.91	0.66	0.66	0.86	0.84
r_{li}	0.96	0.74	0.60	0.67	0.90	0.85	0.67	0.91	–	0.55	–	0.65	0.75
σ_{max}	0.74	0.85	0.74	0.72	0.74	0.79	0.80	0.84	0.97	0.77	0.99	0.68	0.80
σ_{min}	0.67	0.73	0.64	0.68	0.63	0.67	0.68	0.79	0.83	0.70	0.79	0.61	0.70

Global and local interior depth correlations of recovered shapes and correct ratios of estimated signs of surface second derivative.

from the true 3D shapes (denoted as the shapeOF condition). Note that in the shapeOF condition, the same initial values of σ_{max} and σ_{min} were used as the glossy condition. **Tables 3, 4** summarize the errors of the image cues and the estimation performances of the four conditions. Additionally, we tested the algorithm in three more conditions to investigate the effect of the contour constraint, the illumination environment, and the image resolution. These results are shown in **Supplementary Note 1**.

In the noLFAP condition, the shapes were recovered from the glossy surfaces without the light from above prior to check its necessity. In this condition, the initial values of σ_{max} and σ_{min} were all set to $+1$ based on the convex prior possessed by humans (Langer and Bülthoff, 2001; Liu and Todd, 2004). The average values of the correct ratio of the initial values of σ_{max} and σ_{min} for the 12 objects were 0.85 and 0.67. First, the shapes were recovered with the same algorithm that was used with the other conditions. As a result, the estimated σ_{max} and σ_{min} were almost the same as the initial values; 98 and 88% of the estimated σ_{max} and σ_{min} were $+1$. This means that the distinction between convex and

TABLE 3 | Errors of image cues of four conditions.

	Mean absolute error		Correct ratio	
	Orientation	**Anisotropy**	**Initial σ_{max}**	**Initial σ_{min}**
Glossy	11.3°	0.15	0.79	0.70
Mirrored	10.9°	0.13	0.64	0.62
noLFAP	(11.3°)	(0.15)	0.85	0.67
shapeOF	0°	0	(0.79)	(0.70)

Orientation field errors and correct ratios of initial values of σ_{max} and σ_{min}, which are averaged values of 12 objects. Values of orientation field errors of noLFAP condition and correct ratios of initial values of shapeOF condition are parenthesized because these are identical as glossy condition.

concave failed without the light from above prior, although the distinction was successful with it. The average values of the global and local interior depth correlations for the 12 objects were $r_g = 0.74$ and $r_{li} = 0.48$. These estimation performances are not summarized in **Table 4**, because the estimation completely failed.

TABLE 4 | Estimation performances of four conditions.

	Shape recovery accuracy		Correct ratio	
	r_g	r_{li}	Estimated σ_{max}	Estimated σ_{min}
Glossy	0.85	0.76	0.86	0.72
Mirrored	0.84	0.75	0.80	0.70
noLFAP	0.77	0.52	0.80	0.67
shapeOF	0.87	0.88	0.92	0.80

Global and local interior depth correlations of recovered shapes and correct ratios of estimated signs of surface second derivative. These are averaged values of 12 objects.

Next we altered the temperature parameter of the mean field algorithm (see **Appendix C** for details) from $\beta_0 = 10$ to $\beta_0 = 1$ to extend the search range (Parisi, 1988), since the initial values were not reliable in this condition. As a result, we obtained better shape recovery results. The average values of the global and local interior depth correlation for the 12 objects were $r_g = 0.77$ and $r_{li} = 0.52$. The average values of the correct ratio of the estimated σ_{max} and σ_{min} for the 12 objects were 0.80 and 0.67. The estimation performances of objects #1, #8, and #9 were high despite the noLFAP condition. However, most of the recovered shapes look noisy, probably because of the alternation of the temperature parameter, and the estimation performance was lowest in the four conditions. This result suggests again that shape recovery is difficult without the light from above prior. The recovered shapes and the estimated signs of the surface second derivative of the noLFAP condition are shown in **Supplementary Figure 3** (see also **Supplementary Table 1**).

In the shapeOF condition, the shapes were recovered from the surface orientations that were obtained from the true 3D shapes instead of the image orientations to investigate the effect of the orientation field errors on the shape recovery errors. In this condition, the vertical polarity of the glossy surfaces was used to resolve the concave/convex ambiguity. The average values of the global and local interior depth correlations for the 12 objects were $r_g = 0.87$ and $r_{li} = 0.88$. The average values of the correct ratio of the estimated σ_{max} and σ_{min} for the 12 objects were 0.92 and 0.80. The estimation performances of the shapeOF condition were very high, except for objects #9 and #10, and substantially higher than the other conditions. The recovered shapes and the estimated signs of the surface second derivative of the shapeOF condition are shown in **Supplementary Figure 4** (see also **Supplementary Table 2**).

Consistency With Human Shape Perception

Finally, we conducted a psychophysical experiment to investigate the linkage between the shape recovery algorithm and human shape perception. We prepared a glossy surface image that evokes 3D shape misperception (**Figure 6A**) by using another illumination environment (Galileo's Tomb of the Devebec dataset). This illumination environment was taken indoors with a dark ceiling against the light from above prior. **Figure 6B** is an image of the same object rendered under identical illumination

FIGURE 6 | Images used for psychophysical experiment. **(A)** Glossy surface rendered in indoor environment. Red crosses indicate position where misperception likely occurs. **(B)** Glossy surface of identical object as **(A)** rendered in outdoor environment. **(C)** Depth map of true 3D shapes of **(A)** and **(B)**. **(D)** Recovered shape from image in **(A)**. **(E)** Recovered shape from image in **(B)**.

environments as **Figures 4**, **5** (Eucalyptus Grove of the Devebec dataset), which was taken outdoors and is consistent with the light from above prior. **Figure 6C** represents the depth map of the true 3D shapes. The red cross indicates where the surface looks concave from **Figure 6A**, although the surface looks convex from **Figure 6B** and the true surface is convex. In **Figure 6A**, the dark ceiling of the illumination environment caused a negative value of the vertical polarity around the red cross mark despite its convex 3D shape, and humans can perceive the concave shape, assuming that the light comes from above. We carefully made the 3D object's shape so that the image clearly evokes the misperception and the evoked misperceived shape region is consistent with the surrounding information. **Figures 6D,E** indicate the recovered shapes from the images of **Figures 6A,B**. In accordance with the appearance, the recovered shape from **Figure 6A** is concave and that from **Figure 6B** is convex around the red cross mark. The estimation performances (r_g, r_{li}, correct ratio of estimated σ_{max} and correct ratio of estimated σ_{min}) of **Figures 6D,E** were (0.91, 0.76, 0.73, 0.60) and (0.98, 0.99, 0.90, 0.84).

In psychophysical experiments, five subjects were asked whether the local 3D surface around the red crosses in **Figures 6A,B** looks convex or concave. Four perceived the incorrect concave shape from **Figure 6A** as consistent with the recovered shape by the proposed algorithm (**Figure 6D**), and all five perceived a convex shape from **Figure 6B** (see **Appendix B** for detailed results).

DISCUSSION

We developed an algorithm that estimates 3D shapes from a single specular image to investigate a possible mechanism of human 3D shape perception from specular reflections. This algorithm mainly relies on the orientation field suggested by a previous psychophysical study (Fleming et al., 2004). However, since the orientation field cannot resolve the local concave/convex ambiguity, the 3D shape recovery from it alone

was difficult (see the noLFAP condition, **Table 4**). To resolve the concave/convex ambiguity, we added the prior knowledge that objects are illuminated from above. The vertical polarity of the intensity gradient is an image cue to utilize this prior knowledge. We evaluated the developed algorithm with the glossy and mirrored surfaces of 12 complex shapes. The depth correlations between the recovered and the true shapes were as high as around 0.8. To further confirm the necessity of the vertical polarity information, we also conducted a psychophysical experiment with an image that caused human misperception due to the inconsistency with the light from above prior. The human-misperceived and recovered shapes were consistent in most subjects. These findings show that the vertical polarity of the intensity gradient as well as the orientation field are related to 3D shape perception and the combination of both enables 3D shape recovery from a single specular image.

Shape Recovery of Mirrored Surfaces

The shape recovery performance of the mirrored condition was almost as high as the glossy condition (**Table 4**), although the relationship between the vertical polarity and the surface second derivative sign was only proved in the diffuse reflection component (see Materials and Methods). The present result indicates that vertical polarity of the specular component was also useful for the initial second derivative signs for the following reason. The diffuse reflectance component in **Figure 2** shows a relationship where the luminance is high in the upper side and low in the lower side when the surface is convex with respect to the vertical orientation (**Figures 2A,B**) and vice versa (**Figures 2C,D**). The same relationship holds for the mirrored surfaces of **Figure 1**. The luminance tends to be higher in the upper side than the lower side when the surface is convex (**Figures 1A,B**) and vice versa (**Figures 1C,D**). Thus, the vertical polarity of the mirrored surface at low frequencies is related to the surface second derivative sign of the vertical orientation, although the high-frequency component is not related to it. When the vertical polarity is calculated, a relatively low-frequency image component is extracted and further smoothed to remove the high-frequency component of the specular reflection (see Materials and Methods). Therefore, it provides meaningful information about second derivative signs even from mirrored surfaces, although the correct ratio of the initial sign values of the mirrored condition is actually worse than that of the glossy condition (**Table 3**).

Representation of Surface Curvatures

In this study, the sign and magnitude of the surface second derivatives are separately described. Similar representation can be seen in some psychophysical experiments (Koenderink et al., 2014; Dövencioglu et al., 2015), in which subjects classified 3D shapes based on curvature signs. Furthermore, the neural representation of surface curvatures was studied in electrophysiological experiments. Srivastava et al. showed that the neurons in the inferior temporal cortex (the area for object recognition) of macaques are mainly sensitive to the curvature sign, but the neurons in the anterior intraparietal area (the area for motor planning) are sensitive to the curvature

magnitude as well as the sign (Srivastava et al., 2009). This might suggest that the curvature sign's representation is important for object recognition, and its magnitude is also required for motor planning. It is also interesting to note that humans are more sensitive to concavity (negative curvature) than convexity (positive curvature) in change detection and object recognition (Cohen et al., 2005; Leek et al., 2012; Davitt et al., 2014). These and other psychophysical and electrophysiological studies (Yamane et al., 2008; Orban, 2011) provide hints to develop more efficient and human-like shape recovery algorithms.

The estimation of a small surface second derivative sign, σ_{min}, was more difficult than that of a large surface second derivative, σ_{max}, in all four conditions (see right half of **Table 4**). A similar phenomenon can be seen in human shape perception. When subjects classified local shapes based on the curvature signs, saddles were often misclassified as ridges or ruts (convex or concave cylinders; Koenderink et al., 2014; Dövencioglu et al., 2015), suggesting that humans often neglect the small surface curvature of saddle shapes. Since the small surface curvature is less visible in the image, its estimation is intrinsically difficult. In the proposed algorithm, the small second derivative sign is forcibly classified as +1 or −1, but it might be better to treat it ambiguously like the quantum superposition when its classification is difficult.

Note here that the shape recovery from specular reflections has much in common with that from line drawings because lines or specular orientations appear at the high curvature in both cases (Todd, 2004; Cole et al., 2009). In a line drawing study, edge-labeling algorithms classified the orientation edges as either convex or concave (Huffman, 1971; Malik, 1987). This corresponds to the determination of the large surface second derivative sign in our study. It would be interesting to find and utilize the similarities of the shape recovery algorithms from specular reflection and line drawing (Iarussi et al., 2015).

Origin of Shape Recovery Errors

The orientation field error is a major error factor of the proposed algorithm, because the shape recovery performance was very high in the shapeOF condition (**Table 4** and **Supplementary Figure 4**). In this condition, the surface second derivative signs were accurately estimated even though the initial values from the vertical polarity were somewhat incorrect and absent in half of the region. This result indicates that the proposed shape recovery algorithm works well at least under such ideal conditions. Therefore, the error due to the proposed algorithm's methodological imperfection is relatively small. It also indicates that the orientation field is satisfactory for the 3D shape recovery of such curved surfaces examined in this study with the help of the light from above prior. The difference of the shape recovery performances between the glossy and shapeOF conditions reflects errors that originate from the image orientation field. Compared with the orientation field error, the effects of the initial second derivative sign errors are limited because they are expected to be corrected through optimization; orientation field error inevitably affects the resultant shape because it is directly incorporated in the cost function. Actually, the shape recovery performance of the mirrored condition was

comparable to the glossy condition even though the initial second derivative sign errors of the mirrored condition were considerably larger than those of the glossy condition. Of course, too many initial errors cannot be corrected as suggested by the poor shape recovery performance of the noLFAP condition. The orientation field errors probably affect the error corrections of the initial values through optimization.

Limitations and Future Work

The following are the limitations of our shape recovery algorithm. First, since it is based on the relationship between the orientation field and the surface second derivative, large error occurs when this relationship is invalid. For example, if the illumination environment is biased to a specific orientation (e.g., striped illumination), it biases the image orientation (Fleming et al., 2009). The orientation error becomes large where the surface anisotropy is small (Fleming et al., 2009). For example, if the true shape is a plane (i.e., the surface anisotropy is zero), the image orientation reflects not the surface second derivative but only the orientation of the illumination environment and causes shape recovery errors. Second, images under an unnatural illumination environment against the light from above prior could not be properly recovered as it is difficult for humans (Savarese et al., 2004; Faisman and Langer, 2013). Third, the proposed algorithm cannot estimate the depth scale as well as the slant due to the ambiguity about the affine transformation of the recovered shape (Belhumeur et al., 1999). Humans also have difficulty estimating the slant (Koenderink et al., 2001; Khang et al., 2007) and the depth scale (Belhumeur et al., 1999; Khang et al., 2007) from a single image without prior knowledge of the object's shape. Therefore, we evaluated the recovered shapes by depth correlations after the affine transformation so that the slant of the true surface depth becomes zero. We did not evaluate the normal map because it depends on the depth scale. Fourth, because the proposed algorithm assumes that the surface depth is second order differentiable, it cannot properly treat bends, cusps, and self-occlusion inside the object region [occluding edges or limbs (Malik, 1987)] and generates smoother shapes than actual shapes. This property may worsen the shape recovery performance of objects #10, #3, and #4. Note that the limitations listed above (except for the fourth) are closely related to the limitations of human shape perception.

Future work has several promising directions. First, further psychophysical experiments are required to understand human shape perception from specular reflections in detail and will help improve the shape recovery algorithm to better simulate the human shape perception. It would be interesting to use the image-based shape manipulation method based on the orientation field (Vergne et al., 2016) to compare the recovered and human-perceived shapes. It would also be interesting to model and examine human shape perception from the viewpoint of surface-based representations hypothesis (Leek et al., 2005, 2009; Reppa et al., 2015). Second, the proposed shape recovery algorithm will be useful for computer vision methods. By integrating it with a study that estimates material (BRDF) from a single image of a known shape (Romeiro and Zickler, 2010), it might become possible to estimate an unknown shape's material.

By providing more accurate recovered depth information, we expect to enhance the reality of the image-based material editing that is based on shape information (Khan et al., 2006). For further improvement of the shape recovery performance, the proposed shape from the specularity algorithm could be integrated with the shape from shading algorithms (Kunsberg and Zucker, 2014; Barron and Malik, 2015), where it would be helpful to use color information to separate diffuse and specular reflection components (Artusi et al., 2011). Third, it would be interesting to study whether 3D shapes can be recovered from translucent images with specularities. A previous study (Motoyoshi, 2010) argued that an object looks translucent when images are manipulated so that the diffuse reflection component is contrast-reversed, but the specular reflection component is left intact. This result suggests that we must alter how the specular and diffuse reflection components are combined for shape recovery from translucent images, such as reversing the sign of the vertical polarity in the case of translucent images compared with opaque images.

MATERIALS AND METHODS

As a precondition to 3D shape recovery, we assume that the image region is known where the object exists. It may be obtained by an edge detection algorithm or decided by humans. We denote the object region as Ω, the number of pixels in Ω as N_Ω, the boundary region, which is the region between the boundary of Ω and one pixel inside it, as $\partial\Omega$, and the number of pixels in $\partial\Omega$ as $N_{\partial\Omega}$. The resolution of the 3D shape recovery was 256×256 pixels. We set a Cartesian coordinate on the image plane, where the x and y axes represent the horizontal and vertical axes of an image plane and the z axis represents the front direction. We represent the depth of the 3D object surface as z(x,y). The following notations are used: scalars are represented in normal-type letters as x; vectors are represented in lower-case boldface letters as $\mathbf{x}$; matrices are represented in upper-case boldface letters as $\mathbf{X}$.

Images and Extraction of Image Cues

We used the images of 12 different 3D shapes to evaluate the proposed algorithm (**Figures 4**, **5**). The images had $1,024 \times 1,024$ pixel resolution and were colored, although they were downsampled to 256×256 pixels before the 3D shape recovery and became achromatic because the proposed algorithm does not use color information. These images were rendered by Radiance software (http://radsite.lbl.gov/radiance/). The surface reflection property was modeled by the Ward-Duer model (Ward, 1992; Ngan et al., 2005). We set diffuse reflectance ρ_d, specular reflectance ρ_s, and the spread of specular reflection α as $\rho_d = 0.1$, $\rho_s = 0.15$, $\alpha = 0$ for the glossy surfaces (**Figure 4**) and $\rho_d = 0$, $\rho_s = 0.25$, $\alpha = 0$ for the mirrored surfaces (**Figure 5**). For the natural illumination environment, we used a high dynamic range image from the Devebec dataset (http://ict.debevec.org/~debevec/; Eucalyptus Grove). For the quadratic patch images in **Figures 1**, **2**, we set $\rho_d = 0$, $\rho_s = 0.25$, $\alpha = 0$ for the mirrored surfaces in **Figure 1**, and $\rho_d = 0.4$, $\rho_s = 0$ for the matte images in **Figure 2**.

The 3D shapes of objects #1-6 were randomly generated with spherical harmonics. To incrementally increase the complexity of the 3D shapes, the maximum degree of the spherical harmonics was limited to 5 for objects #1-2, 7 for objects #3-4, and 10 for objects #5-6. The weights of the spherical harmonics were determined by a random number and normalized so that the power of each degree is inversely proportional to the degree (pink noise). Then the maximum amplitude of the spherical harmonics was normalized to 0.5. The object's radius of each angle is given by the sum of 1 and the value of the spherical harmonics.

We extracted the orientation field as follows. The image orientation $\theta(x,y)$ is the angle that maximizes the magnitude of response p of the oriented filter (first-derivative operator) as $\theta(x, y) = \text{argmax}_{\theta'} p^2(\theta'(x, y))$. Image anisotropy $\alpha(x,y)$ is defined by the ratio of the minimum and maximum magnitudes of the oriented filter response with respect to its angle (Fleming et al., 2004) as $\alpha(x, y) = 1 - \sqrt{\frac{p^2_{\min}(x,y)}{p^2_{\max}(x,y)}}$. $\alpha = 0$ means that the local image is isotropic, and $\alpha = 1$ means that it only consists of one directional component. The steerable pyramid (Simoncelli et al., 1992; Simoncelli and Freeman, 1995) (matlabPyrTools, https://github.com/LabForComputationalVision/matlabPyrTools) was used to extract the image orientation in accordance with previous studies (Fleming et al., 2004, 2009, 2011). Responses were obtained by steering the filter through 120 equal orientation steps between 0 and 180°. The orientation responses at the finest possible spatial scale (1,024 × 1,024 pixel resolution) were extracted for all the shapes in accordance with a previous study (Fleming et al., 2004). Then the amplitudes, which are the squared responses, were downsampled to 256 × 256 pixels and convolved by a 3 × 3 constant filter for noise reduction. Then the image orientation and the image anisotropy were obtained based on the above equations.

We obtained the vertical polarity of intensity gradient $p_v(x,y)$ by extracting the sign of the oriented filter response of the vertical direction ($\theta = 0°$) as $p_v(x, y) = \text{sgn}(p_{\theta=0°}(x, y))$. The steerable pyramid was used to extract the vertical polarity. The responses of the pyramid level of 256 × 256 resolution were extracted for all the shapes (a relatively low-frequency component compared to the original image resolution, 1024 × 1024 resolution). The response values near the boundary are unreliable because they are affected by the image outside of the object region. Therefore, we overwrote the response values within five pixels from the boundary to zeros and smoothed them by a Gaussian filter whose standard deviation is four pixels to reduce the noise and the high-frequency components of the specular reflection.

We derived the signs of the apparent curvature of the image contour as follows. First, we drew a circle centered at a boundary point with a radius of 128 pixels (1/8 of the image size); second, we determined that the curvature sign value at that boundary point is +1 or −1 when the object region's area within the circle is smaller or larger than the area of the outside object region within the circle; third, for noise reduction, we smoothed the curvature sign values by convolving a constant circular filter of a radius of 16 pixels (1/64 of the image size) and downsampled it to 256 × 256 pixels; then we extracted the signs.

The resultant curvature signs of the image contour are shown in **Supplementary Figure 5**.

Curvature Formulation

We described the surface shape of objects by Hessian matrix H(z) of surface depth z(x,y). Because the Hessian matrix is symmetric, H(z) is diagonalized with rotation matrix R as

$$H(z) = \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial x \partial y} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} = -R(\theta_s) \begin{pmatrix} k_{\min} & 0 \\ 0 & k_{\max} \end{pmatrix} R(-\theta_s), \quad (2)$$

where $k_{\max}$ and $k_{\min}$ are the eigenvalues of the larger and smaller magnitudes. θ_s, which indicates the angle of the small surface second derivative, is called the surface orientation. There is a minus at the beginning of the right-hand side of Equation (2) so that the surface second derivatives become positive in the case of convex shapes (e.g., sphere). In this study, we described the surface curvature by Hessian matrix based on the image coordinate system instead of the standard curvature that is defined on the object surface's intrinsic coordinate system. This difference was previously scrutinized (Fleming et al., 2009). The reason why we adopted the former is that orientation field depends on the Hessian matrix, not on the standard curvatures. For example, in the case of a sphere, the standard curvature is the same at every point on its surface. In contrast, the second derivatives are large near the boundary and small at the center, and correspondingly, the image orientation of the specular reflectance is clear near the boundary and not clear at the center (see Figure 16 of Fleming et al., 2009).

Next we introduce other variables and transform the equation.

First, surface anisotropy α_s is defined as $\alpha_s = 1 - \sqrt{\frac{k^2_{\min}}{k^2_{\max}}}$ (Fleming et al., 2004). $\alpha_s = 0$ denotes that the magnitude of the two surface second derivatives is the same (e.g., a convex sphere, a concave sphere, or a saddle), and $\alpha_s = 1$ means that the small surface second derivative is zero (e.g., a convex cylinder or a concave cylinder). Second, variables are introduced so that the surface second derivative's sign and magnitude are separately described. The sign of the large surface second derivative is represented as $\sigma_{\max} \in \{+1, -1\}$. +1 and −1 correspond to convex and concave shapes. The magnitude of the large surface second derivative is represented as $k_a = |k_{\max}|$. The sign of the small surface second derivative is represented as $\sigma_{\min} \in \{+1, -1\}$. Using these variables, the surface second derivatives are described:

$$k_{\max} = k_a \sigma_{\max}, \quad (3)$$
$$k_{\min} = (1 - \alpha_s) k_a \sigma_{\min}. \quad (4)$$

Relationship Between Vertical Polarity and Surface Second Derivative Signs

With the prior knowledge that the object is illuminated from above, we can derive the relationship among the vertical polarity, p_v, and the surface second derivative signs. In the case of the Lambert reflectance, the surface luminance is proportional to the inner product of the lighting direction and the surface's normal direction. Here we assume that the illumination map is

stronger as it gets closer to just above $(x,y,z) = (0,1,0)$. As a result, the surface luminance becomes stronger as the surface slant $(-\partial z/\partial y)$ is increased. By taking a derivative of this relationship with respect to y and taking the sign, the following equation is obtained:

$$p_v \approx \text{sgn}\left(-\frac{\partial^2 z}{\partial y^2}\right). \tag{5}$$

Here we described it as approximately equal instead of equal because the two assumptions of the Lambert reflectance and lighting from just above do not strictly hold in real situations. For example, for images taken outdoors, the angle of the sun (dominant illumination) changes based on time.

We transform Equation (5) into a more available form. The following equation is derived from Equations (2), (3), and (4) as $-\frac{\partial^2 z}{\partial y^2} = k_a \left(\sigma_{\max} \cos^2 \theta_s + \sigma_{\min} (1 - \alpha_s) \sin^2 \theta_s\right)$. Then we used the approximation of orientation $\theta \approx \theta_s$ and anisotropy $\alpha \approx \alpha_s$:

$$p_v \approx \text{sgn}\left(\sigma_{\max} \cos^2 \theta + \sigma_{\min} (1 - \alpha) \sin^2 \theta\right). \tag{6}$$

We divided object region Ω into two regions: $\cos^2 \theta \geq (1 - \alpha) \sin^2 \theta$ holds in Ω_a, but not in Ω_b. Then the following relationship is obtained:

$$p_v(x, y) \approx \begin{cases} \sigma_{\max}(x, y) & (x, y \in \Omega_a) \\ \sigma_{\min}(x, y) & (x, y \in \Omega_b) \end{cases}. \tag{7}$$

The approximation of Equation (7) was evaluated in our experiment and summarized in the right half of **Table 3**. All of the results of the objects in the glossy and mirror conditions are shown in **Supplementary Figures 1**, **2**.

Formulation of Cost Function

Cost function E consists of two terms: the second derivative constraint given by orientation field C and boundary condition B:

$$E = C + B. \tag{8}$$

We first explain second derivative constraint C and then boundary condition B, which consists of the following three terms: $B = B_0 + B_1 + B_c$.

The second derivative constraint is based on the relationship between the orientation field and the surface second derivatives where the image orientation approximates surface orientation $\theta \approx \theta_s$ and the image anisotropy approximates surface anisotropy $\alpha \approx \alpha_s$ (Fleming et al., 2004, 2009). These relationships are described with error terms as $\theta_s = \theta + \delta\theta$ and $\alpha_s = \alpha + \delta\alpha$. These errors were evaluated in our experiment and summarized in the left half of **Table 3**. For more information, a previous study (Fleming et al., 2009) assessed the orientation error, which depends on the surface anisotropy and the difference between the surface orientation and the illumination map's orientation. By substituting these equations into Equation (2), we obtain

$$R(-\theta) H(z) R(\theta) + k_a \begin{pmatrix} (1 - \alpha) \sigma_{\min} & 0 \\ 0 & \sigma_{\max} \end{pmatrix} = k_a (O(\delta\alpha) + O(\delta\theta)). \tag{9}$$

To simplify this equation, we introduce the coordinate axes (u, v) by rotating the original axes (x, y) by image orientation $\theta(x,y)$. Note that the axes (u, v) depend on each position based on the image orientation in that position. Then this equation is described as

$$\begin{pmatrix} \frac{\partial^2 z}{\partial u^2} + k_a (1 - \alpha) \sigma_{\min} & \frac{\partial^2 z}{\partial u \partial v} \\ \frac{\partial^2 z}{\partial u \partial v} & \frac{\partial^2 z}{\partial v^2} + k_a \sigma_{\max} \end{pmatrix} = k_a (O(\delta\alpha) + O(\delta\theta)), \tag{10}$$

which indicates that the surface strongly bends toward the v direction (the orthogonal direction of the image orientation) by second derivative magnitude k_a with sign $\sigma_{\max}$ and the surface weakly bends toward the u direction by second derivative magnitude $k_a(1-\alpha)$ with sign $\sigma_{\min}$. Second derivative constraint C is based on Equation (10) where the left-hand side is small. The cost is the sum of the squared Frobenius norm of the left-hand side of Equation (10) throughout the object region:

$$C = \frac{1}{2} \sum_{x,y \in \Omega} \left[\left(\frac{\partial^2 z}{\partial u^2} + k_a (1 - \alpha) \sigma_{\min}\right)^2 + \left(\frac{\partial^2 z}{\partial v^2} + k_a \sigma_{\max}\right)^2 \right.$$
$$\left. + 2\left(\frac{\partial^2 z}{\partial u \partial v}\right)^2 \right]. \tag{11}$$

Since this cost function is a quadratic function with respect to z and k_a or with respect to $\sigma_{\max}$ and $\sigma_{\min}$, it is relatively easy to optimize.

Here, because the right-hand side of Equation (10) is proportional to k_a, it would be more appropriate to use a cost function that is the sum of the amplitude of the left-hand side of Equation (10) after multiplied by $1/k_a$. We denote this cost function as C':

$$C' = \frac{1}{2} \sum_{x,y \in \Omega} \left[\left(\frac{1}{k_a}\frac{\partial^2 z}{\partial u^2} + (1 - \alpha) \sigma_{\min}\right)^2 + \left(\frac{1}{k_a}\frac{\partial^2 z}{\partial v^2} + \sigma_{\max}\right)^2 \right.$$
$$\left. + 2\left(\frac{1}{k_a}\frac{\partial^2 z}{\partial u \partial v}\right)^2 \right]. \tag{12}$$

However, cost function C' is more difficult to optimize. Therefore, we use the first cost function C to obtain the solution, and then with the solution as an initial value, we obtain the improved solution with the second cost function C'. The summarized formula and the minimization of the second cost function are described in **Appendix D**.

Boundary conditions B_0 and B_1 were introduced to resolve the solution's ambiguity. B_0 resolves the translation ambiguity along the z axis by making the mean depth value zero at the boundary region:

$$B_0 = \frac{1}{2} \left(\frac{1}{N_{\partial\Omega}} \sum_{x,y \in \partial\Omega} z\right)^2. \tag{13}$$

Another ambiguity exists about affine transformations (Belhumeur et al., 1999). B_1 is introduced so that the solution is not slanted in both the x and y directions:

$$B_1 = \frac{1}{2}\left(\frac{1}{N_{\partial\Omega}}\sum_{x,y\in\partial\Omega}(x - x_{CM})z\right)^2 + \frac{1}{2}\left(\frac{1}{N_{\partial\Omega}}\sum_{x,y\in\partial\Omega}(y - y_{CM})z\right)^2,$$
(14)

where x_{CM} and y_{CM} are the average values of x and y in boundary region $\partial\Omega$. We summarize these boundary conditions as $B_0 + B_1 = \frac{1}{2}\mathbf{z}^T\mathbf{B}\mathbf{z}$, where $\mathbf{z}$ is the column vector of size $N_\Omega \times 1$ that consists of z(x,y) in object region Ω and $\mathbf{B}$ is the coefficient matrix of size $N_\Omega \times N_\Omega$.

Next the constraint from the contour was introduced. Assuming that the 3D surface near the boundary is smooth and differentiable, the second derivative toward the normal direction of the contour at the boundary is minus infinity. Therefore, the surface orientation is parallel to the contour and $\sigma_{max} = +1$. Moreover, a previous study (Koenderink, 1984) proved that the sign of the 3D curvature parallel to the contour ($= \sigma_{min}$) equals the sign of the apparent curvature of the 2D contour. The apparent curvature sign of the image contour, which is calculated and utilized as the initial values of σ_{min} near the boundary, is also incorporated in the cost function:

$$B_c = -\left(\mathbf{h}_{max}^T\boldsymbol{\sigma}_{max} + \mathbf{h}_{min}^T\boldsymbol{\sigma}_{min}\right),$$
(15)

where $\mathbf{h}_{min}$ is a column vector that consists of the contour's curvature sign (**Supplementary Figure 5**), $\mathbf{h}_{max}$ is a column vector that consists of +1 (near the boundary, where the value exists in **Supplementary Figure 5**) and 0 (otherwise) and $\boldsymbol{\sigma}_{max}$ and $\boldsymbol{\sigma}_{min}$ are column vectors that consist of $\sigma_{max}(x,y)$ and $\sigma_{min}(x,y)$. Note that although our constraint from the contour's curvature sign does not depend on the contour's curvature magnitude, using a non-uniform constraint would be interesting based on information theories and empirical findings (Feldman and Singh, 2005; Lim and Leek, 2012).

The cost function is summarized as

$$E = C + B = \frac{1}{2}\mathbf{z}^T\mathbf{A}\mathbf{z} + \mathbf{z}^T\left(\mathbf{D}_{vv}^T\mathbf{K}_a\boldsymbol{\sigma}_{max} + \mathbf{D}_{uu}^T\mathbf{H}\mathbf{K}_a\boldsymbol{\sigma}_{min}\right)$$
$$+ \frac{1}{2}\mathbf{k}_a^T\left(\mathbf{I} + \mathbf{H}^2\right)\mathbf{k}_a - \left(\mathbf{h}_{max}^T\boldsymbol{\sigma}_{max} + \mathbf{h}_{min}^T\boldsymbol{\sigma}_{min}\right),$$
(16)

where $\mathbf{k}_a$ and $\boldsymbol{\alpha}$ are column vectors that consist of $k_a(x,y)$ and $\alpha(x,y)$; $\mathbf{K}_a$ and $\mathbf{H}$ are diagonal matrices with diagonal elements $\mathbf{k}_a$ and $(\mathbf{1}\text{-}\boldsymbol{\alpha})$; $\mathbf{D}$ is a matrix that represents the second order differential operator with respect to subscript variables; $\mathbf{A} = \mathbf{D}_{vv}^T\mathbf{D}_{vv} + \mathbf{D}_{uu}^T\mathbf{D}_{uu} + 2\mathbf{D}_{uv}^T\mathbf{D}_{uv} + \mathbf{B}$; $\mathbf{I}$ is an identity matrix of size $N_\Omega \times N_\Omega$. Optimal 3D shape $\mathbf{z}$ minimizes the cost function. Therefore, the derivative of the cost function with respect to z should be zero. The solution is obtained as

$$\mathbf{z} = -\mathbf{A}^{-1}\left(\mathbf{D}_{vv}^T\mathbf{K}_a\boldsymbol{\sigma}_{max} + \mathbf{D}_{uu}^T\mathbf{H}\mathbf{K}_a\boldsymbol{\sigma}_{min}\right).$$
(17)

Here, matrix A is invertible since A is positive definite, which can be easily shown. First, the eigenvalue of A is non-negative from the definition (Equations 8, 11, 13, and 14). Second, there is no zero eigenvalue because of the boundary condition (Equations 13 and 14). By substituting the solution Equation (17) into Equation (16), the cost function becomes a function of $\boldsymbol{\sigma}_{max}$, $\boldsymbol{\sigma}_{min}$, and $\mathbf{k}_a$:

$$\begin{aligned}E =& -\frac{1}{2}\left(\mathbf{D}_{vv}^T\mathbf{K}_a\boldsymbol{\sigma}_{max} + \mathbf{D}_{uu}^T\mathbf{H}\mathbf{K}_a\boldsymbol{\sigma}_{min}\right)^T\\ & \mathbf{A}^{-1}\left(\mathbf{D}_{vv}^T\mathbf{K}_a\boldsymbol{\sigma}_{max} + \mathbf{D}_{uu}^T\mathbf{H}\mathbf{K}_a\boldsymbol{\sigma}_{min}\right) + \frac{1}{2}\mathbf{k}_a^T\left(\mathbf{I} + \mathbf{H}^2\right)\mathbf{k}_a\\ & -\left(\mathbf{h}_{max}^T\boldsymbol{\sigma}_{max} + \mathbf{h}_{min}^T\boldsymbol{\sigma}_{min}\right).\end{aligned}$$
(18)

The procedure for minimizing the cost function is described in **Appendix C**.

Evaluation of Recovered Depths

We quantify the shape recovery performance by taking the correlation between the recovered depth and the true depth. Note that here we apply the affine transformation so that the slant of the true surface depth becomes zero before taking the depth correlations. The proposed algorithm generates a shape whose slant is zero because of the boundary condition (Equation 14). Therefore, we compared the recovered shape with the true depth after the affine transformation. We summarized the depth correlations without the affine transformation in **Supplementary Note 2**.

We used two depth correlations: global and local interior. The global depth correlation is simply the correlation coefficient of the recovered and true depths throughout the object region. However, the global depth correlation tends to become high as long as the depth around the boundary is small, because the true depth is generally very small around the boundary and modest inside the object region. In other words, it is sensitive to the depth around the boundary and insensitive to the details of the shapes inside the object region. Therefore, we proposed a local interior depth correlation, which was calculated as follows. First, we drew a grid that divided the vertical and horizontal axes of the image region into eight (at 32 pixel intervals). Second, we drew a circle centered at an intersection of the grid with a radius of 32 pixels. Third, we measured a depth correlation in the intersection of the circle and the object area after removing the area near the boundary (within 24 pixels from the boundary). We did not measure a depth correlation if the intersection area was smaller than half of the circle's area. Fourth, we averaged the depth correlation values. As a result, the local interior depth correlation is not affected by the shapes near the boundary and is sensitive to the agreement of the concavity and the convexity inside the object region. Note that we did not evaluate the local interior depth correlation for objects #9 and #11. No depth correlation values were obtained with the above procedure because most of the object region is near the boundary, and the global depth correlation seems sufficient as a measure because there is no fine shape structure inside these object regions.

Psychophysical Experiment

Five unpaid volunteers participated in the experiment (three males and two females; age range, 33–58), all of whom had normal or corrected-to-normal vision and were naïve to its

purpose. The experiment was approved by the Ethics Committee for Human Research of National Institute for Physiological Sciences. The experiment was conducted in accordance with the principles of the Helsinki Declaration. Written informed consent was obtained from all participants.

Stimuli were presented on a 58.1 × 38.6 cm flat screen OLED monitor at a distance of 60 cm in a darkened room. Each image subtended at about a 10° visual angle. The stimulus images are shown in **Figure 6**, although the red crosses in it were not displayed during the experiment. The images of **Figures 6A,B** were rendered by Radiance software with the surface reflection property $\rho_d = 0.1$, $\rho_s = 0.15$, $\alpha = 0$ under illumination environments of the Devebec dataset (Galileo's Tomb for **Figure 6A** and Eucalyptus Grove for **Figure 6B**).

Subjects performed two tasks. Both were two-alternative forced choice tasks with no time limits. First, we presented either the image of **Figure 6A** (Galileo illumination condition) or **Figure 6B** (Eucalyptus illumination condition). Unfilled, 2.7-cm diameter gray circle centered at the red cross position was superimposed in the first task. Subjects were asked whether the local surface indicated by the circle was convex or concave. Next, we presented the same image and the recovered depth map by the proposed algorithm and the true depth map. The image was located in the center, and the two depth maps were located at the image's left and right. The left and right arrangements of the recovered and the true depth maps were random. Subjects were asked whether the recovered 3D shape or the true 3D shape more closely resembled the perceived 3D shape from the image. They sequentially performed two tasks for two conditions: the Galileo illumination condition and the Eucalyptus illumination condition. The order of the conditions was counter-balanced among the subjects (two subjects performed the Galileo illumination condition first and three performed the Eucalyptus illumination condition first). Before the experiment, the subjects performed a practice trial with sphere images rendered under another illumination environment (Campus at Sunset of the Devebec dataset) and were instructed about the depth map's meaning.

DATA AVAILABILITY

All datasets for this study are included in the manuscript and the supplementary files.

AUTHOR CONTRIBUTIONS

TS, AN, MS, MK, and HK designed the study. TS developed the algorithm and performed the computations. AN performed the psychophysical experiments. TS, AN, and MS contributed new analytic tools. TS and HK wrote the manuscript. All authors edited the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom. 2019.00010/full#supplementary-material

REFERENCES

Adato, Y., Vasilyev, Y., Zickler, T., and Ben-Shahar, O. (2010). Shape from specular flow. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 2054–2070. doi: 10.1109/TPAMI.2010.126

Adelson, E. H. (2001). "On seeing stuff: the perception of materials by humans and machines," in *Proceedings SPIE Human Vision and Electronic Imaging VI*, eds B. E. Rogowitz and T. N. Pappas (San Jose, CA: SPIE), 4299, 1–12.

Anderson, B. L., and Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *J. Vis.* 9, 10.1–17. doi: 10.1167/9.11.10

Andrews, B., Aisenberg, D., d'Avossa, G., and Sapir, A. (2013). Cross-cultural effects on the assumed light source direction: evidence from English and Hebrew readers. *J. Vis.* 13:2. doi: 10.1167/13.13.2

Artusi, A., Banterle, F., and Chetverikov, D. (2011). A survey of specularity removal methods. *Comput. Graph. Forum* 30, 2208–2230. doi: 10.1111/j.1467-8659.2011.01971.x

Barron, J. T., and Malik, J. (2015). Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1670–1687. doi: 10.1109/TPAMI.2014.2377712

Belhumeur, P. N., Kriegman, D. J., and Yuille, A. L. (1999). The bas-relief ambiguity. *Int. J. Comput. Vis.* 35, 33–44. doi: 10.1023/A:1008154927611

Blake, A., and Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature* 343, 165–168. doi: 10.1038/343165a0

Breton, P., and Zucker, S. W. (1996). "Shadows and shading flow fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE), 782–789.

Cohen, E. H., Barenholtz, E., Singh, M., and Feldman, J. (2005). What change detection tells us about the visual representation of shape. *J. Vis.* 5, 313–321. doi: 10.1167/5.4.3

Cole, F., Sanik, K., DeCarlo, D., Finkelstein, A., Funkhouser, T., Rusinkiewicz, S., et al. (2009). How well do line drawings depict shape? *ACM Trans. Graph.* 28:28. doi: 10.1145/1531326.1531334

Davitt, L. I., Cristino, F., Wong, A. C. N., and Leek, E. C. (2014). Shape information mediating basic- and subordinate-level object recognition revealed by analyses of eye movements. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 451–456. doi: 10.1037/a0034983

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends Neurosci.* 18, 451–458. doi: 10.1016/0166-2236(95)94496-R

Dövencioglu, D. N., Wijntjes, M. W., Ben-Shahar, O., and Doerschner, K. (2015). Effects of surface reflectance on local second order shape estimation in dynamic scenes. *Vision Res.* 115, 218–230. doi: 10.1016/j.visres.2015.01.008

Faisman, A., and Langer, M. S. (2013). Qualitative shape from shading, highlights, and mirror reflections. *J. Vis.* 13:10. doi: 10.1167/13.5.10

Feldman, J., and Singh, M. (2005). Information along contours and object boundaries. *Psychol. Rev.* 112, 243–252. doi: 10.1037/0033-295X.112.1.243

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Res.* 94, 62–75. doi: 10.1016/j.visres.2013.11.004

Fleming, R. W., Holtmann-Rice, D., and Bülthoff, H. H. (2011). Estimation of 3D shape from image orientations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20438–20443. doi: 10.1073/pnas.1114619109

Fleming, R. W., Torralba, A., and Adelson, E. H. (2004). Specular reflections and the perception of shape. *J. Vis.* 4, 798–820. doi: 10.1167/4.9.10

Fleming, R. W., Torralba, A., and Adelson, E. H. (2009). *Shape from Sheen*. MIT Tech Report, MIT-CSAIL-TR-2009–R-2051.

Gerardin, P., Kourtzi, Z., and Mamassian, P. (2010). Prior knowledge of illumination for 3D perception in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16309–16314. doi: 10.1073/pnas.1006285107

Hanazawa, A., and Komatsu, H. (2001). Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J. Neurosci.* 21, 4490–4497. doi: 10.1523/JNEUROSCI.21-12-04490.2001

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Huffman, D. A. (1971). Impossible objects as nonsense sentences. *Mach. Intell.* 6, 295–323.

Humphrey, G. K., Goodale, M. A., Bowen, C. V., Gati, J. S., Vilis, T., Rutt, B. K., et al. (1997). Differences in perceived shape from shading correlate with activity in early visual areas. *Curr. Biol.* 7, 144–147. doi: 10.1016/S0960-9822(06)00058-3

Iarussi, E., Bommes, D., and Bousseau, A. (2015). Bendfields: regularized curvature fields from rough concept sketches. *ACM Trans. Graph.* 34:24. doi: 10.1145/2710026

Jacquet, B., Hane, C., Koser, K., and Pollefeys, M. (2013). "Real-world normal map capture for nearly flat reflective surfaces," in *Proceedings of the IEEE International Conference on Computer Vision* (Washington, DC: IEEE), 713–720.

Khan, E. A., Reinhard, E., Fleming, R. W., and Bülthoff, H. H. (2006). Image-based material editing. *ACM Trans. Graph.* 25, 654–663. doi: 10.1145/1141911.1141937

Khang, B. G., Koenderink, J. J., and Kappers, A. M. L. (2007). Shape from shading from images rendered with various surface types and light fields. *Perception* 36, 1191–1213. doi: 10.1068/p5807

Koenderink, J., van Doorn, A., and Wagemans, J. (2014). Local shape of pictorial relief. *i-Perception* 5, 188–204. doi: 10.1068/i0659

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception* 13, 321–330. doi: 10.1068/p130321

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., and Todd, J. T. (2001). Ambiguity and the 'mental eye' in pictorial relief. *Perception* 30, 431–448. doi: 10.1068/p3030

Kunsberg, B., and Zucker, S. W. (2014). How shading constrains surface patches without knowledge of light sources. *SIAM J. Imaging Sci.* 7, 641–668. doi: 10.1137/13092647X

Langer, M. S., and Bülthoff, H. H. (2001). A prior for global convexity in local shape-from-shading. *Perception* 30, 403–410. doi: 10.1068/p3178

Leek, E. C., Cristino, F., Conlan, L. I., Patterson, C., Rodriguez, E., and Johnston, S. J. (2012). Eye movement patterns during the recognition of three-dimensional objects: preferential fixation of concave surface curvature minima. *J. Vis.* 12:7. doi: 10.1167/12.1.7

Leek, E. C., Reppa, I., and Arguin, M. (2005). The structure of three-dimensional object representations in human vision: evidence from whole-part matching. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 668–684. doi: 10.1037/0096-1523.31.4.668

Leek, E. C., Reppa, I., Rodriguez, E., and Arguin, M. (2009). Surface but not volumetric part structure mediates three-dimensional shape representation: evidence from part–whole priming. *Q. J. Exp. Psychol.* 62, 814–830. doi: 10.1080/17470210802303826

Li, H., Song, T., Wu, Z., Ma, J., and Ding, G. (2014). "Reconstruction of a complex mirror surface from a single image," in *International Symposium on Visual Computing* (Cham: Springer), 402–412.

Lim, I. S., and Leek, E. C. (2012). Curvature and the visual perception of shape: theory on information along object boundaries and the minima rule revisited. *Psychol. Rev.* 119, 668–677. doi: 10.1037/a0025962

Liu, B., and Todd, J. T. (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision Res.* 44, 2135–2145. doi: 10.1016/j.visres.2004.03.024

Liu, M., Hartley, R., and Salzmann, M. (2013). "Mirror surface reconstruction from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE), 129–136.

Malik, J. (1987). Interpreting line drawings of curved objects. *Int. J. Comput. Vis.* 1, 73–103. doi: 10.1007/BF00128527

Marlow, P., Kim, J., and Anderson, B. L. (2011). The role of brightness and orientation congruence in the perception of surface gloss. *J. Vis.* 11:16. doi: 10.1167/11.9.16

Marlow, P. J., Todorović, D., and Anderson, B. L. (2015). Coupled computations of three-dimensional shape and material. *Curr. Biol.* 25, R221–R222. doi: 10.1016/j.cub.2015.01.062

Mooney, S. W., and Anderson, B. L. (2014). Specular image structure modulates the perception of three-dimensional shape. *Curr. Biol.* 24, 2737–2742. doi: 10.1016/j.cub.2014.09.074

Motoyoshi, I. (2010). Highlight–shading relationship as a cue for the perception of translucent and transparent materials. *J. Vis.* 10:6. doi: 10.1167/10.9.6

Motoyoshi, I., Nishida, S., Sharan, L., and Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature* 447, 206–209. doi: 10.1038/nature05724

Ngan, A., Durand, F., and Matusik, W. (2005). "Experimental analysis of BRDF models," in *Proceedings of the Eurographics Symposium on Rendering* (Konstanz), 117–226.

Nishio, A., Goda, N., and Komatsu, H. (2012). Neural selectivity and representation of gloss in the monkey inferior temporal cortex. *J. Neurosci.* 32, 10780–10793. doi: 10.1523/JNEUROSCI.1095-12.2012

Nishio, A., Shimokawa, T., Goda, N., and Komatsu, H. (2014). Perceptual gloss parameters are encoded by population responses in the monkey inferior temporal cortex. *J. Neurosci.* 34, 11143–11151. doi: 10.1523/JNEUROSCI.1451-14.2014

Norman, J. F., Todd, J. T., and Orban, G. A. (2004). Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychol. Sci.* 15, 565–570. doi: 10.1111/j.0956-7976.2004.00720.x

Orban, G. A. (2011). The extraction of 3D shape in the visual system of human and nonhuman primates. *Annu. Rev. Neurosci.* 34, 361–388. doi: 10.1146/annurev-neuro-061010-113819

Parisi, G. (1988). *Statistical Field Theory*. Redwood City: Addison-Wesley.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature* 331, 163–166. doi: 10.1038/331163a0

Reppa, I., Greville, W. J., and Leek, E. C. (2015). The role of surface-based representations of shape in visual object recognition. *Q. J. Exp. Psychol.* 68, 2351–2369. doi: 10.1080/17470218.2015.1014379

Romeiro, F., and Zickler, T. (2010). "Blind reflectometry," in *Proceedings of the European Conference on Computer Vision* (Berlin: Springer), 45–58.

Sankaranarayanan, A. C., Veeraraghavan, A., Tuzel, O., and Agrawal, A. (2010). "Specular surface reconstruction from sparse reflection correspondences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE), 1245–1252.

Savarese, S., Chen, M., and Perona, P. (2005). Local shape from mirror reflections. *Int. J. Comput. Vis.* 64, 31–67. doi: 10.1007/s11263-005-1086-x

Savarese, S., Fei-Fei, L., and Perona, P. (2004). "What do reflections tell us about the shape of a mirror?" in *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (New York, NY: ACM), 115–118.

Sawayama, M., and Nishida, S. (2018). Material and shape perception based on two types of intensity gradient information. *PLoS Comput. Biol.* 14:e1006061. doi: 10.1371/journal.pcbi.1006061

Shimokawa, T., Nishio, A., Sato, M., Kawato, M., and Komatsu, H. (2018). Computational model for human 3D shape perception from a single specular image. *bioRxiv*, 383174. doi: 10.1101/383174

Simoncelli, E. P., and Freeman, W. T. (1995). "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in Second *International Conference on Image Processing* (Washington, DC: IEEE), 3, 444–447.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Trans. Inf. Theory* 38, 587–607. doi: 10.1109/18.119725

Srivastava, S., Orban, G. A., De Mazière, P. A., and Janssen, P. (2009). A distinct representation of three-dimensional shape in macaque anterior intraparietal area: fast, metric, and coarse. *J. Neurosci.* 29, 10613–10626. doi: 10.1523/JNEUROSCI.6016-08.2009

Sun, J., and Perona, P. (1998). Where is the sun? *Nat. Neurosci.* 1, 183–184. doi: 10.1038/630

Tappen, M. F. (2011). "Recovering shape from a single image of a mirrored surface from curvature constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE), 2545–2552.

Tarini, M., Lensch, H. P., Goesele, M., and Seidel, H. P. (2005). 3D acquisition of mirroring objects using striped patterns. *Graph. Models* 67, 233–259. doi: 10.1016/j.gmod.2004.11.002

Todd, J. T. (2004). The visual perception of 3D shape. *Trends Cogn. Sci.* 8, 115–121. doi: 10.1016/j.tics.2004.01.006

Todd, J. T., Norman, J. F., and Mingolla, E. (2004). Lightness constancy in the presence of specular highlights. *Psychol. Sci.* 15, 33–39. doi: 10.1111/j.0963-7214.2004.01501006.x

Vergne, R., Barla, P., Bonneau, G. P., and Fleming, R. W. (2016). Flow-guided warping for image-based shape manipulation. *ACM Trans. Graph.* 35:93. doi: 10.1145/2897824.2925937

Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Comput. Graph.* 26, 265–272. doi: 10.1145/142920.134078

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360. doi: 10.1038/nn.2202