

单幅图像的深度标签流形学习

叶 华^{1,2}, 谭冠政²

- (1. 中南大学 信息科学与工程学院, 湖南 长沙 410083;
2. 湖南文理学院 洞庭湖生态经济区建设与发展省级协同创新中心, 湖南 常德 415000)

摘 要: 图像中背景与前景对象的空间位置决定了场景在图像中的相对深度, 利用图像的局部特征相似性和流形结构的降维性能, 并应用 salient 区域 DCT 高频系数分布的深度排序索引性能, 定义出图像深度的马尔科夫概率图模型 MRF。通过划分场景对象检测 salient 区域模糊度, 最后估计得出图像场景的相对深度图。通过学习图像数据的流形嵌入对数据流形分布概率密度函数进行迁移, 得出遵循相似流形分布的对象特征类别标记概率密度分布。进一步检测空间变化 salient 区的模糊程度, 融合多尺度梯度幅度的高频离散余弦变换 DCT 系数特征, 依据模糊变化高频特征计算深度标记索引确定深度标签的层级次序, 融合类别标签以生成深度图。这种模型框架下检测单个图像中模糊和未模糊的区域, 可获得图像中场景的相对深度, 而无需了解相机设置或模糊类型的先验参数。在典型的深度图估计数据集中应用 MRF 深度图模型评测图像的深度估计性能, 实验结果给出该方法在检测场景分布和划分场景深度次序上的准确率, 验证了方法的有效性。

关键词: 流形嵌入; 深度排序; 标记流形学习; 离散余弦变换; 模糊检测

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA201847.0626004

Manifold learning of depth label for single image

Ye Hua^{1,2}, Tan Guanzheng²

- (1. School of Information Science and Engineering, Central South University, Changsha 410083, China; 2. Hunan Province Cooperative Innovation Center for The Construction & Development of Dongting Lake Ecological Economic Zone, Hunan University of Arts and Science, Changde 415000, China)

Abstract: The spatial position of the background and foreground determines the relative depth of the scene in the image. Using similarity characteristics of the local region of image and properties of dimensionality reduction of the manifold structure, the depth sorting indexing performance of the DCT high coefficients frequency distribution in the salient region was applied, the probability image map model of Markov Random Field(MRF) was defined to establish a relationship between the local feature and depth of different locations in the image. By segmenting the object, detecting relative blurring of the salient regions, and finally the relative depth map of the scene in the image was estimated. Through learning data embedding of manifold of

收稿日期: 2018-01-07; 修订日期: 2018-02-10

基金项目: 国家自然科学基金(61403136); 湖南省自然科学基金(14JJ5008); 湖南省教育厅项目(16C1087)

作者简介: 叶华(1977-), 女, 博士生, 主要从事深度图估计、红外图像融合、人体行为识别方面的研究。Email: yehuawuhan@163.com

导师简介: 谭冠政(1962-), 男, 教授, 主要从事进化计算算法理论、机器智能方面的研究。Email: tgz@csu.edu.cn

the image, the probability density function of the data manifold distribution was migrated, the probability density function of category labels of object which followed similar manifold distributions was obtained. The blurred extent of salient regions was detected further, the high-frequency coefficient of discrete cosine transform (DCT) of multi-scale gradient amplitudes was fused, then depth mark index was calculated according to the high frequency characteristics of the fuzzy change to determine the hierarchical order of the depth tags, and the category tags were merged to generate a depth map. In this model framework, the blurred and unambiguous areas in a single image were detected to obtain the relative depth of the scene in the image, without knowing the priori settings of the camera or the type of blur. The depth estimation performance of the image was evaluated by using the MRF depth map model in a typical depth map estimation data set. The experimental results show the accuracy of the method in detecting scene distribution and ordering the depth of scene. It verifies the validity of the method.

Key words: manifold embedding; depth ordering; tagging manifold learning; discrete cosine transform(DCT); blur detecting

0 引言

深度信息是获取场景三维信息的重要依据,基于图像的深度图估计是计算机视觉研究的基本问题,与多种应用问题相关,如目标识别与跟踪、机器人视觉、基于图像的可视化、虚拟现实等。由于图像只是实际三维场景的部分显示,对单幅图像进行深度图估计乃至三维重建至今仍然是极具挑战性的。目前的深度估计主流技术是进行两幅以上图像的立体视觉匹配^[1-5],以及单幅图像的模糊图估计。而针对单幅图像进行深度图估计的技术与传统机器学习的图像分割和模糊检测技术^[6-12]紧密相关,往往需要有相机设置或模糊类型的先验参数。

对单目单幅图像进行深度估计,主要解决两大问题,检测特征类别的判别性,定义深度的排序规则。论文的目标是应用流形嵌入对图像特征稀疏化和重新排列,扩大特征的类间区别和缩小其类内间隔,以划分图像场景。利用迁移技术将数据的流形结构迁移^[13-15]到类别标签分布,并检测 salient 区域空间变化的模糊性,引入 DCT 高频系数分布的深度排序索引性能,完成对图像数据的深度赋值。

1 相关工作

深度图估计是计算机视觉中三维重建技术的基础问题,从估计精度的角度划分为绝对深度与相对深度估计,从视差来源的角度划分立体视觉与单目

视觉^[1-5],对单目单幅图像的深度估计研究是深度估计优化与并行算法的基础。单目视觉的相对深度估计相对而言更具挑战性,笔者以此作为研究对象。基于 CRF 深度网络是大型数据集深度图估计研究普遍采用的方法。其中 CNN 深度网络结合 CRF 图模型为代表的深度图估计框架,CNN 模型建立多尺度底层像素特征到高层语义特征的表达^[1-2],结合 CRF 条件概率建模,采用端到端方式,依据相对深度完成语义分割^[3-5],运行效率高。但深度学习方法依赖大量预训练样本,对数据量的需求大。

传统机器学习方法往往建立模糊图与深度映射函数来估计图像的相对深度。其中以边缘模糊值建立深度参考关系的方法是最普遍的^[6]。Tung 通过建立模糊边缘与聚焦面的距离函数以估计深度值,但需要相机参数,多幅图像经过匹配和拟合得到深度图像,操作过程复杂,深度图重建性能很受限。

NAMBOODIRI 开创了只凭借一幅图像估计场景相对深度的方法^[7],利用检测图像边缘,建立点扩散函数估计图像的相对模糊函数,问题是模糊图本身性能的约束,导致噪声大而边缘信息不明确的场景处深度估计精度不高。Liu 沿袭了上述方法,对边缘检测施加了纹理内容约束^[8],使图像场景的深度受纹理等散乱数据的影响降低,但是他假定边缘形式都是阶跃的。

以边缘遮挡关系建立模糊图与深度图关系的方法增加了图像场景深度层次推理的依据,Yu ,Ming 等^[10-12]从多种角度获取图像特征、增加图像场景辨

别性,获得了较稳定图像的的深度估计结果,但都需求图像边缘特征的显著性、鲁棒性。

图像频域特征有显著性、鲁棒性,Golestaneh^[9]研究模糊图与频域特征的关联,不需要精确定位图像场景的边缘,是一种新的深度估计方法,在此基础上,结合图像场景分割将是一种新的发展方向。

文中方法对单目单幅图像进行深度估计依据的是场景模糊变化与深度的关系,主要解决两大问题,检测特征类别的判别性,定义深度的排序规则。利用图像的局部特征相似性和流形结构的降维性能^[13],并应用 salient 区域 DCT 高频系数分布的深度排序索引性能,定义出图像深度的马尔科夫概率图模型 MRF。通过划分场景对象,检测 salient 区域模糊度,最后估计得出图像场景的相对深度图。具体的做法是,首先完成图像数据的流形嵌入。设定像素点为 MRF 模型隐流形局部切空间中心点,对图像中场景的流形空间结构特征作局部切空间排列^[14-15],投影得出对象的全局坐标分布。对数据流形分布概率密度函数进行迁移^[13],得出遵循相似流形分布的对象特征类别标记的概率密度分布,这样,标记分布保留了场景空间结构。最后检测空间变化 salient 区的模糊程度,融合多尺度梯度幅度的高频离散余弦变换 DCT 系数特征,排序并归一化,依据模糊变化高频特征计算深度标记索引,确定深度标签的层级次序融合以生成深度图。这种模型框架下检测单个图像中模糊和未模糊的区域,可获得图像中场景的相对深度,而无需了解相机设置或模糊类型的先验参数,不需要额外的硬件支持,对数据量大小也没有约束。其处理流程如图 1 所示。输入图像 I,对图像中场景的流形空间结构特征作局部切空间排列,投影得出对象的全局坐标分布,这里假定数据遵循 3 种概率密度分布函数形式,对应 3 种灰度。对数据流形分布函数进行迁移,得出遵循相似流形分布的对象特征类别标记分布,那么,图像类别标记分布也有 3 种。至于类别间的深度次序,则由 salient 区域高频特征计算得出深度标记索引,即最深的灰度值接近黑色的区域,对应离相机最远的对象,灰度小的接近白色区域,对应离相机最近的对象。依据深度标记索引,确定图像场景深度标签的层级次序融合类别标签生成深度图。

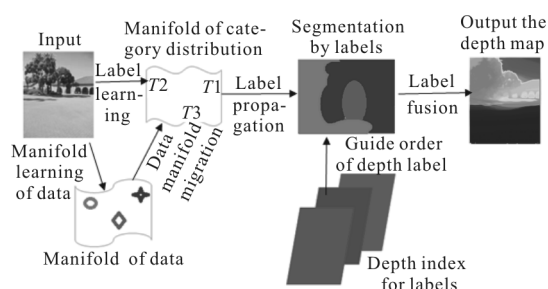


图 1 深度估计算法流程图

Fig.1 Workflow of our proposed depth estimation algorithm

2 深度标记分布流形学习

图像灰度数据是高维数据,假设图像嵌入在流形中是被认可的。笔者利用图像的局部特征相似性和流形结构的降维性能,对图像 salient 区域几何结构进行局部切空间排列,则 salient 区域边界与内点几何关系不变,经过局部切空间排列得到的全局距离关系可以迁移到图像类别标签分布,这只是完成了图像特征类别划分,而深度信息与 salient 程度是正相关的。

2.1 数据流形嵌入

假设样本点采样于嵌入 N 维数据空间的 d 维流形, $d < N, x_i = f(\tau_i) + \zeta_i, (i=1, 2, \dots, N)$ 。L1 损失-局部切空间排列 L1-LTSA 方法先对样本空间进行局部流形拓扑结构近似,再进行全局排列以拟合低维嵌入的流形分布。给定采样于 d 维流形 M 的 D 维样本集 $(1, \dots, d), N$ 为样本个数,L1-LTSA 算法步骤如下。

2.1.1 样本点距离相似性分布

在高维空间距离较远的点互为邻居点的概率小,可以忽略。建立像素点距离相似性关系,只用考虑与其相近的像素点。设样本点 $x_i (i=1, 2, \dots, N)$,取 x_i 的 k -最近邻点组成的邻域矩阵 $X_i = [x_{i1} \ x_{i2} \ \dots \ x_{ik}]$,不包括 x_i 。

2.1.2 样本距离局部切空间排列投影

在 x_i 的邻域 X_i 范围以 $(X_i - \bar{x}_i I_k^T)_{m \times k}$ 构成 x_i 的切空间, $I_k^T = (\underbrace{1, 1, \dots, 1}_k)$, 为 k 维的全 1 向量。投影矩阵则为 $\Theta_j = V_j^T (X_j - \bar{x}_j I_k^T) = (\theta_{j1}, \theta_{j2}, \dots, \theta_{jk})_{k \times m}, \theta_{jk} = \sum_{i=1}^k (x_{ji} - \bar{x}_j)$, $j=1, 2, \dots, k$ 。邻域均值作为切空间中心,与样本点切空间差异较大^[14],则以主成分分析(Principal Component Analysis,

PCA) 获得局部投影变换, 以样本 x_i 作为切空间中中心, 逼近每个样本 x_i 的邻域切空间, 即:

$$\operatorname{argmin} \sum_{j=1}^k \left\| x_j - x_i - VV^T(x_j - x_i) \right\|_2^2 \quad \text{s.t. } VV^T = 1 \quad (1)$$

切空间近似应最优化 V , 也就是使投影后样本值的协方差差值最大,

$$\operatorname{argmin}_V \operatorname{tr}(V^T S_X V) = \operatorname{argmax}_V \operatorname{tr} \frac{1}{k} \sum_{j=1}^k \left\| V^T(x_j - x_i) \right\|_2^2 \quad (2)$$

参考文献[15]中 Kwak 用更具鲁棒性的 L1 损失函数估计切空间, 即

$$\operatorname{argmax} \sum_{j=1}^n \left\| V^T(x_j - x_i) \right\|_1 \quad (3)$$

归一化样本点在切空间的投影距离, 这样, 局部的特征映射到全局, 能保持流形结构。

2.1.3 样本点全局坐标仿射变换

全局坐标设为 $Y_i(i=1, 2, \dots, N)$, 其邻域坐标为 $Y_{ij}=(y_{i_1}, y_{i_2}, \dots, y_{i_k}), j=1, 2, \dots, k$, 局部坐标仿射变换为 $y_{ij}=y_i+L_i\theta_j+\xi_i$, ξ_i 为样本采样产生的噪声。

全局坐标的逼近误差为:

$$E_i = \sum_{j=1}^n \left\| y_{ij} - (L_i\theta_j + y_i) \right\|^2 = \left\| Y_i - (L_i\Theta + Y_i S_i) \right\|^2 = \operatorname{tr}(YMY^T) \quad (4)$$

式中: $Y_i S_i = Y_i$, $S_i \in R^{N \times k}$; $M = \sum_{i=1}^n S_i B_i B_i^T S_i^T$; $B_i = (I - S_i)(I - \Theta^* \Theta)$ 。

归一化最小化局部邻域范围的投影排列误差, 可得全局坐标 Y ,

$$Y = \operatorname{argmin}_Y \sum_{i=1}^n \left\| E_i \right\| = \operatorname{argmin}_Y \sum_{i=1}^n \operatorname{tr}(Y B_i Y_i^T) \quad (5)$$

这可表示为 $\lambda Y = M Y$, 由此, 切空间排列的样本全局值的估计转化为求解线性组的特征向量。

2.2 特征类别标记流形学习

对特征空间 X 进行局部切空间排列得到了全局特征空间 Y , 并使得特征空间嵌入低维流形, 相似的特征点靠得更近, 不相似的特征点分离得更远。对图像像素的特征进行类别标签赋值是对特征空间相似性的评分, 不改变特征空间的拓扑结构, 是对特征空间的归一化与标准化。

由此, 可以从特征空间流形局部拓扑结构迁移得到欧几里得标签空间局部拓扑结构^[6]。用图 $G = \langle V, E, W \rangle$ 建立类别标签分布学习, 其中, V 是顶点集合, 表示样本拓扑结构。 E 是代表数据 x_i 和 x_j 之间关系的

边集, 每个边为 e_{ij} , 权重矩阵 W 的每个元素 W_{ij} 代表边的权重, 表示样本数据的相似性。样本类别标签空间 T 和特征空间 Y 共享相似的局部拓扑结构, 也就是标签空间 T 与经切空间排列的全局坐标空间 Y 共享相似的流形拓扑结构, 建立 Y 的局部邻接关系 W_i 并归一化, 最小化 W 的能量函数 $\varepsilon(W)$,

$$\varepsilon(W) = \sum_{i=1}^k \left\| y_i - \sum_{\substack{j=1 \\ j \neq i}}^k w_{ij} y_j \right\|^2, \quad \sum_{j=1}^k w_{ij} = 1, \quad W_i = [w_{i_1}, \dots, w_{i_k}] \quad (6)$$

这是典型的最小二乘规划问题, 求解过程不赘述。取样本点 y_i 的近邻点 y_j , 忽略较远的节点, N_k 为 y_i 的近邻区域, 这表示一个像素点 y_i 的值对应其 k 近邻点线性加权和, 靠得越近的点, 其权值越大, 与像素点 y_i 的值越相关。离得越远的点, 与像素点 y_i 的值越不相关, 权值近似为 0。标签空间 T 和全局特征空间 Y 共享相似的局部拓扑结构, $T=(\tau_i | i=1, 2, \dots, N)$, 有 N 个元素。通过迁移 Y 的拓扑结构, 数值标签向量 T 的能量函数 $\phi(T)$ 最小化可以推导出标签的流形重构, 即,

$$\phi(T) = \sum_{i=1}^n \left\| \tau_i - \sum_{\substack{j=1 \\ j \neq i}}^k w_{ij} \tau_j \right\|^2, \quad T=(\tau_i | i=1, 2, \dots, N) \quad (7)$$

这符合图像的先验知识, 离得越近的点, 特征类别标签值越相似。

2.3 salient 区域空间变化模糊检测与深度生成

得到像素点的特征类别标签值只获得了像素点特征分类的局部依据, 图像深度值是像素空间位置全局关系的预测, 不仅要考虑像素特征相似性, 还应依据图像高频特征与深度的映射关系。图像经过 DCT 变换发生能量汇聚效应, 大部分能量集中在低频区, 产生了频率系数稀疏化, 能描述细节的高频系数大多近似为零, 不为零的高频系数则对应较为显著的细节特征。因此, 使用搜索窗扫描统计像素点 salient 区域 DCT 高频系数的多少, 可以凸显图像显著性或者模糊性, 这对应着图像的深度信息。加入空间频域特征约束如 DCT 系数特征作为划分深度排列顺序的索引标记, 图像区域高频特征显著意味着区域细节丰富, 模糊度低, 相对镜头空间位置的距离小, 这里设定成像物体位于聚焦面同一侧。频域特征引导深度标签排序等同于在能量函数扩散过程中融合相似区域深度索引标记和像素特征类别标签。具

体的做法是检测空间区域模糊变化,对梯度图像融合多尺度梯度幅度的高频 DCT 系数特征^[9],排序并归一化,以检测单个图像中模糊和未模糊的区域,无需相机参数或图像的模糊类型等先验数据。

2.3.1 计算深度标记索引

取图像的固定大小块区域,维度为 $M \times M$,块中每个像素的 DCT 系数为 $M \times M$ 。为避免分辨率单一而造成图像模糊检测不准确,利用多尺度频域检测,则定义 r 尺度上的高频系数为:

$$H_{ij}^r = \begin{cases} H_{ij}^r(\nu, \omega) & : \nu + \omega \geq M - 1, \\ 0 \leq \nu, \omega \leq M - 1, 1 \leq r \leq m \end{cases} \quad (8)$$

式中: (i, j) 为时域坐标; (ν, ω) 为频域变量,尺度总数为 m 。取分布在右下部分的 DCT 系数作为检测图像模糊特性的高频特征,共可取 $(M^2 + M)/2$ 个。设固定区域维度为 7×7 ,共有 DCT 系数 49 个,则取到 28 个高频系数。统计高频特征系数幅值不为零的系数个数,用它表征图像区域的显著性也就是此区域深度的索引信息。设 M_b 为检测区域 m 个尺度上高频系数总数量, M_h 为检测区域 m 个尺度图像中高频系数不为零的数量,统计第 i 个像素点各层高频系数数量,排序并以最大值作为深度显著强度 M_h/M_b ,则像素点 x_i 在图像中的归一化深度索引值定义为:

$$Id_i = \frac{((M_h/M_b)^m)_i}{\max\{(M_h/M_b)^m\}_k, k=1, 2, \dots, N} \quad (9)$$

这表示了图像中样本点的深度显著度低于最小对比值时,其特征值较小或者距离聚焦面较远,当样本点深度显著度高于最低值时,其相对深度较明显或者空间位置距离聚焦面更近。

2.3.2 深度图生成与评测

图模型广泛应用于描述图像空间结构信息,深度图的 MRF 图模型评测图像块区域中的深度估计信息,这些块具有深度相似性和连通性,以及各区域与邻域区域的深度相似关系,这首先需要划分深度标签平滑赋值的图像 salient 区域范围,并确定该区域深度标签的层级次序。设定图像深度数据为 $D = (d_i, i=1, 2, \dots, N)$,表示 N 个样本点的深度值, S 表示图像的深度块数,特征向量 X 为第 1.3 节得到的样本点类别标签 $T = (\tau_i | i=1, 2, \dots, N)$, \hat{d}_i 由样本点类别标签 $\phi(\tau_i)$ 与深度标签索引 Id_i 相乘得到,由于已完成数据与标签的流形嵌入的迁移,样本点类别标签具

备判别性能,与归一化的深度标签索引值相乘,可以表示样本点深度预测值。模型归一化参数取 Z ,则深度图的 MRF 模型为:

$$P(D|X) = 1/Z \exp\left(-\sum_{i=1}^S h(d_r - \hat{d}_i)/d_i - \sum_{i=1}^S \sum_{j \in \mathcal{N}(i)} h(d_r - \hat{d}_j)/\max(\hat{d}_i, \hat{d}_j)\right) \quad (10)$$

模型的第一项是数据项,评价图像块中样本点的深度预测值;第二项是平滑项,对估计值起到平滑约束,以保证相似块的深度值也是相似的。生成深度预测值后,测试时只需要最大化 MRF 模型的概率密度函数,求解得出模型的最大后验估计概率即可。

3 实验结果及分析

3.1 实验数据集

考虑室外和室内两种现实环境的数据区别很大,适用的处理方法也不尽相同,可以检验方法的适用性。则选择两个深度数据库, NYU 2 dataset(室内场景):该数据集包含 1 449 个 RGBD 图像,每幅图像都有部分遮挡,提供了人工标注和深度数据。 Make3D(室外场景):数据集给出了以多种室外场景划分的子集,每个子集都包括图像和深度图,是现今机器学习评价相应深度估计算法的基本数据集。

3.2 实验数据集基线

以参考文献[10]的方法作为基准方法,因为有源代码而且提供了多种算法的对比效果。

3.3 评测指标

客观评测指标估计图像深度真实值与预测值之间的误差,根据误差值评价算法的性能。设定深度估计值为 D_i ,真实深度值 D_i^* ,选择均方根误差(RMS)作评价指标,评价的误差为对象深度排序的误差,

$$RMS = \sqrt{1/N \sum_{i=1}^N (D_i - D_i^*)^2} \quad (11)$$

3.4 结果分析

图 2 给出了在 Make3D 的几个测试结果,图(a)列是测试图像,为测试深度排序预测误差,图(e)列是文中方法的深度图,用更深的灰度值表示场景离相机聚焦面最远,更浅的灰度值表示离聚焦面最近,由灰度的由深变浅表示深度层次递减。其他方法的深度图结果来自于各个参考文献。图(c)列、图(d)列显示了其他算法对图像中所有场景的分层判定的准确性。

参考文献[10]在图 2 上的深度估计效果有很好

的主观一致性,场景边缘也较完整,但是整体上对纹理特征处理精度不高,实验结果中的场景层次标注受光照等杂散噪声干扰,往往出现不同深度层的区域划分到同一层,第三行的实验结果暴露出的问题最明显,房屋、树木和天空的深度划分完全混淆;相比而言,文中的方法在处理这些问题上都有很好的结果,salient 区域分割处理具有丰富纹理的图像,使其凸显场景结构特征,以 DCT 高频系数特征导引深度层次划分,因其具备频域稳定性,不容易受到噪声干扰,具有较高的鲁棒性。实验结果中的图像场景深度划分与主观一致性非常相符,精度有较大的提升。

实验结果表明:(1) 文中方法对于空间独立的场景能够得到比较正确的深度分层结果,有助推断图像全局场景层次关系的准确性。(2) 对于存在遮挡的场景预测具有局限性(比如对特征差异小的场景进行深度层次关系推断有困难,需要进一步结合语义约束),但是文中方法的 MRF 图模型可以结合图割优化策略,以减少对象与场景分割的误差,使得估计结果与人眼实际感知一致。(3) 文中方法对具有邻接关系的场景分层关系能得出较为准确的预测结果,对不具有邻接关系的场景也能得出相对分层关系,比如第一行的前排树木和处于树木后面的房屋门窗,虽然它们没有邻接关系,但能准确判断它们具有前后关系,因此算法对局部场景的层次关系也具有较高的判断准确性。

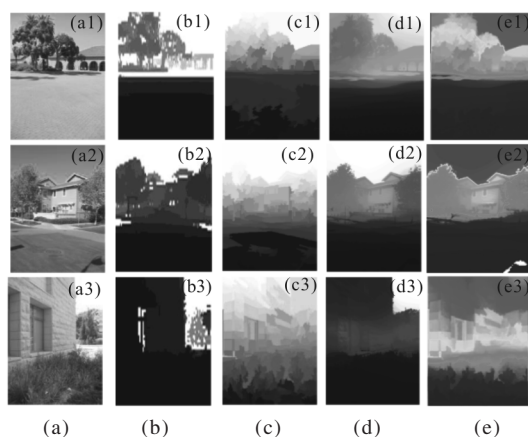


图 2 在 Make3D 数据集的测试。(a) Make3D 数据集图像;(b) Make3D 真实值;(c) Make3D 深度估计;(d)参考文献[10]的 WA 方法;(e)文中的流形标签学习深度估计

Fig.2 Test on Make3D dataset.(a) Make3D images;(b) Make 3D truth; (c) Make3D estimation; (d) reference [10] WA; (e) ours(Label manifold Learning for Depth Estimation)

图 3 显示了在 NYU 2 数据集的实验深度排序精度。表中第一列显示了参考文献[10]的实验结果,第二列是结合文献[10]+a 和全局推理的算法结果,第三列的是结合文献[10]+b 与局部推理,第四列仅使用文献[10]+c 中的颜色特征,第五列实验结果来自于文献[11],第六列文献[12],前六列的数据来自于文献[10]。最后一列是文中方法的实验结果。与文献[10]相比,文中的方法精度提高了大约 19.6%。相比第四列算法,笔者加入 salient 区局部与全局推断的则提高了近 10%。与文献[11]相比提高了 3.6%,与文献[12]相比提高了 0.6%,这显示出利用颜色和纹理的组合特征进行 salient 区域检测,加入组合特征性能提高很明显。

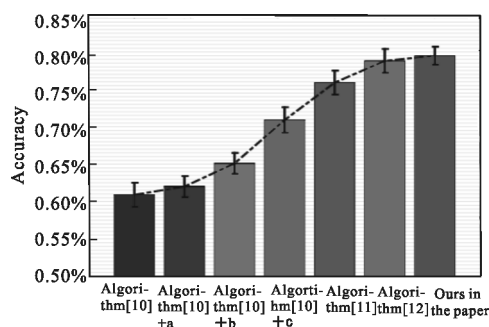


图 3 在 NYU 2 数据集的深度估计精度

Fig.3 Depth estimation accuracy on NYU 2 dataset

4 结论

这项研究提出了用传统照相机拍摄的图像的相对深度估计方法。主要贡献在于:(1) 图像数据流形分布迁移到数据类别标签的流形学习,扩大局部相似性,图像在场景空间结构的信息得到极大的保留。(2) salient 区域被用来约束图像场景,从而消除散乱噪声,保留图像场景的平滑性。(3) 提出一种新的 MRF 图模型,以场景模糊变化估计其深度的排列次序,由 salient 区域的 DCT 高频系数分布,估计该区域在整幅图像中的模糊度,建立深度索引值,避免图像中的散乱数据影响最终的模糊值,以图像 salient 区域的全局深度索引作为指导,把深度排序值融合到图像数据类别标签分布中。实验结果表明,文中方法可以克服先验参数的缺乏对场景相对深度估计的影响。如果结合相机参数,可以从估计的相对深度获得更精确的深度估计图。加上场景语义约束,估计的深度图有助于对象的语义分割,这也是文中研究的下一步目标。

参考文献:

- [1] Xu D, Ricci E, Ouyang W, et al. Multi-scale continuous crfs as sequential deep networks for monocular depth estimation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 161–169.
- [2] Zhang Guoshan, Zhang Peichong, Wang Xinbo. Visual place recognition based on multi-level feature difference map[J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203004. (in Chinese)
- [3] Yang Nan, Nan Lin, Zhang Dingyi, et al. Research on image interpretation based on deep learning[J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203002. (in Chinese)
- [4] Luo Haibo, He Miao, Hui Bin, et al. Pedestrian detection algorithm based on dual-model fused fully convolutional networks(Invited)[J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203001. (in Chinese)
- [5] Yin Yunhua, Li Huifang. RGB-D object recognition based on hybrid convolutional auto-encoder extreme learning machine [J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203008. (in Chinese)
- [6] Tung S S, Hwang W L. Multiple depth layers and all-in-focus image generations by blurring and deblurring operations[J]. *Pattern Recognition*, 2017, 69: 184–198.
- [7] Nambodiri V P, Chaudhuri S. Recovery of relative depth from a single observation using uncalibrated (real aperture) camera [C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1–6.
- [8] Liu S, Zhou F, Liao Q. Defocus map estimation from a single image based on two-parameter defocus model [J]. *IEEE Transactions on Image Processing*, 2016, 25(12): 5943–5956.
- [9] Golestaneh S A, Karam L J. Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017: 596–605.
- [10] Yu C, Liu Y, Wu T, Li K, et al. A global energy optimization framework for 2.1 d sketch extraction from monocular images[J]. *Graphical Models*, 2014, 76(5): 507–521.
- [11] Ming A, Xun B, Ni J, et al. Learning discriminative occlusion feature for depth ordering inference on monocular image[C]//ICIP, 2015.
- [12] Ming A, Wu T, Ma J, et al. Monocular depth ordering reasoning with occlusion edge detection and couple layers inference[J]. *IEEE Intelligent Systems*, 2016, 31(2): 54–65.
- [13] Hou P, Geng X, Zhang M L. Multi-label manifold learning [C]//Thirtieth AAAI Conference on Artificial Intelligence, 2016: 1680–1686. (in Chinese)
- [14] Zhang Z Y, Zha H Y. Principal manifolds and nonlinear dimension reduction via local tangent space alignment [J]. *Journal of Shanghai University*, 2004, 8(4):406–424.
- [15] Kwak N. Principal component analysis based on L1-norm maximization [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(9): 1672–1680.