Optimization algorithm of near-eye light field displays based on human visual characteristics

Jun Ding (丁 俊)**, Mali Liu** (刘玛丽)**, Qing Zhong** (钟 擎)**, Haifeng Li** (李海峰)***,
and Xu Liu** (刘 旭)

*State Key Lab. of Modern Optical Instrumentation, Department of Optical Engineering,
Zhejiang University, Hangzhou 310027, China*
Corresponding author: lihaifeng@zju.edu.cn

The number of layers and the resolution of liquid crystal displays (LCDs) limit the reconstruction fidelity of near-eye light field displays based on multilayer LCDs. Because the eye's resolution capability is different for central vision and peripheral vision, the fidelity can be improved by setting different weights for different areas. First we employ the eye's modulation transfer function (MTF) to acquire the limiting resolution angle. Then, due to the inverse relationship between the limiting angle and the weight values, the weighted function related to retinal eccentricity is calculated. In combination with the linear least-squares algorithm, the peak signal-to-noise ratio (PSNR) of the reconstructed scene is raised. The simulation results indicate that the weighted optimization algorithm can improve the image fidelity and reconstruction accuracy.

OCIS codes: 110.1758, 110.3010, 110.6955.
doi: 10.3788/COL201614.041101.

With the rapid development of the microdisplay field, near-eye light field display devices that provide abundant immersive feelings and visual experience are attracting widespread attention. Compared with the traditional head-mounted near-eye displays, which utilize several complex optical elements such as free-form surface components[1] and either refractive or diffractive elements[2], the multilayer near-eye displays benefit from simpler structures and lower cost.

By decomposing the 4D light field into the tensor product of two liquid crystal display (LCD) masks with the non-negative matrix factorization (NMF) algorithm, Lanman *et al.*[3] optimized an automultiscopic scene with binocular parallax cues. Wetzstein *et al.*[4] adopted tomography techniques to recreate a 3D car by using compact volumes of spatial light modulators, and such attenuation-based displays allowed an accurate depiction of motion parallax. Building on multilayer attenuator and directional backlight architectures, Maimone *et al.*[5] proposed a 3D display design that has the potential to support nearly correct accommodative depth cues. In addition, Maimone *et al.*[6] developed an optical see-through glass composed of a set of LCD panels to provide a 65" diagonal field of view (FOV) and multiple simultaneous focal depths.

In the multilayer display domain, all pixels in the LCD layers are controlled to represent the huge number of 4D light rays. When the number of LCD layers and the spatial resolution are limited, one single pixel unit should support dozens of light rays. To some extent, the average load level per pixel will be heavy and the reconstruction fidelity will be deteriorated. In order to improve the reconstruction results, we proposed a weighted function varying with the retinal eccentricity. That innovation is based on the human visual system, which can make the imaging clear within the central 2° to 4° of the visual field[7] and cause a sharp declining image quality out of the fovea. The weighted function is deduced by the inverse relationship between the weight values and the limiting resolution angle, which is calculated by the modulation transfer function (MTF) curves at different retinal eccentricities. The value of the weighted function represents the perceptual importance of rays in the target light field. We optimize attenuation images in combination with the weighted function and the linear least-squares (LSQLIN) algorithm. The experimental results show an increased peak signal-to-noise ratio (PSNR) of the reconstructed central vision, a better perceptual imagery, and enhanced information utilization. The Letter contributes a novel method for improving the near-eye light field reconstructed performance.

As shown in Fig. 1, a multilayer display device consists of dual-stacked LCDs and a uniform backlight whose
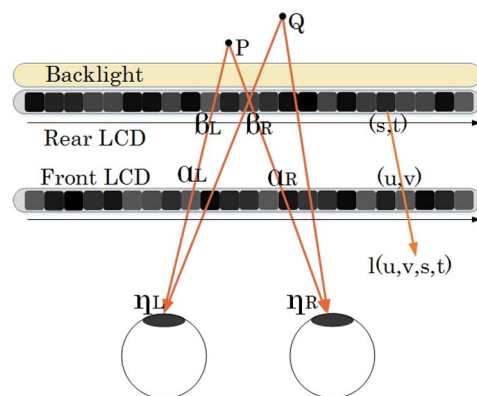

Fig. 1. Configuration of the near-eye light field displays.

intensity is l_{BG}. Based on the light field rendering principle[8], the structure can reconstruct a 3D object by reproducing a 4D light field of the object. For simplicity, only two light rays, η_L and η_R, emitted by one object point P, are discussed. Supposing ray η_L intersects two panels at α_L and β_L, and ray η_R at α_R and β_R. Then the two rays can be reproduced by controlling the transmittance such that

$$\begin{cases} l(\eta_L) = l_{\mathrm{BG}} \cdot f(\alpha_L) \cdot g(\beta_L), \\ l(\eta_R) = l_{\mathrm{BG}} \cdot f(\alpha_R) \cdot g(\beta_R), \end{cases} \tag{1}$$

where $f(\alpha_L)$ and $f(\alpha_R)$ are the transmittance of the front layer at pixel α_L and α_R, $g(\beta_L)$ and $g(\beta_R)$ are the transmittance of the rear layer at pixel β_L and β_R, and $l(\eta_L)$ and $l(\eta_R)$ denote the intensity of the two rays. With this multiplicative attenuation rule, the target point P can be reconstructed. Afterward, point Q in different depths can be rendered in the same way. In the end, a 3D object can be reproduced by computing the optimal transmittance patterns of LCDs.

Furthermore, the two-plane parameterization method[3] is adopted in the analysis. As depicted in Fig. 1, an arbitrary ray is parameterized by the coordinates of its intersections with two LCD panels. Thus, (u, v, s, t) expresses the ray intersecting the front LCD at pixel (u, v) and the rear LCD at pixel (s, t). When the intensity of light field is normalized, Eq. (1) can be re-expressed as

$$l(u, v, s, t) = f(u, v) \cdot g(s, t), \tag{2}$$

where $f(u, v)$ and $g(s, t)$ denote the transmittance of pixels (u, v) and (s, t), respectively, and $l(u, v, s, t)$ represents the normalized intensity.

We can also express the light ray as the summation of two pixels in the logarithm field:

$$\ln(l(u, v, s, t)) = \ln(f(u, v)) + \ln(g(s, t)). \tag{3}$$

The algorithm flow chart for near-eye light field displays is shown in Fig. 2. First we acquire the target light field by perspective projections at different viewpoints. For instance, by setting 3 viewpoints along the horizontal and vertical direction, respectively, 9 patterns are obtained [as depicted in Fig. 2(a)]. Then the two transmittance patterns for the LCD layers corresponding to the target light field mentioned before should be computed to reconstruct the 4D light field. Our aim is to minimize the difference between the emitted and target light field. By casting the optimization method of transmittance patterns as a constrained LSQLIN, we calculate the
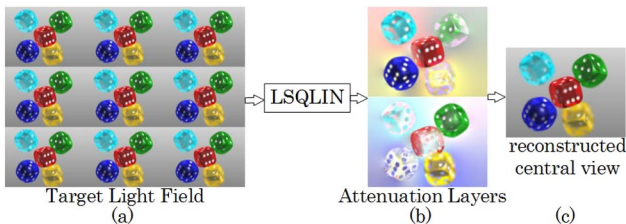
two attenuation layers [as shown in Fig. 2(b)] by the LSQLIN algorithm to reconstruct the light field [e.g., the reconstructed central view in Fig. 2(c)].

The LSQLIN algorithm is discussed clearly in this section. As formulated in Eq. (3), one light ray corresponds to one equation. When all 4D light rays are taken into account, a series of equations can be obtained. This problem can be expressed in the matrix equations

$$\underbrace{\begin{bmatrix} T_{11} & \cdots & T_{1n} & T_{1(n+1)} & \cdots & T_{1(2n)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ T_{m1} & \cdots & T_{mn} & T_{m(n+1)} & \cdots & T_{m(2n)} \end{bmatrix}}_{\text{Transform} \cdot \text{matrix} \cdot \mathbf{T}} \cdot \begin{bmatrix} \ln(f_1) \\ \vdots \\ \ln(f_n) \\ \ln(g_1) \\ \vdots \\ \ln(g_n) \end{bmatrix}$$

$$= \begin{bmatrix} \ln(l_1) \\ \vdots \\ \ln(l_m) \end{bmatrix}, \tag{4}$$

where $[\ln(f_1) \cdots \ln(f_n)]^T$ and $[\ln(g_1) \cdots \ln(g_n)]^T$ represent the log value of the transmittance of the front layer and the rear layer, respectively, and n is the total number of pixels in each panel. $[\ln(l_1) \cdots \ln(l_m)]^T$ is the log value of the normalized intensity of the target light field and m is the total number of light rays. The transform matrix $\mathbf{T}$ is utilized to distinguish which two pixels are selected in the transmittance patterns. For example, if ray l_1 intersects two panels at f_1 and g_1, then T_{11} and $T_{1(n+1)}$ are set to 1, and the rest of the terms are equal to 0 at that row (namely, $[T_{11} \cdots T_{1n} T_{1(n+1)} \cdots T_{1(2n)}]$). At last, we can acquire a formula $\ln(f_1) + \ln(g_1) = \ln(l_1)$ that is similar to the Eq. (3).

Here, the matrix $\mathbf{T}$ is discussed in detail. As depicted in Fig. 3(a), several viewpoints are set up and the target light field can be sampled at these viewpoints. As shown in Fig. 3(b), if the target light field and the dual-layer pixels are vectorized in the same way, namely, rearranging elements from up to down and column by column, then the
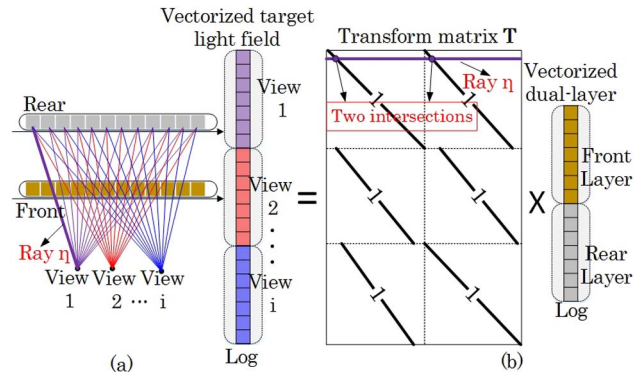

Fig. 2. Algorithm pipeline of the near-eye light field displays.

Fig. 3. Visualization of transform matrix $\mathbf{T}$.

light rays and the LCD pixels can be connected by the transform matrix **T**. When the light rays are traced back to the LCD pixels, the terms in the matrix **T** corresponding to the intersection pixels are set to 1 and the rest of the terms are 0. For example, a light ray η is selected in Fig. 3(a), then its index in the vectorized light field decides which row in the matrix **T** should be processed [shown in Fig. 3(b)], and its intersections in dual-layer LCDs decides which two terms are set to 1 at that selected row in the matrix. In conclusion, matrix **T** is a binary-valued sparse matrix to associate the light field with panel pixels.

However, Eq. (4) indicates that the number of equations is much larger than the number of variables. In other words, the amount of light rays is larger than the amount of pixels. For example, by employing two LCD layers whose resolution is 800×600, if 3×3 perspective views should be optimized, then the number of equations is equivalent to $800 \times 600 \times 3 \times 3$, and number of variables is equal to $800 \times 600 \times 2$. Hence, the LSQLIN solver is utilized to solve these linear overdetermined equations. Because the matrix **T** is sparse and large scale, the MATLAB LSQLIN solver[9] is supplied to make the equations convergent in about 8 to 14 iterations for multilayer LCDs. In the end, the Euclidean distance between the emitted and target light fields can be minimized as

$$\arg \min \|\boldsymbol{T} \cdot \boldsymbol{x} - \boldsymbol{l}\|^2, \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{l}$ denote the transmittance vector and the target light field vector, respectively.

Based on the previous research done by Navarrod *et al.*[10], the point spread function of the human pupil in different fields is detected by a conjugate CCD. By employing a collimated He-Ne laser beam to pass through a hole in a perimeter in transit to the eye, the average MTF for different eccentricities is fitted as

$$\begin{aligned} \text{MTF}(\theta, f) = {} & (1 - C_1 + C_2\theta) \exp[-A_1 \exp(A_2\theta)f] \\ & + (C_1 - C_2\theta) \exp[-B_1 \exp(B_2\theta)f], \end{aligned} \tag{6}$$

where the fitting coefficients can be expressed as

$$A_1 = 0.1743, \quad A_2 = 0.0392, \quad B_1 = 0.0362,$$
$$B_2 = 0.0172, \quad C_1 = 0.215, \quad C_2 = 0.00294, \tag{7}$$

where f is the spatial frequency in cycles per degree and θ is the retinal eccentricity in degrees.

Based on Eqs. (6) and (7), six sets of MTF curves are plotted on a logarithm scale in Fig. 4. Since the contrast threshold of the human visual system is approximately equal to 0.03[11], the cutoff spatial frequency f_{cut} for different eccentricities can be worked out by solving the abscissa values of the intersection of threshold line M = 0.03 and the arbitrary MTF curve. It can be expressed as

$$\text{MTF}(\theta, f_{\text{cut}}) = \text{M}. \tag{8}$$

On account of the complexity of MTF formula, the cutoff frequency is calculated by the Newton iterative method when giving the eccentricity θ (in numerical terms).
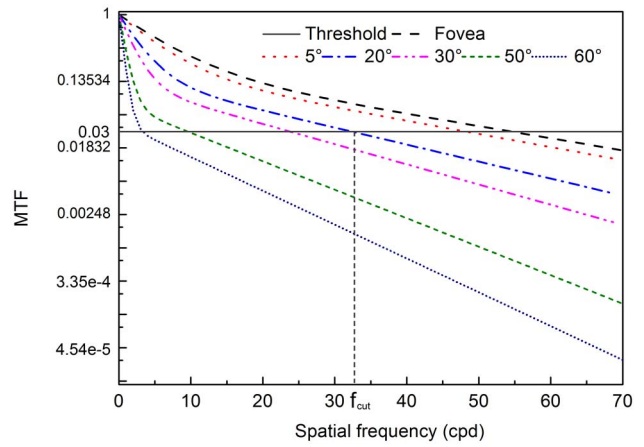

Fig. 4. MTF profiles in different retinal eccentricities.

Therefore, the limiting angle of resolution can be expressed as

$$R(\theta) = 60/f_{\text{cut}}, \tag{9}$$

where $R(\theta)$ is the limiting angle in arc min. Table 1 shows the resulting angle for different FOVs.

As listed in Table 1, the limiting angle is 1.1026 arc min in the central fovea by our method, which agrees well with the pupil's standard resolution (defined as 1 arc min). As is known to all, the limiting angle $R(\theta)$ denotes the minimum field angle for the sake of differentiating adjacent light rays. When the angle decreases, the importance of the light ray increases in this area. On the contrary, an increased angle leads to the reduction of the light ray's importance. Due to the inverse relationship between the limiting angle and the ray's importance, a reasonable weight model is established by Eq. (9), as follows:

$$W(\theta) = R(0)/R(\theta). \tag{10}$$

The purpose of setting the numerator as $R(0)$ is to normalize the weighted function $W(\theta)$.

As shown in Fig. 5, we can plot the scatter distribution of $W(\theta)$ along the FOV θ in terms of Eq. (10). Furthermore, by employing the polynomial fitting method, the result of the weighted curves can be obtained as

$$W(\theta) = p_1 \cdot \theta^2 + p_2 \cdot \theta + p_3,$$
$$p_1 = 0.00013, \quad p_2 = -0.023, \quad p_3 = 0.9994, \tag{11}$$

where p_1, p_2, and p_3 are polynomial coefficients. The root mean square error (RMSE) is equal to 0.01194 after curve

Table 1. Pupil Resolution in Different FOVs

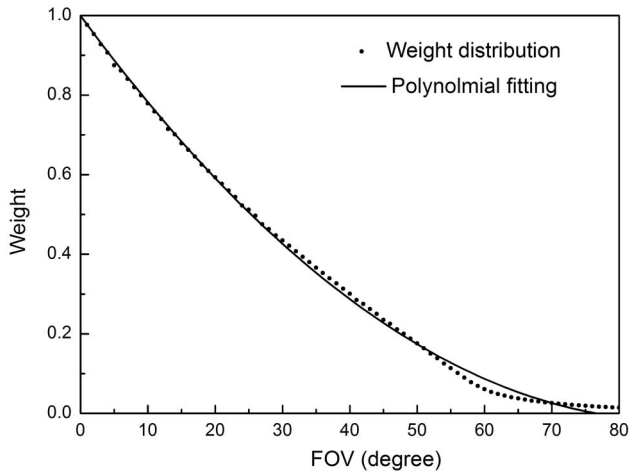| θ	R	θ	R
Fovea	1.1026	30°	2.5371
5°	1.2607	40°	3.6703
10°	1.4158	50°	6.2728
20°	1.8587	60°	18.2725

Fig. 5. Weight distribution in different FOVs.

fitting is completed. Eq. (11) is the final weight model relative to FOV θ.

A reconstructed light field solution has been explored that minimizes the Euclidean distance between the emitted light field and the target light field, synthesizing with the provided weighted matrix and the LSQLIN algorithm. The weight of each ray indicates the perceptual importance, which means that a higher weight corresponds to a more stringent constraint for the ray; on the contrary, a lower weight represents a slacker constraint for the ray. Taking a combination of Eqs. (5) and (11), the Euclidean distance between the emitted light field and the reconstructed light field can be re-expressed by:

$$\arg\min\{\| W\boldsymbol{T} \cdot \boldsymbol{x} - \boldsymbol{l} \|_2^2\} = (\sqrt{w_1} \cdot T_1\boldsymbol{x} - \sqrt{w_1}l_1)^2 + \cdots \\ + (\sqrt{w_m} \cdot T_m\boldsymbol{x} - \sqrt{w_m}l_m)^2. \tag{12}$$

In this section, a series of suitable experimental parameters were provided to simulate the observation results for near-eye displays. Weight optimization was compared with Wetzstein's optimization[4].

Two LCD layers with a resolution of 1280×800 and a pixel size of 117 μm were used for the simulation. Meanwhile, inspired by the parameter design of head-mounted display devices, the distance between the human eye and the front LCD is set to 100 mm, and the layer separation distance is equal to 8 mm. A set of 5×5 viewpoints are created within the area of the pupil, for simplicity, and the pupil size is approximately equal to 4 mm for a normal situation.

By utilizing 3ds Max software, a scene containing a green circular ring and a yellow teapot was set up. Two transmittance patterns for LCD layers were calculated by Wetzstein's optimization and the weight optimization, respectively. Afterward, we could use OpenGL to read the patterns' data into the frame buffer object. With the blend function (glBlendFunc), the observation results were simulated in analogy with the human eye seeing through two transmittance patterns.
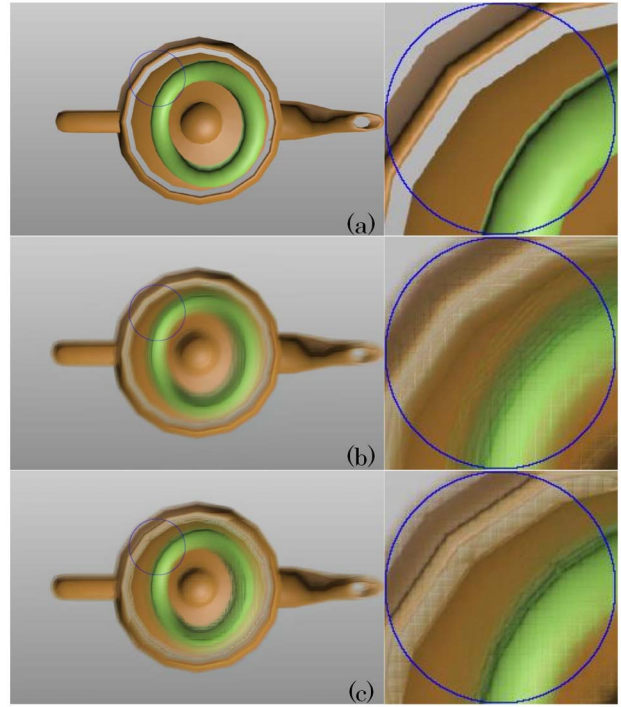

Fig. 6. Reconstructed results by different algorithms: (a) original light field, (b) Wetzstein's optimization, and (c) weight optimization.

As is shown in Fig. 6(a), a perspective view is captured in the target light field. By Wetzstein's optimization, we can obtain the reconstruction performance at the same viewpoint simulated by OpenGL [as shown in Fig. 6(b)]. Similarly, the results can be acquired by weight optimization as well [as depicted in Fig. 6(c)]. Here, the blue circle represents the central vision for this situation. The central region is magnified in the right column, for comparison.

Figure 7 shows a quantitative measure of the central region for the two optimization results, which is figured in the PSNR.
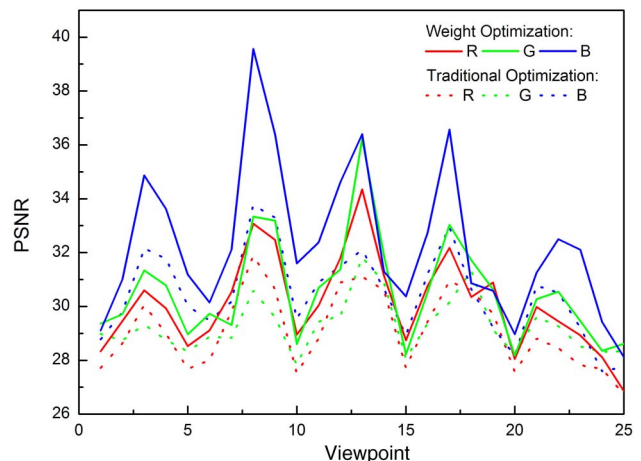

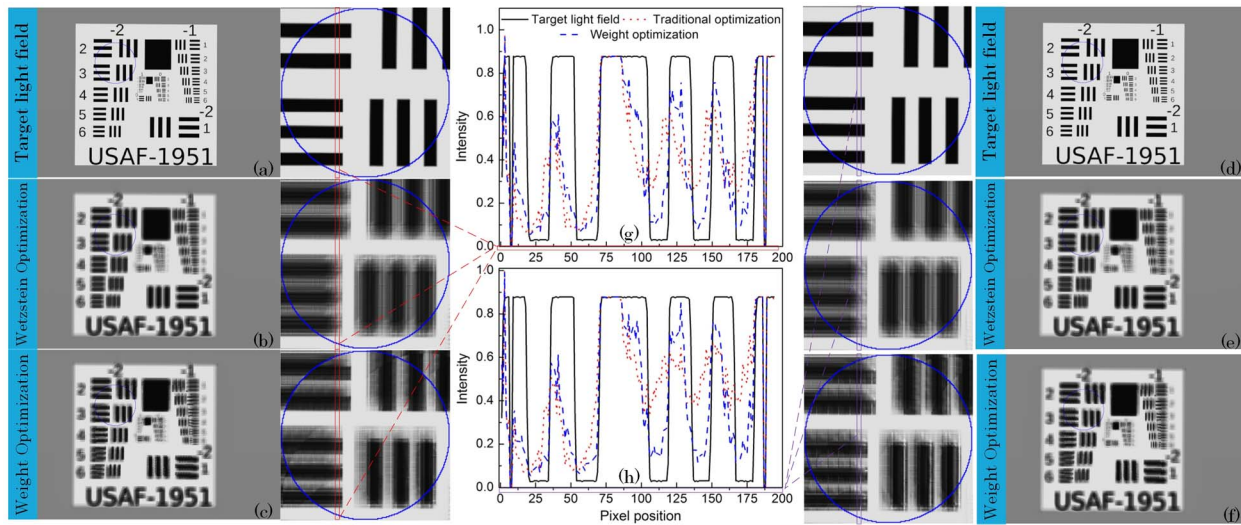Fig. 7. PSNR based on Wetzstein's and the weight optimization.

Fig. 8. Intensity comparison with line patterns. (a) and (d) denote the target light field in different perspective views, (b) and (e) are the construction results by Wetzstein's optimization, (c) and (f) are the construction results by weight optimization, and (g) and (h) are the pixel intensity at the corresponding 1D slices.

As shown in Fig. 7, the solid curves represent weight optimization results and the dotted curves correspond to Wetzstein's methods. It is easy to draw the conclusion that the PSNR values by weight method are higher than the ones by Wetzstein's method, for every channel. Experimental results agree well with our hypothesis that weight optimization supports a better observation performance.

What is more, the line patterns were tested for a more intuitive comparison. As depicted in Fig. 8, the 1951 USAF resolution test chart was set up as the target scene. Figure 8(a) denotes one perspective view in the target light field, Figs. 8(b) and 8(c) are the reconstruction results corresponding to the two optimization methods, respectively. Similarly, Figs. 8(d), 8(e), and 8(f) represent another perspective view. The second and fourth columns are the magnified central regions. When we select a 1D slice in the central regions, the reconstruction intensity can be obtained [namely, Figs. 8(g) and 8(h)) at that slice by different optimization methods. As shown in Figs. 8(g) and 8(h), the x axis corresponds to the pixel position at the slice and the y axis represents the pixel intensity. Note that the result of weight optimization (blue dashed line) more closely matches the target light field (black solid line) than that of Wetzstein's optimization (red dot line).

In conclusion, depending on human vision characteristics, we develop an optimization algorithm enhancing the performance of the near-eye light field displays based on multilayer LCDs. The limiting angle of resolution under different FOVs can be derived by analyzing the corresponding MTF curves; simultaneously, the weight curve is fitted by taking advantage of the inverse relationship between the weight values and the limiting angle of resolution. The 4D light field can be reconstructed in the least-squares sense. The analysis of the weight optimization for the near-eye light field displays indicates a better

reconstructed light field in which the PSNR of the central vision is improved and accordingly better information utilization.

In the future, we would like to explore the eye tracking technique to scan the whole light field by the eye's central vision with high resolution.

References

1. D. Cheng, Q. Wang, Y. Wang, and G. Jin, Chin. Opt. Lett. **11**, 031201 (2013).
2. W. Song, Y. Wang, D. Cheng, and Y. Liu, Chin. Opt. Lett. **12**, 060010 (2014).
3. D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, ACM Trans. Graph. **29**, 163 (2010).
4. G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, ACM Trans. Graph. **30**, 95 (2011).
5. A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs, ACM Trans. Graph. **32**, 153 (2013).
6. A. Maimone and H. Fuchs, in *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (IEEE, 2013), pp. 29.
7. A. Berencsi, M. Ishihara, and K. Imanaka, Hum. Mov. Sci. **24**, 689 (2005).
8. M. Levoy and P. Hanrahan, in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (1996), pp. 31.
9. D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, ACM Trans. Graph. **30**, 186 (2011).
10. R. Navarro, D. R. Williams, and P. Artal, J. Opt. Soc. Am. A **10**, 201 (1993).
11. D. Yu and H. Tan, *Engineering Optics* (China Machine Press, 2011), pp. 210–212.