

Super-Resolution from Corneal Images

Christian Nitschke¹

christian.nitschke@cmc.osaka-u.ac.jp

Atsushi Nakazawa^{1,2}

nakazawa@cmc.osaka-u.ac.jp

¹ Cybermedia Center

Osaka University

Toyonaka, Osaka, Japan

² PRESTO

Japan Science and

Technology Agency (JST)

Kawaguchi, Saitama, Japan

Abstract

The corneal imaging technique enables extraction of scene information from corneal reflections and realizes a large number of applications including environment map reconstruction and estimation of a person's area of view. However, since corneal reflection images are usually low quality and resolution, the outcome of the technique is currently limited. To overcome this issue, we propose a first non-central catadioptric approach to reconstruct high-resolution scene information from a series of lower resolution corneal images through a super-resolution technique. We describe a three-step process, including (1) single image environment map recovery, (2) multiple image registration, and (3) high-resolution image reconstruction. In a number of experiments we show that the proposed strategy successfully recovers high-frequency textures that are lost in the source images, and also works with other non-central catadioptric systems, *e.g.*, involving spherical mirrors. The obtained information about a person and the environment enables novel applications, *e.g.*, for surveillance systems, personal video, human-computer interaction, and upcoming head-mounted cameras (Google Glass [8]).

1 Introduction

In order for us humans to perceive the visual information of the surrounding environment, the cornea and the lens of the eye focus incoming light to form a sharp image on the retina. A part of the light arriving at the corneal surface reflects back into the environment, and an image of this reflection essentially captures a superset of the information arriving at the human visual system. The analysis of corneal reflection images can, therefore, obtain scene information just like a person's perceived view. This so-called *corneal imaging technique* enables the computation of a scene panorama and 3D model, together with the person's field of view and point of gaze [15]. The obtained environment map of incident illumination further enables general applications in vision and graphics, such as face relighting, reconstruction [14, 15, 28] and recognition [16]. Due to its large potential, recent science fiction works showcase several exciting applications [19, 20]. For example, by applying corneal imaging to surveillance video footage that shows a person's facial region, they obtain detailed information about the surrounding scene, revealing a human face or printed fabric patterns the person looks at. In reality, however, even if we manually capture images with a

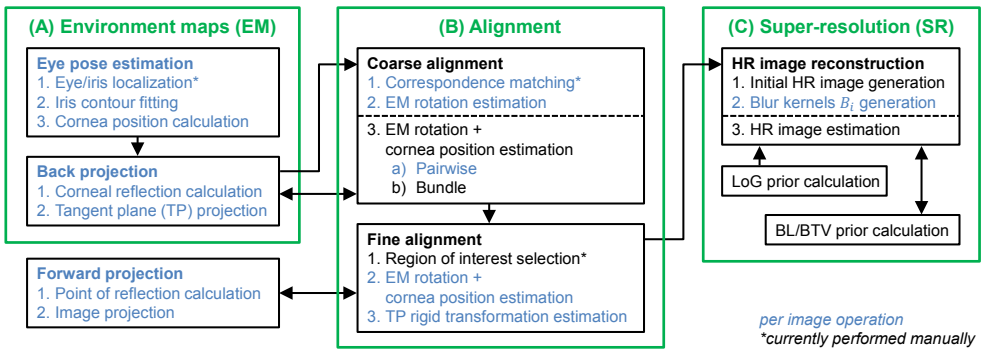


Figure 1: The algorithm consists of three parts: (A) eye pose estimation and initial environment map extraction for each LR image (Sec. 3), (B) registration of environment maps and optimization of system parameters (Sec. 4) and (C) super-resolution estimation for HR image reconstruction (Sec. 5).

high-resolution camera that is placed near to the eye and carefully adjusted to avoid focus and motion blur, the quality of corneal reflections is largely limited due to several factors. These include the low resolution of the eye region and the large field of view of the corneal mirror; the low contrast as the reflectivity of the cornea is less than 1% [14]; the contamination with iris texture reflections [15]; and the distortions from an unknown corneal geometry.

To overcome these issues, this paper introduces a super-resolution (SR) technique to reconstruct a high-resolution scene image from multiple low-resolution corneal reflection images. This enables detailed scene information from a time-series of corneal shots such as in surveillance or personal videos.

A straightforward approach to obtain high-resolution scene information would be to directly apply an SR algorithm to the recorded eye images. However, due to the following reasons, this does not work: First, the corneal image is formed by a non-central catadioptric system, as the result of a specular reflection at the corneal mirror (with only approximately known geometry) and a perspective projection. Thus, while registration is a key requirement for multiple image SR, the corneal images can not be aligned through perspective or other linear transformations. Second, for the same reason, we cannot obtain high-resolution scene information even from a single high-resolution corneal image.

To enable SR for corneal imaging, we propose the following approach: First, we calculate a spherical environment map around the corneal surface from a single image by applying a geometric eye model, 3D eye pose estimation and light path reconstruction. Then, we perform a coarse alignment to register the environment maps from multiple images and optimize eye pose parameters. Next, we set up a local plane that is tangent to the environment map sphere at a user-defined region of interest, project the maps to this plane, and perform a fine alignment. Finally, we apply an SR technique to the combined registered corneal reflections in the tangent plane. Figure 1 shows an overview of the proposed algorithm.

2 Related Work

In the following, we discuss the relation of this work to the fields of corneal reflection analysis and super-resolution.

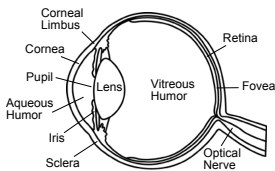


Figure 2: Geometric eye model.

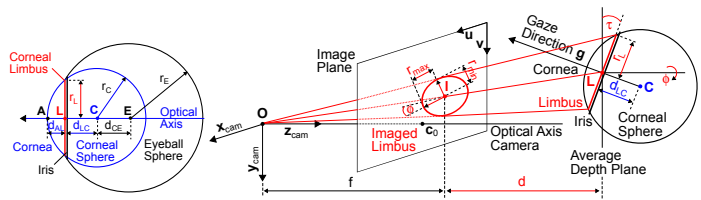


Figure 3: Eye pose estimation.

Corneal reflection analysis: Nishino and Nayar [15] first formalize the imaging characteristics of the eye-camera geometry and introduce the corneal imaging technique. They show how the extracted environment map can be applied for the computation of scene panoramas and subject views, face relighting, and face recognition [14, 16]. Recently, Backes et al. [9] analyze the point spread function of the corneal reflection system and describe a non-blind deconvolution method to recover clear scene information from a single corneal image under focus and motion blur. Johnson and Farid [8] analyze specular highlights in the eyes of different persons depicted in the same image to expose digital forgeries.

Super-resolution: Many algorithms have been proposed in super-resolution literature [1, 12, 13, 26]. The reconstruction based algorithm is a popular technique that performs sub-pixel registration of multiple low-resolution images, followed by iterative optimization to estimate a high-resolution image [18, 22]. Most works assume perspective camera optics, where multiple images can be aligned using linear transformations as the projection of the scene is described by linear systems. There are a few studies on SR under non-linear optics: Nagahara *et al.* [14] use a catadioptric system with a hyper-omnidirectional mirror. They reconstruct a high-resolution scene image from images captured while slightly rotating the system with respect to the scene. Arican *et al.* [1] use a similar optical system and introduce Spherical Fourier Transform (SFT) to reconstruct a high-resolution spherical light map. These works assume central catadioptric systems (having a single focal point), where the relative position between mirror, lens and camera remains fixed [25]. On the other hand, our configuration results in a dynamic non-central catadioptric system that requires handling of non-single focal point optics and per-frame calibration through eye/mirror pose estimation.

3 Environment map reconstruction from corneal image

3.1 Eye pose estimation

The human eyeball can be closely approximated as two overlapping spheres with different radii of curvature. Figure 2 shows a cross-section of the eye and its geometric model that is used to estimate the 3D pose of the eye and calculate corneal reflections. The cornea is modeled as a spherical cap, cut off from the corneal sphere by the limbus plane. The average radius of the limbus r_L is approximately 5.6 mm [24].

The 3D pose of the eye is estimated from the imaged iris contour. When taking an image, a projective transformation maps the 3D iris to an ellipse [1, 23], described by five parameters: the center \mathbf{l} , the major and minor radii r_{\max} and r_{\min} , and the rotation angle ϕ (Figure 3).

Iris contour detection consists of two steps: eye detection to determine if and where an eye exists in the image, and iris contour fitting to estimate the ellipse parameters. There exists a large number of approaches for automatic eye detection and tracking, and a suitable one

should be chosen based on the constraints of the particular system scenario [6]. In this proof-of-concept implementation we use a simple manual strategy, where the user selects at least five points located on the iris contour. After that, iris contour fitting is performed automatically, where initial parameters are obtained by RANSAC-based ellipse fitting through the selected points. The parameters are then refined by an iterative robust fitting approach [11] in the binarized edge image that is obtained through smoothing with a Gaussian filter and extracting edges with an adaptively-thresholded Canny edge detector.

Limbus pose estimation. Assuming weak perspective projection, the 3D limbus center \mathbf{L} is given by $\mathbf{L} = d \cdot \mathbf{l}$, where $d = f \cdot r_L / r_{\max}$ is the distance to the camera, and f is the focal length in pixels. The gaze direction \mathbf{g} , equal to the optical axis of the eye, is obtained as $\mathbf{g} = [\sin \tau \sin \phi \quad -\sin \tau \cos \phi \quad -\cos \tau]^T$, where angle $\tau = \pm \arccos(r_{\min}/r_{\max})$ corresponds to the tilt of the limbus plane with respect to the image plane, and angle ϕ is already known as it is the rotation of the limbus ellipse in the image plane.

Corneal sphere position. We are now able to calculate the center of the corneal sphere \mathbf{C} , located at a distance d_{LC} , approximately 5.6 mm from the limbus center \mathbf{L} along $-\mathbf{g}$. The center \mathbf{C} and the radius r_C , approximately 7.7 mm [24], describe the corneal sphere that enables to model the light-reflection properties of the corneal surface.

3.2 Environment map reconstruction

Taking an image of the eye captures the specular reflection of the scene within the bounds of the visible iris. Assume that light from a scene point \mathbf{P} reflects at a corneal surface point \mathbf{S} into the direction of the camera. Recovering the inverse light path from the imaged reflection allows to transform the corneal image into an environment map (Figure 4 (a),(b)).

Let $\mathbf{s} = (s_u, s_v, 1)^T$ denote the projection of \mathbf{S} , modeled as a ray $\mathbf{S} = t_1 \mathbf{r}_1$ with direction $\mathbf{r}_1 = \mathbf{K}^{-1} \mathbf{s} / \|\mathbf{K}^{-1} \mathbf{s}\|$ at unknown distance t_1 from the camera. It is obtained as the intersection with the corneal sphere by solving the quadratic equation $\|\mathbf{S} - \mathbf{C}\|^2 = r_C^2$ for t_1 as in

$$t_1 = (\mathbf{r}_1 \cdot \mathbf{C}) \pm \sqrt{(\mathbf{r}_1 \cdot \mathbf{C})^2 - \mathbf{C}^2 + r_C^2}. \quad (1)$$

The first intersection at the front side of the cornea is described by the smaller value of t_1 . Knowing \mathbf{S} and the corresponding surface normal $\mathbf{n}_S = \mathbf{S} - \mathbf{C} / \|\mathbf{S} - \mathbf{C}\|$, the normalized direction vector \mathbf{r}_2 of the reflection ray is obtained by calculating the specular reflection as in

$$\mathbf{r}_2 = 2(-\mathbf{r}_1 \cdot \mathbf{n}_S) \mathbf{n}_S + \mathbf{r}_1. \quad (2)$$

The position of scene point \mathbf{P} then lies on the reflection ray extending from \mathbf{S} , defined as $\mathbf{P} = \mathbf{S} + t_2 \mathbf{r}_2$. Since the distance t_2 between eye and scene is usually much larger than the scale of the cornea r_C , it is feasible to register the reflection ray into an environment map, as in $\mathbf{P}_{\text{EM}} = \mathbf{C} + \mathbf{r}_2^1$.

3.3 Local tangent plane projection

In order to undistort the reflection information for visualization and SR, we project a small area around a user-defined point of interest onto a local tangent plane to the environment

¹The maximal distance d_{\max} between both rays is r_C , what is easily verified as follows: Since the rays are parallel and \mathbf{P} is located outside the corneal sphere, the distance d can be calculated as the distance between line $\mathbf{C} + t \mathbf{r}_2$ and point \mathbf{S} , as in $d = r_C |\mathbf{r}_2 \times \mathbf{n}_S| = r_C \sin(\alpha)$, where $\alpha = \angle(\mathbf{r}_2, \mathbf{n}_S)$. From $\alpha \in [0, \pi/2]$ it then follows that $d \in [0, r_C]$. Furthermore, note that in the context of corneal reflections, α is commonly much smaller than the maximum value.

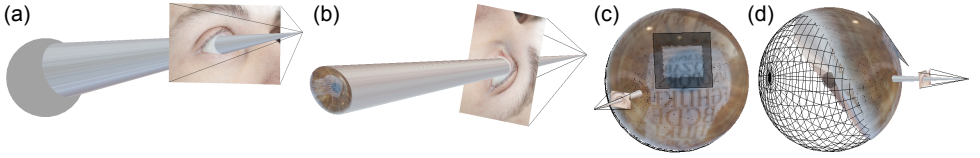


Figure 4: Rendered system model of catadioptric back projection. (a) Camera with image plane and back-projection rays that intersect the corneal sphere. (b) View onto partial inner corneal surface. (c),(d) Environment map and local tangent plane at region of interest.

sphere (Figure 4 (c),(d)). Therefore, we first set up the plane coordinate frame as in

$$\mathbf{z}_{TP} = \mathbf{n}_{TP}, \quad \mathbf{y}_{TP} = \mathbf{z}_{TP} \times \left((-1, 1, -1)^T \circ \mathbf{z}_{TP} \right), \quad \mathbf{x}_{TP} = \mathbf{y}_{TP} \times \mathbf{z}_{TP}, \quad (3)$$

where \mathbf{n}_{TP} is the normal vector of reference point and plane, and (\circ) denotes the Hadamard product (element-wise multiplication). Projecting an environment map point ${}^W\mathbf{P}_{EM}$ onto point ${}^{TP}\mathbf{P}_{TP}$ in the local plane simply involves a rotation and scaling as in

$${}^{TP}\mathbf{P}_{EM} = \begin{pmatrix} TP \\ W \end{pmatrix} R {}^W\mathbf{P}_{EM}, \quad {}^{TP}\mathbf{P}_{TP} = {}^{TP}\mathbf{P}_{EM} / z_{{}^{TP}\mathbf{P}_{EM}} \quad \left| \begin{matrix} TP \\ W \end{matrix} R = \begin{bmatrix} \mathbf{x}_{TP} & \mathbf{y}_{TP} & \mathbf{z}_{TP} \end{bmatrix}^T \right. . \quad (4)$$

3.4 Forward projection from local tangent plane to corneal image

So far, we formulated the back projection of corneal reflections from the image to the environment map and local tangent plane. In addition, we require the forward projection of environmental light into the image. This is more difficult since we need to find the point of reflection \mathbf{S} at the corneal surface without knowing the direction of the light ray from \mathbf{P} . Agrawal *et al.* [4] recently showed that the solution for quadric-shaped mirrors requires solving a 6th-order polynomial equation. The method consists of two steps: At first, a rotation transforms the problem into the plane of reflection. Then, the solution is constructed by enforcing constraints from the equality of reflection angles and the intersection of the reflected ray with \mathbf{P} . For the special case of a spherical mirror, this reduces to a 4th-order equation that can be solved in closed form. We derived this solution and use it for multiple image registration in the next section.

4 Registration of multiple environment maps

In this section we explain how to register reflections from multiple images taken under slight eye and camera movement that occurs for example in continuous video frames or burst mode photography. Under such a scenario, the area of interest is contained in a large number of corneal images, which allows to select only images that are free from focus and motion blur. Regarding the small distance between subsequent corneal sphere positions relative to the distance between eye and environment it is feasible to assume the cornea in each frame to be centered at the world origin. The task of alignment then amounts to finding the pose of the camera w.r.t. the world frame. This is achieved through a multiple-step coarse-to-fine strategy, where the initial estimate for the translation is obtained from the corneal position in eye pose estimation.

4.1 Coarse alignment

Coarse alignment is carried out using $N_F \geq 2$ feature correspondence matches for each LR image that may be found using common local feature detection techniques [24] (Figure 5 (b),(c)). Particular care has to be taken for the issues related to corneal imaging, especially geometric distortion. For this reason, it may be better to detect features in the environment map or in local tangent plane projections rather than the corneal image. In this proof-of-concept implementation we select the correspondences manually.

Step 1. Under the assumptions described above, the transformation between the environment maps is a rotation ${}^W_{C_i}R$ around the origin that aligns the frame of camera i with the world frame, fixed as the frame of camera 1, with ${}^W_{C_1}R = I_3$. The relative rotations are estimated by pairwise registration to the reference frame, through minimizing error functions $e_1(x_i)$, as in

$$\arg \min_{x_i | i=2 \dots N_I} e_1(x_i) = \sum_{f=1}^{N_F} \theta_I(f, 1, i) \quad \left| \quad \theta_I(f, i_1, i_2) = \cos^{-1} \left(\left({}^W_{C_{i_1}}R \right) \mathbf{r}_2(i_1, f) \cdot \left({}^W_{C_{i_2}}R \right) \mathbf{r}_2(i_2, f) \right) \right., \quad (5)$$

where $\theta_I(f, i_1, i_2)$ is the angle between locations of feature f in the environment maps of image i_1 and i_2 , and $x_i = \Phi_i$ is a non-normalized 3-vector representing rotation ${}^W_{C_i}R$.

Step 2. To compensate for the error in eye pose estimation we continue with a joint adjustment of environment map rotations and corneal sphere locations (relative to the cameras). Without information about the scale or absolute location of the scene, however, the problem is ill-posed². We, therefore, first perform pairwise optimization by minimizing error functions $e_1(x_i)$ according to Eq. (5), where now $x_i = (C_i, \Phi_i)$ and C_1 is fixed. The result is then further optimized through bundle-adjusting all images in parallel, where $x = (C_i, \Phi_i)_{i=1}^{N_I}$, by minimizing error function $e_2(x)$ as in

$$\arg \min_x e_2(x) = \sum_{f=1}^{N_F} \sum_{i_1=1}^{N_I} \sum_{i_2=i_1+1}^{N_I} \theta_I(f, i_1, i_2). \quad (6)$$

4.2 Fine alignment

Starting from the result of coarse alignment we perform a fine alignment through image matching in the local tangent plane at a user-defined region of interest (Figure 5 (f)).

Step 1. The region contains a set of uniform sampling points ${}^{\text{TP}}\mathbf{P}_{\text{TP}}(j) | j = 1 \dots N_P$. We again perform pairwise optimization of $x_i = (C_i, \Phi_i)$, with C_1 fixed, but now by minimizing the sum of absolute differences (SAD) through error functions $e_3(x_i)$ as in

$$\arg \min_{x_i | i=2 \dots N_I} e_3(x_i) = \sum_{j=1}^{N_P} \left| I({}^{\text{TP}}\mathbf{P}_{\text{TP}}(i, j)) - I({}^{\text{TP}}\mathbf{P}_{\text{TP}}(1, j)) \right| \quad \left| \begin{array}{l} I({}^{\text{TP}}\mathbf{P}_{\text{TP}}(i, j)) = I(\mathbf{p}_I(i, j)) \\ \mathbf{p}_I(i, j) = \text{FP}(i, ({}^{\text{TP}}R^{\text{T}}) {}^{\text{TP}}\mathbf{P}_{\text{TP}}(j)) \end{array} \right. \quad (7)$$

Here, $\mathbf{p}_I(i, j)$ is the forward projection of point ${}^{\text{TP}}\mathbf{P}_{\text{TP}}(j)$ into the image of camera i , and $I()$ is the interpolated intensity value in the image. The notation ${}^{\text{TP}}\mathbf{P}_{\text{TP}}(i, j)$ indicates that the point is assigned with the intensity value from camera i .

Step 2. Finally, we perform a 2D subpixel rigid registration in the tangent plane to correct the remaining misalignment. Similar to the last step, this is done through pairwise

²The catadioptric image formation depends on the length of the combined light path between scene, mirror and camera. If the relation between mirror and camera is not fixed, then it is not clear whether a change in the image results from a change in the relation to the scene or to the camera.

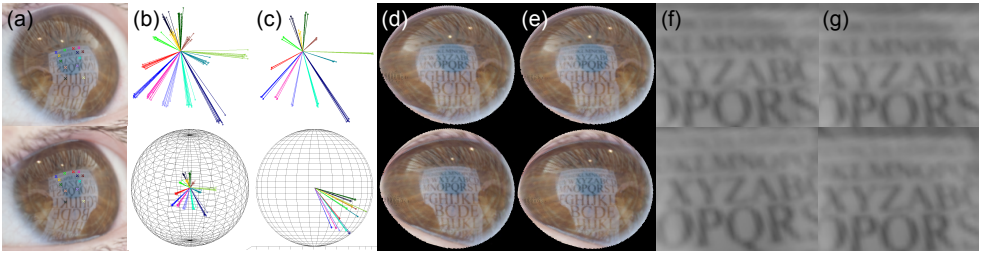


Figure 5: Alignment. (a) Cropped corneal regions of two eye images taken from similar views, with 13 feature correspondences. (b) Back-projected features for the whole dataset of 10 images, before alignment, and (c) after coarse alignment. (d) Environment map (rendered from viewpoint of corneal center), before alignment, and (e) after coarse alignment. (f) Region of interest in local tangent plane, before alignment, and (g) after fine alignment.

optimization of planar rotation and translation components $x_i = (\mathbf{t}_i, \varphi_i)$, by minimizing the SAD through error functions $e_4(x_i)$ as in

$$\arg \min_{x_i | i=2 \dots N_I} e_4(x_i) = \sum_{j=1}^{N_P} \left| I \left(\left(\begin{matrix} \text{TP}_i^1 \\ \text{TP}_i \end{matrix} \mathbf{A} \right)^{\text{TP}} \mathbf{P}_{\text{TP}}(i, j) \right) - I \left(\begin{matrix} \text{TP}_i^1 \\ \text{TP}_i \end{matrix} \mathbf{P}_{\text{TP}}(1, j) \right) \right| \quad \left| \begin{matrix} \text{TP}_i^1 \mathbf{A} = \begin{bmatrix} \text{R}(\varphi_i) & \mathbf{t}_i \\ \mathbf{0} & 1 \end{bmatrix} \end{matrix} \right|. \quad (8)$$

The result are aligned low-resolution points in the local tangent plane (Figure 5 (g)).

5 Reconstruction based super-resolution

Using the recovered optimized parameters we back project the region of interest for N_I corneal images to the local tangent plane and apply the rigid alignment. This creates sets of low-resolution (LR) points $\mathbf{I}_{\text{LR}i} | i = 1 \dots N_I$ that represent non-uniform samples (observations) of an unknown high-resolution (HR) image, defined on a uniform grid in the plane. The blur kernel matrices \mathbf{B}_i describe the downsampling of the vectorized high-resolution image \mathbf{I}_{HR} for each vector $\mathbf{I}_{\text{LR}i}$ as in $\mathbf{I}_{\text{LR}i} = \mathbf{B}_i \mathbf{I}_{\text{HR}}$.

We perform MAP (maximum a posteriori) based super-resolution (SR) using gradient descent optimization to estimate the optimal HR image that minimizes the error between the synthesized and observed LR images as in

$$\arg \min_{\mathbf{I}_{\text{HR}}} E(\mathbf{I}_{\text{HR}}) = \sum_{i=1}^{N_I} \|\mathbf{B}_i \mathbf{I}_{\text{HR}} - \mathbf{I}_{\text{LR}i}\|^2 + \alpha c(\mathbf{I}_{\text{HR}}). \quad (9)$$

The matrix

$$\mathbf{B}_{i[N_{\text{LR}i} \times N_{\text{HR}}]} \quad \left| \quad \mathbf{B}_{i,j,k} = \frac{1}{\sum_{k=1}^{N_{\text{HR}}} \mathbf{B}_{i,j,k}} e^{-\frac{\|\mathbf{p}_{\text{HR}k} - \mathbf{p}_{\text{LR}i,j}\|^2}{2\sigma^2}} \quad (10)$$

is a PSF matrix that maps the HR image to the i -th LR image, defined by a Gaussian blur kernel, where \mathbf{p}_{HR} and \mathbf{p}_{LR} are locations of HR and LR points, respectively.

Function c describes the prior knowledge of the high-resolution image, where the positive constant α represents the contribution of the prior to the overall cost. Several functions have been proposed, and we use the norm of the HR image filtered by either a Laplacian of

Gaussian filter (LoG), a bilateral filter residual (BL) [24] or a bilateral total variation filter (BTV) [9, 24]. As for LoG, function c is defined as in

$$c_{\text{LoG}}(\mathbf{I}_{\text{HR}}) = \left\| \mathbf{C}^T \mathbf{I}_{\text{HR}} \right\|^2 \quad \left| \quad C_{[N_{\text{HR}} \times N_{\text{HR}}]}, C_{j,k} = F(j,k), \right.$$

$$F_{\text{LoG}}(j,k) = \frac{\left(\left\| \mathbf{p}_{\text{HR}k} - \mathbf{p}_{\text{HR}j} \right\|^2 - 2\sigma_1^2 \right) G_1(j,k)}{2\pi\sigma_1^6 \sum_{k=1}^{N_{\text{HR}}} G_1(j,k)} \quad \left| \quad G_1(j,k) = e^{-\frac{\left\| \mathbf{p}_{\text{HR}k} - \mathbf{p}_{\text{HR}j} \right\|^2}{2\sigma_1^2}}. \quad (11)$$

where \mathbf{C}^T is a filter matrix containing the weights $F(j,k)$ for mapping input pixels k to output pixels j . Similar, the definition for BL, as in

$$c_{\text{BL}}(\mathbf{I}_{\text{HR}}) = \left\| \mathbf{C}^T \mathbf{I}_{\text{HR}} - \mathbf{I}_{\text{HR}} \right\|^2,$$

$$F_{\text{BL}}(j,k) = \frac{G_1(j,k) G_2(j,k)}{\sum_{k=1}^{N_{\text{HR}}} G_1(j,k) G_2(j,k)} \quad \left| \quad G_2(j,k) = e^{-\frac{(I(\mathbf{p}_{\text{HR}k}) - I(\mathbf{p}_{\text{HR}j}))^2}{2\sigma_2^2}}, \quad (12)$$

And for BTV, as in

$$c_{\text{BTV}}(\mathbf{I}_{\text{HR}}) = \sum_{l=-r}^r \sum_{m=0}^r \alpha^{|l|+|m|} \left\| \mathbf{S}_{xl}^T \mathbf{S}_{ym} \mathbf{I}_{\text{HR}} - \mathbf{I}_{\text{HR}} \right\|^2, \quad (13)$$

where $\mathbf{S}_{xl}^T \mathbf{S}_{ym}$ defines a matrix that shifts \mathbf{I}_{HR} by l and m in x and y direction, respectively.

6 Experiments

We took data sets of 10 partial face images from different subjects at different locations with a Nikon D60 (55 mm, f5.6) at a low resolution of 968×648 (Figure 6 (a),(b)). The average diameter for the projected limbus is approximately 150 pixels (Figure 6 (c)). We further took data sets for a spherical mirror with a radius of 7.9 mm that is similar in size to the corneal sphere, where we estimate the position of the mirror from its occluding contour [24].

Figure 6 shows the results. HR image size is set to 125×125 pixels, corresponding to a viewing angle of about 35 deg. The HR area contains approximately 1300 back-projected LR points from each corneal image. Row (e) shows natural neighbor interpolation results (from the points of all 10 registered images) that are used to initialize the previously described SR approaches, including ML (maximum likelihood) (row (g)) and MAP (rows (h)-(j)). From the SR results, we confirm that the MAP approaches (in particular BL and BTV) perform better and produce scene images with a quality high enough to recognize small characters, human faces and fine structures, that cannot be restored by natural neighbor interpolation or blind deconvolution (rows (e),(f)). On the other hand, the ML results exhibit high-frequency artifacts, most likely caused by slight intensity differences in the LR images, as the wave patterns occur in the direction of the aligned LR images. Nevertheless, the scene priors in the MAP approaches can successfully eliminate these artifacts.

We further confirm that the described approach also works for a spherical mirror. Figure 7 shows a single image and the corresponding environment map, revealing details from very compressed regions. Figure 6, right column, shows high quality scene images recovered from mirror images. The results suggest the applicability of the approach with other non-central catadioptric systems such as specular and liquid surfaces in everyday environments.



Figure 6: SR result for different scenes (from datasets of 10 eye/mirror images). (a) Scene images taken from the same position as the eyes, using a fisheye lens. (b) Single original image. (c) Cropped eye region. (d)-(g) Local plane projection of recovered environment map at region of interest: (d) Single LR image, natural neighbor interpolated. (e) Combined aligned LR images, natural neighbor interpolated. (f) Blind deconvolution of (e). (g)-j) SR result: (g) ML, before convergence. (h) MAP using LoG, (i) BL and (j) BTV prior, after convergence. (k) Cropped pictures of the region of interest from scene images in (a). (right column) Result from a spherical mirror at the size of the human cornea.

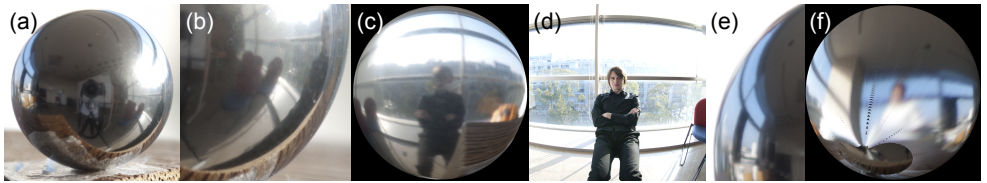


Figure 7: Environment map from a spherical mirror (with radius 7.9 mm), captured using 3872×2592 resolution (with a projected mirror diameter of approximately 850 pixels). (a) HR image of the mirror. (b) A close up on the bottom right suggests a sitting person. (c) The environment map clearly reveals the person, shown in a fisheye image (d). (e) It is very challenging to identify the content at the top left of the mirror. (f) Interestingly, the environment map reveals the silhouette of a person and details of the room.

In these experiments, we experimentally find PSF parameters around $\sigma = 2.5 \sim 3.0$. The projection to the local tangent plane may cause non-uniform (per pixel varying) blur effects, however, the current approach using a uniform PSF seems to work due to the locality of the projection. While each corneal image shows several contamination with iris textures, the SR process eliminates most of these as they follow a different reflection model than the scene reflections and are, thus, not correctly aligned and reconstructed. Nevertheless, quality may be further increased by performing iris texture separation [80].

7 Conclusion

This paper proposed a novel super-resolution technique for corneal reflection images that also works with other non-central catadioptric systems, *e.g.*, involving spherical mirrors. Through experimental results of indoor and outdoor scenes, we confirmed the effectiveness of the approach for recovering high-resolution textures in the surrounding scene. Since this solves the quality degradation problem in corneal imaging techniques, we believe our contribution can become a foundation for future applications in this research category.

Acknowledgement. This work was supported by the JST PRESTO program.

References

- [1] A. Agrawal, Y. Taguchi, and S. Ramalingam. Beyond Alhazen’s problem: Analytical projection model for non-central catadioptric cameras with quadric mirrors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2993–3000, 2011.
- [2] Z. Arican and P. Frossard. Joint registration and super-resolution with omnidirectional images. *IEEE Trans. Img. Proc.*, 20(11):3151–3162, 2011.
- [3] M. Backes, T. Chen, M. Dürmuth, H. P. A. Lensch, and M. Welk. Tempest in a teapot: Compromising reflections revisited. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 315–327, 2009.

- [4] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Img. Proc.*, 13(10):1327–1344, 2004.
- [5] Google. Project Glass on Google+. <https://plus.google.com/111626127367496192147>. Last accessed on July 26, 2012.
- [6] D. W. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):478–500, 2010.
- [7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, 2nd edition, 2003. ISBN 0521540518.
- [8] M. K. Johnson and H. Farid. Exposing digital forgeries through specular highlights on the eye. In *Proc. International Conference on Information Hiding (IH)*, pages 311–325, 2007.
- [9] A. K. Katsaggelos, R. Molina, and J. Mateos. Super resolution of images and video. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 3(1):1–134, 2007.
- [10] P. L. Kaufman and A. Alm. *Adler’s Physiology of the Eye: Clinical Application*. Mosby, Inc., St. Louis, MO, 10th edition, 2003.
- [11] P. Li, X. Liu, L. Xiao, and Q. Song. Robust and accurate iris segmentation in very noisy iris images. *Image Vision Comput.*, 28(2):246–253, 2010.
- [12] P. Milanfar. *Super-Resolution Imaging*. CRC Press, Boca Raton, 2010.
- [13] H. Nagahara, Y. Yagi, and M. Yachida. Resolution improving method from multi-focal omnidirectional images. In *Proc. IEEE International Conference on Image Processing (ICIP)*, pages 654–657, 2001.
- [14] K. Nishino and S. K. Nayar. Eyes for relighting. In *Proc. ACM SIGGRAPH*, pages 704–711, 2004.
- [15] K. Nishino and S. K. Nayar. Corneal imaging system: Environment from eyes. *Int. J. Comput. Vision*, 70(1):23–40, 2006.
- [16] K. Nishino, P. N. Belhumeur, and S. K. Nayar. Using eye reflections for face recognition under varying illumination. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 519–526, 2005.
- [17] J. D. van Ouwerkerk. Image super-resolution survey. *Image Vision Comput.*, 24(10):1039–052, 2006.
- [18] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Signal Proc. Mag.*, 20(3):21–36, 2003.
- [19] Palm Pictures. *Ghost In the Shell (DVD)*, 1998.
- [20] Paramount Pictures. *CSI NY Season 1, Ep. 10: "Night, mother" (DVD)*, 2005.
- [21] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. Nonlinear Phenom.*, 60(1–4):259–268, 1992.

- [22] R. R. Schultz and R. L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Process.*, 5(6):996–1011, 1996.
- [23] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1952.
- [24] R. S. Snell and M. A. Lemp. *Clinical Anatomy of the Eye*. Blackwell Publishing, Malden, 2nd edition, 1997.
- [25] P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, and J. Barreto. Camera models and fundamental concepts used in geometric computer vision. *Found. Trends. Comput. Graph. Vis.*, 6:1–183, 2011.
- [26] J. Tian and K.-K. Ma. A survey on super-resolution imaging. *Signal Image Video Process.*, 5(3):329–342, 2011.
- [27] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 839–846, 1998.
- [28] N. Tsumura, M. N. Dang, T. Makino, and Y. Miyake. Estimating the directions to light sources using images of eye for reconstructing 3D human face. In *Proc. IS&T/SID Color Imaging Conference (CIC)*, pages 77–81, 2003.
- [29] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(3):177–280, 2008.
- [30] H. Wang, S. Lin, X. Liu, and S. B. Kang. Separating reflections in human iris images for illumination estimation. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 1691–1698, 2005.
- [31] K.-Y. Wong, D. Schnieders, and S. Li. Recovering light directions and camera poses from a single sphere. In *Proc. European Conference on Computer Vision (ECCV)*, pages 631–642, 2008.