Light Filed Render and Optimization for Measurable 3D Depth Perception Interaction

ABSTRACT - The main benefit of 3D display over 2D display, beyond the obvious ability to create a more lifelike character with perception of depth, is that different depth-scenes of sight presentation on 3D display device can be controlled by adjusting the disparity structure of multiview content for a realistic visual experience [1]. However, most of the existing approaches are often reproducing a sub-optimal multiview content for inferior 3D display productions. Lei et al. [2] shifted each view image respectively by computing disparity separately, but the errors of disparity acquisition make it extremely hard to maintain a smooth motion parallax. Recently, a dense light filed controlling method [3] solved the rough-shifting disparity problem by adjusting the slope of linear structures in epi-polar plane image (EPI). However, the view images in the each re-building dense light filed are usually severely twisted due to the discrete rendering for each EPI, and cannot be measurable. As a result, no turnkey solution exists for reproducing high quality 3D light filed for accurate depth perception interaction. In this paper, through minimizing a global spatially regularized energy functional in a novel non-convex optimization framework, we introduce a novel approach to render and optimization light filed playing up to users' depth perception requirements through a voice interaction system, based on a RGBD inputting image, aiming at controlling measurable visual presentations with fidelity 3D scenes.

KEYWORDS - *measurable depth perception interaction; light filed rendering; epi-polar plane image; non-convex optimization*

MAJOR TECHNICAL DESCRIPTION - schematic of the technique is shown in Figure 1 below. A 3D light filed, created from a set of multiview images, is presented as $\boldsymbol{LF(y,x,V)}$. $\boldsymbol{V}$ denotes the discrete number of views and $\boldsymbol{V_{RGB}}$ represents the reference image $\boldsymbol{I_{RGB}}$'s sequence view number. A planar $\boldsymbol{x\text{–}V}$ cut represents epi-polar plane image (EPI). $\boldsymbol{l(y',x',V_{RGB})}$, which passes through the point $\boldsymbol{p(y',x',V_{RGB})}$ on $\boldsymbol{I_{RGB}}$, denotes the linear structure in EPI. And the slopes $\boldsymbol{k}$ of all the linear structure is calculated as: $\boldsymbol{k} = f \cdot baseline/z - k'$, where $baseline$ is the baseline of adjacent cameras (in Figure 2(a)), f denotes the camera focal length and z represents the depth for $\boldsymbol{I_{RGB}}$'s each pixel. k' is acquired according to the users' depth perception requirements. And the depth scenes finally can be accurately and measurably obtained (in Figure 2(b)):

$$\text{depth scenes} = \left(1 - \frac{d_e}{d_e + S\cdot k}\right) D_{e\_s} \tag{1}$$

When $\boldsymbol{k}$ changes from negatives to positives, the depth scenes varies from "into the display screen" to "out of the display screen" (quantitative values for our display system are shown in Figure 3(a)). By propagating the color of $\boldsymbol{p}$ to all the points on $\boldsymbol{l(y',x',V_{RGB})}$, the initial light filed $\widehat{\boldsymbol{LF}}$ is rendered. During $\widehat{\boldsymbol{LF}}$ created time, when each EPI of the initial light filed rendered through the direction in Figure 3(b), we record the last $\boldsymbol{k}$, current $\boldsymbol{k}$, last rendered pixel and current rendered pixel respectively to determine the holes' fill directions and pixels' cover strategies.

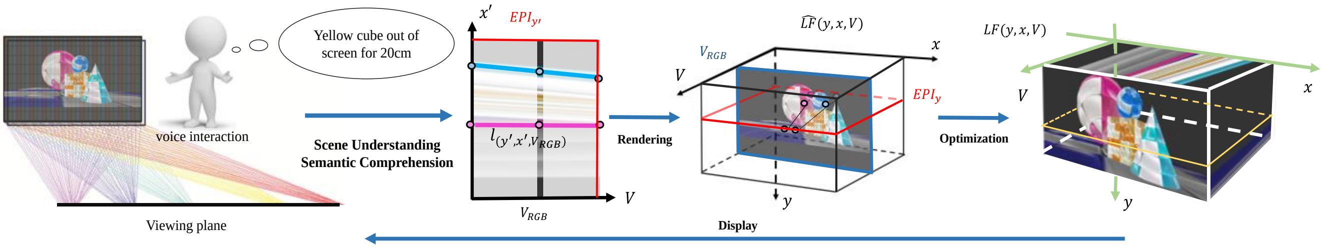

Fig 1. The schematic of the technique.

We use a regulariser comprising a weighted Huber norm over the gradient of the final light filed $\boldsymbol{LF}$, $\omega\left\|\nabla_{y,x}LF(y,x,V)\right\|_{\varepsilon}$, which ensures un-twisted view images in $\boldsymbol{LF}$ for high quality 3D display, where $\omega \propto \nabla \boldsymbol{I_{RGB}}(y',x')$ $((y',x') \in \boldsymbol{l(y',x',V_{RGB})}$ and $(y,x) \in \boldsymbol{l(y',x',V_{RGB})})$. The resulting energy functional therefore contains a non-convex photometric error data term and a convex regulariser:

$$min_{LF}E = \iiint \left\{ \left(\boldsymbol{LF(y,x,V)} - \widehat{\boldsymbol{LF}}\boldsymbol{(y,x,V)}\right)^2 + \omega\left\|\nabla_{y,x}\boldsymbol{LF(y,x,V)}\right\|_{\varepsilon} \right\} dydxdV \tag{2}$$

, which can be solved by a similar scheme in [4].

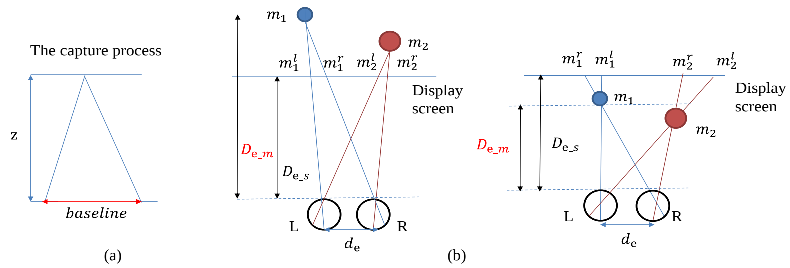

Fig 2. (a). The capture process when acquired the RGBD images; (b). Depth scenes' calculation: D_{e_s} and D_{e_m} represent the distances between viewer's eyes to screen and to displayed object m_1, respectively. And S is the display scale between the display image size and the original view image size.

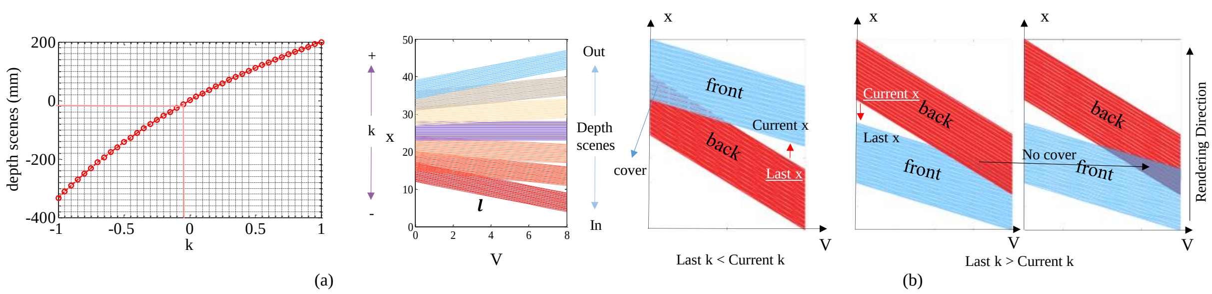

Fig 3. (a). The relationship between $\boldsymbol{k}$ and depth scenes; (b). Holes' fill directions and pixels' cover strategies.