第 43 卷 第 19 期/2023 年 10 月/光学学报 研究论文

光学学报

# 基于聚焦堆栈视差维超分辨的视差估计方法

王昱凯，刘畅*，邱钧

北京信息科技大学应用数学研究所，北京 100101

**摘要** 聚焦堆栈数据受到视差维分辨率低的限制，导致由聚焦堆栈数据估计的视差的精度低、鲁棒性差。从聚焦堆栈数据的视差维频谱优化出发，引入聚焦堆栈视差维滤波器，提出基于视差维滤波的聚焦堆栈视差维超分辨方法，实现高精度的稠密视差估计。通过聚焦堆栈的频谱分析，选取巴特沃斯滤波作为视差维滤波器，实现聚焦堆栈数据高保真的视差维超分辨。利用视差维超分辨后的稠密聚焦堆栈，基于 Robust focus volume regularization（RFV）算法实现稠密、高精度视差估计。模拟数据与实际数据实验结果表明：视差维滤波能够实现高效的视差维超分辨和高精度的稠密视差估计。

**关键词** 聚焦堆栈；视差维滤波；视差维超分辨；视差估计

**中图分类号** TP391 **文献标志码** A **DOI**：10.3788/AOS230727

## 1 引 言

聚焦堆栈是在不同成像平面聚焦成像形成的图像序列，可以看作是光场数据[1]的投影域表达形式，可用于视差估计[2-8]、光场重建[9-14]和扩展景深成像[15-17]等计算成像任务。获取聚焦堆栈数据的方式可分为两类：一类由光场数据进行数字重聚焦[18-20]获取；另一类是设置成像面的序列由成像系统直接获取。具有稠密视差维采样的聚焦堆栈数据可提高视差重建能力，支撑高精度的光场重建、三维显示和显微成像等相关应用。

由光场数据可计算出任意成像平面的重聚焦图像，进而获取稠密视差维采样的聚焦堆栈数据。Ng[21]利用四维（4D）光场数据在不同的视差层进行合成，实现数字重聚焦，并在频域中对 4D 光场频谱进行切片[22]，得到数字重聚焦的频域方法。Arroyo[23]对传统相机的成像过程进行建模，并对聚焦堆栈图像的散焦区域进行拟合，得到图像散焦的变化规律，据此可模拟不同视差对应的聚焦图像。Le 等[24]对 4D 光场频谱进行解构，得到傅里叶视差层（FDL），并推导其与聚焦堆栈的空间维频谱之间的关系，用于合成各视差层的聚焦图像。Pérez 等[25]提出了一种基于中心切片定理的超分辨方法，获取了更多的光场频谱切片，然后通过傅里叶逆变换，还原出更多的聚焦堆栈图像。虽然由光场数据可获取任意成像平面的重聚焦图像，但光场数据量巨大，这为数据采集、存储和传输带来极大挑战。

由成像系统采集聚焦堆栈数据，需要对成像平面进行精密控制，聚焦堆栈视差维的采样率较低，从而限制视差估计和光场重建等聚焦堆栈数据计算成像的应用[26-28]。对聚焦堆栈数据进行视差维优化[29-33]和超分辨能够提升聚焦堆栈对场景的刻画能力，这对提升聚焦堆栈计算成像性能具有重要的实际意义。Xiao 等[34]2021 年提出了一种基于深度学习的方法，由极平面图像（EPI）推导新视图以提升聚焦堆栈的重建效果，并获取图像数量更多的聚焦堆栈。Xiao 等[35]在 2022 年提出了一种 COLF-GAN 网络，对一组稀疏的聚焦堆栈进行训练，并利用训练好的模型实现视差维超分辨。

本文从聚焦堆栈数据的视差维优化出发，通过分析聚焦堆栈频谱选取视差维滤波器，实现聚焦堆栈的高保真视差维超分辨。利用视差维超分辨后的稠密聚焦堆栈，基于聚焦测度算子实现稠密、高质量的视差估计。基于视差维滤波的聚焦堆栈超分辨方法降低了聚焦堆栈数据的采集成本，实现了稠密视差估计，同时为实现更高质量的光场重建和三维显示等计算成像任务提供数据支撑。

## 2 基于聚焦堆栈视差维超分辨的视差估计方法

### 2.1 聚焦堆栈的频谱特性

聚焦堆栈数据可表达为光场数据在不同视差层上的投影，可表示为

**收稿日期**：2023-03-29；**修回日期**：2023-04-22；**录用日期**：2023-04-24；**网络首发日期**：2023-05-08

**基金项目**：国家自然科学基金（62171044，61931003）、北京市自然科学基金（4222004）、北京信息科技大学“勤信人才”培育计划（QXTCPB202105）

**通信作者**：*liu. chang. cn@ieee. org

$$F(s,x,y)=\iint L\left[u,v,x+s(u-u_0),y+s(v-v_0)\right]\mathrm{d}u\mathrm{d}v,\tag{1}$$

式中：$F(s,x,y)$为聚焦堆栈；$s$为视差；$L(u,v,x,y)$为4D光场；$(u,v)$为视点坐标；$(x,y)$为成像平面坐标；$(u_0,v_0)$为中心视点。由于在实际操作中一般取$(u_0,v_0)=(0,0)$，故式(1)可简化为

$$F(s,x,y)=\iint L(u,v,x+su,y+sv)\mathrm{d}u\mathrm{d}v。\tag{2}$$

对聚焦堆栈数据进行三维傅里叶变换可得到聚焦堆栈频谱，如图1(b)所示，即

$$\bar{F}(\omega_s,\omega_x,\omega_y)=\iiint F(s,x,y)\exp(-\mathrm{i}\omega_s s)\times\exp(-\mathrm{i}\omega_x x)\exp(-\mathrm{i}\omega_y y)\mathrm{d}s\mathrm{d}x\mathrm{d}y。\tag{3}$$

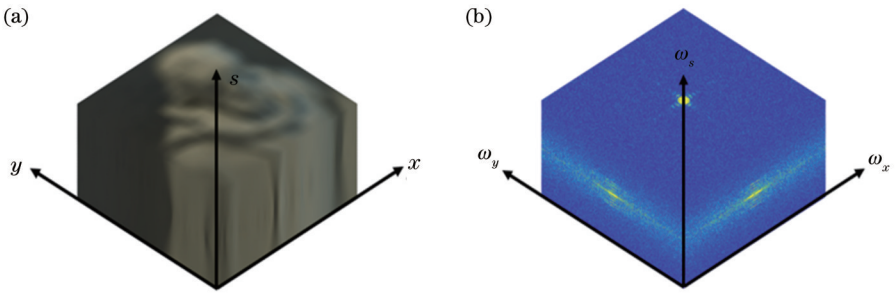


图1　聚焦堆栈数据及其频谱。(a)聚焦堆栈数据；(b)聚焦堆栈频谱

Fig. 1　Focal stack data and its spectra. (a) Focal stack data; (b) focal stack spectrum

将式(2)代入式(3)得

$$\bar{F}(\omega_s,\omega_x,\omega_y)=\iiint\iint L(u,v,x+su,y+sv)\times\exp(-\mathrm{i}\omega_x x-\mathrm{i}\omega_y y-\mathrm{i}\omega_s s)\mathrm{d}u\mathrm{d}v\mathrm{d}x\mathrm{d}y\mathrm{d}s。\tag{4}$$

式(4)可改写为

$$\bar{F}(\omega_s,\omega_x,\omega_y)=\iiint\iint\iint L(u,v,x_s,y_s)\delta(x_s-x-su,y_s-y-sv)\times\exp(-\mathrm{i}\omega_x x-\mathrm{i}\omega_y y-\mathrm{i}\omega_s s)\mathrm{d}x_s\mathrm{d}y_s\mathrm{d}u\mathrm{d}v\mathrm{d}x\mathrm{d}y\mathrm{d}s,\tag{5}$$

式中：$\delta(\cdot)$为狄拉克函数。由于

$$\iiint\delta(x_s-x-su,y_s-y-sv)\exp(-\mathrm{i}\omega_x x-\mathrm{i}\omega_y y-\mathrm{i}\omega_s s)\mathrm{d}x\mathrm{d}y\mathrm{d}s\xrightarrow[q=y_s-y-sv]{p=x_s-x-su}\iiint\delta(p,q)\exp\left[-\mathrm{i}\omega_x(x_s-p-su)-\mathrm{i}\omega_y(y_s-p-sv)-\mathrm{i}\omega_s s\right]\times\left|J(p,q,s)\right|\mathrm{d}p\mathrm{d}q\mathrm{d}s=$$

$$\exp(-\mathrm{i}\omega_x x_s-\mathrm{i}\omega_y y_s)\int\exp\left[-\mathrm{i}s(\omega_s-u\omega_x-v\omega_y)\right]\mathrm{d}s=\exp(-\mathrm{i}\omega_x x_s-\mathrm{i}\omega_y y_s)\delta(\omega_s-u\omega_x-v\omega_y),\tag{6}$$

式中：$J(p,q,s)$为雅可比行列式，则

$$\bar{F}(\omega_s,\omega_x,\omega_y)=\iint\iint L(u,v,x_s,y_s)\delta(\omega_s-u\omega_x-v\omega_y)\times\exp(-i\omega_x x_s-i\omega_y y_s)\mathrm{d}x_s\mathrm{d}y_s\mathrm{d}u\mathrm{d}v=\iint\hat{L}(u,v,\omega_x,\omega_y)\delta(\omega_s-u\omega_x-v\omega_y)\mathrm{d}u\mathrm{d}v,\tag{7}$$

式中：$\hat{L}(u,v,\omega_x,\omega_y)$为4D光场的空间维频谱，即子孔径图像的频谱。由此可见，聚焦堆栈频谱的分布为所有子孔径图像的频谱组合而成，每个子孔径图像的频谱在聚焦堆栈频谱中都过原点的平面，且平面的法向量与视点有关。

### 2.2　聚焦堆栈视差维滤波与滤波器的设计

令$F_{\Delta s}(s,x,y)$表示视差维采样的聚焦堆栈，视差采样间隔为$\Delta s$，如图2(a)所示，则

$$F_{\Delta s}(s,x,y)=F(s,x,y)\sum_{n=-\infty}^{+\infty}\delta(s-n\Delta s),\tag{8}$$

依据采样定理，

$$\bar{F}_{\Delta s}(\omega_s,\omega_x,\omega_y)=\frac{1}{\Delta s}\sum_{n=-\infty}^{+\infty}\bar{F}\left(\omega_s-n\frac{2\pi}{\Delta s},\omega_x,\omega_y\right),\tag{9}$$

式中：$\bar{F}_{\Delta s}(\omega_s,\omega_x,\omega_y)$为$F_{\Delta s}(s,x,y)$的三维频谱，如图2(b)所示。

由此可见，视差维等间隔采样的聚焦堆栈频谱是沿着视差维度周期性变化的。重建连续视差维采样的聚焦堆栈，需要进行视差维度滤波以获得单个周期聚焦堆栈频谱。理想滤波器的函数可表示为

$$\bar{W}(\omega_s,\omega_x,\omega_y)=\begin{cases}1, & |\omega_s-\omega_0|\leqslant\Delta\omega\\ 0, & |\omega_s-\omega_0|>\Delta\omega\end{cases}。\tag{10}$$

式(10)为矩形窗函数，其傅里叶逆变换为sinc函

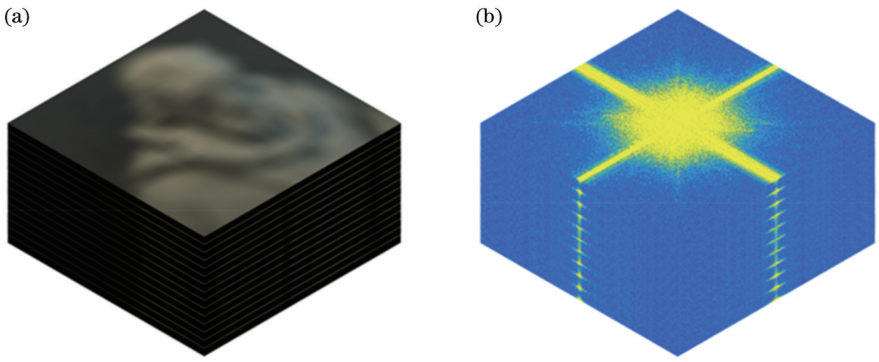


图 2　低视差分辨率的聚焦堆栈与频谱。(a)低视差分辨率的聚焦堆栈;(b)低视差分辨率的聚焦堆栈频谱

Fig. 2　Focal stack and spectrum with low disparity resolution. (a) Focal stack with low disparity resolution; (b) focal stack spectrum with low disparity resolution

数,该函数两边的余波会导致重建结果出现振铃[36]现象。在选取滤波器时,滤波函数越接近于矩形窗函数,振铃现象越明显。巴特沃斯低通滤波器的频率响应曲线最大限度平坦、没有纹波,而在阻频带则逐渐下降为0。选取巴特沃斯滤波器作为视差维滤波器抑制频谱混叠。

应用于聚焦堆栈滤波的巴特沃斯滤波器的函数可表示为

$$W_{\mathrm{B}}\left(\omega_s,\omega_x,\omega_y\right)=\frac{1}{\sqrt{1+\left(\dfrac{\omega_s-\omega_0}{\Delta\omega}\right)^{2K}}},\tag{11}$$

式中:$K$为滤波器阶数。若$K$过大,则在通频带内,巴特沃斯滤波器的频率响应曲线接近平坦;而在阻频带内迅速下降为0,此时的滤波器接近于理想滤波器。若$K$过小,则滤波器过于平滑,此时的滤波器无法完整地过滤出一个完整的频谱,这样会丢失较多的高频信息,且滤波结果依然存在较多的混叠。选取巴特沃斯滤波器的原因是当$K$选取适当时,滤波器能够保留更多的高频信息,减少混叠,且可以避免振铃现象。

### 2.3　聚焦堆栈视差维超分辨和视差估计

综合上述讨论,实现聚焦堆栈视差维滤波的流程图如图4所示,选取巴特沃斯滤波作为视差维滤波器。使用获取的聚焦堆栈进行视差估计。

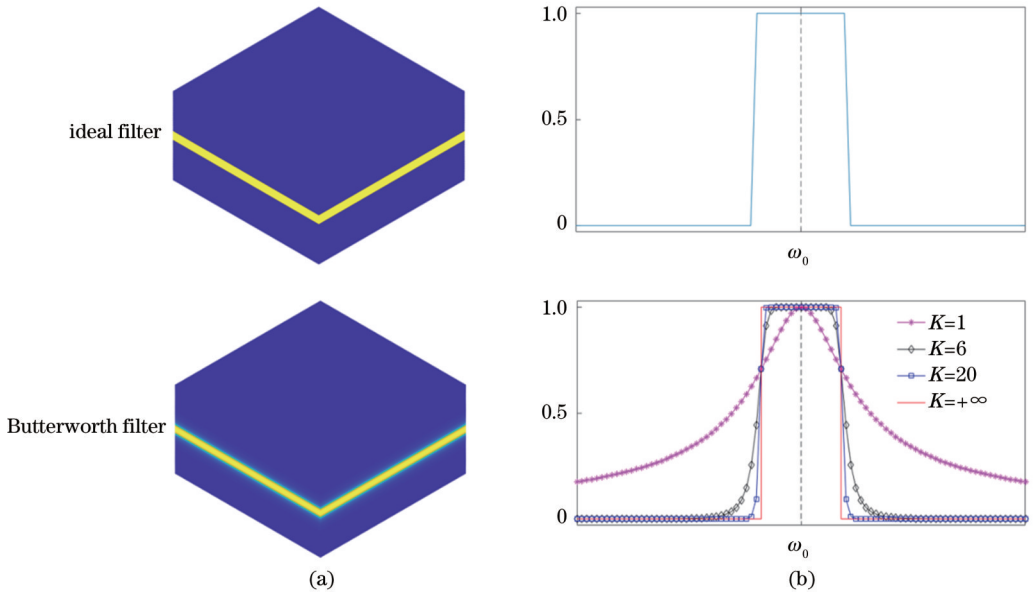


图 3　不同滤波器的频率响应曲线。(a)理想滤波器和巴特沃斯滤波器;(b)频率响应曲线

Fig. 3　Frequency response curves of different filters. (a) Ideal filter and Butterworth filter; (b) frequency response curve

## 3　实验与分析

采用模拟的聚焦堆栈数据与实测聚焦堆栈数据验证所提方法。3.1节为模拟数据实验,先合成低视差采样率的聚焦堆栈,然后采用视差维滤波获取视差维稠密的聚焦堆栈数据,再进行视差估计。3.2节为实测数据实验,对实拍的聚焦堆栈进行视差维超分辨和视差估计,以验证方法的有效性。在视差估计的具体

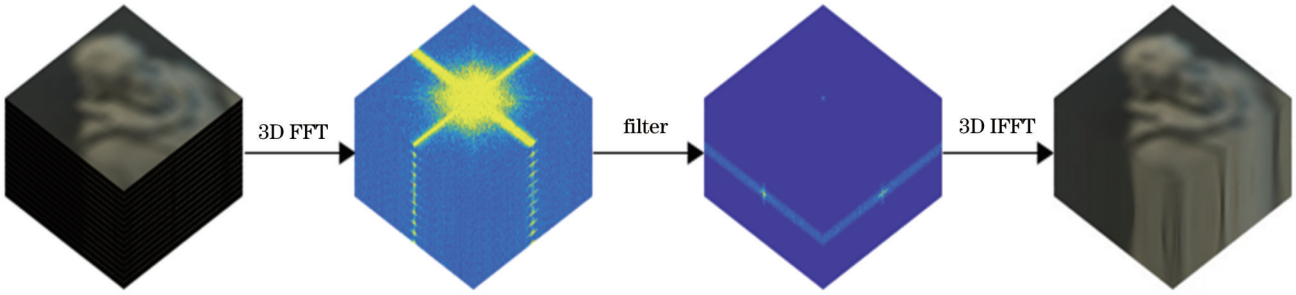


图 4　聚焦堆栈视差维滤波过程

Fig. 4　Process of focal stack disparity dimension filtering

计算中，选用 Robust focus volume regularization (RFV)算法[37]。

### 3.1　模拟数据实验

#### 3.1.1　视差维超分辨与视差估计结果

本实验选用的 cotton、pens、platonic、table 场景数据源自 Heidelberg collaboratory for image processing (HCI)数据集，视差范围分别为 −1.6～1.5、−1.7～2.0、−1.7～1.5、−2.0～1.6，角度分辨率为 9×9，空间分辨率均为 512×512。先利用由光场投影生成重聚焦图方法生成含有 16 张图像的聚焦堆栈数据，其对应的视差范围为 −4.5～4.5，相邻图像的视差间隔为 $\Delta s=0.6$。然后，利用所提的视差维超分辨方法得到含有 151 张图像的聚焦堆栈，相邻图像的视差间隔为 $\Delta s=0.06$，实验设置中，选取滤波器参数 $K=6$。视差维超分辨后的聚焦堆栈中的部分图像如图 5 所示。

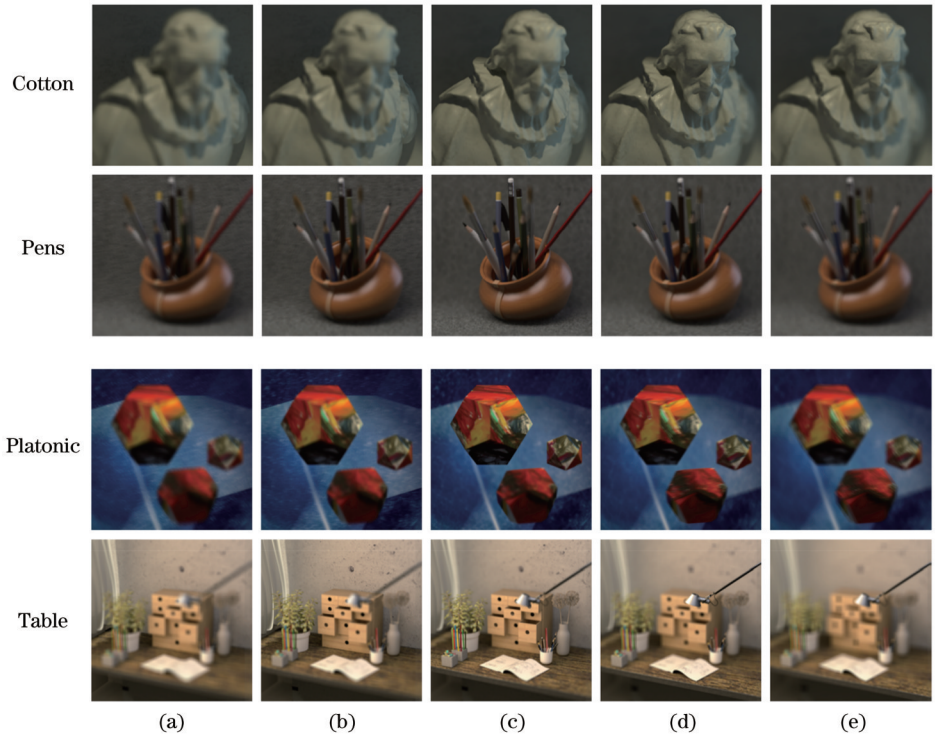


图 5　聚焦堆栈视差维超分辨结果。(a) $s=-1.68$；(b) $s=-0.84$；(c) $s=0$；(d) $s=0.84$；(e) $s=1.68$

Fig. 5　Focal stack disparity dimensional super-resolution results. (a) $s=-1.68$; (b) $s=-0.84$; (c) $s=0$; (d) $s=0.84$; (e) $s=1.68$

分别对超分辨前与超分辨后的聚焦堆栈进行视差估计，其结果如图 6 所示。其中，视差维超分辨后的聚焦堆栈数据，cotton、pens、platonic、table 场景分别选择第 49～101、47～110、47～101、42～103 张图像进行实验，对应视差分别为 −1.62～1.50、−1.74～2.04、−1.74～1.50、−2.04～1.62。视差维超分辨前的聚焦堆栈数据，分别选择第 5～11、5～12、5～11、5～12 张图像，对应视差分别为 −2.1～1.5、−2.1～2.1、−2.1～1.5、−2.1～2.1。

由图 6 可知，第 2 行超分辨前聚焦堆栈的视差估计结果存在较多的与第 4 行视差图真实值明显不符的像素点，而第 3 行所提方法超分辨后的聚焦堆栈的视差估计结果更加平滑，且更加接近于真实值。为进一步对比视差估计的优化效果，分别计算第 2 行、第 3 行视

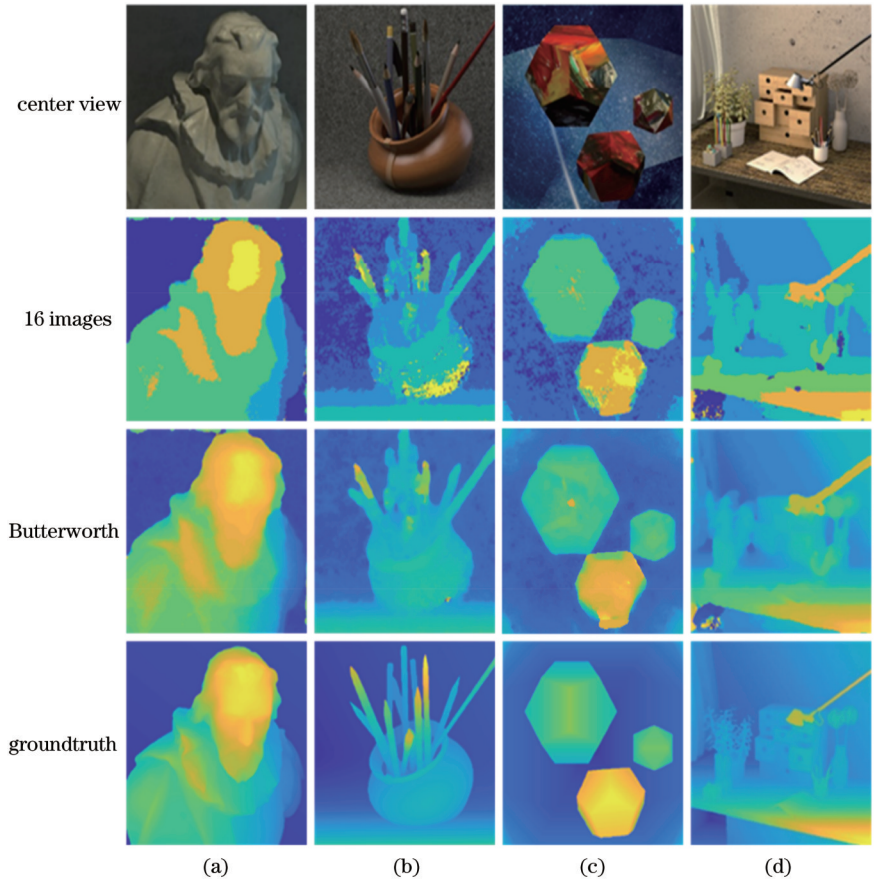


图 6　利用不同分辨率聚焦堆栈估计的视差图。(a)Cotton 场景;(b)Pens 场景;(c)Platonic 场景;(d)Table 场景

Fig. 6　Disparity map estimation using different resolution focal stacks. (a) Cotton scene; (b) Pens scene; (c) Platonic scene; (d) Table scene

差图与真实值之间的峰值信噪比(PSNR)、结构相似度(SSIM),计算结果如表 1 所示。

表 1　超分辨前后的聚焦堆栈视差估计结果与真实值对比

Table 1　Comparison of focal stack disparity estimation results before and after super-resolution and groundtruth

| Data | PSNR /dB | | | | SSIM | | | |
|---|---|---|---|---|---|---|---|---|
| | Cotton | Pens | Platonic | Table | Cotton | Pens | Platonic | Table |
| 16 images | 54.8714 | 52.5809 | 54.4988 | 55.5002 | 0.9835 | 0.9632 | 0.9823 | 0.9833 |
| Butterworth | 56.8403 | 55.9989 | 57.1102 | 57.4376 | 0.9864 | 0.9813 | 0.9852 | 0.9876 |

由表 1 可知,超分辨后聚焦堆栈的视差估计结果与真实值之间的 PSNR 与 SSIM 有明显提高。故使用所提方法得到的聚焦堆栈进行视差估计的结果比超分辨前的结果更加精确。

### 3.1.2　与其他生成稠密视差维聚焦堆栈方法的比较

将所提方法与其他生成稠密视差维聚焦堆栈方法进行比较,包括由光场数据投影生成聚焦堆栈、由 FDL 生成聚焦堆栈。分别生成 151 张图像的聚焦堆栈,图 7、图 8 为其中的部分图像。FDL 生成聚焦堆栈方法是通过选取 30 层的 FDL 数据来生成聚焦堆栈。

对上述两种方法的结果进行视差估计,并计算两种方法的视差图与真实值之间的 PSNR、SSIM,其结果分别如图 9、表 2 所示。由表 1、表 2 的第 3、4 行数据可知,所提方法得到的聚焦堆栈的视差估计结果与真实值间的 PSNR、SSIM 与其他两种方法的差异不大,其他两种方法对一些纹理不复杂的场景视差估计结果甚至更优,如 cotton 场景。故在模拟数据实验中,所提方法能够得到高保真的聚焦堆栈视差维超分辨数据,实现稠密的视差估计,对一些纹理简单的场景,所提方法的视差估计结果最精确。

### 3.1.3　巴特沃斯滤波器不同 $K$ 值对视差维超分辨结果的影响

选取 $K=0.6$、$K=60$ 时的巴特沃斯滤波器对 16 张图像的聚焦堆栈进行视差维超分辨,获得 151 张图像的聚焦堆栈。图 10(a)为 Cotton 场景不同参数下的超分辨结果。使用得到的聚焦堆栈进行视差估计,并使用 PSNR、SSIM 指标对结果进行评价,其结果分别如图 11、表 3 所示。沿着图 10(a)的直线切片得到图 10(b),局部放大图 10(b)中的方框得到图 10(c)。

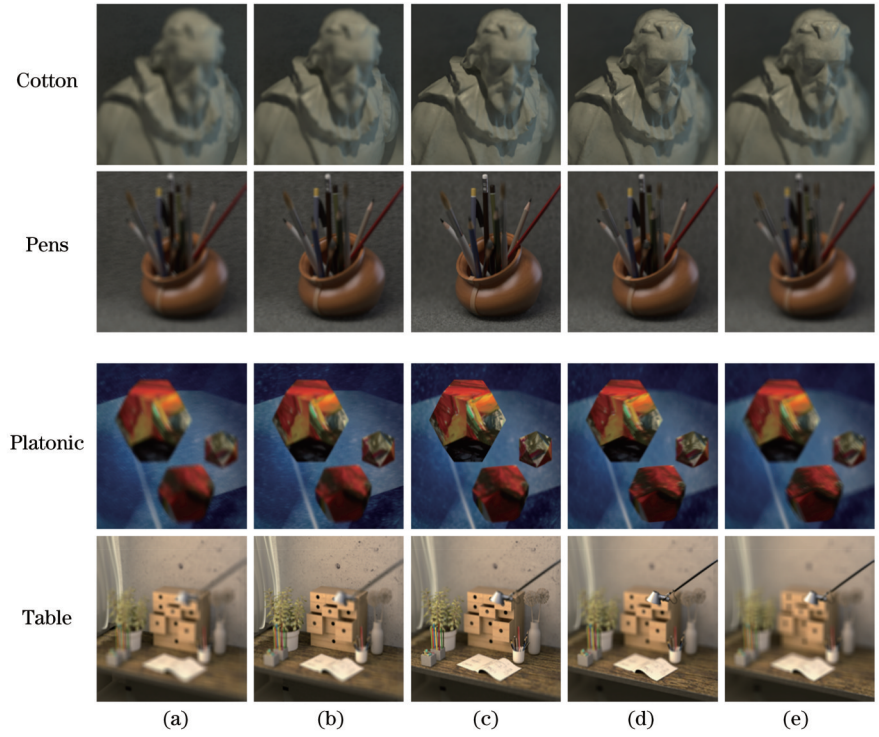


图 7　由光场投影生成聚焦堆栈。(a) $s=-1.68$;(b) $s=-0.84$;(c) $s=0$;(d) $s=0.84$;(e) $s=1.68$

Fig. 7　Generating focal stack from light field projection. (a) $s=-1.68$; (b) $s=-0.84$; (c) $s=0$; (d) $s=0.84$; (e) $s=1.68$

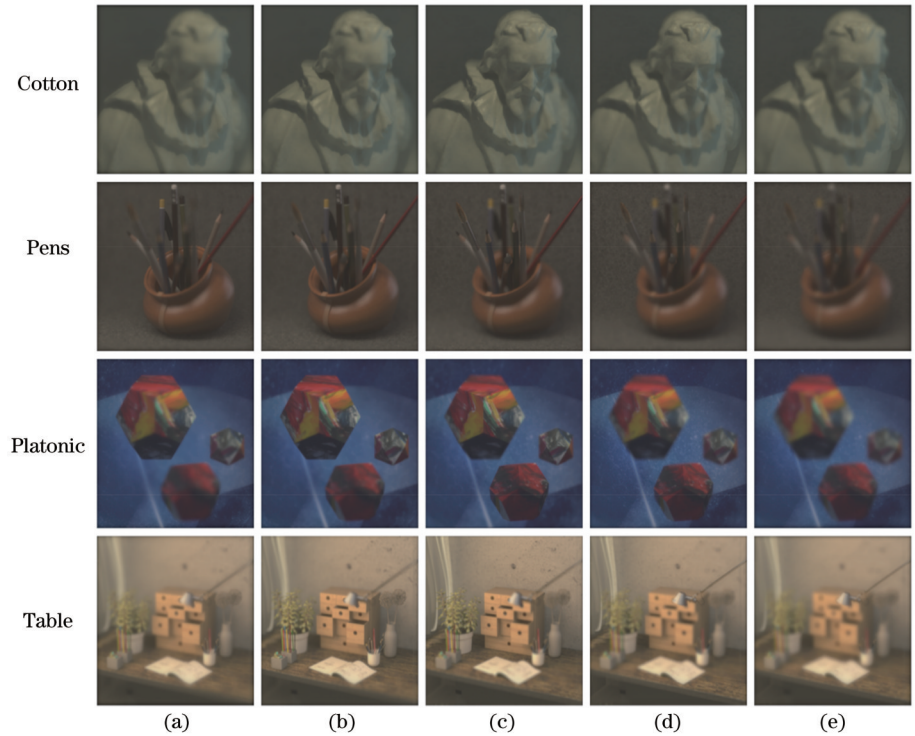


图 8　FDL 生成聚焦堆栈。(a) $s=-1.68$;(b) $s=-0.84$;(c) $s=0$;(d) $s=0.84$;(e) $s=1.68$

Fig. 8　Generating focusing stack from FDL generation. (a) $s=-1.68$; (b) $s=-0.84$; (c) $s=0$; (d) $s=0.84$; (e) $s=1.68$

实验结果表明,选取 $K$ 值可以实现高保真的聚焦堆栈视差维超分辨,进而实现高质量的稠密视差估计。当 $K$ 值过小时(图 10 第 1 行),滤波后的频谱会丢失一些高频信息,且依然存在较多的混叠,故聚焦堆栈图像像素随视差的变化不均匀。而 $K$ 值过大时(图 10 第 3 行),像素变化出现波动,这是振铃现象在聚焦堆栈视差维度的体现。由表 3 数据和表 1 第 3、4 行数据可知,$K=0.6$ 与 $K=60$ 时聚焦堆栈的视差估计结果与真实值的 PSNR 与 SSIM 明显没有 $K=6$ 时的 PSNR 与 SSIM 大。由此可见,$K$ 值过大或过小均会导致视差估

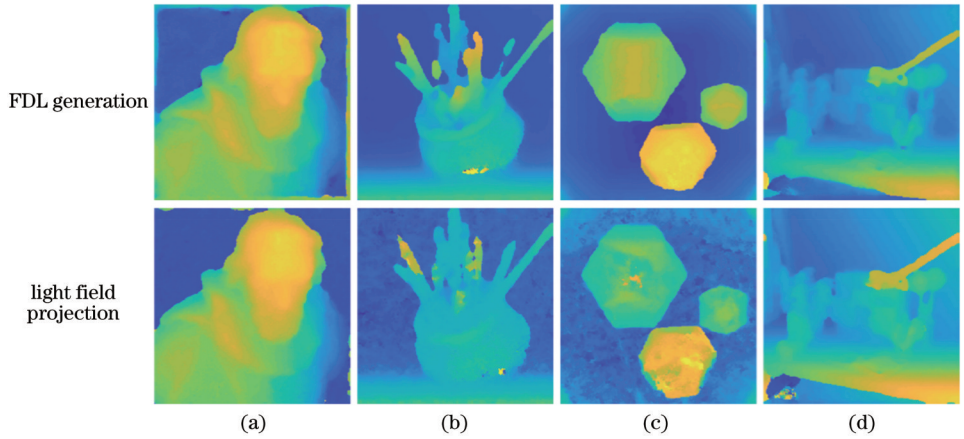


图 9　光场投影方法与FDL方法的视差估计结果。(a)Cotton场景;(b)Pens场景;(c)Platonic场景;(d)Table场景

Fig. 9　Disparity estimation results of light field projection method and FDL generation method. (a) Cotton scene; (b) Pens scene; (c) Platonic scene; (d) Table scene

表 2　不同方法下的视差图与真实值对比

Table 2　Comparison between disparity maps under different methods and groundtruth

| Method | PSNR /dB | | | | SSIM | | | |
|---|---|---|---|---|---|---|---|---|
| | Cotton | Pens | Platonic | Table | Cotton | Pens | Platonic | Table |
| Light field projection | 56.4263 | 53.8427 | 56.3264 | 57.5501 | 0.9849 | 0.9718 | 0.9823 | 0.9878 |
| FDL generation | 53.5724 | 56.1407 | 57.8347 | 58.3013 | 0.9751 | 0.9840 | 0.9883 | 0.9905 |

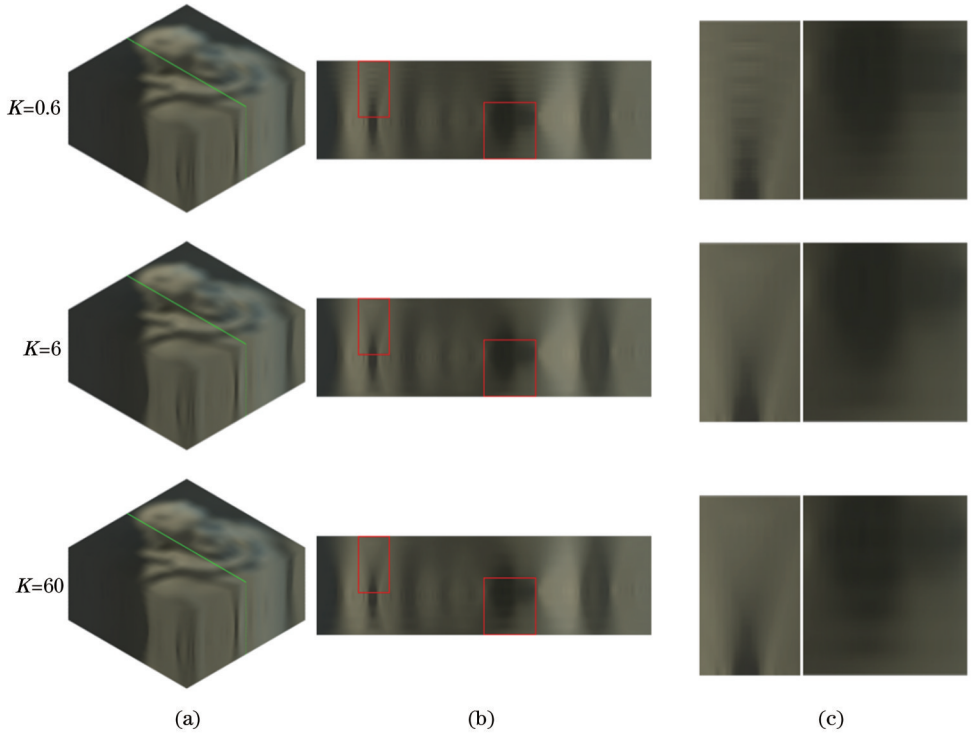


图 10　不同$K$值的巴特沃斯滤波视差维超分辨结果。(a)聚焦堆栈;(b)切片;(c)局部放大

Fig. 10　Disparity dimension super-resolution results of Butterworth filter with different $K$ values. (a) Focal stack; (b) section; (c) partial enlarged view

计的效果变差。

### 3.2　实测数据实验

本实验选取的 Bottles 与 Metals 场景数据源自 Sameer 等[38]的数据集,每个场景数据包含 31 张图像的视差维稠密聚焦堆栈实测数据,其分辨率为 640×360。选取其中的第 1、7、13、19、25、31 张图像,组成含 6 张图像的视差维稀疏聚焦堆栈数据,并选择不同参数的巴特沃斯滤波器进行超分辨获得 31 张图像的视差维稠

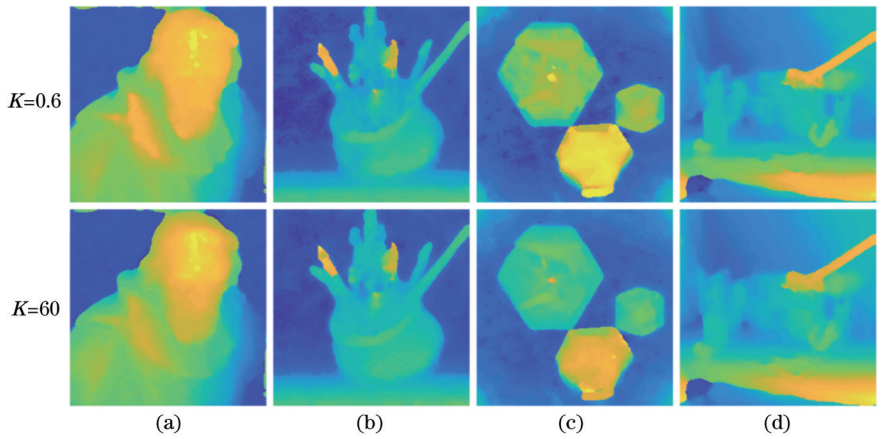


图 11　不同 $K$ 值巴特沃斯滤波视差维超分辨后估计的视差图。(a)Cotton 场景;(b)Pens 场景;(c)Platonic 场景;(d)Table 场景

Fig. 11　Disparity maps estimated after disparity dimensions super-resolution of Butterworth filter with different $K$ values. (a) Cotton scene; (b) Pens scene; (c) Platonic scene; (d) Table scene

表 3　不同 $K$ 值巴特沃斯滤波视差维超分辨后估计的视差图与真实值对比

Table 3　Comparison of estimated disparity maps after disparity dimensions super-resolution of Butterworth filter with different $K$ and groundtruth

| $K$ | PSNR/dB | | | | SSIM | | | |
|---|---|---|---|---|---|---|---|---|
| | Cotton | Pens | Platonic | Table | Cotton | Pens | Platonic | Table |
| 0.6 | 54.2536 | 55.3009 | 56.1361 | 56.5317 | 0.9736 | 0.9789 | 0.9810 | 0.9833 |
| 60 | 55.5668 | 55.1568 | 56.8884 | 56.6549 | 0.9815 | 0.9782 | 0.9843 | 0.9837 |

密聚焦堆栈,聚焦堆栈数据中的部分图像如图 12、图 13 所示。使用 PSNR、SSIM 计算超分辨结果与 Bottles 与 Metals 场景的实测数据误差,其结果如图 14 所示。

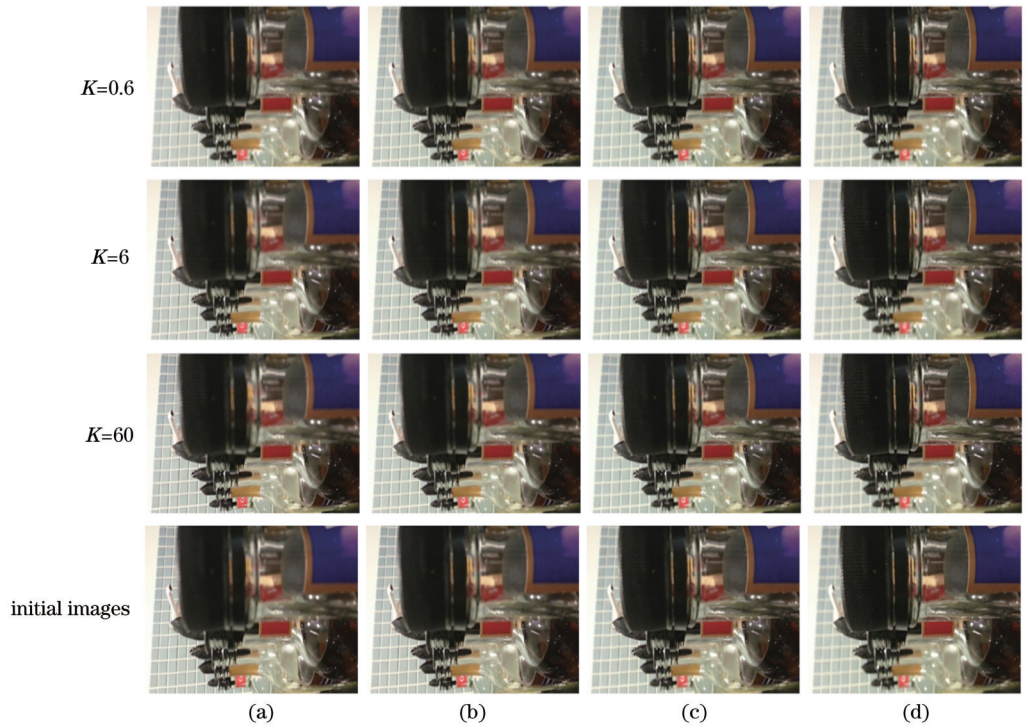


图 12　Bottles 场景在不同 $K$ 值下的视差维超分辨结果。(a) $n=6$;(b) $n=12$;(c) $n=18$;(d) $n=24$

Fig. 12　Disparity dimension super-resolution results of Bottles scene at different $K$ values. (a) $n=6$; (b) $n=12$; (c) $n=18$; (d) $n=24$

由图 13、图 14 可知,选取合适的 $K$ 值可以实现高保真稠密视差维聚焦堆栈重建。$K=60$ 时得到的聚焦堆栈第 7、13、19、25 张图像的 PSNR 与第 7、25 张图像的 SSIM 比 $K=6$ 时的结果略小,$K=0.6$ 时得到的

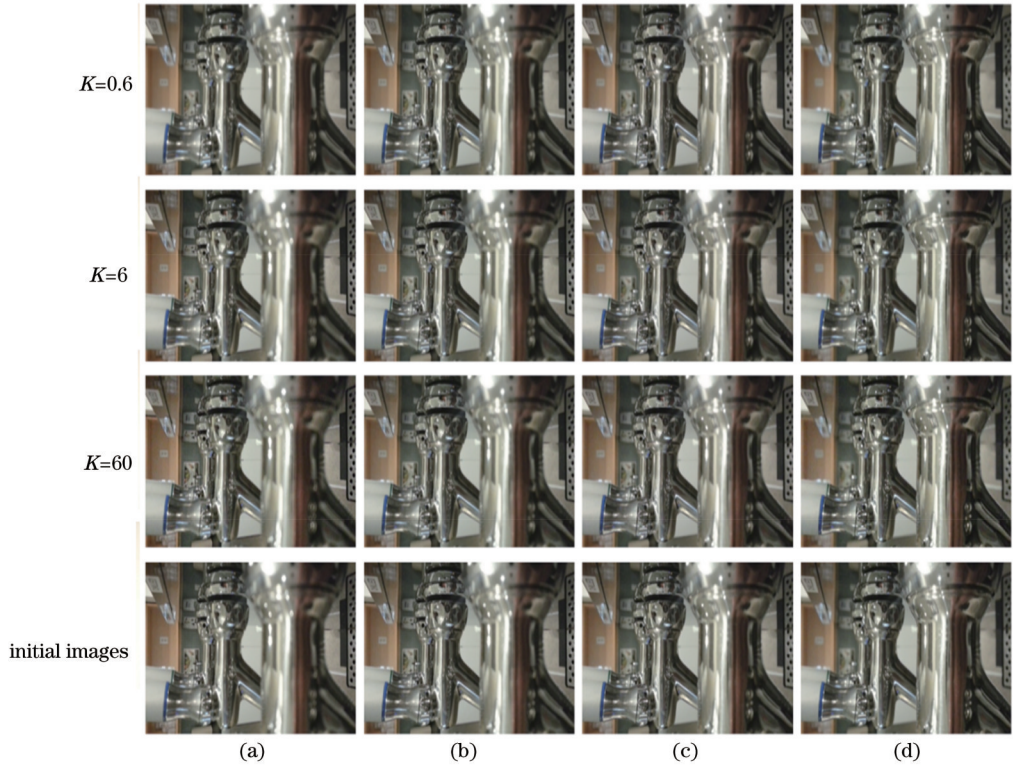


图 13　Metals场景在不同$K$值下的视差维超分辨结果。(a) $n=6$;(b) $n=12$;(c) $n=18$;(d) $n=24$

Fig. 13　Disparity dimension super-resolution results of Metals scene at different $K$ values. (a) $n=6$; (b) $n=12$; (c) $n=18$; (d) $n=24$

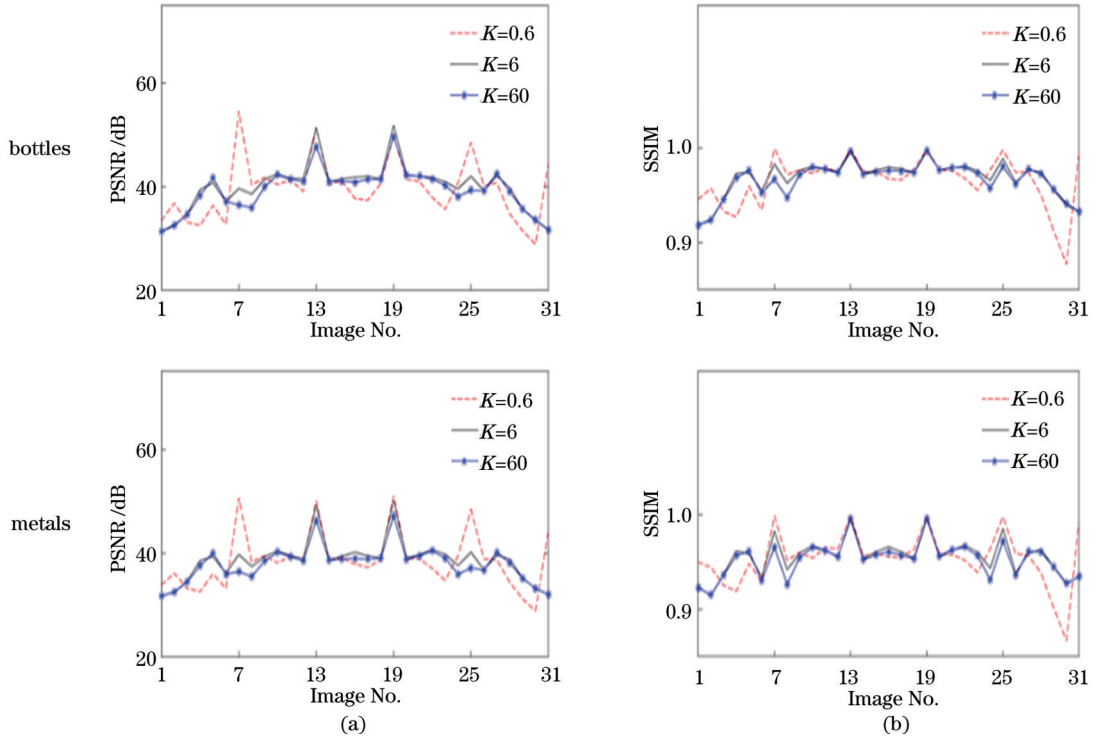


图 14　不同$K$值巴特沃斯滤波视差维超分辨定量评价结果。(a) PSNR;(b) SSIM

Fig. 14　Quantitative evaluation results of disparity dimension super-resolution of Butterworth filtering with different $K$ values. (a) PSNR; (b) SSIM

聚焦堆栈第 1 至 7、19 至 31 张之间部分图像的 PSNR 与 SSIM 明显低于 $K=6$ 时的结果。故 $K$ 值过大或过小均会导致实测聚焦堆栈的超分辨效果变差。

对 $K=6$ 时的超分辨结果与原始数据进行视差估计，其结果如图 15 所示。之后选取视差图的某一行(或列)的视差变化曲线进行对比，其中 Bottles 场

景选择第 400 列，Metals 场景选择第 175 行[图 16(a)]，其对比结果如图 16(b)所示。由图 16(b)可知，与利用视差维稀疏聚焦堆栈相比，利用视差维超分辨后的视差维稠密聚焦堆栈进行视差估计的结果得到大幅提高，与利用实测的视差维稠密聚焦堆栈进行视差估计的视差变化曲线接近，且能够反映 Bottles 场景中瓶子与 Metals 场景中曲面金属的视差变化情况。对实测数据而言，所提方法可实现高保真的聚焦堆栈视差维超分辨，可用于实现稠密、高质量的视差估计。

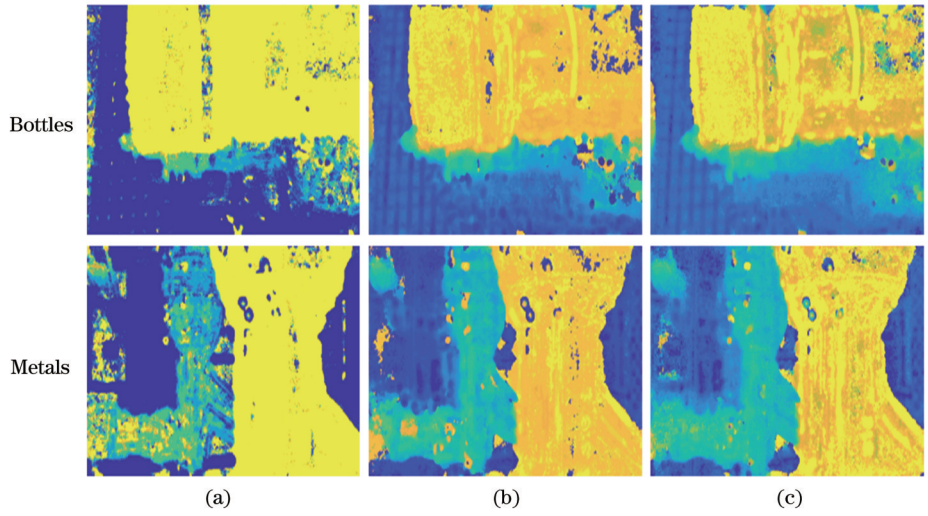


图 15　实测数据的视差估计结果。(a)利用视差维稀疏聚焦堆栈进行视差估计；(b)利用视差维超分辨后的视差维稠密聚焦堆栈进行视差估计；(c)利用实测的视差维稠密聚焦堆栈进行视差估计

Fig. 15　Disparity estimation results of measured data. (a) Disparity estimation using sparse focal stack in disparity dimension; (b) disparity estimation using dense focal stack of disparity dimension super-resolution; (c) disparity estimation using actual disparity dimension dense focal stack

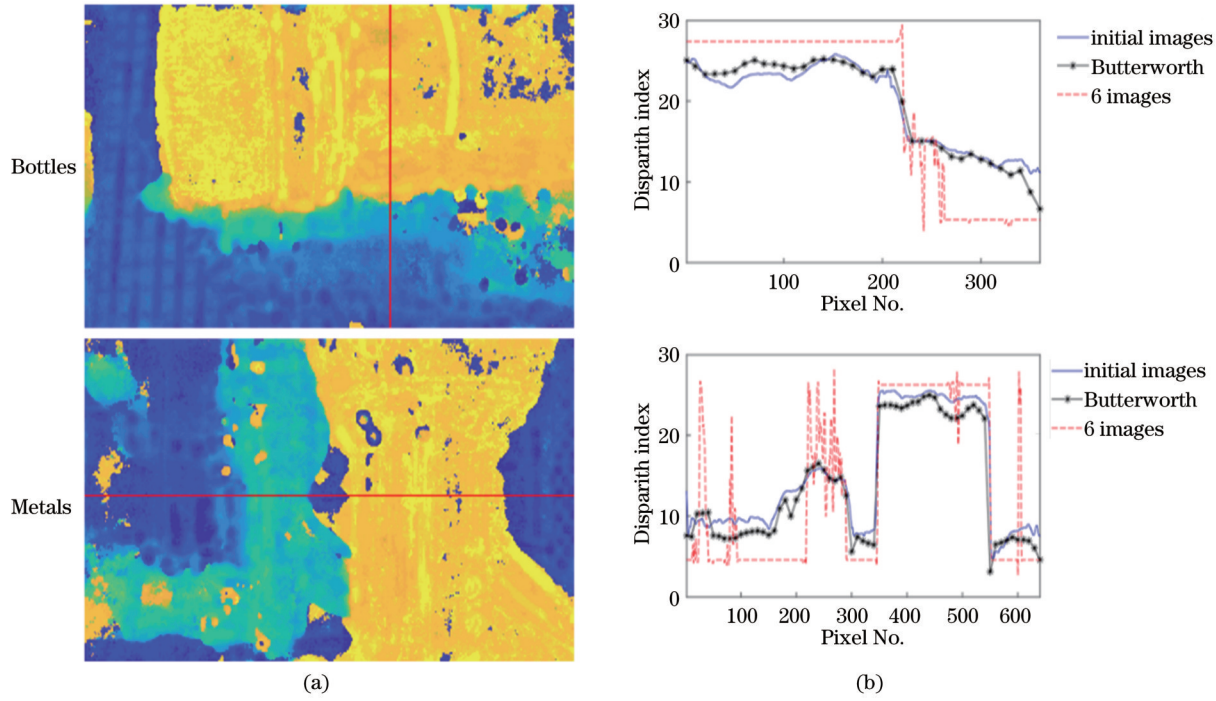


图 16　实测场景视差变化曲线对比。(a)视差图；(b)视差变化曲线

Fig. 16　Comparison of disparity curves in measured scenes. (a) Disparity map; (b) disparity curves

## 4　讨　　论

所提的聚焦堆栈视差维超分辨和视差估计方法基于聚焦堆栈的频谱分析，视差维傅里叶变换要求聚焦堆栈在视差维度等间隔采样。在实际数据采集中，聚焦堆栈在视差维是非等间隔采样的，引入数据插值和

数据拟合的方法处理视差维非等间隔采样的聚焦堆栈数据，推进所提方法应用于聚焦堆栈的实际计算成像技术。

在聚焦堆栈视差维滤波的具体计算中，选取巴特沃斯滤波器。相较于其他滤波器，通频带内的频率响应曲线最大限度平坦、没有纹波，可以保留更多的高频信息，且能够抑制更多的混叠。$K$值是巴特沃斯滤波器最重要的参数，给出的巴特沃斯滤波器的$K$值选取是通过实验得到经验值，根据场景频谱分析和视差采样率，从理论上给出$K$值的选取准则是未来的研究方向。

## 5 结　论

受数据通量和实际采集系统的限制，聚焦堆栈数据在视差维度的低采样率限制了基于聚焦堆栈的视差估计、全聚焦成像和光场重建等计算成像性能。基于聚焦堆栈频谱提出聚焦堆栈视差维滤波的视差维超分辨方法，为基于聚焦堆栈的计算成像技术提供数据支撑。针对由聚焦堆栈估计视差的问题，对视差维稀疏的聚焦堆栈频谱进行视差维滤波得到高视差分辨率的聚焦堆栈，最终达到优化视差估计的目的。实验结果表明：所提方法能够实现聚焦堆栈数据高保真的视差维超分辨，并用于稠密、高精度的视差估计，且有望进一步提升基于聚焦堆栈数据的全聚焦成像、光场重建和场景重构等计算成像性能。

## 参 考 文 献


[1] Levoy M, Hanrahan P. Light field rendering[M]. New York: ACM Press, 1996: 31-42.

[2] Rajagopalan A N, Chaudhuri S, Mudenagudi U. Depth estimation and image restoration using defocused stereo pairs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(11): 1521-1525.

[3] Strecke M, Alperovich A, Goldluecke B. Accurate depth and normal maps from occlusion-aware focal stack symmetry[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2529-2537.

[4] Lin H T, Chen C, Kang S B, et al. Depth recovery from light field using focal stack symmetry[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 3451-3459.

[5] Huang Z Y, Fessler J A, Norris T B, et al. Light-field reconstruction and depth estimation from focal stack images using convolutional neural networks[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 4-8, 2020, Barcelona, Spain. New York: IEEE Press, 2020: 8648-8652.

[6] Ma Z Y, He K M, Wei Y C, et al. Constant time weighted Median filtering for stereo matching and beyond[C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2014: 49-56.

[7] Zhang Q, Xu L, Jia J Y. 100 times faster weighted median filter (WMF)[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 2830-2837.

[8] Guo X J, Li Y, Ma J Y, et al. Mutually guided image filtering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(3): 694-707.

[9] Levin A, Durand F. Linear view synthesis using a dimensionality gap light field prior[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 1831-1838.

[10] Liu C, Qiu J, Jiang M. Light field reconstruction from projection modeling of focal stack[J]. Optics Express, 2017, 25 (10): 11377-11388.

[11] Sakamoto T, Kodama K, Hamamoto T. A study on efficient compression of multi-focus images for dense Light-Field reconstruction[C]//2012 Visual Communications and Image Processing, November 27-30, 2012, San Diego, CA, USA. New York: IEEE Press, 2013.

[12] Mousnier A, Vural E, Guillemot C. Partial light field tomographic reconstruction from a fixed-camera focal stack[J]. Computer Science, 2015, 22(8): 347-356.

[13] Yin X W, Wang G J, Li W T, et al. Iteratively reconstructing 4D light fields from focal stacks[J]. Applied Optics, 2016, 55 (30): 8457-8463.

[14] Blocker C J, Chun Y, Fessler J A. Low-rank plus sparse tensor models for light-field reconstruction from focal stack data[C]// 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), June 10-12, 2018, Aristi Village, Greece. New York: IEEE Press, 2018.

[15] Xiao Y H, Wang G J, Hu X W, et al. Guided, fusion-based, large depth-of-field 3D imaging using a focal stack[J]. Sensors, 2019, 19(22): 4845.

[16] Zhang C, Bastian J, Shen C H, et al. Extended depth-of-field via focus stacking and graph cuts[C]//2013 IEEE International Conference on Image Processing, September 15-18, 2013, Melbourne, VIC. New York: IEEE Press, 2014: 1272-1276.

[17] Jacobs D E, Baek J, Levoy M. Focal stack compositing for depth of field control[R]. Stanford: Stanford Computer Graphics Laboratory Technical Report, 2012.

[18] Ng R, Levoy M, Bredif M, et al. Light field photography with a hand-held plenoptic camera[R]. Stanford: Stanford Computer Science Tech Report CSTR, 2005.

[19] Veeraraghavan A, Raskar R, Agrawal A, et al. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing[J]. ACM Transactions on Graphics, 26(3): 69-es.

[20] Levoy M. Light fields and computational imaging[J]. Computer, 2006, 39(8): 46-55.

[21] Ng R. Digital light field photography[D]. Stanford: Stanford University, 2006.

[22] Ng R. Fourier slice photography[J]. ACM Transactions on Graphics, 2005, 24(3): 735-744.

[23] Arroyo S D P. Modeling and applications of the focus cue in conventional digital cameras[D]. Tarragona: Universitat Rovira i Virgili, 2013.

[24] Le Pendu M, Guillemot C, Smolic A. A Fourier disparity layer representation for light fields[J]. IEEE Transactions on Image Processing, 2019, 28(11): 5740-5753.

[25] Pérez F, Pérez A, Rodríguez M, et al. Fourier slice super-resolution in plenoptic cameras[C]//2012 IEEE International Conference on Computational Photography (ICCP), April 28-29, 2012, Seattle, WA, USA. New York: IEEE Press, 2012.

[26] 魏菲, 刘畅, 邱钧. 基于光场傅里叶视差层表达的场景视差重建[J]. 光学学报, 2022, 42(16): 1610001.
Wei F, Liu C, Qiu J. Scene parallax reconstruction based on Fourier parallax layer expression of light field[J]. Acta Optica Sinica, 2022, 42(16): 1610001.

[27] 史立根, 邱钧, 刘畅, 等. 基于YCbCr光场数据的视差重建算

法[J]. 激光与光电子学进展, 2022, 59(2): 0211002.
Shi L G, Qiu J, Liu C, et al. Disparity reconstruction algorithm based on YCbCr light field data[J]. Laser & Optoelectronics Progress, 2022, 59(2): 0211002.
[28] 李亚宁, 王雪, 周果清, 等. 四维光场表达模型综述[J]. 激光与光电子学进展, 2021, 58(18): 1811012.
Li Y N, Wang X, Zhou G Q, et al. Overview of 4D light field representation[J]. Laser & Optoelectronics Progress, 2021, 58(18): 1811012.
[29] Li Y N, Wang X, Zhu H, et al. Deep anti-aliasing of whole focal stack using slice spectrum[J]. IEEE Transactions on Computational Imaging, 2021, 7: 1328-1340.
[30] Chang A C, Sung T P, Shih K T, et al. Anti-aliasing for light field rendering[C]//2014 IEEE International Conference on Multimedia and Expo (ICME), July 14-18, 2014, Chengdu, China. New York: IEEE Press, 2014.
[31] Xiao Z L, Wang Q, Zhou G Q, et al. Aliasing detection and reduction scheme on angularly undersampled light fields[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2103-2115.
[32] Isaksen A, McMillan L, Gortler S J. Dynamically reparameterized light fields[C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques-SIGGRAPH '00, July 23-28, 2000, New Orleans, LA, USA. New York: ACM Press, 2000: 297-306.
[33] Nava F P, Luke J P. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera[C]//2009 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, May 4-6, 2009, Potsdam, Germany. New York: IEEE Press, 2009.
[34] Xiao Z L, Shi J L, Jiang X R, et al. A learning-based view extrapolation method for axial super-resolution[J]. Neurocomputing, 2021, 455: 229-241.
[35] Xiao Z L, Liu H A, Jin H Y. COLF-GAN: learning to axial super-resolve focal stacks[J]. IET Image Processing, 2022, 16(7): 1949-1958.
[36] Levin A, Fergus R, Durand F, et al. Image and depth from a conventional camera with a coded aperture[J]. ACM Transactions on Graphics, 26(3): 70-es.
[37] Ali U, Mahmood M T. Robust focus volume regularization in shape from focus[J]. IEEE Transactions on Image Processing, 2021, 30: 7215-7227.
[38] Sameer A, Keir M. Ceres solver[EB/OL]. [2022-11-12]. https://code.google.com/p/ceres-solver/.

# Disparity Estimation Method Based on Focal Stack Disparity Dimension Super-Resolution

Wang Yukai, Liu Chang*, Qiu Jun

*Institute of Applied Mathematics, Beijing Information Science and Technology University, Beijing* 100101, *China*

## Abstract

**Objective** Focal stack is a projection domain representation model of the four-dimensional (4D) light field, which can be used in disparity estimation, light field reconstruction, extended depth of field imaging, and other fields. The accuracy and robustness of computational imaging based on focal stack data depend on the disparity dimensional resolution of the focal stack data. There are two ways to obtain focal stack images. The first one is to directly capture them on multiple disparity planes by imaging equipment, and the second one is to use digital refocusing methods to generate multiple images of different disparity layers. As capturing focal stack images by the imaging equipment needs to set the focal length and other parameters beforehand, and the focal stack with high quality and high disparity resolution can only be obtained by strictly controlling the imaging plane during the capturing process, while the digital refocusing method requires 4D light field data, resulting in computational redundancy. In view of the problem of insufficient resolution of disparity dimension in focal stack data, a method of focal stack super-resolution in disparity dimension was proposed. According to the disparity dimension spectrum optimization of the focal stack data, we proposed the focal stack disparity dimension filter and the disparity dimension super-resolution method of the focal stack data to estimate the disparity with high accuracy and robustness.

**Methods** The focal stack spectrum contains the disparity dimension spectrum, whereby focal stack data can be processed in the disparity dimension. In this paper, based on the disparity dimension spectrum optimization of focal stack data, a focal stack disparity dimension filter was introduced, and a focal stack disparity dimension super-resolution method based on disparity dimension filtering was then proposed to achieve high-precision and dense disparity estimation. Through the spectral analysis of the focal stack, the Butterworth filter was selected as the disparity dimension filter to achieve high-fidelity disparity dimension super-resolution of the focal stack data. Dense and high-precision disparity estimation was achieved based on the robust focus volume regularization (RFV) algorithm by using the dense focal stack after disparity dimension super-resolution.

**Results and Discussions** In the simulated data experiment, a focal stack containing 16 images was first generated by the light field projection method, and then a focal stack containing 151 images was obtained by super-resolution through the proposed method (Fig. 5). The RFV algorithm was applied for disparity estimation (Fig. 6). In the experiment,

Butterworth filter parameter was set to be $K=6$. By comparing the disparity estimation results of other data, including the focal stack before disparity-dimensional super-resolution (Table 1), the focal stack obtained with the light field projection method, and the Fourier parallax layer (FDL) generation method (Fig. 9 and Table 2), the peak signal to noise ratio (PSNR) and structural similarity (SSIM) values by the proposed method were larger than those before disparity-dimensional super-resolution and were close to those by the focal stack obtained with the light field projection method and the FDL generation method. Finally, we utilized the Butterworth filter with different $K$ values to obtain the focal stack for disparity estimation (Fig. 11) and then compared the PSNR and SSIM values (Table 3) of the disparity estimation results, and it was found that the real values at $K=0.6$ and $K=60$ were both smaller. In the measured data experiment, six images in the focal stack containing 31 images were selected to form the focal stack with sparse disparity dimension, and then the Butterworth filter with different $K$ values was used for disparity dimension super-resolution (Figs. 12 and 13). By comparing the obtained focal stack with the original data (Fig. 14), the PSNR and SSIM values of some focal stack images in the super-resolution results at $K=0.6$ and $K=60$ were significantly smaller than those at $K=6$. Then we implemented the disparity estimation (Fig. 15) and selected the profiles of the disparity map for comparison (Fig. 16). It can be seen that the disparity profiles obtained by the proposed method were smoother than that before super-resolution and were closer to the disparity obtained from the original data.

**Conclusions** The results of simulated data experiments and real data experiments show that the method of focal stack disparity dimension super-resolution proposed in this paper can effectively improve the disparity resolution of focal stacks and provide data for applications such as disparity estimation. The experimental results of simulated data show that the disparity estimation result of the focal stack obtained by the proposed method is more accurate and robust than the result before super-resolution, and it can obtain high-fidelity and high-disparity resolution focal stack data and realize dense disparity estimation. The dense disparity estimation is achieved based on the RFV algorithm by using the dense focal stack after the disparity dimension super-resolution. The experimental results of simulated and real data show that disparity dimension-based filtering can achieve efficient disparity dimension super-resolution, as well as high-precision and dense disparity estimation.

**Key words** focal stack; disparity dimension filtering; disparity dimension super-resolution; disparity estimation