Dual-sensor foveated imaging system

2 authors:

Hong Hua
The University of Arizona
173 PUBLICATIONS **3,037** CITATIONS

SEE PROFILE

Sheng Liu
Goertek USA
25 PUBLICATIONS **592** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Connected components, Eye tracking View project

Computational light field display View project

Dual-sensor foveated imaging system

Hong Hua* and Sheng Liu

College of Optical Sciences, University of Arizona, 1630 E. University Boulevard, Tucson, Arizona 85721, USA

*Corresponding author: hhua@optics.arizona.edu

Conventional imaging techniques adopt a rectilinear sampling approach, where a finite number of pixels are spread evenly across an entire field of view (FOV). Consequently, their imaging capabilities are limited by an inherent trade-off between the FOV and the resolving power. In contrast, a foveation technique allocates the limited resources (e.g., a finite number of pixels or transmission bandwidth) as a function of foveal eccentricities, which can significantly simplify the optical and electronic designs and reduce the data throughput, while the observer's ability to see fine details is maintained over the whole FOV. We explore an approach to a foveated imaging system design. Our approach approximates the spatially variant properties (i.e., resolution, contrast, and color sensitivities) of the human visual system with multiple low-cost off-the-shelf imaging sensors and maximizes the information throughput and bandwidth savings of the foveated system. We further validate our approach with the design of a compact dual-sensor foveated imaging system. A proof-of-concept bench prototype and experimental results are demonstrated.   © 2008 Optical Society of America

OCIS codes: 110.0110, 220.4830, 330.7338, 330.1800, 230.4685.

1. Introduction

Real-time acquisition of high-resolution, wide field of view (FOV), and high dynamic range (HDR) images is essential for many vision-based applications, ranging from the macroscale imaging in surveillance [1], teleconferencing [2], and robot navigation [3], to micrometer or submicrometer-scale imaging in biomedical imaging [4] and vision-guided microfabrication [5]. Although detector technologies have improved dramatically over the past decades, many imaging systems are limited by their performances on FOV, resolution, speed, and dynamic range. Most conventional imaging techniques spread a finite number of pixels evenly across the entire FOV, and thus their imaging capabilities are constrained by a well-known inherent trade-off between the FOV and the resolving power; the higher the resolving power, the smaller the FOV. Consider a surveillance camera with a typical National Television System Committee (NTSC) CCD detector of 640×480 pixels, as an example. A diagonally 60° FOV only offers an angular resolution of 6.4 arc min per pixel at the maximum. If mounted onto a 20 ft ceiling, the camera can only deliver a human face image of about 20×16 pixels, which are neither sufficient for object tracking nor face identification.

This resolution-FOV trade-off problem may be relieved with the advent of higher-resolution sensors and the development of sensor tiling or image mosaicing schemes [4,5]. For instance, in an adaptive scanning optical microscope system [5], a 40 mm observing field is formed by mosaicing 25×25 images, each having a resolution of 512×512 pixels. Thus the full-field image will be 12800×12800 pixels. However, processing and analysis of such high-resolution images is computationally demanding. Furthermore, capturing 25×25 images to reconstruct a full-field image at a tolerable frame rate (e.g., 10 fps) not only requires an image sensor with a very high frame rate, but also demands a fast scanning mechanism. Moreover, with the increasing interests in teleoperation and telemedicine, real-time transmission of these images over a network can be very challenging.

Rather than demanding more expensive (both economically and computationally), higher resolution, and faster image sensors as well as associated processing technologies. In this paper, we present an approach to a foveated imaging system design,

aiming to improve the information throughput of low-cost off-the-shelf detectors. In Section 3, we will present a comprehensive analysis and exploration on how the spatially variant properties of the human visual system (HVS), including the well-explored domain of spatially variant resolution [6–8] and the less-explored spatial variances of contrast and color sensitivities, may be utilized to maximize the information throughput and reduce data bandwidth requirements of an imaging system. In Section 4, we will describe a schematic design of a compact, actively foveated dual-sensor imaging system, which mimics several aspects of the HVS. Finally, a bench prototype and experiments are presented in Section 5, which demonstrates that 96.4% bandwidth savings can be achieved at the stage of image capturing in a dual-sensor foveated imaging system.

2. Related Work

Foveation techniques, inspired by the foveation properties of the human eye, can be generally characterized as a method that dynamically tracks a user's attended area of interest (A-AOI) by means of a gaze tracker, or other mechanism, and treats the A-AOI differently from the peripheral area. A great amount of effort has been made to explore foveation techniques in imaging and display applications, which are sometimes referred to as gaze-contingent or eye-tracked multiresolution techniques. Examples include spatially variant compression ratio in image and video processing [6–8], variable levels of detail in three-dimensional (3D) rendering [9–11], variable pixel resolution in imaging or display systems [12,13], and adaptive correction of optical aberrations in imaging or display systems [14,15]. More application examples and reviews can be found in [9,16].

The prior work on foveation techniques falls into one of three categories, as follows. The first category of work relates to experimental research on perception and cognition to understand visual processing and perceptual artifacts produced by simulated multiresolution images or display systems [17–22]. For instance, researchers investigated perceptual artifacts in foveated multiresolution displays, such as perceptible image blur and image motion, which have the potential to distract users [17,20–22]. Ideally, one would like to maximize the bandwidth savings of foveation techniques, while minimizing perception artifacts and performance costs.

The second category is the algorithmic approach in which foveation techniques are applied primarily to image processing [2,7], video encoding [6,8], and graphics rendering [9–11] to achieve real-time video communication through low-bandwidth networks and to save data transmission bandwidth and processing resources. For instance, Geisler and Perry demonstrated a three times greater compression ratio in multiresolution images/videos [8]; Ienaga et al. described a stereoscopic video transmission system with embedded high-resolution fovea images to accelerate the image processing and

to reduce the bandwidth requirement [2]; several researchers applied perceptually driven foveation treatments to 3D graphics rendering such a level-of-detail management and polygon reduction in 3D models based on the view eccentricity [9–11]. Murphy and Duchowski demonstrated that the rendering speed can be significantly accelerated by only rendering high levels of detail of a 3D scene around a user's gaze direction in virtual environments [10]; Luebeke et al. demonstrated that 2–6 times fewer polygons can produce the similar rendering effect in perceived model resolution [11].

The third category of work takes a hardware approach, in which various imaging sensors or displays with spatially varying resolution are developed to reduce requirements for high-resolution detectors and displays or high-quality and complex optical systems [3,12–15,23–25]. For instance, Sandini et al. described the implementation of a retina-like complimentary metal-oxide semiconductor (CMOS) sensor characterized by spatially variant resolution similar to that of the human retina [12]. They demonstrated that 35 times fewer pixels were needed in the space-variant resolution sensor as compared with a constant high-resolution image of 1100×1100 pixels. Martinez and Wick presented the designs of foveated imaging systems in which a liquid crystal (LC) spatial light modulator (SLM) was used to dynamically correct the aberrations of a simple-structure wide FOV optics at the foveated region of interest (FRoI) [14,15]. Godin described a multiprojector system to achieve a dual-resolution 3D display [24]. Rolland presented the design of a head-mounted display with high-resolution insets [13]. Several groups of researchers demonstrated the use of a camera pair with different angular resolutions in robot navigation systems to achieve foveated images [3,25].

Although both the algorithmic and hardware approaches can achieve foveated images, the hardware approach is advantageous in the sense that it achieves foveated imaging at a more fundamental level. It captures foveated images with spatially variant resolution by designing novel imaging sensors and systems, rather than postprocessing full-resolution images by applying foveated filtering and encoding techniques. It potentially requires less complex optical and electronic designs and reduces the requirements on data bandwidth and computational resources, while the postprocessing approach is mostly limited to the benefits of information storage and transmission.

3. Foveated Imaging with Spatially Varying Resolution and Chromaticity

In the HVS, only a narrow region around the fovea offers exceptional resolving power, contrast, and color sensitivities, while these properties fall off rapidly with an increasing retinal eccentricity. The essential goal of designing a foveated imaging system is to improve the sampling efficiency and to reduce the requirements for data bandwidth as well as compu-

tational resources by mimicking such spatially varying properties of the HVS.

Rather than pursuing specially designed sensors with retina-like properties [12,23], we use multiple low-cost sensors to approximate the spatially variant properties of the HVS, including the well-explored domain of spatially variant resolution [6–8] and the less-explored spatial variances of contrast and color sensitivities. Our approach significantly improves the information throughput of low-cost sensors and reduces the total number of required pixels. The reduction in total pixel counts directly translates to less storage space and communication bandwidth, which offers great opportunities to break the technical barrier in designing wide FOV systems with much simpler optical designs, to achieve system miniaturization, and to develop distributed sensing networks with limited bandwidths.

This section details the analytical methods for achieving spatially varying resolution and spatial-chromatic foveation. The analysis predicts that over 80% of bandwidth saving can be easily achieved with a dual-resolution dual bit depth (DRDB) foveated imaging system, compared to a traditional rectilinear sampling approach.

A. Spatially Variant Sampling and Multiresolution Imaging Systems

We characterize the sampling scheme of an imaging system with its relative spatial resolution distribution as a function of field angles. The distribution, denoted as $F(e_x, e_y)$, is the normalized reciprocal of the angular resolution at a given visual field (e_x, e_y), where e_x and e_y are the horizontal and vertical eccentricities measured in degrees, respectively. The angular resolution here is measured by arcminutes per pixel. The amount of raw data acquired by the system can be calculated as the integration of its relative resolution response across the visual field, given as

$$B = \int_{-\theta_X/2}^{\theta_X/2} \int_{-\theta_Y/2}^{\theta_Y/2} F^2(e_x, e_y) de_x de_y, \qquad (1)$$

where θ_X and θ_Y are the full FOVs covered by the system along the horizontal and vertical directions, respectively. The amount of raw data characterizes the bandwidth requirement.

The sampling efficiency of a given system may be evaluated against a reference system whose spatial resolution distribution represents the just-adequate resolvability required for a particular task. Systems with the same or better resolvability than that of their reference system are considered to be perceptually equivalent. Without loss of generality, we use the visual acuity response of the human eye, $F_{HVS}(e_x, e_y)$, as the reference system throughout the paper. Compared against the reference system providing the same FOV and the same peak resolution, the sampling efficiency, E, of a given system is defined as

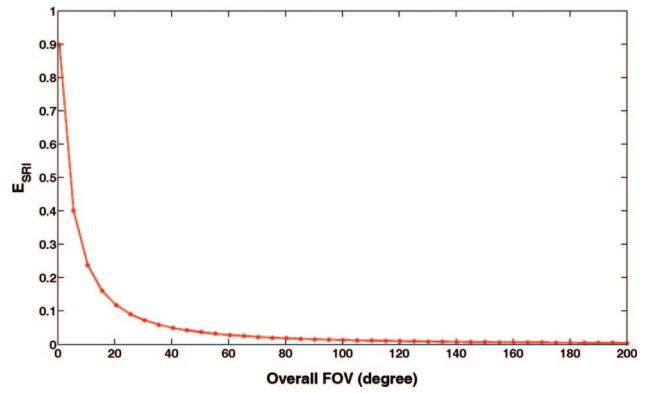
Fig. 1. (Color online) Sampling efficiency of a SRI system as a function of the overall FOV.

$$E = \frac{\int_{-\theta_X/2}^{\theta_X/2} \int_{\theta_Y/2}^{\theta_Y/2} F_{HVS}^2(e_x, e_y) de_x de_y}{\int_{-\theta_X/2}^{\theta_X/2} \int_{-\theta_Y/2}^{\theta_Y/2} F^2(e_x, e_y) de_x de_y}. \qquad (2)$$

E also characterizes the relative information throughput of the system. For a single-resolution imaging (SRI) system with uniform sampling characterized by $F_{SRI}(e_x, e_y) \equiv 1$, Eq. (2) is simplified as $E_{SRI} = \frac{1}{\theta_X \theta_Y} \int_{-\theta_X/2}^{\theta_X/2} \int_{\theta_Y/2}^{\theta_Y/2} F_{HVS}^2(e_x, e_y) de_x de_y.$

In Fig. 1 we plot the sampling efficiency of a SRI system as a function of the overall FOV. The relative visual acuity response of the eye was modeled as $F_{HVS}(e_x, e_y) = 2.3/(2.3 + \sqrt{e_x^2 + e_y^2})$, adopted from [8,19]. A 4:3 aspect ratio was assumed for the overall field. As shown in Fig. 1, the efficiency monotonically decreases with an increasing FOV. For instance, the efficiency is as low as 3% for a 60° system and is less than 0.5% for a 200° system. The amount of redundant raw data produced by a 60° SRI system is almost 35 times more than that of the perceptually equivalent reference system, and over 200 times more for a 200° SRI system.

Aiming to improve the sampling efficiency of a SRI system depicted in Fig. 1, we explored an approach, where multiple low-cost sensors are utilized to approximate the spatial-variant properties of the retina. Let us consider approximating the eccentricity response of the visual acuity with two levels of spatial resolution as shown in Fig. 2(a). One sensor, referred to as the foveated imager, covers the central region of the overall visual field, while the other sensor covers the peripheral region. The foveated imager uniformly maintains a maximum resolution within the range of $\pm \theta_{XF}$ and $\pm \theta_{YF}$ degrees in the horizontal and vertical directions, respectively. In order to avoid discernible artifacts, the relative resolvability of the peripheral imager is set to be $F_{DRI}(\theta_{XF}, \theta_{YF})$, where $F_{DRI}(\theta_{XF}, \theta_{YF}) = \max(F_{HVS}(\theta_{XF}, 0), F_{HVS}(0, \theta_{YF}))$ for rectangular visual fields. Compared with a SRI system, the bandwidth saving of a dual-resolution imaging (DRI)
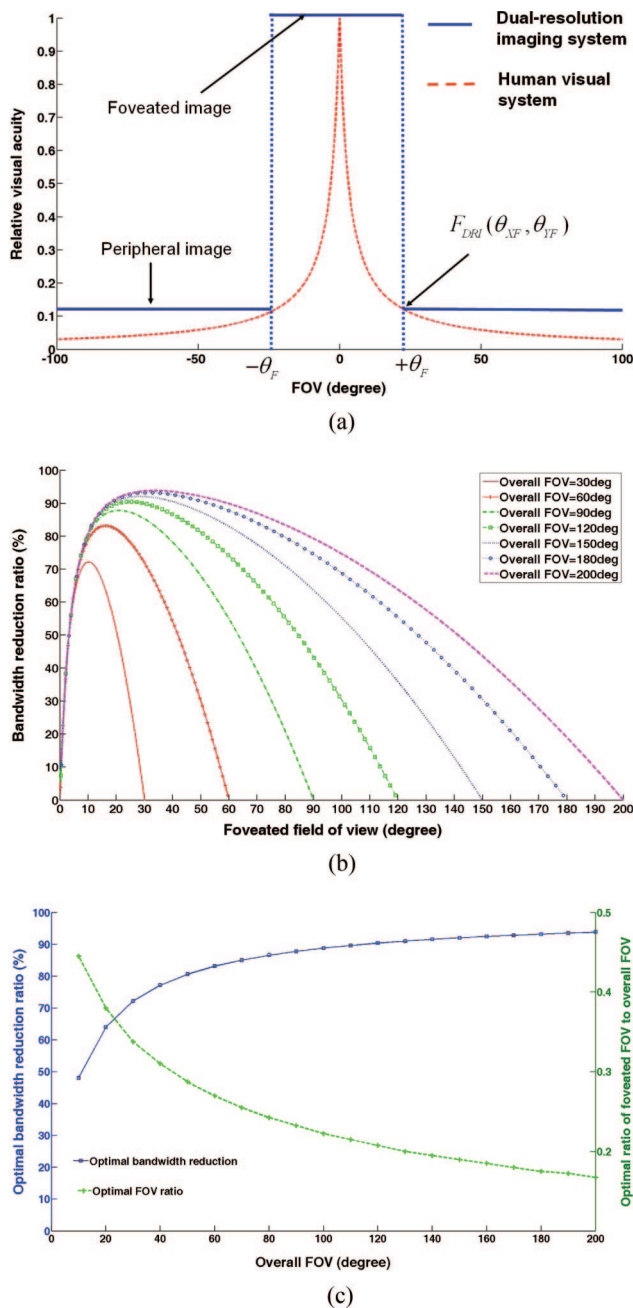
(a)

(b)

(c)

Fig. 2. (Color online) Optimization of a dual-resolution foveated imaging system: (a) eccentricity of visual acuity and dual-resolution sampling scheme; (b) bandwidth saving ratio as a function of peripheral and foveated FOVs; (c) maximum bandwidth saving in percentile (blue curve with square marker) and optimal ratio of foveated FOV to overall FOV (green curve with + marker).

system is given as

$$
\begin{aligned}
S_{\mathrm{DRI}} &= \frac{B_{\mathrm{SRI}} - B_{\mathrm{DRI}}}{B_{\mathrm{SRI}}} \\
&= \frac{\theta_X\theta_Y - (\theta_X\theta_Y - \theta_{XF}\theta_{YF})F_{\mathrm{DRI}}^2(\theta_{XF},\ \theta_{YF}) - \theta_{XF}\theta_{YF}}{\theta_X\theta_Y}.
\end{aligned}
\tag{3}
$$

In Fig. 2(b), we plot the bandwidth saving ratio of a DRI system as a function of the foveated FOV and the overall visual field. A 4:3 aspect ratio was assumed for both the overall and foveated FOVs. For a given overall FOV, the bandwidth saving ratio can be maximized by varying the foveated FOV. Figure 2(c) plots the maximum bandwidth saving ratio as a function of the overall FOV. In the same figure we further plot the optimal ratio of the foveated FOV to the overall FOV.

Based on Fig. 2(c), by correctly choosing the foveated FOV relative to the overall FOV, we can maximize the bandwidth reduction and information throughput of a dual-sensor foveated imaging system while maintaining indiscernible image quality degradation. Consider a system requiring a total of 120° diagonal FOV and 1 arc min of peak angular resolution. A SRI system with uniformly 1 arc min resolution requires a single sensor with 5760×4320 pixels. In contrast, in a dual-sensor foveated system, the relationship in Fig. 2(c) suggests that a 25° foveated FOV maximizes the spatial sampling efficiency of a 120° system. The combination of the 25° foveated FOV in 1 arc min resolution with a 95° peripheral FOV in 4.25 arc min of resolution yields a bandwidth reduction ratio of about 90.4%, which directly translates to almost ten times fewer pixels than those for a SRI design.

The above sampling approach can be extended to more than two resolution levels using a similar scheme. Maximizing the overall sampling efficiency of multisensors will involve optimizing the FOV coverage of these sensors and correctly selecting their resolution levels so that the resulting system presents a multiresolution image with indiscernible quality degradation while minimizing the amount of redundant information.

B. Contrast/Color Sensitivity and Image Bit Depth

Similar to the property of spatially variant resolution, the contrast and color sensitivities of an imaging system can be encoded with the greatest precision at the center of the area of interest and with decreasing precision as the distance from the center (the eccentricity) increases. Several researchers have suggested that exploring peripheral chromatic degradation can potentially lead to significant bandwidth saving [9,16]. In this subsection, we explore the spatial variance of the contrast sensitivity function of the HVS to examine the minimal bit depth and color channels, which yield images with indiscernible peripheral chromatic degradation.

The contrast sensitivity function (CSF) of the HVS at different retinal eccentricities can be adequately described by the formula [8,19]

$$
\mathrm{CSF}(e,\ \nu) = \frac{1}{CT(e,\ \nu)} = \mathrm{CSF}(0,\ \nu)\exp\!\left(-\alpha\nu\,\frac{e_2 + e}{e_2}\right),
\tag{4}
$$

where $e = \sqrt{e_x{}^2 + e_y{}^2}$ is the retinal radial eccentricity in degrees, ν is the spatial frequency in cycles per degree (cpd), α is the spatial frequency decay constant ($\alpha \cong 0.045$), e_2 is the retinal eccentricity, where spatial resolution falls to half of what it is in the center of the fovea ($e_2 \cong 2.3°$), and $\mathrm{CSF}(0, \nu)$ models the contrast sensitivity of the fovea as a function of spatial frequency (please see [26] for detailed explanations).

The decreasing contrast sensitivity with retinal eccentricity suggests that the bit depth of an image may be reduced monotonically as the eccentricity increases without creating perceptible banding effects. Let us consider the DRI system discussed in Subsection 3.A. Here we assume the foveated imager, with a bit depth of N, maintains a uniform spatial resolution of μ_f arc min within the foveated FOV ($\pm\theta_{XF}, \pm\theta_{YF}$). The peripheral imager maintains a uniform spatial resolution of μ_p arc min with a lower bit depth, M. To avoid perceivable resolution degradation, μ_p is determined by $\mu_p \leq \mu_f / F_{\mathrm{DRI}}(\theta_{XF}, \theta_{YF})$ as discussed in Subsection 3.A. Furthermore, to minimize artifacts caused by the reduction of bit depth, we set up a constraint for the reduced bit depth, M, of the peripheral imager, written as

$$\Delta C_{i,j}(N, M)\big|_{\max} < CT\left(\theta_F, \ \nu \leq \frac{60}{\mu_f + \mu_p}\right)$$

$$(i = 0, 1, \ldots 2^N - 1, j = 0, 1, \ldots 2^M - 1), \quad (5)$$

where $\Delta C_{i,j}(N, M)\big|_{\max}$ is the maximum contrast between the original and remapped pixel values when a pixel value at the ith gray level in a N-bit image is remapped to the jth gray level in a M-bit sensor. This constraint ensures that a user is unable to discern the bit reduction at the boundaries of the two imagers when fixating on the center of the foveal image.

Figure 3(a) plots the HVS contrast sensitivity as a function of spatial frequency and eccentricity. Under the conditions of $\mu_f = 1$ arc min and $N = 8$ bits per color channel, in Fig. 3(b) we plot $\Delta C_{i,j}(N, M)\big|_{\max}$ as a function of the bit depth, M, of the peripheral imager. On the same graph, we also show the contrast modulation thresholds for $\pm10°$, $\pm15°$, and $\pm20°$ of foveated FOVs for 6 cpd spatial frequency. For instance, the contrast modulation threshold at $\pm10°$ of eccentricity is 0.0385, and the bit depth for the peripheral imager may be reduced down to 4 without noticeable boundary artifacts.

Compared to a SRI system with a uniform resolution and single bit depth, the percentage of bandwidth saving of a DRDB design is given as
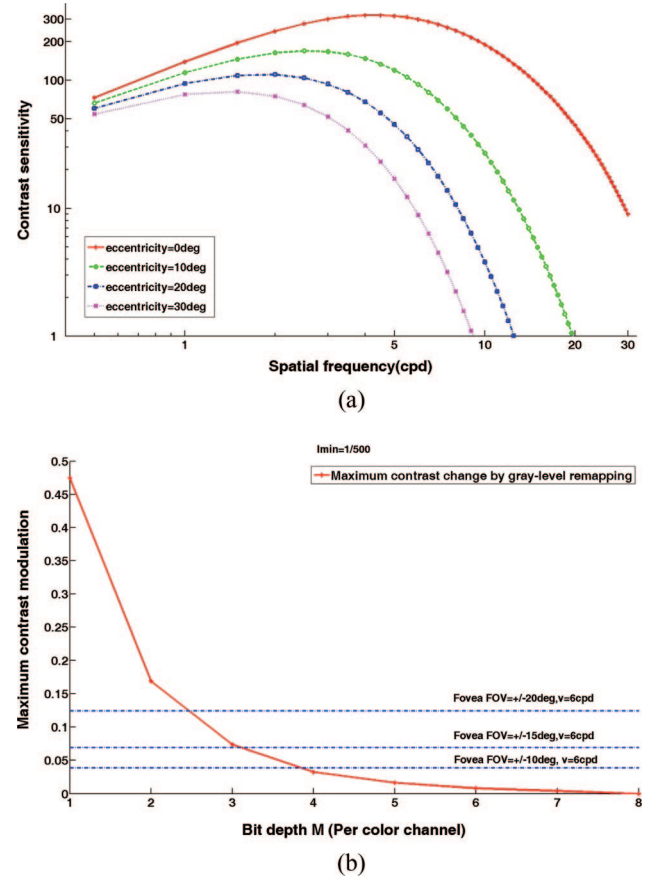

(a)

(b)

Fig. 3. (Color online) Contrast sensitivity and image bit depth: (a) the HVS contrast sensitivity as a function of spatial frequency and eccentricity; (b) maximum contrast modulation change as a function of reduced bit depth.

For a given overall FOV, the foveated FOV can be selected to maximize the sampling efficiency based on the relationship in Fig. 2. As a result, the magnitude of bandwidth saving essentially depends on the overall FOV and the bit depth of the foveal and peripheral imagers. Figure 4 plots the bandwidth saving ratio of the DRDB system as a function of the overall FOV and reduced peripheral bit depth M, where we assumed a 4:3 aspect ratio for the visual fields and eight bits per color channel for the foveated imager ($N = 8$). As an example, to design a system of 120° diagonal FOV, the bandwidth saving ratio is increased to 95.65% with a 25° foveated FOV in eight bits per color channel, and the remaining 95° peripheral FOV in five bits per color channel. This result directly translates to 23 timesless storage space and communication bandwidth. The plots in Fig. 4 further suggest that the benefit of bit depth reduction diminishes dramatically for M less than 5.

$$S_{\mathrm{DRDB}} = \frac{B_{\mathrm{SRI}} - B_{\mathrm{DRDB}}}{B_{\mathrm{SRI}}} = \frac{2^N \cdot \theta_X \theta_Y - 2^M \cdot (\theta_X \theta_Y - \theta_{XF}\theta_{YF}) F_{\mathrm{DRI}}{}^2(\theta_{XF}, \ \theta_{YF}) - 2^N \cdot \theta_{XF}\theta_{YF}}{2^N \cdot \theta_X \theta_Y}. \quad (6)$$
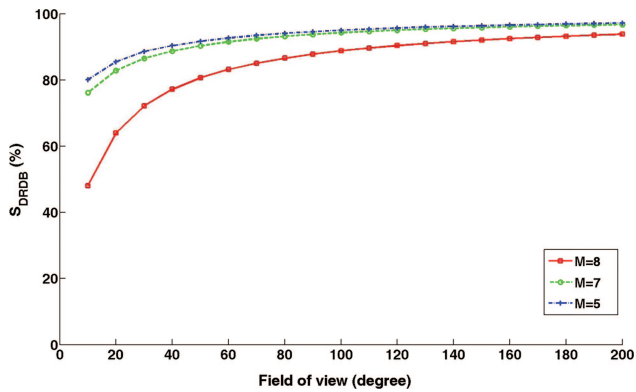
Fig. 4. (Color online) Bandwidth saving ratio of a DRDB system as a function of overall FOV and reduced bit depth.

Fig. 5. (Color online) Simulated foveated images: (a) original SRI (2400 × 1800 24 bit pixels); (b) DRDB image (691 × 518 24 bit pixels in the center and 766 × 574 15 bit pixels in the peripheral); (c) DRDB image (691 × 518 24 bit pixels in the center and 766 × 574 8 bit gray-scale pixels in the peripheral); (d)–(f) enlarged views of a narrow region of images (a)–(c).

Besides reducing the image bit depth, the bandwidth savings can be further improved by exploring the color sensitivity of the HVS and reducing the number of color channels for the peripheral imager, for instance, from a full-color system to a gray scale as the eccentricity increases [9]. For example, a 120° dual-sensor system described above can be reduced from a 15 bit color imager for the peripheral to a five bit gray scale imager. The resultant bandwidth saving ratio is 95.66% compared with a single-sensor design with perceptually equivalent quality. However, further human factor research is needed to construct a chromaticity eccentricity function that can be utilized to create a spatially varying color system that is just imperceptibly different from a full-color image.

From the point of view of bandwidth reduction, reducing color channels on top of spatially varying resolution and bit depth appears to only provide marginal benefits, while the combination of applying spatially varying resolution and reducing bit depth per color channel for the peripheral imager yields much more effective reduction. However, from the implementation point of view, reducing color channels requires much less hardware modification and achieves similar efficiency to bit depth reduction. In applications where peripheral color sensation is not critical for the aimed tasks, spatially varying resolution methods can be sensibly combined with spatially varying color schemes.

C. Simulation Results

Based on the eccentricity properties discussed in Subsections 3.A and 3.B, a multisensor system can be designed with maximum sampling efficiency and bandwidth savings. Figures 5(a)–5(c) present a set of simulation results to demonstrate the visual effects of resolution, bit depth, and color channel reduction. Figure 5(a) is the original 24 bit full color image with 2400 × 1800 pixels. This image provides approximately 50° diagonal FOV with an angular resolution of about 1 arc min when viewed on a 30 in. monitor at a distance of 32 in. It is worth noting that, in general, a uniform angular resolution in the object space does not 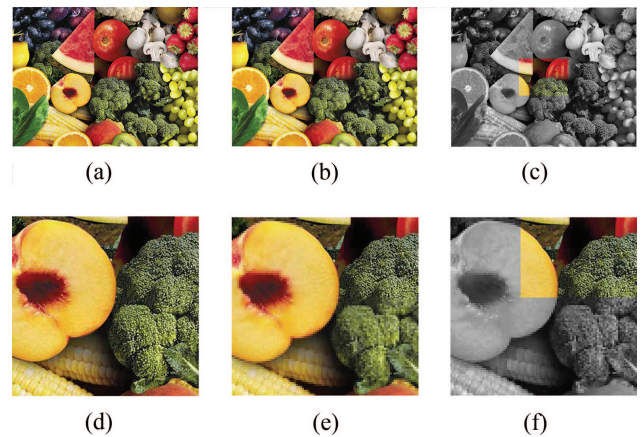map into a uniform pixel density on the image plane. Although sensors with spatially varying pixel densities may be custom-made, in implementations throughout the paper, we choose to use regular sensors with a uniform pixel density in which the worst angular resolution of the pixels satisfies the angular resolution requirements.

Figure 5(b) simulates a DRI with a 24 bit center region of 691 × 518 color pixels and a peripheral area of 766 × 574 pixels with 5 bits per color channel. The center region covers about 14.4° of the overall FOV in 1 arc min resolution when viewed on the same monitor. Each enlarged image pixel of the peripheral area is equivalent to 3 × 3 pixels on the monitor, which corresponds to an angular resolution of 3.3 arc min. Figure 5(c) simulates a similar image with a 24 bit center region of 691 × 518 pixels but an eight bit black/white (B/W) peripheral area of 766 × 574 enlarged pixels. Due to the reduced size of the images shown in Fig. 5, it is hard to tell their resolution differences. Figures 5(d)–5(f) show enlarged views of a narrow region of the images corresponding to Figs. 5(a)–5(c), respectively.

Compared with the original image in Fig. 5(a), the total number of pixels for both Figs. 5(b) and 5(c) was reduced from 4.32 × 10⁶ to 0.798 × 10⁶. The overall bandwidth saving ratios of Figs. 5(b) and 5(c) are 91.71% and 91.73%, respectively. From the point of view bandwidth reduction, they are almost equivalently efficient. If these simulated images are viewed under the same condition mentioned above, and the user fixates at the center of the images, Figs. 5(b) and 5(c) should provide the about same spatial resolvability perceptually equivalent to Fig. 5(a), while the color shift in Fig. 5(b) would not be as noticeable as Fig. 5(c).

4. Design of a Dual-Sensor Foveated Imaging System

When the foveation approach discussed in Section 3 is combined with a dynamic tracking and scanning

mechanism, we can make the highest resolution, best contrast, and best color images available, where the user is mostly interested in, while using considerably less hardware bandwidth and computational resources. In this section, we present a design of a foveated imaging system based on the dual-sensor approach. The system operates by capturing two images of the scene simultaneously with two relatively low-resolution detectors: one covers the entire scene and the other covers a narrow area of interest. The two imagers, sharing the same entrance pupil, are optically coupled together through a two-axis microelectro-mechanical systems (MEMS) scanner, which is driven by fovea tracking algorithms. Depending on the light levels of the entire scene, and local features of interest, the imagers may be adaptively modulated to acquire HDR images by implementing the pixel-by-pixel modulation approach described in [27]. The ability to dynamically steer the fovea region toward a detected stimulus via instantaneous eye movements is referred to as adaptive foveation; the ability to adapt the pupil opening in response to a large magnitude of illumination level variations is referred to as adaptive illumination sensing.

A. Conceptual Design

Conceptually, our targeted system mainly consists of five components: a foveated imager, a peripheral imager, a fast 2D scanner, adaptive aperture modulators, and fovea tracking and high-dynamic range control algorithms. The foveated imager mimics the fovea pit of the eye and provides the fine high-contrast details and color sensation of a narrow foveated region for target recognition. The peripheral imager mimics the peripheral vision of the eye and captures an extended field with relatively low resolution. It provides the peripheral context for stimulus detection. The scanning system mimics the eye movements and provides the dynamic capability for target tracking. The adaptive aperture modulator consists of two elements. The first modulator can mimic the effect of pupil size adaptation to the scene luminance variation. It dynamically adjusts the active aperture of the imaging system based on the overall light levels detected from the peripheral and foveated imagers. The second modulator is to enhance the dynamic range of the imaging system by using a method by Gao *et al.* [27], in which the image exposure of a scene is controlled on a pixel-by-pixel basis to deal with scenes with a wide dynamic range locally [27]. Fovea tracking algorithms can be implemented to dynamically detect the salient region of interest and stimulus events from the peripheral image.

The dual-sensor design differs from the existing methods of foveated imaging in several aspects [12,14,15,23]. First of all, the dual-sensor approach offers real-time peripheral awareness and potentially the capability of event detection and tracking. Second, compared to the single-sensor foveated imaging approach by Wick *et al.* [14,15], or the retina-like sensor approach [12,23], the dual-sensor method

yields high information throughputs for low-cost detectors without the need for custom-made special sensors. It is worth noting that the single-sensor foveated systems by Wick *et al.* capture both the foveated and peripheral areas on the same detector. The angular coverage per pixel remains approximately constant across the entire field if a regular sensor with a constant pixel density is utilized. The low-resolution appearance of their systems is due to optical aberrations [14,15]. It has the advantage of using simple optics for imaging but compromises the detector efficiency. Furthermore, a retina-like sensor with space-variant pixel density imposes a higher cost and lacks the ability to achieve dynamic foveation unless the spatial-variant pixel arrangements can be dynamically controlled. Finally, the proposed design can potentially incorporate adaptive aperture modulators for advanced HDR imaging [27].

B. Schematic Design

The schematic design of an actively foveated imaging system is shown in Fig. 6. To achieve compactness, the foveated and peripheral imagers share the same objective lens, which forms the first intermediate image plane. The objective lens itself contains a physical stop near its front focal plane for the entire imaging system. A beam splitter placed after the first intermediate image plane separates the paths for the two imagers. A scanner lens collects the light transmitted through the beam splitter. It forms an intermediate pupil plane at which a two-axis analog MEMS mirror is placed to achieve a compact design. Other types of scanning mirror with lower cost may be used if compactness is less critical. By tilting the mirror instantaneously toward the direction of interest, rays from the interested FOV are redirected toward the optical axis. To achieve compactness, the same scanner lens collects a narrow field of the light reflected by the mirror and forms a high-resolution foveated image for the region of interest. In the meantime, a camera with a relay lens is placed on the reflected path of the beam splitter to form a low-resolution peripheral image of the entire visual field. An SLM may be placed at the first intermediate image plane to facilitate the
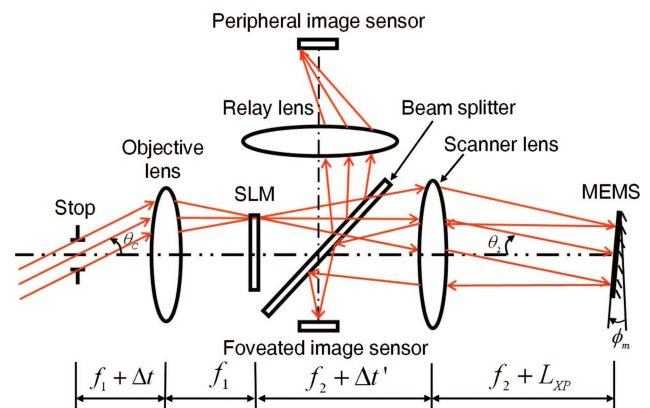

Fig. 6. (Color online) Schematic design of a dual-sensor foveated imaging system.

capture of HDR images by applying the mask approach [27]. A second SLM may replace the stop for active aperture control over the entire FOV.

The first-order parameters of the foveated imager path are annotated in Fig. 6. The effective focal lengths of the objective and scanner lenses are denoted as f_1 and f_2, respectively. The entrance pupil is offset by a distance of Δt from the front focal plane of the objective lens. The front focal point of the scanner lens is offset by a distance of $\Delta t'$ from the back focal point of the objective lens. All distances and angles utilize the well-accepted sign conventions [28].

When the mirror is scanned by an angle of ϕ_m from its reference orientation, the central angle of the visual fields instantaneously imaged by the foveated imager is given by

$$\theta_c = \arctan\left[\frac{-f_1 f_2 \tan(2\phi_m)}{f_1^2 + \Delta t \Delta t'}\right], \tag{7}$$

where θ_c corresponds to the scene point that is imaged onto the center of the foveated detector. The half FOV of the foveated imager, θ_r, is given by

$$\theta_r = \arctan\left[\frac{D_F(f_1^2 + \Delta t \Delta t')}{2f_1^2 f_2}\right], \tag{8}$$

where D_F is the diagonal dimension of the foveated detector. Further, $m_p = -f_1 f_2 / (f_1^2 + \Delta t \Delta t')$ is the magnification between the diameters of the mirror and the entrance pupil.

Providing that the MEMS mirror is capable of mechanically scanning within a range of $\pm\phi_{max}$ degrees, the maximum visual field covered by scanning the MEMS is given by

$$\theta_F|_{max} = 2 \arctan\left[\frac{f_1 f_2 \tan(2\phi_{max} + \theta_r)}{f_1^2 + \Delta t \Delta t'}\right]. \tag{9}$$

At a given visual field angle in the object space, the corresponding image height on the peripheral imager is given by

$$y_F = \frac{f_1^2 f_2 \tan(2\phi_m - \theta_2)}{f_1^2 - \Delta t \Delta t'}, \tag{10}$$

where a 1:1 relay from the intermediate image plane is assumed, and θ_2 is the chief ray angle exiting the scanner lens and is given as $\theta_2 = \arctan(-m_p \tan\theta)$.

The mirror is conjugate to the entrance pupil, and its location is characterized by the distance L_{XP} from the back focal point of the scanner lens,

$$L_{XP} = \frac{f_2^2 \Delta t}{f_1^2 + \Delta t \Delta t'}. \tag{11}$$

As the projection area of the mirror surface on the intermediate pupil plane varies with its scanning angle, to ensure that the effective F-number of the foveated imager is independent of the mirror scanning angles, the entrance pupil diameter, D_{EP}, should satisfy the condition of

$$D_{EP} \leq \frac{D_{MEMS}}{|m_p|}, \tag{12}$$

where D_{MEMS} is the effective diameter of the mirror device.

There are several advantages associated with the above schematic design. First of all, the design ensures that the foveated and peripheral imagers share the same entrance pupil. Satisfying the condition of single-viewpoint imaging allows an easy mosaic of the two images independent of the scene depth. Second, as shown in Eq. (9), with an appropriate choice of focal lengths for the objective and scanning lenses, the design can effectively relax the requirements for the mirror scanning angle. Fabricating a large-size, fast steering mirror with large tip-tilt capability is quite challenging with current technologies. Third, it offers flexible pupil controls and can be enhanced with adaptive HDR image capturing methods [27]. Finally, the design is compact and low-cost. The folding design of the scanner lens helps reduce the overall length of the foveated beam. The system is exempt from using high-resolution detectors and complex optics while achieving high resolving power across a relatively wide visual field.

5. Experimental Prototype and Results

A. Prototype Design

Based on the schematic layout in Fig. 6, we built a bench prototype, as shown in Fig. 7(a), using off-the-shelf optical components to demonstrate the potential system capability. In the prototype, we utilized two low-cost CCD sensors, one 1/4 in. full-color CCD (640×480 pixels) and one 1/4 in. B/W CCD (640×480 pixels), for foveated and peripheral imaging, respectively. Considering the factors of compactness, scanning range, and cost, we implemented the scanning element with a two-axis analog MEMS chip by Texas Instruments as shown in Fig. 7(b). The chip has an elliptical active area measured between 3.2 mm and 3.6 mm along its minor and major axes, respectively. Its mechanical scanning range is $\pm 5°$ around two orthogonal axes, and it has a mechanical resonant frequency of 130 Hz. The limited scanning range of the mirror requires $|m_p| > 1$ between the objective and scanner lenses to dynamically scan the foveated image across a wide peripheral visual field. The mirror dimensions along with the magnification requirements set up an upper limit to the entrance pupil diameter of the system [Eq. (12)].

An off-the-shelf 28 mm Erfle eyepiece assembly was utilized as the objective lens. It produces a low aberration intermediate image plane, where we can
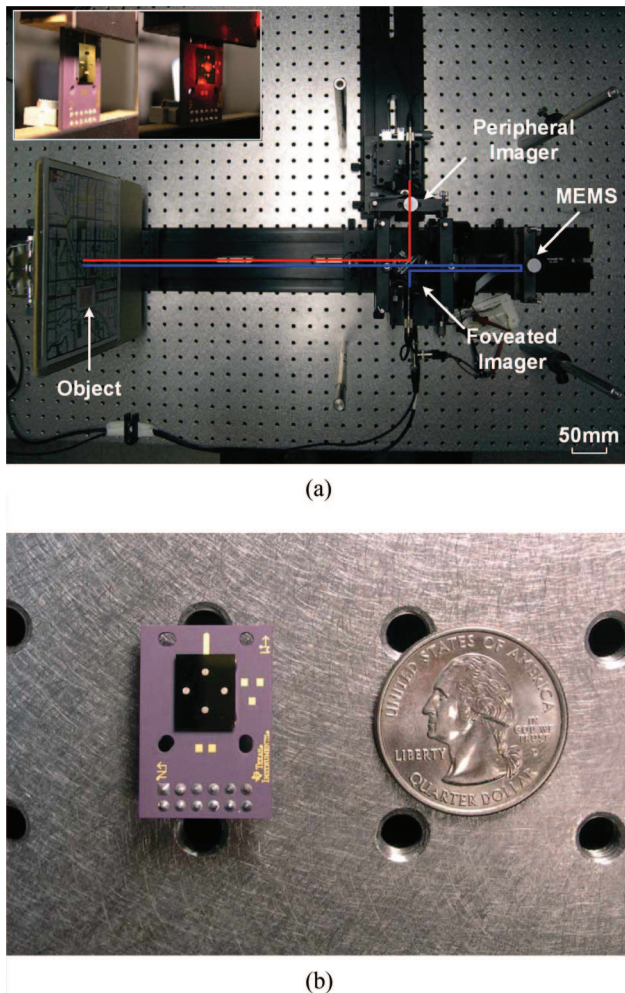
Fig. 7. (Color online) (a) Bench prototype of a dual-sensor foveated imaging system and the pupil conjugation on the MEMS chip (inset); (b) the MEMS chip.

potentially insert a SLM for pixel-level HDR image capturing. Another 28 mm lens was used together with the objective lens for the peripheral imaging path. As a result, the overall peripheral FOV is 44.6° diagonally with an angular resolution of 3.3 arc min per pixel at the best. The actual resolution of the peripheral imager may be further reduced by using lower-cost simpler optics.

Based on the plot in Fig. 2(c), a foveated imager with a total of 12° FOV will maximize the bandwidth saving ratio while maintaining imperceptible image degradation in the peripheral. A 60 mm doublet was used as the scanner and imaging lens. Its combination with objective lens yields a magnification power of about 2.1. The instantaneous FOV of the foveated imager is 8.5° diagonally with an angular resolution of 0.48 arc min per pixel at the best.

Combining with the optics above, the MEMS mirror is capable of sweeping the 8.5° foveated FOV across the full peripheral field by scanning the MEMS mirror within ±4°. Compared with a single-sensor design that maintains uniformly high resolution across the entire FOV, the bandwidth saving

ratio of the above dual-sensor design is 96.4%, which is higher than the optimal value in Fig. 2. However, due to the use of off-the-shelf components, the compromise is the perceptible image degradation in the peripheral when gazing at the center of the foveated image.

B. Experimental Results

Image registration is a crucial postprocessing step to create a seamless mosaic of the foveated and peripheral images as the MEMS scans the foveated FOV across the overall visual field. It requires deliberate calibration of the MEMS device and the distortion coefficients of the imaging systems. We calibrated the MEMS tilt angle responses versus the voltages applied to the x- and y-axes of the chip. In both directions, the tilt angles demonstrated excellent linearity for voltages within ±4.5 V. The distortion coefficient of the peripheral imager, independent of MEMS scanning angles, was calibrated by applying a well-established camera calibration technique [29]. Meanwhile, the distortion of the foveated imager depends on MEMS scanning and cannot be calibrated by a readily available method. We have developed a calibration technique that accurately models and calibrates the distortion of the foveated images as a function of MEMS scanning angles. By unwarping the distorted images captured at a given scanning angle, we are able to create a seamless foveated image [Fig. 8(a)] with a 44.6° diagonal FOV. The high-resolution foveated region with 8.5° FOV is shown in Fig. 8(b). In this example, a printed map of Tucson City was used as the scene object placed at the near range of the depth of field (~300 mm away from the entrance pupil).

The MEMS mirror was driven by applying voltages from −3.5 V to +3.5 V with 0.5 V increments in both orthogonal directions and scans a range from −4° to +4° with approximately 0.6° increments, which allowed us to steer the 8.5° foveated FOV across the entire 44.6° peripheral FOV for dynamic foveation. A sequence of high-resolution image samples was captured at different MEMS tilting angles. The mosaic of these foveated images with the peripheral background is shown in Fig. 9(a). Two enlarged views for a region near the center and the top-left corner are shown in Figs. 9(b) and 9(c), respectively. They dem-
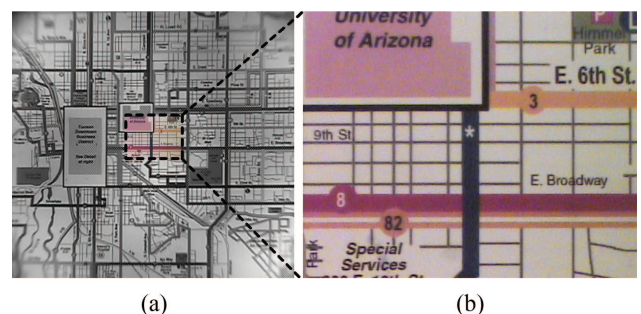

Fig. 8. (Color online) Registration and mosaic of (a) a peripheral image with (b) a foveated image.
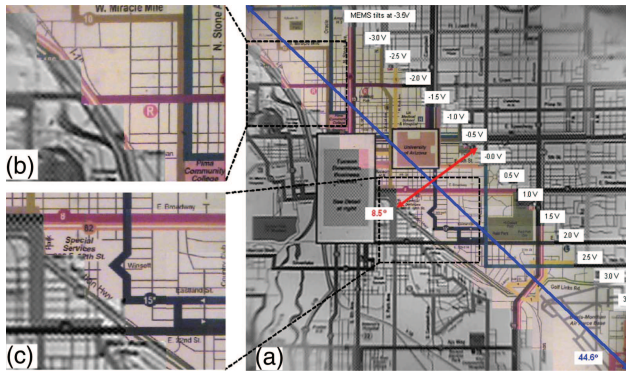
Fig. 9. (Color online) Mosaic of a sequence of foveated and peripheral images.

onstrated that the foveated foreground images were well aligned with the corresponding regions on the peripheral background with simple postprocessing steps.

C. Image Quality Characterization

We used the slanted edge target on an ISO12233 standard resolution chart to characterize the modulation transfer function (MTF) of the prototype. The Imatest software (http://www.imatest.com) was utilized for MTF analysis. At each of the MEMS scanning angles, the slanted edge on the resolution chart was positioned to the center of the foveated FOV, and the corresponding image was analyzed. The measured MTF performance of the bench prototype is shown in Fig. 10 as a function of the field angle θ_c of the foveated image. Here MTF50 measures the spatial frequency, where MTF drops to 0.5. In both tangential and sagittal directions, the MTF50 performances remain constant around 55 cycles/mm within $\pm 15°$ of the visual field. The degradation beyond that range is mainly due to the use of off-the-shelf lenses.
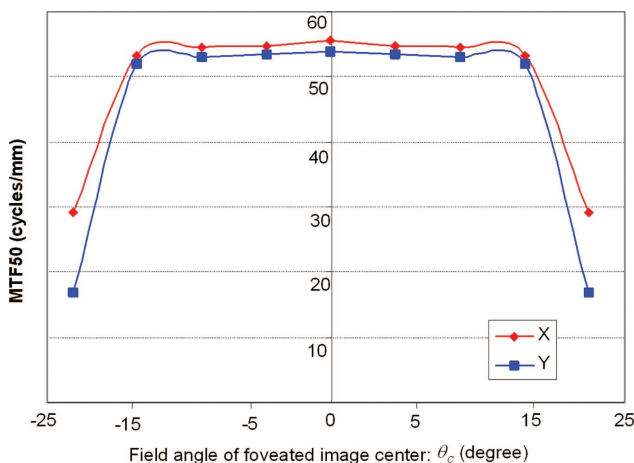

Fig. 10. (Color online) Image quality characterization: MTF50 versus the field angle of foveated image center.

6. Conclusion and Future Work

Acquiring high-resolution wide FOV and HDR images in real time is essential for many vision-based applications. The efficiency of the conventional rectilinear sampling schemes is as low as 3% for a 60° FOV system. In this paper, we presented an approach for the design of a foveated imaging system, aiming to maximize the sampling efficiency and information throughput of low-cost off-the-shelf image sensors, by exploring the spatially variant properties of the HVS. We demonstrated that a dual-sensor system with a 120° total FOV can achieve a 90.4% bandwidth reduction ratio, requiring ten times less pixels than a uniformly sampled single resolution system. The bandwidth reduction ratio can be further increased to 95.7% by applying spatial variances of contrast and color sensitivities, which directly translates to 23 times less communication bandwidth. We presented an analytical method to maximize the bandwidth saving of a multisensor foveated system. We further described the schematic design of a compact dual-sensor system and demonstrated the implementation of a bench prototype and experimental results. The prototype captures foveated images with 96.4% of bandwidth reduction compared to a single resolution system. In the future, we plan to custom-design an optimized system with an overall FOV of over 120° and optimum information throughput ratio based on the analysis in Section 3. We will further develop fovea tracking algorithms that dynamically detect the salient region of interest from the peripheral image and adaptively steer the fovea region toward a detected stimulus via the scanner unit. Such a capability can be particularly beneficial to applications such as surveillance and robot navigation.

References

1. G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, and C. S. Regazzoni, "Dual camera intelligent sensor for high definition 360 degrees surveillance," IEE Proc. Vision Image Signal Process. **152,** 250–257 (2005).
2. T. Ienaga, K. Matsunaga, K. Shidoji, K. Goshi, Y. Matsuki, and H. Nagata, "Stereoscopic video system with embedded high spatial resolution images using two channels for transmission," in *Proceedings of ACM Symposium on Virtual Reality Software and Technology* (2001), pp. 111–118.
3. A. Ude, C. Gaskett, and G. Cheng, "Foveated vision systems with two cameras per eye," in *Proceedings of IEEE International Conference on Robotics and Automation* (IEEE, 2006), pp. 3457–3462.
4. C. Wang, P. Shumyatsky, F. Zeng, M. Zevallos, and R. R. Alfano, "Computer-controlled optical scanning tile microscope," Appl. Opt. **45,** 1148–1152 (2006).
5. B. Potsaid, Y. Bellouard, and J. T. Wen, "Adaptive scanning optical microscope (ASOM): a multidisciplinary optical microscope design for large field of view and high resolution imaging," Opt. Express **13,** 6504–6518 (2005).
6. M. Böhme, M. Dorr, T. Martinetz, and E. Barth, "Gaze-contingent temporal filtering of video," in *Proceedings of ACM*

Symposium on Eye Tracking Research & Applications (2006), pp. 109–116.

7. W. Zhou and A. C. Bovik, "Embedded foveation image coding," IEEE Trans. Image Process. **10,** 1397–1410 (2001).

8. W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," Proc. SPIE **3299,** 294–305 (1998).

9. A. T. Duchowski and A. Çöltekin, "Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging," ACM Trans. Multimedia Comput. Commun. Appl. **3,** 1–21 (2007).

10. H. Murphy and A. T. Duchowski, "Hybrid image-/model-based gaze-contingent rendering," in *Proceedings of ACM Symposium on Applied Perception in Graphics and Visualization* (2007), pp. 1–8.

11. D. Luebke and B. Hallen, "Perceptually driven simplification for interactive rendering," in *Proceedings of the 2001 Eurographics Workshop on Rendering* (2001), pp. 223–234.

12. G. Sandini, P. Questa, D. Scheffer, and A. Mannucci, "A retina-like CMOS sensor and its applications," in *Proceedings of IEEE Workshop on Sensor Array and Multichannel Signal Processing* (IEEE, 2000), pp. 514–519.

13. J. P. Rolland, A. Yoshida, L. D. Davis, and J. H. Reif, "High-resolution inset head-mounted display," Appl. Opt. **37,** 4183–4193 (1998).

14. D. V. Wick, T. Martinez, S. R. Restaino, and B. R. Stone, "Foveated imaging demonstration," Opt. Express **10,** 60–65 (2002).

15. T. Martinez, D. V. Wick, and S. R. Restaino, "Foveated, wide field-of-view imaging system using a liquid crystal spatial light modulator," Opt. Express **8,** 555–560 (2001).

16. E. M. Reingold, L. C. Loschky, G. W. McConkie, and D. M. Stampe, "Gaze-contingent multiresolutional displays: an integrative review," Hum. Factors **45,** 307–328 (2003).

17. L. C. Loschky and G. S. Wolverton, "How late can you update gaze-contingent multiresolutional displays without detection?," ACM Trans. Multimedia Comput. Commun. Appl. **3** (2007).

18. W. S. Geisler, J. S. Perry, and J. Najemnik, "Visual search: the role of peripheral information measured using gaze-contingent displays," J. Vision **6,** 858–873 (2006).

19. L. C. Loschky, G. W. McConkie, H. Yang, and M. E. Miller, "The limits of visual resolution in natural scene viewing," Visual Cognition **12,** 1057–1092 (2005).

20. M. Dorr, M. Böhme, T. Martinetz, and E. Barth, "Visibility of temporal blur on a gaze-contingent display," in *Proceedings of the ACM Symposium on Applied Perception, Graphics & Visualization* (2005), pp. 33–36.

21. D. J. Parkhurst and E. Niebur, "Variable-resolution displays: a theoretical, practical, and behavioural evaluation," Hum. Factors **44,** 611–629 (2002).

22. L. C. Loschky and G. W. McConkie, "Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display," Q. J. Exp. Psychol. A **8,** 99–117 (2002).

23. R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Z. Zhang, "A foveated silicon retina for two-dimensional tracking," IEEE Trans. Circuits Syst. II **47,** 504–517 (2000).

24. G. Godin, P. Massicotte, and L. Borgeat, "High-resolution insets in projector-based stereoscopic displays: principles and techniques," Proc. SPIE **6055,** 60550F (2006).

25. S. J. D. Prince, J. H. Elder, Y. Hou, and M. Sizinstev, "Pre-attentive face detection for foveated wide-field surveillance," in *IEEE Workshops on Application of Computer Vision* (IEEE, 2005), pp. 439–446.

26. Peter G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality* (SPIE Optical Engineering Press, 1999).

27. C. Gao, N. Ahuja, and H. Hua, "Active aperture control and sensor modulation for flexible imaging," in *Proceedings of International Conference on Computer Vision and Pattern Recognition* (2007).

28. J. E. Greivenkamp, *Field Guide to Geometrical Optics* (SPIE Press, 2004).

29. Camera calibration toolbox, http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.