



# Purkinje Images: Conveying Different Content for Different Luminance Adaptations in a Single Image

Sami Arpa<sup>1,2,3</sup>, Tobias Ritschel<sup>2,4</sup>, Karol Myszkowski<sup>2</sup>, Tolga Çapın<sup>5</sup> and Hans-Peter Seidel<sup>2</sup>

<sup>1</sup>Bilkent University, Ankara, Turkey  
sami.arpa@epfl.ch

<sup>2</sup>MPI Informatik, Saarbrücken, Germany  
karol@mpii.mpg.de, hpseidel@mpi-sb.mpg.de

<sup>3</sup>School of Computer and Communication Sciences, EPFL, Lausanne, Vaud, Switzerland

<sup>4</sup>Saarland University, Saarbrücken, Germany  
ritschel@mpi-inf.mpg.de

<sup>5</sup>TED University, Ankara, Turkey  
tolga.capin@tedu.edu.tr

---

## Abstract

*Providing multiple meanings in a single piece of art has always been intriguing to both artists and observers. We present Purkinje images, which have different interpretations depending on the luminance adaptation of the observer. Finding such images is an optimization that minimizes the sum of the distance to one reference image in photopic conditions and the distance to another reference image in scotopic conditions. To model the shift of image perception between day and night vision, we decompose the input images into a Laplacian pyramid. Distances under different observation conditions in this representation are independent between pyramid levels and pixel positions and become matrix multiplications. The optimal pixel colour can be found by inverting a small, per-pixel linear system in real time on a GPU. Finally, two user studies analyze our results in terms of the recognition performance and fidelity with respect to the reference images.*

**Keywords:** photopic vision, scotopic vision, Purkinje illusion, perceptually based rendering

**ACM CCS:** I.3.3 [Computer Graphics]: Picture/Image Generation—Viewing algorithms

---

## 1. Introduction

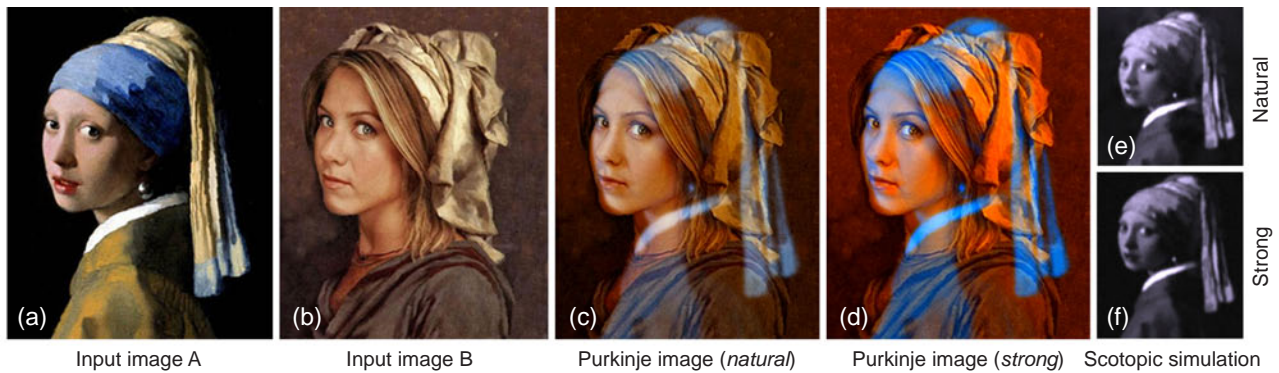
Visual experience is not a mechanical recording but rather involves interpretation of scenes in a meaningful way [Arn04]. For one image several percepts are possible, since the human visual system (HVS) combines various cues which gain significance only under certain viewing conditions [Wan95]. Artists such as S. Dali and M. C. Escher have used these cues to attribute multiple visual meanings to their works. In computer graphics, approaches such as autostereograms [TC90], hybrid images [OTS06] or camouflage images [CHM\*10] combine multiple percepts into a single image.

In this work, we propose a novel type of images—Purkinje images—that provide different percepts, depending on the observer's luminance adaptation. We optimize the output Purkinje image, so that at a given luminance level the corresponding percept is apparent, while the cross-talk with other percept, which should be seen at different viewing conditions, is minimized. The separation between the percepts with respect to day and night viewing

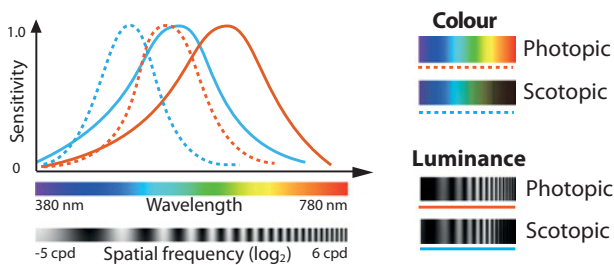
conditions is achieved by accounting for colour and spatial vision properties in the HVS. While our primary objective is recreation and artistic explorations, our approach can serve in more practical applications such as designing novel test images for detecting colour vision deficiencies based on natural images, e.g. for children. Since many animals are dichromats our framework can help in fabricating camouflage clothing with reduced visibility for animals and improved visibility for colour-blind humans.

## 2. Background

Human vision operates in a wide range of luminance values ( $10^{-6}$  to  $10^8$  cd/m<sup>2</sup>) where rod photoreceptors are active in dark conditions up to  $\sim 3$  cd/m<sup>2</sup> while cone photoreceptors become sensitive for luminance over 0.1 cd/m<sup>2</sup>. Purely rod- and cone-mediated vision is called *scotopic* and *photopic*; the range of mixed cone and rod activity is called *mesopic*.



**Figure 1:** Our work combines two images A (a) and B (b) to produce a single image (c or d) that is perceived as image B in daylight (photopic vision) and as image A at night (scotopic vision). We can adjust between a natural look (c) with a medium separation and a strong separation with a stylized look (d). (Scotopic only visible in print when shown in dark lighting conditions below  $\sim 0.01$   $\text{cd}/\text{m}^2$ ; Images (e and f) show simulations based on [TSF02].) Photos: (a) Mauritshuis, [www.mauritshuis.nl](http://www.mauritshuis.nl); (b) [www.worth1000.com](http://www.worth1000.com).



**Figure 2:** Sensitivity of the HVS to spatial luminance frequency is different for photopic (solid orange line) and scotopic (solid blue line) vision as specified by their respective contrast sensitivity functions (CSFs), which have been derived for  $0.01$  and  $100$   $\text{cd}/\text{m}^2$  adaptation luminance [Bar89]. The luminous efficiency functions  $V(\lambda)$  for photopic (dashed orange line) and  $V'(\lambda)$  for scotopic (dashed blue line) vision also demonstrate the shift of peak sensitivity as a function of spectral wavelength  $\lambda$  [RKAJ08]. Note that all functions shown in this graph are normalized to emphasize on the relative positions of their respective sensitivity peaks, which we exploit in this work.

In a vast majority of graphics techniques dealing with colour and contrast information, photopic vision is tacitly assumed. For example, the commonly used luminance  $Y$  is derived using the photopic luminous efficiency function  $V(\lambda)$  [Wal45] (refer to Figure 2), which is a weighted sum of the sensitivities of L, M and S-type of cones for all visible light wavelengths  $\lambda$ . Three-dimensional colour spaces such CIE  $XYZ$ , where  $Y$  denotes the photopic luminance, or more display device-oriented ones, such as RGB are used to specify light intensity and colour as registered by cones (refer to Reinhard *et al.* [RKAJ08, ch. 8] for more details on those and other commonly used colour spaces as well as the corresponding conversion matrices). In this work, we distinguish between photopic, scotopic and mesopic vision as it is typically assumed in the high dynamic range imaging (HDR) [RWD\*10] and colour appearance [RKAJ08] literature. As modelling of scotopic and mesopic vision is

less common, we summarize recent developments that are relevant for this work.

**Scotopic Vision:** is characterized by loss of colour vision, reduced visual acuity and temporal aspects of dark adaption, which are less relevant for this work [DD00].

Since only one receptor type is active in scotopic vision, a 1D function fully characterizes the rod response, and the scotopic luminance  $Y_{\text{scot}}$  (the counterpart of the photopic luminance  $Y$ ) is used for this purpose.  $Y_{\text{scot}}$  is derived from the scotopic luminous efficiency function  $V'(\lambda)$  (Figure 2), which shows a shift of sensitivity (the so-called Purkinje shift) from longer  $\lambda$  (shades of red) towards shorter  $\lambda$  (shades of blue).

The sensitivity to luminance patterns of varying spatial frequencies can be characterized by the contrast sensitivity function (CSF) as shown in Figure 2. When the CSFs for rod- and cone-mediated vision are compared, over 10-fold reduction in the sensitivity in dark conditions can be observed [Wan95, figure 7.21]. Also, the peak of sensitivity shifts from 8 cpd in daylight vision to 1 cpd in night conditions. In tone mapping, the visual acuity is typically modelled by low-pass filtering of the original image [FPSG96, WRP97, PFFG98, DD00, TSF02], where the cut-off frequency is a function of adaptation luminance as measured by Shlaer [Shl37].

**Mesopic Vision:** combines characteristics of scotopic and photopic vision as a function of adaptation luminance. A recent International Commission on Illumination (CIE) recommendation for standardization [CIE10], which is based on exhaustive visual performance experiments, assumes mesopic luminance as a linear combination of scotopic and photopic luminance. A linear combination of, possibly differently derived, rod and cone responses is typically used in image quality metrics [MKRH11], colour appearance models [RKAJ08] and tone mapping operators [FPSG96, DD00, WRP97]. Remarkable realism of tone mapping for spectral images has been shown by Kirk and O'Brien [KO11], who employed a biologically inspired model that predicts the offset in the L, M and S-cone channels due to rod response [CPSZ08]. An advanced colour appearance and tone mapping approach in a single framework

has been proposed by Pattanaik *et al.* [PFFG98], where different contrast transducers are considered for cone- and rod-mediated signals, prior to their combination into an achromatic signal.

### 3. Related Work

In this section, we discuss various approaches of combining multiple percepts in one output image.

**One Percept from Multiple Images:** In image fusion and multi-exposure high dynamic range techniques [RWD\*10] differently exposed photographs of the same, ideally static scene, are combined to avoid under- and overexposure. Image photomontage typically favours locally coherent image content coming from a single input image, and focuses mostly on suppressing the visibility of boundaries between different input images [PGB03]. Digital image compositing mostly relies on linear interpolation [PD84], which leads to contrast and sharpness reduction in the composite image due to compression of the colour distribution around its mean. As a remedy, Grundland *et al.* [GVWD06] employ linear stretching of each colour channel around its mean, which leads to a better preservation of average colour and contrast of the component images. Our goals are different as we want to obtain the percept of two fully distinct images, which are possibly reproduced at their full resolutions. Similar to Grundland *et al.* [GVWD06] we optionally employ the image saliency [IKN98] to favour the most informative regions from the component images when the scotopic and photopic views cannot be cleanly separated.

**Multiple Percepts in One Image:** As we discussed in Section 2, the HVS exhibits variations in visual contrast sensitivity depending on spatial frequency. Setlur and Gooch [SG04] use this fact to create facial images with different emotional states regarding the CSF separation between peripheral vision and central vision [Liv02, ch. 5]. Differently, Oliva *et al.* [OTS06] in their ‘hybrid images’ take advantage of visual sensitivity as a function of viewing distance. They decompose luminance of two images into frequency bands, and depending on viewing distance they choose different frequency bands for each image and combine them. In this manner, the resulting hybrid image has two different interpretations when it is viewed at different distances. Their approach is achromatic and does not account for shifts in colour and frequency perception with luminance adaptation. A similar procedure has been applied by Didyk *et al.* [DRE\*11] in the context of disparity sensitivity function (DSF) estimation, which led to two different depth percepts as a function of viewing distance with respect to a stereo 3D display. Image mosaics [Hau01, KP02] and collages [HZZ11, GTZM10] are stylizations, in which icon-sized tile images are tightly packed into a larger container image to appear similar to the large image when observed from distance. Complete photographs or their arbitrary-shaped cutouts (e.g. consistent with meaningful objects for the Arcimboldo-like effect) can be used as the tile images, whose details are readily visible from short distances. Relief images [AM10] employ a height-field surface, which through diffuse shading depicts two unique images when illuminated from two directions. We also embed two different percepts in one image, but our separation method is based on luminance adaptation.

**Hiding Percepts in the Dominant Image:** In camouflage images, features of the dominant image, especially its edges [TZHM11]

and texture [CHM\*10], are used to depict selected elements of the hidden image, whose recognition might require more effort due to sketchy presentation. Mitra *et al.* [MCL\*09] generate emergence images of 3D objects, which locally appear as noise and become meaningful to the human observer when viewed as a whole, while remain difficult to interpret by machines. In steganography, a hidden message in the dominant image can be revealed using a decoder tool such as a Cardan Grille or a Magic Lens [PHN\*12], where refractive lenslet arrays are placed over seemingly unstructured images to reveal a hidden image. Copyright protection by watermarking is another example of hiding an (ideally imperceptible) image. Our goal is different as we fully reveal one of the two images in one specific lighting condition.

**Concurrent Multiple Image Display:** In stereo 3D systems, specialized hardware (anaglyph or polarization glasses, time multiplexing) enables separation between the left and right eye views which ideally do not suffer from any crosstalk [vBPJ\*11]. Techniques of crosstalk reduction used in such systems, have similar goals to ours, but they affect only local image regions and typically favour one image content over another [vBPJ\*11]. Recently, Kim *et al.* [KCZT12] demonstrated dual viewing for regular LCD displays. Their technique has a number of limitations as it works for strictly selected and narrow view directions, smooth shading can be achieved only through dithering and a pair of different images must be shown through spatial or temporal multiplexing. In our approach, each pixel is optimized to be meaningful in two different viewing contexts, and smooth shading is easy to achieve.

**Separation by Adaptation Luminance:** Mantiuk *et al.* [MRH09] show that by displaying long-wavelength light (red and amber) cones might be adapted to much higher luminance levels that ensures display legibility, while rods remain adapted to scotopic conditions. Such rod–cone separation has important implications on the design of displays in a vehicle cockpit, which should not affect night vision nor cause dazzling glare. Here, the rod–cone separation is considered for different gaze directions, while we want to achieve such separation at the same spatial location by explicitly adapting to different lighting levels.

**Separation by Spectral Sensitivity:** One of the earliest uses of multiperceptual images is the Ishihara standard test for colour blindness [Ish17]: Here, colour-blind observers perceive an arrangement of isoluminant circles while subjects with normal colour vision experience coloured letters. Although there is a large body of work on image ‘daltonization’, where the goal is to improve the appearance of a single image for a specific colour vision deficiency [KOF08], in this work we show how to optimize for simultaneous viewing by people with normal and defective vision (Section 4.3).

Many mammal species such as dogs, cats or cows are dichromats, i.e. have only two cone types [Jac93], which has similar consequences as colour blindness of humans. Moreover, the spectral sensitivity is often shifted towards short wavelengths, which improves the dusk, dawn and night vision, but at the same the sensitivity to red and orange is significantly reduced as it is the case for deer [JDN\*94, figure 6] (refer also to Section 4.3). This is exploited in designing camouflage clothing for deer hunters [Bur07], and our approach can be directly used for reducing visibility of clothing for any case, where the spectral sensitivity of photoreceptors is known.

## 4. Approach

**Overview:** In this section, we introduce our framework for creating a Purkinje image. Input is two images  $I_p$  and  $I_s$  while the output is a conventional image  $I$ , that is similar to  $I_p$  in photopic conditions and similar to  $I_s$  in scotopic conditions  $I = \text{argmin} \hat{I} \Delta_p(\hat{I}, I_p) + \Delta_s(\hat{I}, I_s)$ , where  $\Delta_p$  is the perceived distance of two images in photopic and  $\Delta_s$  in scotopic conditions. Note, how this formulation includes enforcing the result  $I$  to be perceived as the photopic image  $I_p$  in photopic conditions, without being distorted by the scotopic one as this would increase the response of  $\Delta_p$ . In the same way, result  $I$  tends to resemble the scotopic input  $I_s$  under scotopic conditions, without being affected by the photopic one detected by an increase in  $\Delta_s$ .

Most methods combining multiple images assume them to be aligned [Wol98]. This step was performed manually for our results using the approach of Schaefer *et al.* [SMW06].

Optionally the user provides two saliency maps [IKN98]. Our work addresses two challenges: first, choosing distance functions  $\Delta_p$  and  $\Delta_s$  (Section 4.1), and secondly, efficiently optimizing for  $I$ , given  $I_p$  and  $I_s$  (Section 4.2).

**Viewing Conditions:** A Purkinje image's scotopic content is only perceivable on printed paper sheets of sufficient size (A4) or colourful ink screens in sufficiently dark conditions, where luminance is less than  $\sim 0.01$  cd/m<sup>2</sup>. The black level of current computer displays is still too bright and cone receptors are still effective even in a fully dark environment. A practical test for dark adaptation, which can take 5–30 min, is whether colour perception is still present: Its absence indicates sufficient rod vision. In the following, we will simulate night vision similarly to [TSF02], but encourage the reader to dark-adapt and verify the results using sufficiently large (e.g., A4) paper prints.

### 4.1. Distances

Perception of images in photopic and scotopic conditions differs in two main regards: colour vision and spatial frequency sensitivity (see Section 2, in particular, Figure 2). We need a representation of the input images  $I_p$  and  $I_s$ , and the solution image  $I$  that enables the manipulation on signals that independently represent colour components as well as image patterns of different spatial frequencies. This is achieved by a Laplacian decomposition of  $I_p$  and  $I_s$  for each colour channel. Effectively, at one pyramid level and one spatial location, this decomposition denotes how strong the spatial frequency of each colour component is present. Another important issue for the optimization is the measure of distances  $\Delta_p$  and  $\Delta_s$  that must be calibrated relative to each other. A fixed distance value must be perceived as equal in both viewing conditions. A careful selection of colour spaces for scotopic and photopic conditions, as well as proper scaling of signals at pyramid levels allows to account for the HVS sensitivity in both conditions and consequently, the Euclidean distance can be used as  $\Delta_p$  and  $\Delta_s$ .

**Colour Spaces:** The differences in colour perception between scotopic and photopic vision (Section 2) require adequate colour spaces that provide meaningful distance measures leading to a balanced, simultaneous minimization of both  $\Delta_p$  and  $\Delta_s$ . This requires

perceptual uniformity within each candidate colour space, so that similar magnitude of colour change leads to similarly perceived differences irrespectively of the initial pixel intensity. Furthermore, perceptual uniformity should be maintained between scotopic and photopic conditions.

Since our goal is not a high-fidelity reproduction of the scene appearance as in tone mapping [RWD\*10], we refer to approximations used in image compression and encoding literature. Similar to compression we assume that in  $I_p$  and  $I_s$  we deal with gamma corrected R'G'B' colour channels (the primed quantities denote non-linear signals), which can be transformed into the luma  $Y'$  and chroma  $C_B$  and  $C_R$  channels (we assume the ITU-R BT.709/sRGB primaries [RKAJ08, ch. 8]):

$$\begin{pmatrix} Y' \\ C_B \\ C_R \end{pmatrix} = \begin{pmatrix} 0.22 & 0.71 & 0.07 \\ -0.11\alpha & -0.38\alpha & 0.50\alpha \\ 0.50\alpha & -0.45\alpha & -0.46\alpha \end{pmatrix} \begin{pmatrix} R' \\ G' \\ B' \end{pmatrix}. \quad (1)$$

The parameter  $\alpha$  controls chroma preservation. Chroma is preserved when  $\alpha$  is around 1, and not preserved when  $\alpha$  is close to 0. A good choice is  $\alpha = 0.3$  that is used in all our results unless noted otherwise (e.g.  $\alpha = 0.3$  in Figure 1c and  $\alpha = 0$  in Figure 1d).

Note that luma  $Y'$  is an approximation of lightness used in colour appearance, and that the same amount of luma distortion, e.g. through lossy-compression, has a similar perceptual effect, irrespectively of the absolute luma value. This way we can benefit, in our optimization, from the error measure that is perceptually uniform, and we avoid non-linearities in the visual distance measure.

One can observe that the weights used to derive  $Y'$  based on R'G'B' (the first row of the  $3 \times 3$  matrix in Equation 1) are identical to those used to derive photopic luminance  $Y$  from linear RGB components. Following this analogy, we introduce the scotopic luma  $Y'_{\text{scot}}$  based on weights derived in [PFFG98] for the scotopic luminance  $Y_{\text{scot}}$  (here the weights are converted to the RGB colour space, refer also to Section 2):

$$Y'_{\text{scot}} = (0.16 \quad 0.62 \quad 0.52) \cdot (R' \quad G' \quad B')^T. \quad (2)$$

Since the derivation of  $Y' C_B C_R$  and  $Y'_{\text{scot}}$  from R'G'B' is based on the linear transformations, the formulation of the optimization problem, which we discuss in Section 4.2, remains equally simple as it would be the case for a linear colour space. Also, chroma  $C_B$  and  $C_R$  contain mostly chromatic information, which is important to model different CSFs for chromatic and achromatic channels.

**Discussion of Colour Space Selection:** While the CIELAB or CIELUV colour space would be a better choice in terms of perceptual uniformity, we aim a closed-form solution of our per-pixel optimization problem (Section 4.2). It conveniently enables interactive design of various image alignments with the real-time result preview for the photopic condition (please refer to the Supporting Information Video S1). In the case of CIELAB or CIELUV, the problem would be convex and the solution needs to be found using a more costly non-linear solver such as gradient descent.

**Achromatic and Chromatic CSFs:** Two different aspects of sensitivity to spatial image content in scotopic and photopic conditions can be beneficial in the visual separation between the  $I_p$  and  $I_s$  content: the shift of sensitivity peaks in the CSFs for scotopic and photopic luminance, and different cut-off frequencies, when signals cannot be perceived (as discussed in Section 2). For the achromatic CSF, we use the model of Barten [Bar89], which is parametrized by spatial frequency  $\rho$  and adaptation luminance  $Y$ . We always normalize the maximum gain in CSF to 1.0 both for scotopic and photopic vision (Figure 2), as our goal is that both  $I_p$  and  $I_s$  are equally well visible when presented in their designated conditions. Note that our goal is different from reproducing the relative loss of sensitivity in scotopic with respect to photopic conditions for the same pattern as measured by Wandell [Wan95, figure 7.21]. We assume also the cut-off frequency of 8 cpd for both chroma channels  $C_B$  and  $C_R$ .

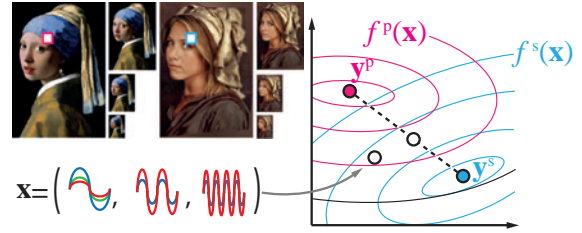
**Discussion of CSF Filtering:** While CSFs have originally been measured for luminance and opponent colours, which strictly speaking means that the relevant filtering should be performed in a linear colour space [PFFG98], many HVS models, which are successfully used in image processing (refer to a similar discussion by Bolin and Meyer [BM95]), JPEG/MPEG compression, and image quality metrics [Dal92] perform CSF filtering over non-linear functions of luminance akin to lightness, or directly on luma and chroma as in our work. For supra-threshold contrast signals, the sensitivity differences for various spatial frequencies are reduced [GS75], but as we show in the first user study (the relevant stimuli are shown in Supporting Information) ignoring the CSFs leads to inferior visual separation between  $I_p$  and  $I_s$ . This indicates that low contrast signals play an important role in our optimization.

## 4.2. Optimization

**Cost Function:** Input to the optimization are the Laplacian pyramids of the reference images. The optimization is performed for all spatial positions and all levels independently which are summed to produce a single result image. In the following, we consider optimizing a particular pixel at a particular level. For an  $R'G'B'$  colour  $\mathbf{x} \in \mathbb{R}^3$  on level  $i$ , the cost

$$f_i : \mathbb{R}^3 \rightarrow \mathbb{R} \quad f_i(\mathbf{x}) = \frac{\|s_i^p C_i^p(\mathbf{x} - \mathbf{y}_i^p)\|^2 + \|s_i^s C_i^s(\mathbf{x} - \mathbf{y}_i^s)\|^2}{\|s_i^p C_i^p(\mathbf{x} - \mathbf{y}_i^p)\|^2 + \|s_i^s C_i^s(\mathbf{x} - \mathbf{y}_i^s)\|^2} \quad (3)$$

is the sum of squares of the perceived distances between the choice  $\mathbf{x}$  and the photopic reference  $\mathbf{y}_i^p$  and the scotopic reference  $\mathbf{y}_i^s$  ( $\mathbf{y}_i^p, \mathbf{y}_i^s \in \mathbb{R}^3$ ), calculated using Equations (1) and (2), weighted by the optional saliency scalars  $s_i^p$  and  $s_i^s \in \mathbb{R}$ . The matrices  $C_i^p \in \mathbb{R}^{3 \times 3}$  and  $C_i^s$  are used to transform input  $R'G'B'$  at level  $i$  into  $Y' C_B C_R$  and  $Y'_{scot}$  channels, and at the same time rescale the per-channel signal by the CSFs of level  $i$  as follows:  $C_i^p$  is obtained by rescaling each row in the transformation matrix from Equation (1) by respective contrast sensitivity for the achromatic and chromatic channels with the photopic adaptation luminance  $Y$ . Similarly,  $C_i^s$  is built from the rescaled coefficients in Equation (2), where the achromatic CSF with the scotopic adaptation luminance  $Y_{scot}$  is considered. The contrast sensitivity values used for such rescaling are derived for the central spatial frequency  $\rho_i$  at level  $i$ , which is expressed in cycles per degree and depends on the angular resolution of the input image



**Figure 3:** Input Laplacian pyramids (left) that map a spatial location (red and blue inset) to an RGB amplitude of all perceivable frequencies, i.e. a point  $\mathbf{x}$  in a 3n-dimensional colour-frequency space. The photopic and scotopic cost functions  $f^p$  and  $f^s$  (right) describe the cost of using a value  $\mathbf{x}$  instead of the reference values  $\mathbf{y}^p$  and  $\mathbf{y}^s$ .

$n_{ppd}$  given in pixels per degree as  $\rho_i = n_{ppd}/2^i$ . Note that  $C_i^s$  has a rank of only 1 due to the insensitivity for chroma, while  $C_i^p$  has full rank 3.

For regularization, we add a small constant  $3 \times 3$  matrix having value 0.01 for its each element to both  $C_i^p$  and  $C_i^s$  to prefer the average of the references as the solution if multiple solutions are equally good.

**Minimization:** The 3-variate cost function  $f_i$  is the sum of two non-uniformly scaled and rotated quadratic functions (Figure 3). Dropping the dependency on the level  $i$  and assuming the saliency was multiplied into the cost matrices  $C^p$  and  $C^s$ , the cost of colour  $\mathbf{x}$  equals

$$\|C^p(\mathbf{x} - \mathbf{y}^p)\|_2^2 + \|C^s(\mathbf{x} - \mathbf{y}^s)\|_2^2 = \quad (4)$$

$$\left(C^p(\mathbf{x} - \mathbf{y}^p)\right)^T C^p(\mathbf{x} - \mathbf{y}^p) + \left(C^s(\mathbf{x} - \mathbf{y}^s)\right)^T C^s(\mathbf{x} - \mathbf{y}^s). \quad (5)$$

Seeking to minimize this expression, we take its derivative and set it to zero. As

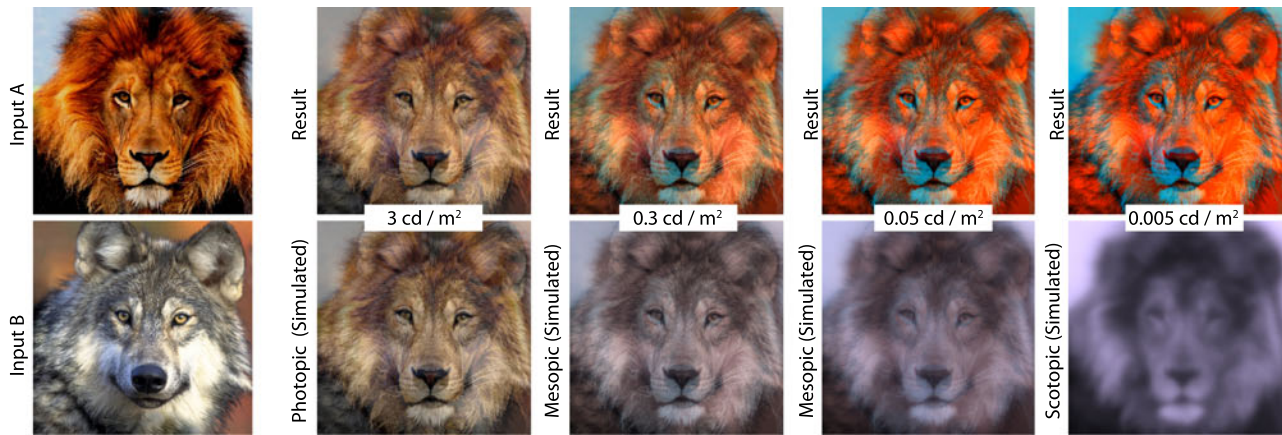
$$d\left((M\mathbf{x})^T M\mathbf{x}\right)/d\mathbf{x} = M^T M\mathbf{x} \quad (6)$$

holds for all matrices  $M$ , we can write the derivative as

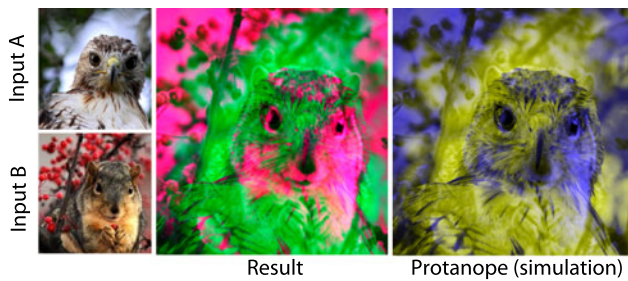
$$\underbrace{\left(C^{pT} C^p + C^{sT} C^s\right)}_A \mathbf{x} = \underbrace{C^{pT} C^p \mathbf{y}^p + C^{sT} C^s \mathbf{y}^s}_b. \quad (7)$$

Consequently, the optimal colour is found by inverting the  $3 \times 3$  matrix  $A$  of normal equations in a closed form and multiplied with  $b$  for every pixel on each level.

The optimal value  $\mathbf{x}$ , however, might not be reproducible by every output device, e.g. a printer. Instead, we wish to restrict the solution to a reproducible subset  $\mathcal{R} \subseteq \mathbb{R}^3$ . To this end, we perform a few iterations of gradient descent, where the gradient is a known linear mapping  $A$ . In every step of size  $\lambda$ , we reproject the new solution onto the subset of reproducible solutions, as in



**Figure 4:** Purkinje images of two inputs (left) for mesopic conditions of different adaptation luminance (right).



**Figure 5:** An image perceived more like a squirrel (left, bottom) for subjects with normal vision (middle) and more like a bird (left, top) to protanope daltonian vision (right).

$\mathbf{x}^{(i+1)} = r(\mathbf{x}^{(i)} + \lambda \mathbf{A}\mathbf{x}^{(i)})$ , where  $r(\mathbf{x}) \in \mathbb{R}^3 \rightarrow \mathcal{R}$  maps a colour to its closest reproducible colour, and typically  $\lambda = 0.1$ .

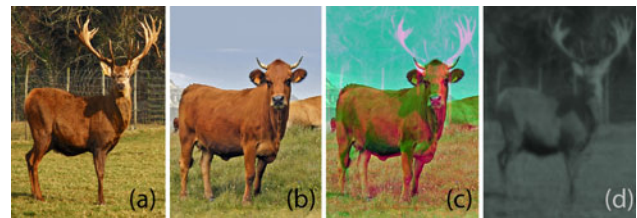
An (unoptimized) implementation of this solver computes the result of a one-megapixel image (10 Laplacian levels) parallel over all pixels and levels in 20 ms on a Nvidia Geforce Quadro 6000 GPU.

### 4.3. Variants

We now extend our approach to mesopic conditions and non-standard observers.

**Mesopic:** To simulate mesopic vision, we replace the matrix  $C^s$  in Equation (3) with a linear blend  $C^m(a) = (1 - a)C^p + aC^s$  controlled by a factor  $a \in (0, 1)$  that describes the luminance adaptation (Figure 4). We compute  $a$  using a sigmoid that is 0 at 3 and 1 at 0.005  $\text{cd}/\text{m}^2$  log-average luminance.

**Daltonians:** Daltonian human vision has only a single colour dimension, i.e. a rank deficiency [MVB99]. Replacing the scotopic matrix  $C^s$  in Equation (3) with the one reported by Mollon *et al.* [MVB99], our approach creates images appearing different to observers with normal and daltonian vision. While most Purkinje images serve curiosity and recreation, this extends the standard test for colour vision defects [Ish17] to arbitrary natural images. However,



**Figure 6:** Combining a cow (a) and deer (b) to be perceived as a cow by a human (c) and as a deer by a deer (d, simulated).

the response difference between daltonian and normal vision is less than the one of photopic and scotopic vision and consequently the separation is more challenging to optimize (Figure 5).

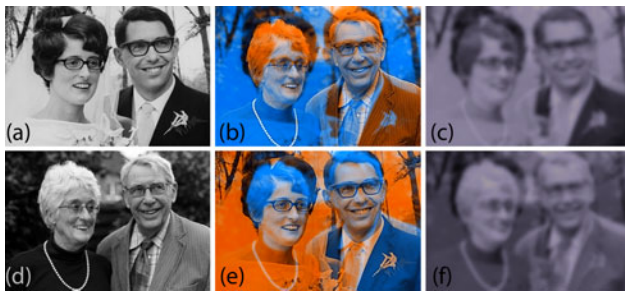
**Animals:** Most animal colour vision suffers a rank deficiency as well [JDN\*94]. This can be used to design patterns that appear different to humans and animals for camouflage purposes [Bur07]. The visual acuity of a deer is about 80% lower than humans, which is due to the reduced number of cones. We use the spectral responses of two types of cones for deers [JDN\*94, figure 6] to create a simple opponent colour space for deer to be used in our optimization (Figure 6).

## 5. Study

In this section, we report the results of two perceptual studies which quantify the recognition power and the resulting image fidelity. A full description of the experiments, including detailed statistics is presented in the Supporting Information.

### 5.1. Recognition study

All 12 participants (six M/six F, 22–27 years) were volunteers, naïve to the specific purpose of the experiments and had normal or corrected-to-normal visual acuity. Stimuli were presented printed on A4 matte paper in two rooms: typical office lighting conditions



**Figure 7:** Purkinje images are not symmetric. An input (a+d) leads to (b) and (c), while the input (d+a) leads to (e) and (f). The input is a ‘before/after-image’ and the couple looks older (younger) at day (night) depending on the order. Photos: (a and d) Sander Koot, www.sanderkoot.nl.



**Figure 8:** Combination of three images A, B and C (First column). The bold letter indicates the dominant percept.

(100 cd/m<sup>2</sup>) and a dark room (0.01 cd/m<sup>2</sup>). Subjects were adapted to the conditions of the respective room. Our goal was to test if subjects indicate the correct dominant percept under a certain condition when presenting them a test image and two reference images and asking a 2AFC question: ‘Which of the two reference images is more similar to the test image?’. The stimuli were 30 Purkinje images (Figure 9; 30 ordered pairs from 15 combinations of six different female faces). We generated four alternatives for each Purkinje image: Three versions of our optimization (Pixel, Laplacian, small  $\alpha = 0$  and high  $\alpha = 1$  chroma weight) and trivial blending without optimization. A single test image and reference pair was shown to the subject in each trial all. The order of test images and the arrangement of reference pairs were randomized. We performed two sessions for each subject: one in the lit and one in the dark room.

We would like to test, if the proportion of correct answers in each category significantly differs from a random probability (all statements in this paragraph are significant  $p < 0.01$ ; pairwise binomial, unless noted otherwise). The results show that our optimization improves the rate of correct answers averaged over all pairs and conditions, from 53% (close to chance level) for linear blending to 96% (close to always-correct). Using the Laplacian decomposition

leads more correct answers (97%) than not using it (94.5%), which effectively means that misclassification is almost halved from 5.5% to 3%. We find that for the Laplacian solver answers are more correct for photopic (99%) than for scotopic (95%) but find the reverse without (91% photopic vs. 98% scotopic). This effect of size 8% might be explained by limitations of our metric that over- or underestimates perceived error in one condition for each approach. Of 28 combinations of faces, eight did not achieve a significant improvement under scotopic conditions and nine did not achieve a significant improvement under photopic conditions. Two of 28 images did not produce an improvement in both conditions. When aggregating over all 14 combinations for each of the six images the difference of all techniques is significant: again 53% for blending, 97% for Laplacian and 95% without it.

## 5.2. Fidelity study

Having shown that our method provides superior recognition rates, we next tested if the fidelity is roughly equivalent as well, since it is not expected that combining two images in one improves image fidelity. To assess the result fidelity, the same image set was shown to 12 other participants (five M/seven F, 22–27 years, experienced in digital photography and colour imaging) who were asked to rate each image under photopic conditions, on a 0–10 Likert scale with regard to the following aspects: (i) overall fidelity, (ii) colour fidelity, (iii) absence of pollution, such as ghosting and (iv) similarity to the original image (Table 1). Note that after showing our method provides superior recognition in Section 5.1, we would now like to know if the fidelity is roughly equivalent as well. It is not expected that combining two images in one improves image fidelity.

**Overall Fidelity:** All variants of our approach are significantly preferred over linear blending ( $p < 0.01$ , paired  $t$ -test) with a medium effect size for linear blending and Laplacian and a small effect size for others ( $g = \text{Hedge's } g$ ). From a small or medium but significant effect, we can conclude that our method is mostly equivalent in terms of overall composition to an alternative which has inferior recognition. We also see moderately significant differences ( $p < 0.1$ ) between the variants of our algorithm, where Laplacian turns out best.

**Colour Fidelity:** The differences in colour fidelity are smaller and not significant. This is expected since in all methods colours are distorted. The scores are significantly different ( $p < 0.01$ ) between our chroma-enhanced and our Laplacian variant, however, effect size for this difference is small.

**Absence of Pollution:** Laplacian Purkinje images have significantly less visual pollution ( $p < 0.01$ ) than linear blending with a large effect size. We can also observe a significant difference between using Laplacian compositing or not. The effect size is medium for both differences. This yields the conclusion that pollution is equivalent between our approach and a baseline which has inferior recognition.

**Similarity:** Both our Laplacian and chroma-enhanced variant significantly outperform linear blending in terms of similarity ( $p < 0.01$ ). Enhancing chroma or not using Laplacian significantly decreases the average scores of our method. We conclude that eliminating high frequencies from the scotopic image increases the



**Figure 9:** The matrix of female faces used in the perceptual experiment. Photos: (a) Antonio Serebryakov; (b) Grigoriy Shipakov; (c) Alina Troeva; (d) Grafik, [www.grafikfoto.ru](http://www.grafikfoto.ru); (e) Dmitriy Ragin and (f) Lyudmila Vilchevskaya.



**Table 1:** Preference scores (rate) and significance of the preference effect between different approaches.

	Overall				Colour				Pollution				Similarity				
	Bl.	Pi.	La.	Ch.	Bl.	Pi.	La.	Ch.	Bl.	Pi.	La.	Ch.	Bl.	Pi.	La.	Ch.	
Rate	3.7	3.8	4.8	4.3	4.3	4.1	4.8	4.6	3.8	3.5	4.7	4.4	3.3	3.6	5.5	4.5	
Blend		0.31	0.01	0.01			0.13	0.25			0.57	0.01	0.02		0.02	0.01	0.01
Pixel	0.31		0.01	0.09	0.13		0.01	0.01	0.57		0.01	0.01	0.02		0.01	0.01	0.01
Lapl.	0.01	0.01		0.06	0.25	0.01		0.98	0.01	0.01		0.73	0.01	0.01		0.01	0.01
Chroma	0.01	0.09	0.06		0.16	0.01	0.98		0.02	0.01	0.73		0.01	0.01	0.01		

fidelity of photopic image in terms of similarity to the original image.

## 6. Discussion and Conclusion

Purkinje images are interesting as they challenge observers to discover multiple interpretations and meanings. We presented a framework to create such ambiguity depending on luminance adaptation conditions.

Our approach has several applications. First, we think it is simply fun to discover hidden messages in images, which is a value on its own. Artists like Dali or Escher have created images of multiple meanings. Secondly, the ability to demonstrate the Purkinje shift on arbitrary images is valuable for computer graphics and perceptual computing education. For medical testing, artificial colour patches are used while our approach allows for real images which are better suited for infants, handicapped people and do not require verbalization. Finally, our technical solution could be used to optimize for images that appear similar under different conditions (traffic signs in day and night, metro plans for colour blinds), for images that undergo distortions in reproduction (gamut mappings), when fabricating clothes that show different patterns under different conditions or to different species (as already done for hunting camouflage).

Our main limitation is—similar to hybrid [OTS06] or emerging images [CHM\*10]—that many combinations of images cannot be successfully combined without human intervention. Purkinje images are also not symmetric (Figure 7). In general, images have to be aligned and must only differ in spatially and chromatically small, but important details. This is possible for multiple images as well (Figure 8). Users can additionally provide automatic or painted saliency maps to guide the separation (not used in any result in this paper). Video S1 shows, how the efficiency of the GPU solver allows to both deform the input images and to paint saliency maps with interactive feedback in the form of a Purkinje image and a simulation for scotopic vision.

Selecting a good pair still depends on the user's intervention. In future work, a more sophisticated way of pair selection could be explored to exclude user control including feedback on the solution quality and choice of images. Furthermore, we would like to account for the temporal characteristics. Purkinje images are just one instance from a large family of multi-perceptual images or multi-perceptual content that should be treated in a unified multi-perceptual framework.

## Acknowledgements

We would like to thank Ahmet Oğuz Akyüz for supplying the experiment set-up, Krzysztof Templin for proofreading and the anonymous reviewers for helpful comments.

## References

- [AM10] ALEXA M., MATUSIK W.: Reliefs as images. *ACM Transactions on Graphics* 29, 4 (2010), 1–7.
- [Arn04] ARNHEIM R.: *Art and Visual Perception: A Psychology of the Creative Eye/New Version*. University of California Press, Berkeley, CA, USA, 2004.
- [Bar89] BARTEN P.: The square root integral (SQRI): A new metric to describe the effect of various display parameters on perceived image quality. In *Proceedings of OE/LASE* (Los Angeles, CA, USA, 1989), pp. 73–82.
- [BM95] BOLIN M., MEYER G.: A frequency based ray tracer. In *Proceedings of SIGGRAPH* (New York, NY, USA, 1995), ACM, pp. 409–418.
- [Bur07] BURRELL J.: Multi-spectral imaging with differential visualizability in discrete visualization domains. US patent 20090017267 A1. 2007.
- [CHM\*10] CHU H.-K., HSU W.-H., MITRA N. J., COHEN-OR D., WONG T.-T., LEE T.-Y.: Camouflage images. *Proceedings of SIGGRAPH Asia, ACM Transactions on Graphics* 29, 4 (2010), 51.
- [CIE10] CIE: *Recommended System for Mesopic Photometry Based on Visual Performance*, Vol. 192. IOS, Vienna, Austria, 2010.
- [CPSZ08] CAO D., POKORNY J., SMITH V. C., ZELE A. J.: Rod contributions to color perception: Linear with rod contrast. *Vision Research* 48, 26 (2008), 2586–2592.
- [Dal92] DALY S. J.: Visible differences predictor: An algorithm for the assessment of image fidelity. In *Proceedings of SPIE/IS&T Electronic Imaging* (Los Angeles, CA, USA, 1992), pp. 2–15.
- [DD00] DURAND F., DORSEY J.: Interactive tone mapping. In *Proceedings of EGWR* (Vienna, Austria, 2000), Durand, F., & Dorsey, J. (Eds.), Springer, pp. 219–230.

- [DRE\*11] DIDYK P., RITSCHER T., EISEMANN E., MYSZKOWSKI K., SEIDEL H.-P.: A perceptual model for disparity. *Proceedings of SIGGRAPH, ACM Transactions on Graphics* 30, 4 (2011).
- [FPSG96] FERWERDA J. A., PATTANAIK S., SHIRLEY P., GREENBERG D. P.: A model of visual adaptation for realistic image synthesis. In *Proceedings of SIGGRAPH* (New Orleans, LA, USA, 1996), pp. 249–258.
- [GS75] GEORGEON M., SULLIVAN G.: Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physics* 252 (1975), 627–656.
- [GTZM10] GOFERMAN S., TAL A., ZELNIK-MANOR L.: Puzzle-like collage. *Computer Graphics Forum* 29, 2 (2010), 459–468.
- [GVWD06] GRUNDLAND M., VOHRA R., WILLIAMS G. P., DODGSON N. A.: Cross dissolve without cross fade: Preserving contrast, color and salience in image compositing. *Proceedings of Eurographics, Computer Graphics Forum* 25, 3 (2006), 577–586.
- [Hau01] HAUSNER A.: Simulating decorative mosaics. In *Proceedings of SIGGRAPH* (Los Angeles, CA, USA, 2001), pp. 573–580.
- [HZZ11] HUANG H., ZHANG L., ZHANG H.-C.: Arcimboldo-like collage using internet images. *Proceedings of SIGGRAPH Asia, ACM Transactions on Graphics* 30, 6 (2011), 155:1–155:8.
- [IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–59.
- [Ish17] ISHIHARA S.: *Test for Colour-Blindness*. Handaya Hongo Harukicho, Tokyo, 1917.
- [Jac93] JACOBS G. H.: The distribution and nature of colour vision among the mammals. *Biological Reviews* 68 (1993), 413–471.
- [JDN\*94] JACOBS G., DEEGAN J., NEITZ J., MURPHY B., MILLER K., MARCHINTON R.: Electrophysiological measurements of spectral mechanisms in the retinas of two cervids: White-tailed deer (*odocoileus virginianus*) and fallow deer (*dama dama*). *Journal of Comparative Physiology A* 174 (1994), 551–557.
- [KCZT12] KIM S., CAO X., ZHANG H., TAN D.: Enabling concurrent dual views on common LCD screens. In *Proceedings of SIGCHI* (Austin, TX, USA, 2012), pp. 2175–2184.
- [KO11] KIRK A. G., O'BRIEN J. F.: Perceptually based tone mapping for low-light conditions. *Proceedings of SIGGRAPH, ACM Transactions on Graphics* 30, 4 (2011), 42:1–42:10.
- [KOF08] KUHN G. R., OLIVEIRA M. M., FERNANDES L. A. F.: An efficient naturalness-preserving image-recoloring method for dichromats. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1747–1754.
- [KP02] KIM J., PELLACINI F.: Jigsaw image mosaics. *ACM Transactions on Graphics* 21, 3 (2002), 657–664.
- [Liv02] LIVINGSTONE M.: *Vision and Art: The Biology of Seeing*. Harry N. Abrams, New York, USA, 2002.
- [MCL\*09] MITRA N. J., CHU H.-K., LEE T.-Y., WOLF L., YESHURUN H., COHEN-OR D.: Emerging images. *Proceedings of SIGGRAPH Asia, ACM Transactions on Graphics* 28, 5 (2009), 163.
- [MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *Proceedings of SIGGRAPH, ACM Transactions on Graphics* 30, 4 (2011), 40:1–40:14.
- [MRH09] MANTIUK R., REMPEL A. G., HEIDRICH W.: Display considerations for night and low-illumination viewing. In *Proceedings of APGV* (Crete, Greece, 2009), pp. 53–58.
- [MVB99] MOLLON J. D., VIÉNOT F., BRETTEL H.: Digital video colourmaps for checking the legibility of displays by dichromats. *Color Research and Application* 24, 4 (1999), 243–252.
- [OTS06] OLIVA A., TORRALBA A., SCHYNS P. G.: Hybrid images. *Proceedings of SIGGRAPH, ACM Transactions on Graphics* 25, 3 (2006), 527–532.
- [PD84] PORTER T., DUFF T.: Compositing digital images. In *Proceedings of SIGGRAPH* (Minneapolis, MN, USA, 1984), pp. 253–259.
- [PFFG98] PATTANAIK S. N., FERWERDA J. A., FAIRCHILD M. D., GREENBERG D. P.: A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of SIGGRAPH* (Orlando, FL, USA, 1998), pp. 287–298.
- [PGB03] PÉREZ P., GANGNET M., BLAKE A.: Poisson image editing. In *Proceedings of ACM SIGGRAPH* (San Diego, CA, USA, 2003), pp. 313–318.
- [PHN\*12] PAPAS M., HOUIT T., NOWROUZEZAHRAI D., GROSS M., JAROSZ W.: The magic lens: Refractive steganography. *Proceedings of SIGGRAPH Asia, ACM Transactions on Graphics* 31, 6 (2012), 186.
- [RKAJ08] REINHARD E., KHAN E. A., AKYUZ A. O., JOHNSON G. M.: *Color Imaging Fundamentals and Applications*. AK Peters, Natick, MA, USA, 2008.
- [RWD\*10] REINHARD E., WARD G., DEBEVEC P., PATTANAIK S., HEIDRICH W., MYSZKOWSKI K.: *High Dynamic Range Imaging* (2nd edition). Morgan Kaufmann Publishers, Burlington, MA, USA, 2010.
- [SG04] SETLUR V., GOOCH B.: Is that a smile?: Gaze dependent facial expressions. In *Proceedings of NPAR* (Annecy, France, 2004), ACM, pp. 79–151.
- [Shl37] SHLAER S.: The relation between visual acuity and illumination. *The Journal of General Physiology* 21 (1937), 165–188.

- [SMW06] SCHAEFER S., MCPHAIL T., WARREN J.: Image deformation using moving least squares. *Proceedings of SIGGRAPH, ACM Transactions on Graphics* 25, 3 (2006), 533–540.
- [TC90] TYLER C. W., CLARKE M. B.: The autostereogram. In *Proceedings of SPIE Stereoscopic Displays and Applications* (Santa Clara, CA, USA, 1990), vol. 1258, pp. 182–196.
- [TSF02] THOMPSON W. B., SHIRLEY P., FERWERDA J. A.: A spatial post-processing algorithm for images of night scenes. *Journal of Graphics Tools* 7, 1 (2002), 1–12.
- [TZHM11] TONG Q., ZHANG S.-H., HU S.-M., MARTIN R. R.: Hidden images. In *Proceedings of NPAR* (Vancouver, Canada, 2011), pp. 27–34.
- [vBPJ\*11] VAN BAAR J., POULAKOS S., JAROSZ W., NOWROUZEZAHRAI D., TAMSTORF R., GROSS M.: Perceptually-based compensation of light pollution in display systems. In *Proceedings of APGV* (Toulouse, France, 2011), pp. 45–52.
- [Wal45] WALD G.: Human vision and the spectrum. *Science* 101 (1945), 653–658.
- [Wan95] WANDELL B. A.: *Foundations of Vision*. Sinauer Associates, Sunderland, MA, USA, 1995.
- [Wol98] WOLBERG G.: Image morphing: A survey. *The Visual Computer* 14, 8 (1998), 360–372.
- [WRP97] WARD G., RUSHMEIER H., PIATKO C.: A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics* 3, 4 (1997), 291–306.

### Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

### Video S1 Purkinje Images