

采用 KPCA 和 BP 神经网络的单目车载红外图像深度估计

孙韶媛^{1,2}, 李琳娜^{1,2}, 赵海涛³

- (1. 东华大学 信息科学与技术学院, 上海 201620;
2. 东华大学 数字化纺织服装技术教育部工程研究中心, 上海 201620;
3. 华东理工大学 信息科学与工程学院, 上海 200237)

摘要: 提出一种基于监督学习得到深度估计模型的单目车载红外图像深度估计方法。首先用核主成分分析法(KPCA)筛选红外图像特征。将最初提取的红外图像特征用核函数非线性映射到一个线性可分的高维特征空间,再完成主成分分析(PCA),得到降维后的红外图像特征。然后以 BP 神经网络为模型基础,对红外图像特征和深度值进行训练,训练后的深度估计模型可对单目车载红外图像的深度分布进行估计。实验结果证明,利用该模型估计的单目车载红外图像的深度信息与原红外图像的深度信息一致。

关键词: 深度估计; 红外图像; KPCA; BP 神经网络

中图分类号: TN219 **文献标志码:** A **文章编号:** 1007-2276(2013)09-2348-05

Depth estimation from monocular vehicle infrared images based on KPCA and BP neural network

Sun Shaoyuan^{1,2}, Li Linna^{1,2}, Zhao Haitao³

- (1. College of Information Science and Technology, Donghua University, Shanghai 201620, China; 2. Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China;
3. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: A depth estimation algorithm from monocular vehicle infrared image based on depth estimation model by supervised learning was proposed. Firstly, kernel-based principle component analysis (KPCA) was used to select infrared image features. Original features extracted from infrared image were project nonlinearly to a high dimensional and linear separable feature space using kernel function. Principle component analysis (PCA) was performed to get dimension reduction infrared image features. Then the infrared image features and depth values were trained using BP neural network. A depth estimation model was obtained which can estimate the depth distribution of monocular vehicle infrared image. The experimental results show that most of the depth estimated by the model is consistent with the original depth information of infrared image.

Key words: depth estimation; infrared image; KPCA; BP neural network

收稿日期: 2013-01-11; 修订日期: 2013-02-12

基金项目: 国家自然科学基金(61072090); 上海市浦江人才计划(12PJ1402200); 中央高校基本科研业务费专项资金(12D10418)

作者简介: 孙韶媛(1974-), 女, 副教授, 博士, 主要从事彩色夜视、信息融合方面的研究。Email: shysun@dhu.edu.cn

0 引言

车载红外系统越来越多地安装在汽车上以辅助夜间安全驾驶。与可见光图像不同,红外图像反映的是物体的温度分布,没有色彩和阴影,普遍缺乏空间立体感。图像的空间立体感可以让我们辨别景物的前后关系,识别路径及目标位置,在对场景的理解中至关重要。若要赋予红外图像空间立体感显示效果,首先需要估计图像中景物的深度信息。从景物的二维图像中恢复出三维信息是计算机视觉中的一个重要问题^[1],单、双目深度估计和基于图像序列的深度估计都是其主要方法。在彩色图像领域,大部分深度估计的研究围绕着双目和基于图像序列的深度估计^[2-3],依赖于图像间的特征差异。而单目深度估计算法中较为经典的是“由阴影恢复形状(shape from shading)”^[4],参考文献[5]依据图像景物构成和小孔成像模型来估计深度,斯坦福大学 Andrew Ng 团队利用 Markov Field 训练的模型对单幅图像深度估计^[6];参考文献[7]先手动标定场景中简单的类别,利用大量的数据学习上述类别,再构建 Bayes 模型对新图像进行分类,最后恢复深度信息。上述深度估计研究都是在彩色图像范围内,可利用色彩和阴影等图像特征,而车载红外图像是单目无色彩和阴影的图像,与彩色图像相比存在缺乏纹理信息、对比度低、图像模糊等问题^[8]。论文针对车载红外图像的深度估计问题提出一种基于监督学习得到深度估计模型的单目深度估计方法。对于单幅灰度的红外图像,无法获取景物的色彩和阴影信息,只能提取基于灰度值的图像特征,所提取特征是否能与深度线索较相关,是深度估计问题中的关键。笔者曾在参考文献[9]用逐步线性回归和独立主成分分析筛选出与深度信息较相关的十个图像特征,但这种特征筛选方法没有考虑到图像的空间结构和特征向量间的非线性关系。文中用基于核主成分分析法(KPCA)将原始的图像特征向量非线性映射到高维特征空间,然后在高维空间进行 PCA 分析得到降维后的特征向量。再以 BP 神经网络为模型基础对降维后的红外图像特征向量和其对应的深度图进行训练,训练后的深度估计模型可对单目车载红外图像的深度分布进行估计。

1 红外图像特征提取及降维

1.1 红外图像特征提取

红外图像反映的是物体表面的温度分布,但图像中不同景物的温度分布可能类似,造成红外图像局部信息不足。为了能更全面地描述每个像素的特征,采用更大区域范围的“多尺度”和“空间上下文”信息描述图像中每个像素的特征,包括三个尺度上的纹理能量、梯度能量、梯度方差、梯度均值、清晰度和灰度值共 345 维特征,即红外图像中的每一个像素点由一个 345 维的特征向量来描述^[9-10]。但该特征向量是依据前人对彩色图像单目深度线索研究的结果,并不完全适用于红外图像。为寻找适合红外图像的深度线索,控制训练的计算量,应对上述 345 维特征进行筛选降维。

1.2 基于 KPCA 的特征降维

1.2.1 KPCA 原理

PCA 是用几个较少的综合指标来代替原来较多的指标,这些较少的指标既能反映原来较多指标的有用信息,又相互独立,但该方法对大量的非线性数据无效。作为一种由线性到非线性之间的桥梁,核方法通过核函数把低维线性不可分的数据映射到高维空间,使之成为线性可分数据。在高维空间中使用 PCA 对其进行降维,称为核主成分分析(KPCA)^[11]。KPCA 方法的基本思想是通过某种隐式方式将输入空间映射到某个高维空间(常称为特征空间),并且在特征空间中实现 PCA。下面简单介绍 KPCA 的原理^[12-15]。

假设 $X=[x_1, x_2, \dots, x_N]$ 为训练样本, $x_i \in R^d$, d 为 x_i 的维数, N 为训练的样本数。定义 $\Phi(x) \in F$, $\Phi: R^d \rightarrow F$, F 为特征空间。定义 $\Phi(X)$, m^Φ 和 e 为:

$$\Phi(X)=[\Phi(x_1), \dots, \Phi(x_N)] \quad (1)$$

$$m^\Phi = \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \quad (2)$$

$$e=[1, 1, \dots, 1] \in R^{1 \times N} \quad (3)$$

F 的协方差矩阵是:

$$\Sigma = \frac{1}{N} \sum_{i=1}^N (\Phi(x_i) - m^\Phi e)(\Phi(x_i) - m^\Phi e) \quad (4)$$

核主成分分析就是求解特征值问题 $\lambda_k v_k = \Sigma v_k$, v_k 是第 k 个特征值 λ_k 对应的特征向量。若 K 表示公

式(5)所示的核矩阵,那么根据公式(6)和(7)可推出公式(8)和(9)。

$$K=(K(x_i, x_j))_{ij}=(\Phi(x_i), \Phi(x_j))_{ij} \quad (5)$$

$$\tilde{K}=(\tilde{K}(x_i, x_j))_{ij}=(\Phi(x_i)-m^{\phi}e, \Phi(x_j)-m^{\phi}e)_{ij} \quad (6)$$

$$v_k=(\Phi(X)-m^{\phi}e)\alpha_k\frac{1}{\sqrt{\lambda_k}} \quad (7)$$

$$\tilde{K}=\left(1-\frac{1}{N}e^Te\right)K\left(1-\frac{1}{N}e^Te\right) \quad (8)$$

$$\lambda_k\alpha_k=\tilde{K}\alpha_k \quad (9)$$

让 $\hat{\Phi}(z)=\Phi(z)-m^{\phi}$, β_k 表示映射 $\Phi(z)$ 中的第 k 个分量的映射,那么,

$$\beta_k=\hat{\Phi}(z)^Tv_k=\sum_{i=1}^N\frac{\alpha_{ki}}{\sqrt{\lambda_k}}\tilde{K}(z, x_i) \quad (10)$$

1.2.2 基于 KPCA 的图像特征降维

基于上述 KPCA 的基本原理,图像特征降维处理过程如下:

(1) 将 132 行、234 列的红外图像的所有 $N(N=132\times 234=42\ 768)$ 个像素的 345 维的特征向量 f_i 写成一个 345 行 N 列的数据矩阵 $F=[f_1, f_2, \dots, f_N]$, $f_i \in \mathbb{R}^{345}$ 。

(2) 选定核函数,根据公式(5)和(8)计算核矩阵 K 和 \tilde{K} 。

(3) 根据公式(9)得到 λ_k 和 α_k 。

(4) 用 α_k 前的 $p(p<345)$ 个分量用公式(10)做投影,即 $\beta_k=\sum_{i=1}^N\frac{\alpha_{ki}}{\sqrt{\lambda_k}}\tilde{K}(z, x_i)$, $k=1, \dots, p$ 。也就是说在 KPCA

空间降维后用 $\beta^i=[\beta_1^i, \dots, \beta_p^i]$ 来表示 f_i , 即为原始红外图像特征向量 f_i 经 KPCA 降维后的特征向量。

2 基于 BP 神经网络的红外图像深度估计模型

2.1 BP 神经网络

由于红外图像特征与深度值之间是一种非线性映射,而选择非线性拟合函数往往比较困难,但神经网络提供的一种在训练线性拟合函数的同时学习其非线性程度的方法能够回避这个困难。BP 神经网络由于其计算效率高,局部计算简单等因素被广泛应用。BP 神经网络是一种多层前馈神经网络,主要特点是信号前向传递,误差反向传播。在前向传递中,输入信号从输入层经隐含层逐层处理,直至输出层。每一层的神经元状态只影响下一层神经元状态。如

果输出层得不到期望输出,则转入反向传播,根据预测误差调整网络权值和阈值,从而使 BP 神经网络预测输出不断逼近期望输出。BP 神经网络首先需要通过训练使网络具有联想记忆和预测能力。这里不详细述 BP 神经网络的原理,只介绍实现步骤:

(1) 首先对 BP 神经网络初始化,根据降维后的红外图像特征的维数,确定输入层的节点数。

(2) BP 神经网络的隐含节点数对它的预测精度有较大的影响。一般隐含节点数的选择首先参考公式 $l=\sqrt{m+n}+a$ 来确定大致范围,式中 n 为输入层节点数, m 为输出层节点数, a 为 0~10 之间的常数, l 为隐含层的节点数。因为隐含层节点数取输入层节点数和输出层节点数之间的值时,尤其当靠近输入节点数时,网络收敛速度较快,笔者首先设隐含层节点数为 n ,通过最小均方误差结果选择最佳节点数。

(3) 确定输出层的节点数为 1。

(4) 初始化连接输入层、隐含层和输出层神经元之间的权值,给定学习速率和神经元激励函数。

(5) 用整幅红外图像及相应深度图像构建训练集合,对集合用 BP 神经网络训练,得出 BP 神经网络深度模型。训练好的 BP 神经网络模型可以对新的红外图像的深度分布进行估计。

2.2 深度估计模型训练

训练样本为如图 1 所示的单幅红外图像和它对应的深度图。对每个像素点提取 345 维特征向量 f_i 后进行 KPCA 降维,得到降维后的特征向量 β^i ,深度值是该红外图像对应深度图的灰度值 $depth_i$ 。这样就得到了训练样本集合 $\{\beta^i, depth_i\}$, $i=1, \dots, N$ 。

KPCA 可以选择不同的核函数,如线性核函数,多项式核函数,径向基核函数和多层感知核函数。核函数不同,红外图像的深度特征筛选结果也各不相同。笔者采用最小均方误差确定采用每种核函数用 KPCA 降维后的最佳特征维数和 BP 神经网络的隐含层数。

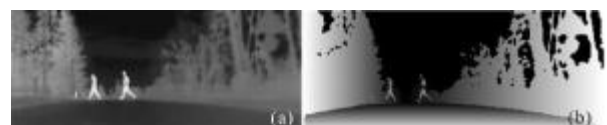


图 1 源红外图像及其对应的深度图

Fig.1 Source infrared image and its depth map

图 2 为预测误差和特征个数的关系与预测误差和隐含层节点数的关系。

图 2(a)是采用核函数多项式核函数 $K(x, x_i) = [(x \cdot x_i) + 1]^p$ 预测误差和特征个数的关系, KPCA 特征降维后的个数为 29 的时候 MSE 最小, 因此特征向量的个数设置为 29。即 BP 神经网络的输入层有 29 个节点。由图 2(b)可以看出, 在隐含层节点数为 21 时 MSE 最小, 因此将隐含层节点数设为 21。

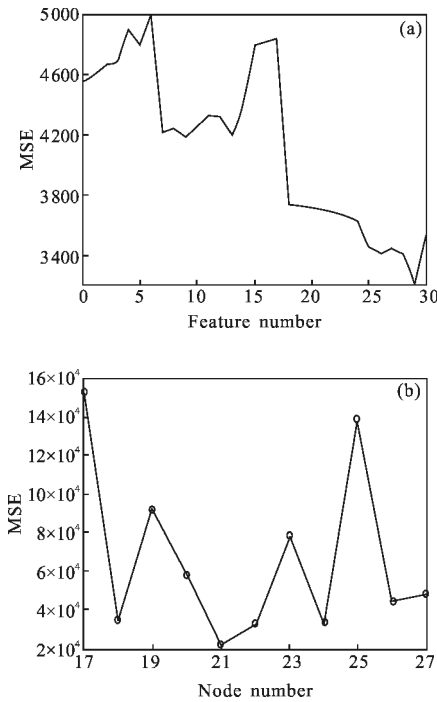


图 2 预测误差和特征个数的关系与预测误差和隐含层节点数的关系

Fig.2 Relations between prediction error and features number and relations between prediction error and nodes number of hidden layer

图 3 是基于四种不同核函数训练出的深度估计模型对图 1 中的源红外图像进行深度估计的结果, 可以看出基于多项式核函数和高斯径向基核函数的估计结果更准确。



(a) 基于线性核函数的深度图 (b) 基于多项式核函数的深度图
(a) Depth map based on linear kernel function (b) Depth map based on polynomial kernel function



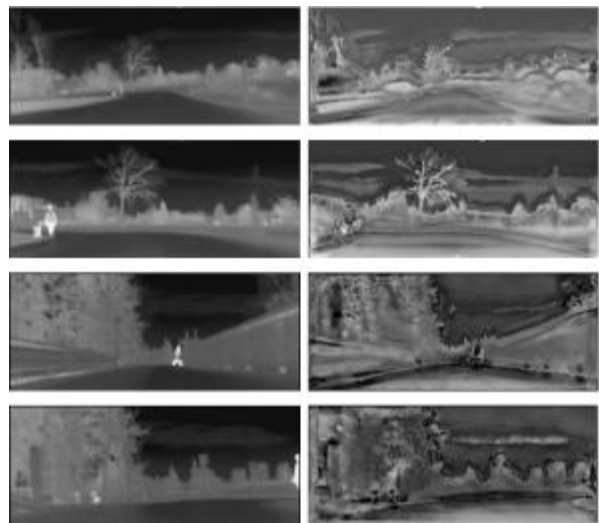
(c) 基于高斯径向基核函数 (d) 基于多层感知核函数
(c) Depth map based on Gaussian RBF kernel function (d) Depth map based on multilayer perceptron kernel function

图 3 基于不同核函数的红外图像的深度图

Fig.3 Depth maps of infrared image based on different kernel functions

3 测试图像深度估计实验及结果分析

选取了与训练图像场景相近的车载红外图像作为测试图像, 用 2.2 节训练出的深度估计模型来估计图像深度。图 4(a)为测试用车载红外图像, 图 4(b)分别为对应的深度估计结果。在用测试图像检验 2.2 节四种不同核函数训练出的深度估计模型时发现与图 3 中用训练所用的红外图像来检验的结果并不十分一致, 并非所有的图像都是如图 3 中那样是多项式核函数和高斯径向基核函数的估计结果更准确。图 4(a)中第一行和第二行的红外图像是用基于多项式核函数的深度模型估计出的结果更准确, 而第三行和第四行的红外图像是用基于多层感知核函数的深度模型估计出的结果更准确。这说明尽管基于 KPCA 和神经网络的深度估计模型的方法可行, 但还需进一步研究与该模型的准确性相关的各种因素, 这也是笔者今后的研究方向。另外, 从图 3 的所



(a) (b)

图 4 测试用车载红外图像和深度估计结果

Fig.4 Testing vehicle infrared image and results of depth estimation

有结果来看,图像中景物深度估计的结果准确性也是相对的,部分道路和天空的深度估计是不准确的。这与测试图像与训练图像的场景相似程度有关,模型的泛化性能还需提高。

4 结 论

文中针对车载红外图像的深度估计问题提出一种基于监督学习得到深度估计模型的单目深度估计方法。用 KPCA 做红外图像特征降维,用 BP 神经网络训练学习图像特征和深度之间的关系并得到深度估计模型。用此模型可以对其他的红外图像做深度估计。实验结果表明:利用该模型能相对准确地估计单目车载红外图像的深度信息,部分场景的深度估计结果不够准确,模型的泛化能力还需进一步提高。虽然实验结果没有达到完全精确,但该方法对估计无色彩无阴影的红外图像的深度提供了一种较可行的思路,笔者将对深度估计模型的准确性和泛化能力做进一步的研究。

参考文献:

- [1] Pentland A P. A new sense for depth of field [J]. IEEE Transactions on PAMI, 1987, 9(4): 523-531.
- [2] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal of Computer Vision, 2002, 47(1): 7-42.
- [3] Wu Wenji, Xiao Zhaolin, Wang Qing, et al. Depth estimation based on confocal image sequence[J]. Computer Applications and Software, 2012, 29(12): 43-46. (in Chinese)
吴文杰, 肖照林, 王庆, 等. 基于共聚焦图像序列的深度估计方法[J]. 计算机应用与软件, 2012, 29(12): 43-46.
- [4] Horn B K P. Obtaining shape from shading information[C]//The Psychology of Computer Vision. New York: McGraw Hill, 1975: 115-155.
- [5] Li Le, Zhang Maojun, Xiong Zhihui, et al. Depth estimation from a single still image of street scene based on content understanding[J]. Robot, 2011, 133(1): 174-180. (in Chinese)
李乐, 张茂军, 熊志辉, 等. 基于内容理解的单幅静态街景图像深度估计[J]. 机器人, 2011, 133(1): 174-180.
- [6] Ashutosh Saxena, Chung Sung H, Ng Andrew Y. 3-D depth reconstruction from a single still image [J]. International Journal of Computer Vision, 2008, 76(1): 53-69.
- [7] Hoiem D, Efros A, Hebert M. Automatic Photo pop-up [J]. ACM Transaction on Graphics, 2005, 24(3): 577-584.
- [8] Bai Junqi, Chen Qian, Wang Xianya, et al. Contrast enhancement algorithm of infrared image based on noise filtering model [J]. Infrared and Laser Engineering, 2010, 39(4): 777-780. (in Chinese)
白俊奇, 陈钱, 王娴雅, 等. 红外图像噪声滤波对比度增强算法[J]. 红外与激光工程, 2010, 39(4): 777-780.
- [9] Xi Lin, Sun Shaoyuan, Li Linna, et al. Depth estimation from monocular infrared images based on SVM model [J]. Laser & Infrared, 2012, 42(11): 1311-1315. (in Chinese)
席林, 孙韶媛, 李琳娜, 等. 基于 SVM 模型的单目红外图像深度估计[J]. 激光与红外, 2012, 42(11): 1311-1315.
- [10] Gu X J, Leung H, Sun S Y, et al. Colorizing single-band thermal night vision images [C]//Proceedings of IEEE Conference on Image Processing, 2010: 301-304.
- [11] Bernhard S, Alexander S, Klaus-Robert M. Kernel principal component analysis [C]//Artificial Neural Networks-ICANN'97, Lecture Notes in Computer Science, 1997, 1327: 583-588.
- [12] Zheng W, Lai J, Yuen P C. Penalized preimage learning in kernel principal component analysis [J]. IEEE Transactions on Neural Networks, 2010, 21(4): 551-570.
- [13] Liu Q, Cheng J, Lu H, et al. Distance based kernel PCA image reconstruction [C]//Proceedings of the 17th International Conference on Pattern Recognition, 2004: 670-673.
- [14] Mika S, Schölkopf B, Smola A, et al. Kernel PCA and denoising in feature spaces [C]//Advances in Neural Information Processing Systems, 1998: 536-542.
- [15] Wong W K, Zhao H. Eyeglasses removal of thermal image based on visible information [J]. Information Fusion, 2013, 14(2): 163-176.