基于光场分析的多线索融合深度估计方法

杨德刚[1)] 肖照林[1)] 杨 恒[1),2)] 王 庆[1)]

[1)](西北工业大学计算机学院 西安 710072)

[2)](中国兵器工业集团有限公司第 205 研究所 西安 710065)

摘 要 受人类视觉深度感知机理的启发,结合最近光场分析理论的进展,文中提出了一种融合多种深度线索的全局一致深度估计方法.该方法首先利用摄像机阵列获取光场数据,然后通过合成孔径成像方法获得指定深度的重投影光场,并从中提取表征深度变化不同维度的模糊与视差线索.接着文中采用光场分析方法对比了模糊与视差线索的适用情况,设计了多深度线索的融合算法,以实现不同深度线索的优势互补.进而为了获得精确一致的深度结果,文中在马尔可夫随机场模型的基础上,提出了一种自适应平滑约束的全局能量函数.最终,文中利用图割算法最小化全局能量函数,获得了平滑的高精度深度估计结果.文中分别在虚拟数据和真实数据上测试了所提出的方法,与单一深度线索和局部深度估计方法相比,文中方法能结合不同深度线索的优势并获得更加鲁棒的深度结果.

关键词 相机阵列;光场分析;多线索融合;深度估计

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2015.02437

Depth Estimation from Light Field Analysis Based Multiple Cues Fusion

YANG De-Gang[1)] XIAO Zhao-Lin[1)] YANG Heng[1),2)] WANG Qing[1)]

[1)](*School of Computer Science, Northwestern Polytechnical University, Xi'an* 710072)

[2)](*China Weapon Industry No.205 Research Institute, Xi'an* 710065)

Abstract Inspired by the depth perception mechanism of the human visual system, this paper proposes a novel globally consistent depth estimation method from defocus blur and stereo disparity depth-cues. According to the recent progress of light field theory, we can simulate the focus and defocus depth cue by using synthetic aperture photograph technique. Firstly, the defocus blur and stereo disparity depth cues are extracted from light field data sets, which is acquired by using a camera array system. Then based on the characteristic of the two depth cues, we design a fusion algorithm for the light field depth estimation. To acquire the globally consistent structural depth result, we introduce an adaptive weighted smoothing function in Markov random field framework. Finally, the global energy is minimized by graph cut algorithm, which leads a consistent and precise depth estimation. We test the proposed method on both virtual data and real data, the experimental results have shown that our method can take advantage of the two depth-cues and obtain more robust depth estimation.

Keywords camera array; light field analysis; multiple cues fusion; depth estimation

---

收稿日期:2014-05-06;在线出版日期:2014-11-02.本课题得到国家自然科学基金(61103060,61272287)、国家"八六三"高技术研究发展计划项目基金(2012AA011803)和教育部博士点基金(20116102110031)资助.**杨德刚**,男,1987 年生,硕士,主要研究方向为计算机视觉、光场理论及应用.E-mail:yiiwood@gmail.com.**肖照林**,男,1984 年生,博士研究生,主要研究方向为计算机视觉和计算摄影学.**杨 恒**,男,1981 年生,博士,主要研究方向为计算机视觉.**王 庆**,男,1969 年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为计算机视觉和模式识别.

1 引 言

三维场景几何结构获取是计算机视觉的基础问题之一.人类认识和表达世界是以几何结构信息为基础的,获取场景的几何结构能使人更好地理解视觉内容,同时可以为其他计算机视觉任务提供更加有效的数据支持.深度结构信息的获取包括主动式和被动式两大类方法[1],主动式深度获取方法主要包括激光扫描、TOF 相机、结构光法、阴影法等,相关技术的研究比较成熟,但是适用范围有限.被动式深度获取方法直接采集复杂场景的图像数据,用数据驱动的方式重建复杂场景的几何结构,具有与主动式方法不同的研究意义和实用价值.其中基于图像的被动式深度估计及三维结构恢复方法以其显著的优点得到了迅速发展.该类方法主要包括从运动、亮度、模糊、轮廓、纹理等视觉线索恢复场景深度[2].虽然不同方法进行深度估计的角度不同,但是具体到某种方法,大都仅使用了人类视觉系统中模糊、视差、阴影、纹理、透视等单一的深度线索[3],使得对场景的深度估计存在精度或鲁棒性不足的问题.尤其在处理多深度复杂场景时,采用单一深度线索估计的深度误差更大[4].

视觉生理学和心理学的研究表明,人类视觉的深度感知融合了模糊、视差、运动、方向、颜色等多种深度线索.由不同的脑区分层抽象出不同的概念和深度判断,最终组织成感知深度的一种深层层次结构[5].Parker 等也发现,不同的视觉线索由大脑皮层不同区域响应,然后大脑将这些响应通过融合形成深度感知[6].在计算机视觉领域,深度估计最常用的是模糊与视差线索,计算模糊线索需要同一视点不同聚焦变化采集场景不同模糊度的数据,而视差线索的获取则需要从不同视点对目标场景进行图像采集,为了模拟人脑的深度感知机理融合模糊与视差深度线索,需要同时采集更加丰富的场景数据.而计算摄影学(Computational Photography)中的光场相机(Light Field Camera)[7-10]采集的数据恰能满足这一要求.光场相机对三维场景中的光线能同时进行位置和角度采样,进而形成光场数据集[11-12].以这些稠密的采样数据集为基础,光场成像技术可以实现全聚焦(All-in-focus)图像合成、景深扩展(Extension of Depth of Field)和数字重聚焦(Digital Refocus)[10-11,13]等一系列特殊成像效果.光场数据中的模糊和视差是场景深度在不同维度上的表现,因此可以将其作为场景三维重建过程的有效深度线索.基于上述观点,本文提出一种融合模糊与视差线索的深度估计方法.

本文首先利用摄像机阵列采集光场数据,然后对获得的光场进行重投影,并从重投影后的光场中提取表示光线角度变化的视差线索,随后对光场进行角度积分获得共聚焦图像,并从中提取模糊线索.由于两种线索具有一定的互补性[14],针对不同线索的优势,本文设计了一种自适应权重方法对两种线索进行融合.为了获得全局一致的结构化深度结果,本文引入了马尔可夫随机场(Markov Random Field)模型,并设计了鲁棒的邻域平滑项,最终通过图割算法(Graphcut)最小化目标能量函数估计场景的深度.在虚拟数据和真实数据上测试结果显示,相比于单一深度线索和局部深度估计方法,本文所提方法能结合不同深度线索的优势获得更加鲁棒的深度估计结果.

2 相关工作

在计算机视觉深度结构重建领域,大量学者的研究实践表明,模糊线索和视差线索是两种非常有效的深度线索,这两种线索对应深度信息在两个典型维度的表现,在恢复场景深度方面被广泛使用.

单目视觉下的深度估计主要依靠模糊、阴影纹理变化及梯度、运动极差等线索,目前使用较多的是模糊线索.Krotkov[15]和 Pentland[16]首先提出了从聚焦恢复深度(Depth From Focus,DFF)和从散焦恢复深度(Depth From Defocus,DFD)的理论模型,利用图像点模糊程度估计场景深度.Saxena 等人[3]建立了基于分层多尺度马尔可夫随机场的有监督学习深度估计方法,利用像素之间的上下文一致性估计其深度.Burge 等人[17]从视觉心理学出发分析了模糊产生的生理机制,提出一种基于贝叶斯模型的图像散焦估计方法.Hasinoff 等人[18]则提出了共焦一致性(Confocal Constancy)假设获得图像中每个像素点的散焦度量.常用的模糊度量值有灰度方差、局部梯度、局部直方图熵值等.Nayar 等人[19]则提出了基于改进拉普拉斯算子的 SML(Summed Modified Laplacian)模糊度量方法.由于 DFF/DFD 方法需要多张关于目标场景不同聚焦参数的图像,单相机必须进行多次拍摄,这使其无法估计动态场景的深度信息.而作为光场数据获取系统的一种,摄像机阵列不仅具有一般多摄像机系统的多视点、高信噪比等优点,而且可以虚拟一个大孔径相机以获

取更高的景深分辨率.现有合成孔径重聚焦理论[20]仅需一次拍摄就能获得场景任意候选焦深的图像[21].基于相机阵列光场数据的深度估计方法可以弥补传统模糊法中孔径小及需要多次拍摄的不足,适用于较远场景以及动态场景的深度估计.

多视点深度估计的一个主要线索是三维场景经过多个相机成像在不同图像之间形成的视差,在光场数据中即不同角度光线之间的差异,视差的计算过程也即多视点图像之间的特征匹配过程.常用的视差匹配值的计算方法有 SSD(Summed Squared Difference)、SAD(Summed Absolute Difference)、灰度相关法、秩方法、归一化互相关法(Normalized Cross Correlation,NCC)、自适应窗口法(Adaptive Support-Window)[22] 等.虽然这些局部代价集成(Cost Aggregation)方法计算效率较高,但是对匹配噪声难以抑制,影响深度估计的准确性.在视差匹配中,由于真实凸透镜相机存在孔径限制,超出景深的区域会出现一定程度的模糊,从而使得匹配精度降低,影响深度估计的准确性.另外,如果场景中出现重复纹理或无纹理区域,视差线索的可靠性也将下降.

利用深度线索直接进行深度估计的结果往往噪声较大,为了利用邻域平滑约束获得全局一致的深度估计结果,可以利用马尔可夫随机场模型结合深度线索和上下文平滑约束,将深度估计表示为能量泛函最小化过程,并利用图割法[23]、置信传播法[24]等优化方法获得深度结果.传统马尔可夫随机场深度估计模型中,平滑权重参数只能通过一些启发式的方法进行设置,针对这一情况,Scharstein 等人[25]将条件随机场(Conditional Random Field)引入到视差立体匹配中,利用训练集自动地学习平滑参数,提高结构优化效果.Li 等人[26]在最大间隔(Max-Margin)框架下,利用结构支持向量机(Structured Support Vector Machine)学习非参损失函数进行深度估计.

根据不同深度目标在光场中的差异,Wanner 等人[27]提出了利用结构张量提取光场结构线索.为了在无混叠噪声的情况下降低光场的角度采样率,Liang[28]提出了与深度相关的视点插值算法提取深度线索.Kim 等人[29]利用稠密的 3D 光场提取深度线索,先估计较可靠的目标边缘的深度,然后再重建平滑的内部区域.

研究表明,模糊与视差线索是场景深度变化在不同维度上的表现.因此,借鉴人类视觉的深度感知机制,研究人员从不同角度提出了一些利用多线索融合解决深度估计的方法.Frese 等人[30]利用不同焦距的摄像机阵列获取场景深度,该方法融合了视差线索和模糊线索.Li 等人[31]提出了一种从不同聚焦参数的双视图像估计模糊核和视差图的方法.在相机阵列合成孔径成像的基础上[20],Vaish 等人[32]设计了对遮挡鲁棒的目标函数重建被遮挡目标表面结构.Wang 等人[33]对三维显示中影响人眼深度感知的模糊与视差线索的关系进行了量化分析.这些方法虽然都涉及利用模糊和视差线索提取场景的深度信息,但是如何将这两种线索自适应融合,构建统一的联合计算模型是一个尚未解决的问题.本文正是从这一角度出发,通过统一的光场框架分析模糊与视差线索的适用特性,设计融合算法实现两种深度线索优势互补,最终在马尔可夫随机场模型中结合鲁棒的邻域平滑约束构建多深度线索深度估计的统一计算模型,并以最大后验概率得到深度求解的目标能量函数,采用图割算法进行能量最小化获得一致的深度估计结果.

3　光场多深度线索分析

光线在空间中的传播过程可以用七维函数来表示,称之为全光函数(Plenoptic Function)[34].全光函数 $P(x,y,z,\theta,\phi,\lambda,t)$ 中 (x,y,z) 表示空间中任意一点的三维坐标,(θ,ϕ) 表示光线的方向,λ 为光线的波长,t 为时间.全光函数主要用于研究光线与空间几何物体的交互特性,一旦确定了全光函数,我们就可以对光的各种变换进行模拟.但是全光函数的采样、存储等是非常困难的,因此必须进行简化,通常只考察静态场景以去掉时间维,同时假设每个颜色通道内光波长的影响可忽略,这样就得到了五维全光函数 $P(x,y,z,\theta,\phi)$.更进一步,如果去掉光线传播过程中衰减的影响,全光函数可以降至四维,Levoy 和 Hanrahan[35]将其命名为光场(Light Field),Gortler 等人[36]则称之为流明图(Lumigraph).由于光场包含丰富的场景结构信息,在获得光场数据后,利用计算摄影学的相关算法不仅可以恢复场景几何结构,还可以克服传统成像系统的一些局限性.

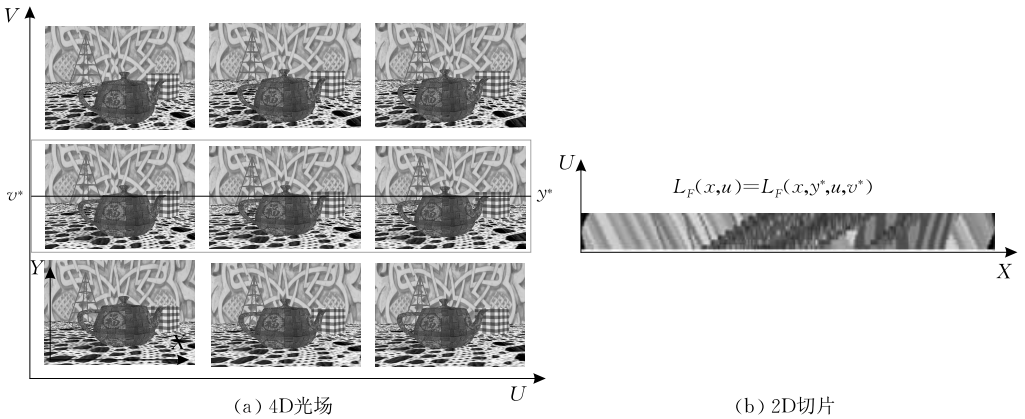
光场有多种不同的参数化形式,在 Levoy 等人[35]提出的光场理论中,利用两个平行平面分别记录光线的位置信息和角度信息,即得到光场的双平面参数化模型 $L_F(x,y,u,v)$,其中 (x,y) 和 (u,v) 分别为光线穿过位置平面和角度平面的坐标.为了方便地进行可视化和直观理解,通常将光场 $L_F(x,y,u,v)$ 的位置坐标 y 和角度坐标 v 设定为一个固定的值,这

样就得到了 4D 光场简化后的 2D 切片 $L_F(x,u)$，如图 1 所示. 其图示称为对极平面图(Epipolar Plane Images，EPI)[27]或光线空间图(Ray-Space Diagram，RSD)[11].

当在相机内部以镜头平面和成像平面参数化光场时，整个成像过程可以表示为光场对角度的积分：

$$I(x,y)=\iint L_F(x,y,u,v)G(\theta)\mathrm{d}u\mathrm{d}v \quad (1)$$

其中：$I(x,y)$为获得的图像；$G(\theta)$表示角度的加权值；θ是光线(x,y,u,v)与参数平面法向的夹角.



(a) 4D光场 (b) 2D切片

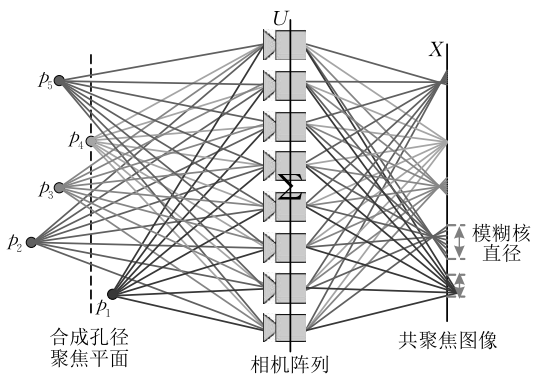
图 1 4D 光场及其 2D 切片

相机阵列是对整个空间光场的离散采样，每个相机的成像是光场在不同角度范围积分的结果：

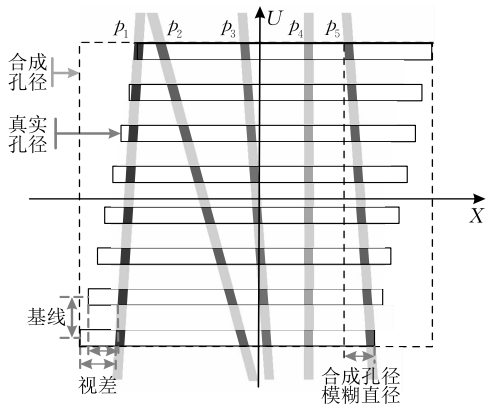
$$I_k(x,y)=\int_{u_{k,l}}^{u_{k,h}}\int_{v_{k,l}}^{v_{k,h}}L_F(x,y,u,v)G(\theta)\mathrm{d}u\mathrm{d}v \quad (2)$$

其中，$u_{k,l}$、$u_{k,h}$和$v_{k,l}$、$v_{k,h}$表示第 k 个相机的角度采样范围. 当相机排列较分散时，光场的角度采样率不足，会导致合成孔径成像获得的共聚焦图像出现混叠噪声.

摄像机阵列采集到光场数据后，利用合成孔径成像理论可以将整个相机阵列虚拟成一个大孔径相机，并可以通过重聚焦获得不同焦深的光场. 如图 2 所示为摄像机阵列对指定深度 d 重聚焦得到的光场，图 2(a)为不同深度目标的光线重聚焦后光线角

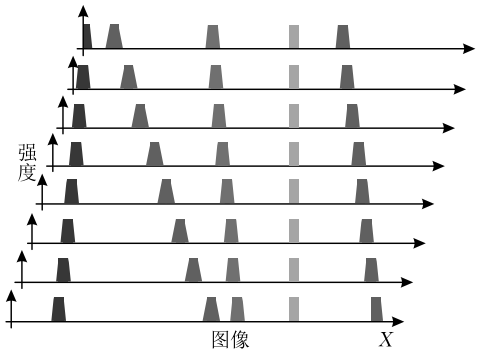


(a) 相机阵列重聚焦



(b) 重聚焦光场

(c) 共聚焦图像



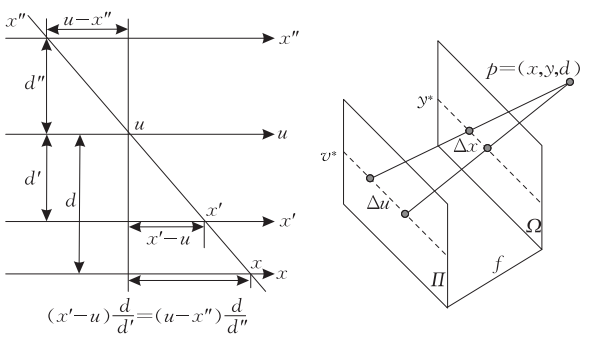
(d) 重投影图像

图 2 光场重聚焦及模糊与视差线索

度的分布，图 2(b)为由虚拟合成孔径平面和共聚焦成像平面参数化的光场. 设深度 d 处的合成共聚焦图像为 $I_k^d(x,y)$，则由重聚焦光场获得共聚焦图像的过程即是对 4D 光场进行离散角度积分的过程：

$$I_s^d(x,y)=\sum_k I_k^d(x,y)G(\theta_k) \tag{3}$$

这里 θ_k 为第 k 个相机光场采样的角度. 此时，如果目标处在合成孔径候选焦深处，则其在共聚焦图像中成清晰的像，如图 2(c)所示，而不在焦深处的目标，其共聚焦图像会有相应的模糊，模糊程度随着离焦深的距离增加而增大. 从图 2(b)中可以看到，相机阵列不同相机是对光场不同角度的采样，这样合成大孔径相机可以覆盖更大的角度采样范围，极大地提高模糊线索的景深分辨率. 但是如果只利用模糊线索，则积分过程中的视差信息会丢失，因此本文不仅在合成大孔径共聚焦图像中提取模糊线索，还在原始的重投影图像中提取视差线索，即对某一位置计算不同相机重投影图像的匹配值. 如图 2(d)所示，如果目标在这一合成孔径候选焦深处，则不同角度的光线对应于相同的位置，即垂直于位置轴，因此不同相机重投影图像目标位置处的匹配代价较小，否则匹配代价较大. 计算多个候选焦深同一位置的匹配代价，可以获得视差深度线索. 对较弱的纹理区域，不同深度视差匹配值呈单峰变化，视差线索对深度的分辨能力较高；而对强纹理区域如果候选深度与真实深度稍有不同，其视差匹配代价就会很大，与其他远离真实深度的候选深度视差匹配代价差异较小. 同样地，对重复纹理区域，不同候选深度视差匹配代价呈多峰形式，因此视差线索对强纹理区域和重复纹理区域的匹配鲁棒性不高.



(a) 改变光场位置参数平面　　(b) 位置和角度的相对变化

图 3　改变光场的位置参数化平面及光线簇位置和角度的相对变化关系

根据重聚焦理论，改变合成孔径重聚焦焦深相当于改变了光场的位置参数化平面，与新位置平面相交于 x' 的光线与原位置平面相交于 $(x'-u)\dfrac{d}{d'}$，如图 3(a)所示，这时光场变为

$$L_{F'}(x',y',u,v)=L_F\left(u+(x'-u)\frac{d}{d'},v+(y'-v)\frac{d}{d'},u,v\right) \tag{4}$$

即新的光场表示只是原光场的位置参数缩放和平移的结果. 利用这一特性，在要获得多个候选重聚焦深度的重投影图像时，可以加快计算速度. 图 3(b)为参数化光场中点 p 的光线簇在位置轴和角度轴的相对变化关系，根据文献[27]得

$$\frac{\Delta x}{\Delta u}=-\frac{f-d}{d} \tag{5}$$

其中：f 为参数化双平面之间的距离；d 为点 p 的深度. 当利用重投影图像平面参数化光场时，相当于对原始光场进行了图 3(a)中 x'' 位置的平移，新的位置轴偏移与原始位置轴偏移符号相反，式(5)变为

$$\frac{\Delta x}{\Delta u}=\frac{f-d}{d}\Rightarrow\Delta u=\frac{d}{f-d}\Delta x \tag{6}$$

即在重聚焦光场中深度小于重聚焦深度 f 的点，其光线簇斜率为正，并且深度越小斜率也越小；深度大于重聚焦深度 f 的点，其光线簇斜率为负，并且深度大斜率也越大. 即与重聚焦深度差越大，光线簇相对于角度轴 u 倾斜角越大，从而共聚焦图像越模糊，其相应的视差匹配代价也越大. Wanner 等人[27]即直接利用光场中不同深度目标的光线簇斜率与深度的关系及一致性约束建立全局目标函数，优化求解获得场景深度，然而这种直接利用光场特性进行深度估计的方法其适用范围有限.

对不同的候选焦深重聚焦，可以获得一组共聚焦图像和重投影图像. 图 4 所示为不同重聚焦深度计算模糊线索的示意图，在共聚焦图像中，与重聚焦深度 d 接近的目标比较清晰，而远离深度 d 的目标则比较模糊，因此可以根据不同重聚焦深度模糊线索值的变化，选择模糊值最小的深度作为目标点的估计深度. 对强纹理区域或边缘，由于在焦深和远离焦深其模糊线索相差较大，因此模糊线索可靠性较高. 视差线索中不同光场采样角度之间的匹配也即重投影图像的匹配，如图 5 所示为不同重聚焦深度的光场，可以看到不同光场采样角度之间匹配代价的变化，只有重聚焦深度 d_2 与目标深度相同，匹配代价才最小，否则匹配代价较大，因此可以通过选择最小匹配代价求得目标深度. 模糊线索度量光场积分后空间位置的变化，而视差线索则是对光场不同角度匹配情况的度量，两种线索代表光场中表征深度变化的两个维度，具有一定的互补性.

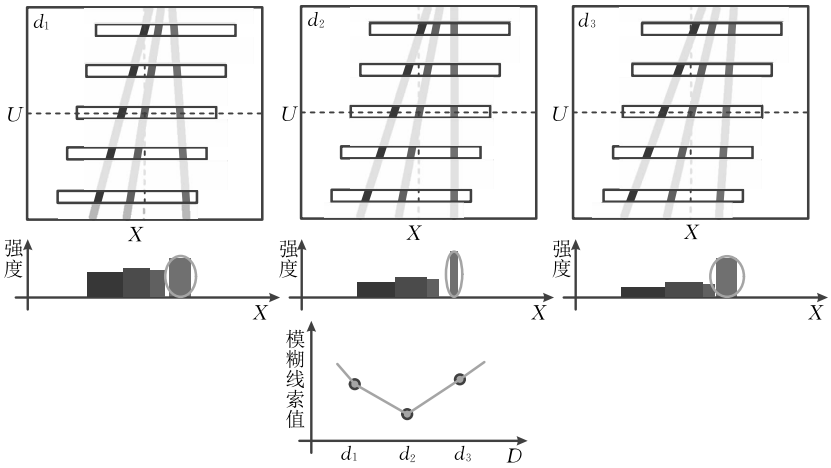


图 4 不同重聚焦深度的模糊线索

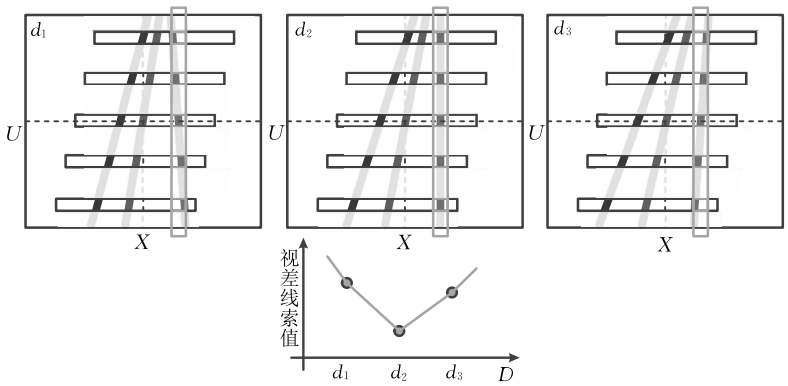


图 5 不同重聚焦深度的视差线索

4 多线索融合全局一致深度估计

基于上述对光场数据中包含的模糊与视差线索的分析,结合各个深度线索的适用情况,本文提出了自适应融合模糊与视差线索的全局一致深度估计方法,算法的整体框架如图 6 所示. 首先利用重聚焦理论获得 L 个候选深度 $d_i\{i=1,\cdots,L\}$ 的重投影图像 $I_k^{d_i}(x,y)$,并提取每个位置表示光线角度变化的视差线索. 同时利用重投影图像合成共聚焦图像 $I_s^{d_i}(x,y)$,并从中提取表示空间变化的模糊线索,然后对获得的不同深度模糊与视差线索,利用自适应权重进行融合. 为了获得全局一致的结构化深度结果,本文在马尔可夫随机场模型中结合鲁棒的成对结点势函数,通过图割算法进行最大后验概率推理获得场景深度.

对于任意候选深度 d 的重投影图像 $I_k^d(x,y)$,可以利用式(3)计算这一深度共聚焦图像 $I_s^d(x,y)$,度量共聚焦图像中目标点模糊度的算法有很多,由于本文的目标是验证多深度线索融合与优化方法,因此采用传统高效的共聚焦图像的灰度方差[15]作为模糊线索的度量:

$$blur_p(d)=1-\frac{1}{z_{blur}}\sum_{q\in N_p}(I_{s,q}^d-\bar{u}_p^d)^2 \qquad (7)$$

其中 $\bar{u}_p^d$ 表示深度 d 处共聚焦图像中以点 p 为中心的邻域内像素的灰度均值,N_p 表示点 p 的邻域,z_{blur} 为归一化因子. 如图 7 所示,点 p 的邻域灰度方差越大表示其越清晰,模糊线索值就越小;相反,灰度方差越小表示其越模糊,模糊线索值就越大.

在立体匹配中,有多种代价集成方法可以计算像素邻域的视差匹配值,本文采用不同重投影图像中点 p 邻域的 SSD 作为视差线索:

$$disp_p(d)=\frac{1}{z_{disp}}\sum_{q\in N_p}\sum_{k=1}^{N}\|I_{k,q}^d-\bar{I}_q^d\|_2 \qquad (8)$$

其中 $\bar{I}_q^d$ 表示深度 d 的所有重投影图像 $I_k^d\{k=1,\cdots,N\}$ 在位置 q 的颜色向量平均值,z_{disp} 为归一化因子. 如果点 p 的真实深度与候选深度 d 越接近,其在不同重投影图像中的邻域对应就越准确,相应的视差

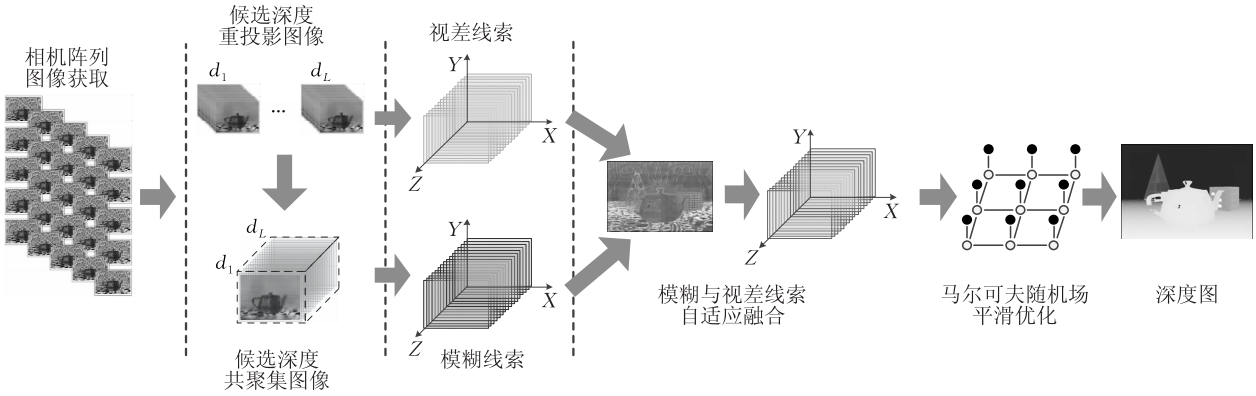


图 6　模糊与视差融合全局一致深度估计框架

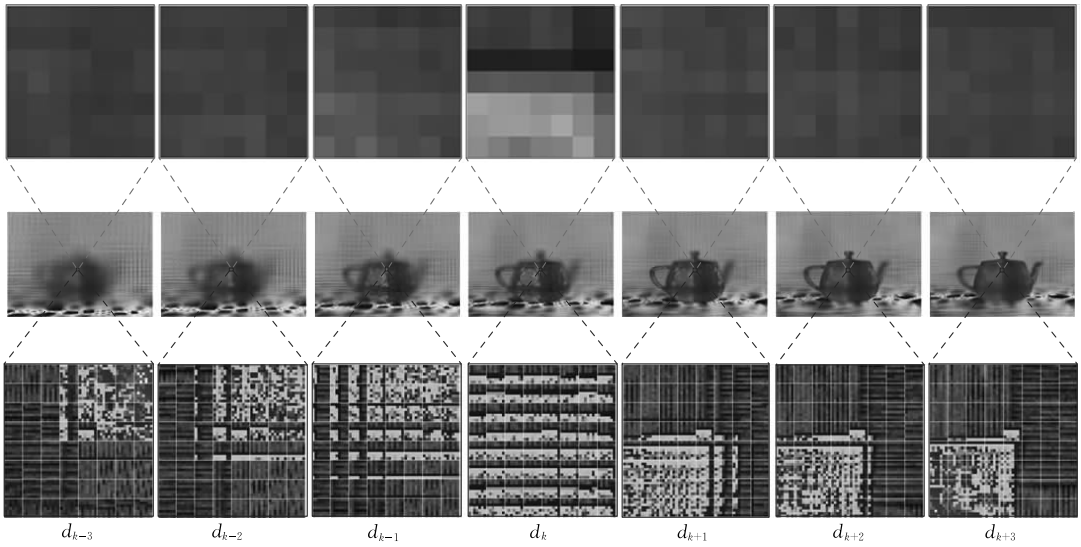


图 7　点 p 在连续 7 个聚焦深度共聚焦图像及重投影图像中的邻域窗口

线索值就越小，即匹配度越高.

基于第 3 节的分析，模糊线索对强纹理及边缘深度判别可靠性较高；而视差线索对弱纹理区域效果较好，对强纹理和重复纹理处则不够鲁棒. 两种不同深度线索表示深度信息在光场数据集的不同特征，彼此具有一定的互补性. 因此在融合过程中，为了达到优势互补，本文采用了类似文献[37]的自适应融合方法：

$$V_p(d)=\lambda_p blur_p(d)+(1-\lambda_p)disp_p(d) \quad (9)$$

其中 λ_p 是自适应权重. 如果深度线索值在不同的候选深度差别较大，那么该线索对深度的区分性就较强，应该分配其获得更大的融合权重. 因此，本文设计自适应的权重分配计算方式为

$$\lambda_p=\frac{C_{blur}(p)}{C_{blur}(p)+C_{disp}(p)} \quad (10)$$

$$C_{blur}(p)=\frac{1}{\sum_{i=1}^{L}\exp\left(\frac{-(blur_p(d_i)-blur_{p,\min})^2}{2\sigma_{blur}^2}\right)} \quad (11)$$

$$C_{disp}(p)=\frac{1}{\sum_{i=1}^{L}\exp\left(\frac{-(disp_p(d_i)-disp_{p,\min})^2}{2\sigma_{disp}^2}\right)} \quad (12)$$

其中：共有 L 个候选深度，$blur_{p,\min}$ 和 $disp_{p,\min}$ 分别表示点 p 在不同候选深度中模糊线索和视差线索的最小值；σ_{blur} 和 σ_{disp} 是相应的参数. 框架图 6 中融合阶段的灰度图表示计算得到的自适应权重.

当获得了点在不同候选深度的判别值后，可以通过直接求解：

$$\hat{d}_p=\arg\min_{d_i} V_p(d_i) \quad (13)$$

获得每个点的深度，也可以对融合前的模糊与视差线索分别利用式(13)直接求解，但是直接求解方法获得的深度结果噪声比较大，因为没有利用每个点的邻域信息，为了利用邻域平滑获得全局一致的深度结果 $\hat{\boldsymbol{D}}$，本文利用马尔可夫随机场表示深度估计问题，其后验概率密度为

$$P(\boldsymbol{D}|\boldsymbol{V})\propto\prod_{p}\Phi(d_p,\boldsymbol{V}_p)\prod_{p,q}\Psi(d_p,d_q) \quad (14)$$

其中：$\Phi(d_p,\boldsymbol{V}_p)=\mathrm{e}^{-E_p(d_p,\boldsymbol{V}_p)}$ 表示深度 d_p 与融合线索 $\boldsymbol{V}_p$ 的匹配似然；$\Psi(d_p,d_q)=\mathrm{e}^{-\lambda E_{pq}(d_p,d_q)}$ 为马尔可夫随机场中邻域节点之间的平滑先验，λ 表示邻域约束强度. 最大化后验概率进行深度估计：

$$\hat{\boldsymbol{D}}=\arg\max_{\boldsymbol{D}}P(\boldsymbol{D}|\boldsymbol{V})=\arg\max_{\boldsymbol{D}}\log(P(\boldsymbol{D}|\boldsymbol{V})) \quad (15)$$

化简得到相等的最小化能量函数形式：

$$\begin{aligned}\hat{\boldsymbol{D}}&=\arg\min_{\boldsymbol{D}}E(\boldsymbol{D})\\&=\arg\min_{\boldsymbol{D}}E_{\text{data}}(\boldsymbol{D})+\lambda E_{\text{smooth}}(\boldsymbol{D})\\&=\arg\min_{\boldsymbol{D}}\sum_p E_p(d_p,V_p)+\lambda\sum_{p,q}E_{pq}(d_p,d_q)\end{aligned} \quad (16)$$

其中数据项为模糊与视差线索的融合结果：

$$E_{\text{data}}(\boldsymbol{D})=\sum_p E_p(d_p,\boldsymbol{V}_p)=\sum_p \boldsymbol{V}_p(d_p) \quad (17)$$

平滑项表示局部邻域的平滑约束，本文考虑的两个目标是：(1) 在图像坐标空间与颜色空间距离相近的点要有相似的深度；(2) 对深度突变处避免过平滑导致结果错误. 基于以上的两点考虑，本文借鉴立体匹配[22]中的思想，定义任意点与其邻域点的平滑约束权重

$$w(p,q)=\exp\left(-\frac{\|\boldsymbol{I}_p-\boldsymbol{I}_q\|_2}{r_c}-\frac{\|\boldsymbol{G}_p-\boldsymbol{G}_q\|_2}{r_g}\right) \quad (18)$$

其中：$\boldsymbol{I}_p$、$\boldsymbol{I}_q$ 分别表示点 p、q 的颜色向量；$\boldsymbol{G}_p$、$\boldsymbol{G}_q$ 表示点 p、q 的坐标向量. 当两个相邻点距离较近并且颜色较相似时，两个点的深度平滑约束较强，否则平滑约束较弱. 平滑项计算公式为

$$E_{\text{smooth}}(D)=\sum_p\sum_{q\in N_p}w(p,q)|d_p-d_q| \quad (19)$$

基于式(19)得到的平滑约束，既能平滑相似点深度值，又能在深度突变处自适应地减弱约束. 对式(16)的目标能量函数，本文用图割[23,38-39]算法进行最小化获得最终的深度结果，如算法 1 所示.

算法 1. 多线索融合深度估计.

输入：原始光场数据 $\boldsymbol{L}_F^0$

输出：深度 $\hat{\boldsymbol{D}}$

```
FOR k=1 To L DO
  L_F^k=Refocus(L_F^0,d_k)            //合成孔径重聚焦
  FOR ∀p∈ν DO
    blur_p(d_k)=Blur(L_F^k,p)         //式(7)
    disp_p(d_k)=Disparity(L_F^k,p)    //式(8)
  END FOR
END FOR
FOR ∀p∈ν DO
  计算融合权重 λ_p                    //式(10)～(12)
END FOR
FOR ∀{p,q}∈ε DO
  计算平滑约束权重 w(p,q)             //式(18)
END FOR
构建马尔可夫随机场模型 P(D|V)        //式(14)
用图割算法求解能量最小化得到深度 D̂   //式(16)
```

5 实验结果与分析

为了验证提出算法的有效性，本文分别在虚拟数据和真实数据上进行了实验，并与现有流行的深度估计方法进行了比较. 通过对结果精细程度和计算量的权衡，本文选择在场景目标的深度范围内均匀获得 100 个重投影深度.

5.1 虚拟数据实验

虚拟数据是对一个绘制好的三维场景，利用一个 8×8 的虚拟相机阵列对该场景进行拍摄，从而获得 64 幅不同视角的原图像，每幅图像的分辨率为 780×538. 如图 8(a)所示的简单虚拟场景是由处于

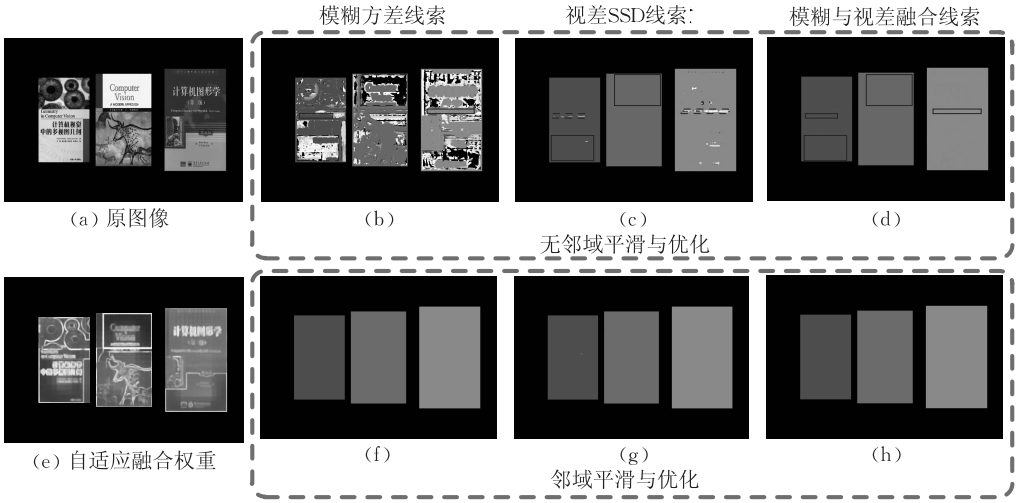


图 8 简单虚拟场景的深度估计结果

3 个不同深度的书组成，对 12 个不同深度重投影获得的光场如图 9，可见在目标所在的焦深，由于目标同一位置不同角度光线来自于空间同一点，其光场不随角度而变化，这时获得的共聚焦图像目标位置模糊判别值小，同样相机阵列不同角度相机之间的视差匹配代价也较小.

图 9　12 个不同重聚焦深度的光场，3 本书所在深度分别为 d_{k_1}、d_{k_2} 和 d_{k_3}

本文计算模糊线索的窗口大小为 7×7，图 7 最上面一行显示了一个复杂虚拟场景某点 p 在连续 7 个聚焦深度共聚焦图像中模糊方差线索的计算窗口，从图中可以看到，在点 p 的真实深度 d_k 处，其附近邻域像素梯度较大，式(7)的模糊方差判别值较小，其他聚焦深度点 p 的邻域则较为模糊. 视差 SSD 深度线索需要计算某一聚焦深度不同相机重投影图像中同一点的匹配代价，图 7 最下一行为点 p 在相应聚焦深度的视差邻域窗口，窗口大小为 7×7，可以看到在点 p 所在的深度 d_k 处，其在不同相机之间邻域较相似，式(8)视差匹配代价较小.

图 10 为点 p 在图 7 所示窗口中计算所得的模糊方差、视差 SSD 及融合线索值曲线，可见在点 p 的真实深度处，两个线索值均为最小，但是模糊线索值的差异更大，即对深度的区分性更强，所以融合权重更大，融合后的线索也保持了模糊线索的强区分性，从图 7 中可见点 p 处在边缘处，这就验证了第 3 节的分析，即在边缘处模糊线索的可靠性要高于视差线索.

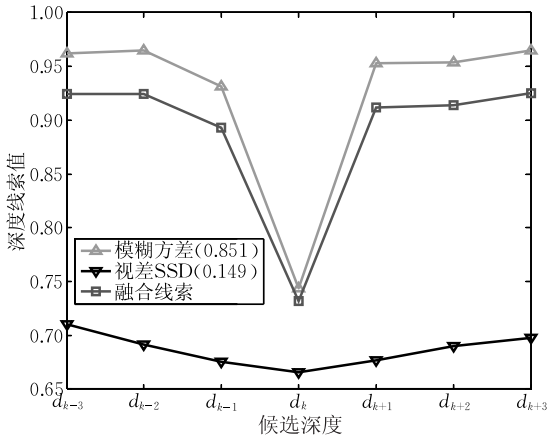


图 10　图 7 点 p 处计算的模糊方差、视差 SSD 及融合线索值（模糊方差与视差 SSD 的融合权重分别为 0.851 和 0.149）

图 8 为简单虚拟场景目标区域的深度估计结果，其中(b)、(c)、(d)没有邻域平滑与优化，即由式(13)直接取每个点的最小判别值所在深度为估计结果，其中模糊、视差及融合线索的准确率分别为 59.0%、98.9%和 99.8%，即自适应融合线索深度准确率相比于单一线索有所提高，图中虚线及实线方框分别表示结果较差和结果较好的区域，可以看到融合后的结果确实利用了不同线索的优势，这证明本文的自适应融合方法是有效的.

图 8(e)为用灰度表示的自适应融合权重图，可见边缘和强纹理处模糊线索权重较大，这些也正是模糊线索可靠性较高的区域；而弱纹理处视差线索权重较大，与第 3 节分析的视差线索适用区域一致. 另外由于此虚拟场景较简单，所以不需要邻域平滑准确度就已经很高，当对各个线索利用马尔可夫随机场进行平滑优化时，无论原始单一深度线索还是融合线索，其深度估计结果准确率都几乎达到 100%，并没有差异.

复杂虚拟场景的深度估计结果如图 11 所示. 由于场景比较复杂，对融合线索直接求解而无邻域平滑与优化的结果噪声还是比较大，但是在有些区域还是获得了模糊与视差中较好的结果. 图 11(f)、(g)、(h)是利用马尔可夫随机场邻域平滑与优化的结果，可见相比于直接求解，利用邻域约束获得的深度估计结果中噪声得到了较好的抑制，图 11(i)自适应融合权重和图 11(h)深度结果表明本文的方法较好地利用了不同线索的优势，而且目标边缘深度突变处深度结果明显好于直接求解结果，表明本文提出方法的有效性. 图 11(e)为场景的真实深度，(f)、(g)、(h)与真实深度的相对差异显示在(j)、(k)、

(1),从图中可见本文方法获得的深度结果大部分深度误差较小,而视差 SSD 的结果由于图像边缘处光场的非均匀采样,导致匹配代价计算不准确,使边缘处结果深度误差较大.但是经过多深度线索融合与马尔可夫随机场平滑约束的传播,视差线索边缘的错误并没有对最终的深度结果造成影响,由此表明本文所提出的模型与单一深度线索方法相比具有较强的鲁棒性.

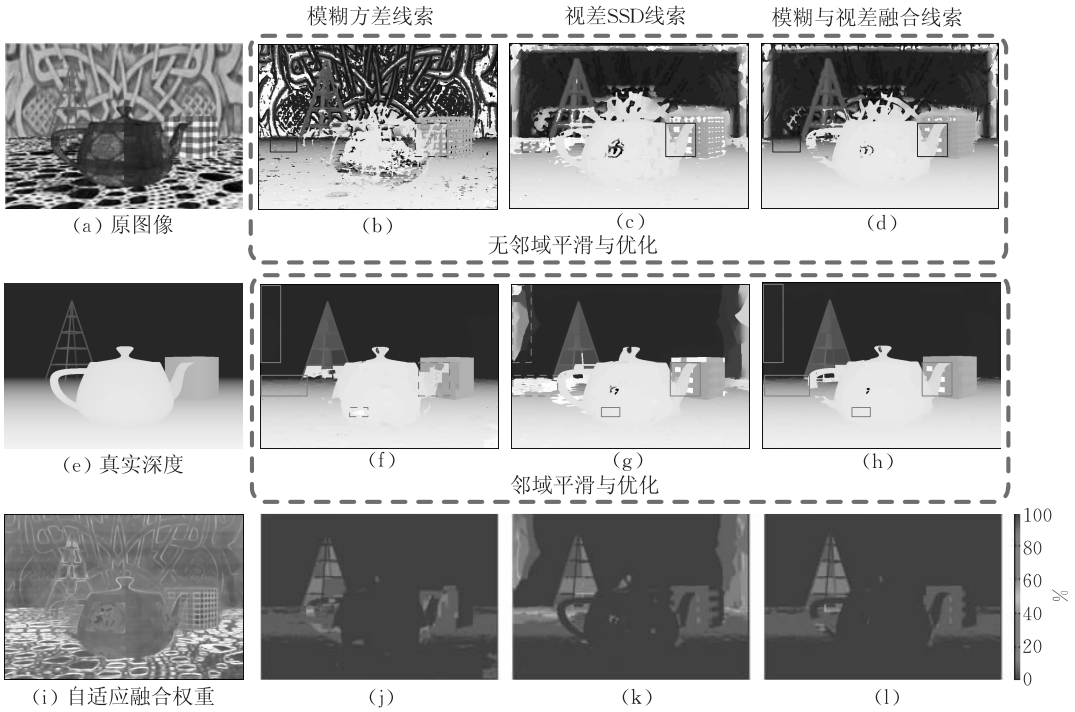


图 11 复杂虚拟场景的深度估计结果

图 12 为使用和不使用平滑优化的深度错误率随允许误差的变化曲线,可以看到利用马尔可夫随机场平滑优化后的结果明显好于未平滑优化的结果,证明邻域平滑是有效的.同时融合多深度线索也较单一线索能够获得更准确的深度结果.

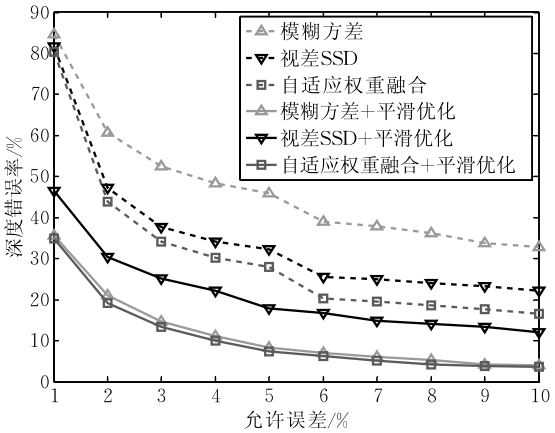


图 12 复杂虚拟场景的深度错误率曲线

5.2 真实数据实验

由于本文的方法需要相机阵列获得的光场数据用于深度估计,已公开的标准深度估计数据集中,并没有符合要求的相机阵列数据.因此为了获取真实场景的光场数据,本文搭建了由 64 台工业相机组成的相机阵列,采用的相机型号为 1H046C.

两组真实场景深度估计结果如图 13 和图 14 所示,其中图 13(b)、(c)和图 14(b)、(c)利用模糊与视差线索直接求解获得的结果噪声都比较大,而通过马尔可夫随机场进行邻域约束的传播,深度噪声得到了较好的抑制,获得的结果更加平滑准确(如图 13(f)、(g)、图 14(f)、(g)).图 13 和图 14 的结果表明相比于单一深度线索,融合深度线索在很大程度上能够利用不同深度线索的优势,利用融合深度线索进行深度求解获得的结果更加可靠,对原始深度线索中的噪声不敏感.实验中为了验证光场角度采样率对深度估计结果的影响,图 14 的场景只用了 9 台相机进行拍摄,其模糊线索的深度估计结果较差,原因主要是进行模糊线索计算的共聚焦图像由式(3)得到,即是光场的积分过程,相机数量少导致光场角度采样率严重不足,获得的共聚焦图像信号混叠严重,影响了模糊线索的可靠性.不过,由于视差线索只需进行不同相机重投影图像之间的匹配计算,所以不受光场角度采样率不足的影响,结果较好.

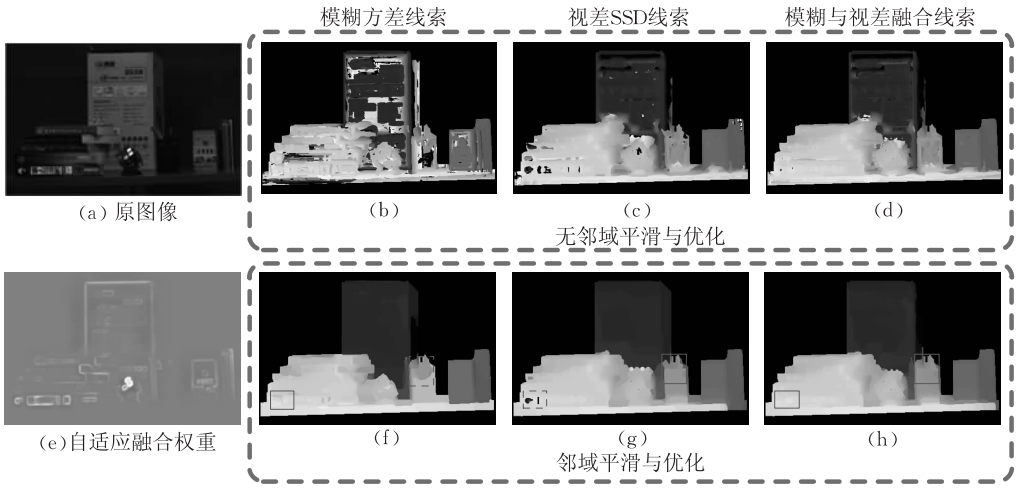


图 13　8×8 相机阵列获取真实场景的深度估计结果

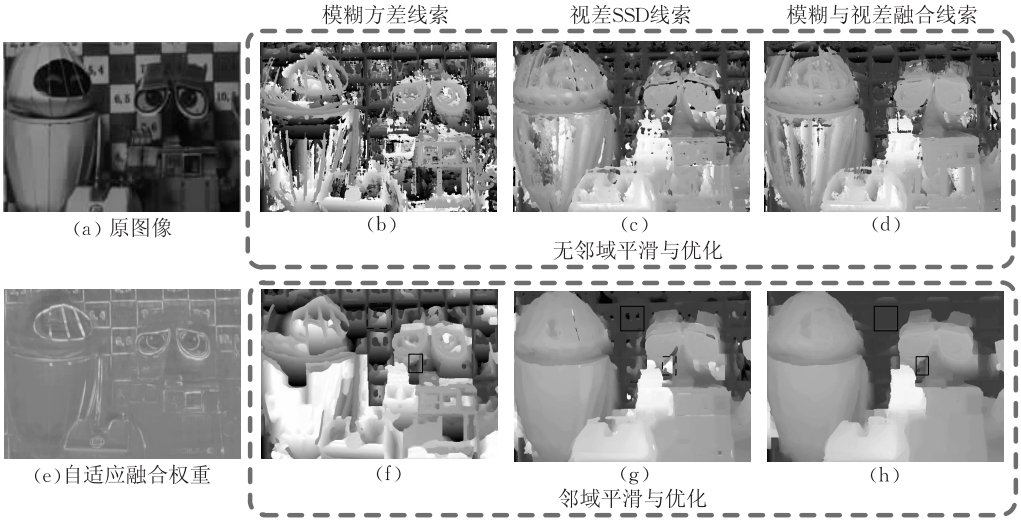


图 14　3×3 相机阵列获取真实场景的深度估计结果

6　结　论

本文提出了一种基于光场数据多线索融合的全局深度估计方法. 根据对相机阵列光场数据中包含的模糊与视差深度线索的分析，本文首先利用合成孔径重聚焦理论获得指定深度重投影图像，并对重投影图像进行光场积分得到共聚焦图像. 随后，分别从共聚焦图像和重投影图像提取模糊与视差线索. 进而，本文对模糊与视差线索进行了自适应融合，并利用马尔可夫随机场平滑优化对深度的计算，以获得鲁棒的深度估计结果. 实验结果验证了深度线索融合方法能够利用不同线索的优势，自适应的邻域平滑方法使深度估计结果更加鲁棒. 然而从实验结果也可以看出，本文的方法在深度突变剧烈及不同深度纹理相似度较高的区域结果仍存在一些不足，尤其对于存在遮挡的区域，本文方法未能很好地解决深度信息丢失的问题. 另外，本文能量函数中的整体平滑项权重 λ 及相邻点之间平滑权重计算中的 r_c、r_g 是通过多次试验获得的较优参数值. 后续工作中，本文作者将研究更加鲁棒的平滑约束方法和处理遮挡对深度线索的影响，并采用基于统计学习的方法自适应地学习相关参数值.

参　考　文　献

[1] Herbort S, Wöhler C. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. 3D Research, 2011, 2(3): 1-17

[2] Shu Bo, Qiu Xian-Jie, Wang Zhao-Qi. Survey of shape from image. Journal of Computer Research and Development, 2010, 47(3): 549-560(in Chinese)
(束搏，邱显杰，王兆其. 基于图像的几何建模技术综述. 计算机研究与发展，2010，47(3)：549-560)

[3] Saxena A, Chung S H, Ng A Y. 3-D depth reconstruction

from a single still image. International Journal of Computer Vision, 2008, 76(1): 53-69

[4] Saxena A, Schulte J, Ng A Y. Depth estimation using monocular and stereo cues//Proceedings of the International Joint Conference on Artificial Intelligence. Hyderabad, India, 2007: 2197-2203

[5] Bullier J. Integrated model of visual processing. Brain Research Reviews, 2001, 36(2-3): 96-107

[6] Parker A J. Binocular depth perception and the cerebral cortex. Nature Reviews Neuroscience, 2007, 8(5): 379-391

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

[8] Liang C K, Shih Y C, Chen H H. Light field analysis for modeling image formation. IEEE Transactions on Image Processing, 2011, 20(2): 446-460

[9] Bishop T E, Favaro P. The light field camera: Extended depth of field, aliasing, and superresolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(5): 972-986

[10] Georgiev T, Lumsdaine A. Focused plenoptic camera and rendering. Journal of Electronic Imaging, 2010, 19(2): 021106-021106-11

[11] Ng R. Digital Light Field Photography[Ph. D. dissertation]. the Stanford Computer Science Department, Stanford University, Palo Alto, California, USA, 2006

[12] Georgiev T, Zheng K C, Curless B, et al. Spatio-angular resolution tradeoffs in integral photography//Proceedings of the 17th Eurographics Symposium on Rendering (EGSR). Nicosia, Cyprus, 2006: 263-272

[13] Bishop T E, Favaro P. Full-resolution depth map estimation from an aliased plenoptic light field//Proceedings of the Asian Conference on Computer Vision. Queenstown, New Zealand, 2010: 186-200

[14] Held R T, Cooper E A, Banks M S. Blur and disparity are complementary cues to depth. Current Biology, 2012, 22(5): 426-431

[15] Krotkov E. Focusing. International Journal of Computer Vision, 1988, 1(3): 223-237

[16] Pentland A P. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987, 9(4): 523-531

[17] Burge J, Geisler W S. Optimal defocus estimation in individual natural images. Proceedings of the National Academy of Sciences, 2011, 108(40): 16849-16854

[18] Hasinoff S W, Kutulakos K N. Confocal stereo. International Journal of Computer Vision, 2009, 81(1): 82-104

[19] Nayar S K, Watanabe M, Noguchi M. Real-time focus range sensor. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(12): 1186-1198

[20] Vaish V. Synthetic Aperture Imaging Using Dense Camera Arrays[Ph. D. dissertation]. the Stanford Computer Science Department, Stanford University, Palo Alto, California, USA, 2007

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

[22] Yoon K J, Kweon I S. Adaptive support-weight approach for correspondence search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 650-656

[23] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1222-1239

[24] Sun J, Zheng N N, Shum H Y. Stereo matching using belief propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7): 787-800

[25] Scharstein D, Pal C. Learning conditional random fields for stereo//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007: 1-8

[26] Li Y, Huttenlocher D P. Learning for stereo vision using the structured support vector machine//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8

[27] Wanner S, Goldluecke B. Globally consistent depth labeling of 4D light fields//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 41-48

[28] Liang C. Analysis, Acquisition, and Processing of Light Field for Computational Photography [Ph. D. dissertation]. National Taiwan University, Taiwan, China, 2009

[29] Kim C, Zimmer H, Pritch Y, et al. Scene reconstruction from high spatio-angular resolution light fields. ACM Transactions on Graphics, 2013, 32(4): 73

[30] Frese C, Gheta I. Robust depth estimation by fusion of stereo and focus series acquired with a camera array//Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems. Heidelberg, Germany, 2006: 243-248

[31] Li F, Sun J, Wang J, et al. Dual-focus stereo imaging. Journal of Electronic Imaging, 2010, 19(4): 043009-043009-12

[32] Vaish V, Levoy M, Szeliski R, et al. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 2331-2338

[33] Wang J, Barkowsky M, Ricordel V, et al. Quantifying how the combination of blur and disparity affects the perceived depth//Proceedings of IS&T/SPIE Electronic Imaging. San Francisco, USA, 2011: 78650K-78650K-10

[34] Adelson E H, Wang J Y A. Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2): 99-106

[35] Levoy M, Hanrahan P. Light field rendering//Proceedings of the ACM Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, USA, 1996: 31-42

[36] Gortler S J, Grzeszczuk R, Szeliski R, et al. The lumigraph//Proceedings of the ACM Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, USA, 1996: 43-54

[37] Heo Y, Lee K, Lee S. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(5): 1094-1106

[38] Kolmogorov V, Zabin R. What energy functions can be minimized via graph cuts?. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(2): 147-159

[39] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(9): 1124-1137

YANG De-Gang, born in 1987, M. S. His research interests include computer vision and light field theory and application.

XIAO Zhao-Lin, born in 1984, Ph. D. candidate. His research interests include computer vision and computational photography.

YANG Heng, born in 1981, Ph. D. His research interest is computer vision.

WANG Qing, born in 1969, Ph. D., professor, Ph. D. supervisor. His research interests include computer vision and pattern recognition.

Background

Depth estimation is one of fundamental problem in computer vision. The depth information can be extracted from particular image sets, which imply different kinds of depth cues. Only a single depth cue is considered in most existing methods, such as motion disparity, defocus blur, shading relations and so on. Because the complexity of the real world, single depth cue methods cannot generate robust and accurate depth estimation, even with the state of the art optimization algorithm, for example graph cuts and belief propagation.

In this paper, we propose a novel globally consistent depth estimation method based on both disparity and blur depth cues. To acquire these depth cues, we build a dense camera array, which can be simultaneously triggered by a controlling system. The proposed method can extract more robust estimation of scene depth. Most importantly, a unified adaptive fusion framework of multiple depth cues is introduced for depth recovery in this paper.

This work is supported by the following grants:

(1) "Depth Estimation and Optimization from Multiple Depth Cues Based on Camera Array" (No. 61272287), the National Natural Science Foundation of China.

(2) "Complicated Structure Recovery from Confocal Image Sequence" (No. 61103060), the National Natural Science Foundation of China.

(3) "Multiple Sensors Based Physical and Interactive Attributes Sensing and Modeling for Natural Phenomenon Simulation" (No. 2012AA011803), the National High Technology Research and Development Program (863 Program), Ministry of Science and Technology, China.

(4) "Dynamic Depth Estimation from Confocal Image Sequence Based on Camera Array" (No. 20116102110031), Specialized Research Fund for the Doctoral Program of Higher Education, Ministry of Education, China.

Our research group has been engaged in research of 3D structure reconstruction since 2007. Recently, we are aiming at multi-view depth recovery and light field imaging. This work is an extension of the following research work: 4D Scene Modeling and Rendering from Multiple Time-variant and Unordered Images (No. 2007AA01Z314, the National High Technology Research and Development Program (863 Program), Ministry of Science and Technology, China), Large Scale Light Probing, Real-time Modeling and Rendering (No. 2009AA01Z332, the National High Technology Research and Development Program (863 Program), Ministry of Science and Technology, China) and High Reliable Multi-view Correspondence and Optimization from Multiple Unordered Images (No. 60873085, the National Natural Science Foundation of China). Our lab has published many papers in highly-ranked international conferences and journals in the fields of 3D structure reconstruction, depth estimation, light field analysis, etc.

This work is directly based on the research of synthetic aperture photography, light field theory, depth from focus and stereo matching. This work is aiming at exploring joint model of depth estimation by fusing multiple depth cues.