

基于 SVM 模型的单目红外图像深度估计

席 林, 孙韶媛, 李琳娜, 邹芳喻
(东华大学信息科学与技术学院, 上海 201620)

摘要:提出一种通过非线性学习模型来估计单目红外图像深度的算法。该算法首先通过逐步线性回归和独立成分分析(ICA)寻找对于红外图像深度相关性较强的特征,然后以具有核函数的非线性支持向量机(SVM)为模型基础,采用监督学习的方法对红外图像深度特征进行回归分析并训练,在训练过程中通过已知数据回归后的最小均方误差对模型参数进行修正,训练后的模型可对单目红外图像的深度分布进行估计。实验结果证明,利用该模型能较一致地估计单目红外图像的深度信息。

关键词:红外图像;深度估计;SVM;核函数;单目深度

中图分类号:TP391.41 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2012.11.024

Depth estimation from monocular infrared images based on SVM model

XI Lin, SUN Shao-yuan, LI Lin-na, ZOU Fang-yu

(College of Information Science and Technology, DongHua University, Shanghai 201620, China)

Abstract: A depth estimation algorithm from monocular infrared image based on nonlinear learning model is proposed in this paper. Firstly, some features corresponding to the depth of infrared image are extracted using stepwise regression and ICA. Secondly, based on the nonlinear SVM model with kernel function, we regress and train the depth features of infrared image using supervised learning method. During the training process, the model parameters are fixed by calculating MMSE from regression analysis of the known data. Finally, we get the depth estimation model which can estimate the depth distribution of monocular 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: infrared image; depth estimation; SVM; kernel function; monocular depth

1 引言

图像的深度估计从本质上讲是一个深度感知的问题。单目深度估计、双目深度估计和基于图像序列的深度估计都是深度估计的主要方法。在彩色图像领域内,大部分深度估计的研究工作是围绕着双目视觉和基于图像序列的深度估计^[1]展开的,这两种方法都依赖于图像间的特征差异。而对于单目深度估计,在早期传统的算法中较为经典的是“由阴影恢复形状(shape from shading)”^[2],该算法以空间立体几何为理论基础,依据光源照射到物体表面所产生的明暗变化来恢复物体深度,但因为该算法需要先验知识如反

射模型和光源方向等使得应用的局限性增加。之后,斯坦福大学的 Andrew Ng 团队通过利用马尔科夫场训练的模型对单幅图像进行深度估计^[3],达到了很好的效果;卡耐基梅隆大学的 Aloysha Efros 团队则在训练前手动标定场景中简单的类别,然后利用大量的数据学习这些类别,并最终通过构建贝叶斯模型对新图像进行分类从而恢复深度信息^[4]。

基金项目:国家自然科学基金项目(No. 61072090)资助。

作者简介:席 林(1989 -),男,硕士,主要研究方向为红外图像处理,模式识别等。

收稿日期:2012-04-13; **修订日期:**2012-05-07

上述前人的研究工作都是在彩色图像中进行深度估计的,而对于红外图像,因其有无彩色、缺乏纹理信息、对比度低、图像模糊等缺点^[5],以至于目前还没有对于单目红外图像的深度估计方法。本文提出了一种基于支持向量机(SVM)学习的红外深度估计算法。通过逐步线性回归和独立成分分析(ICA)对深度特征进行筛选,将筛选后的特征通过具有核函数的支持向量机进行学习训练得出模型,该模型即可用于红外图像的深度估计。图1是本文的算法思路框图。

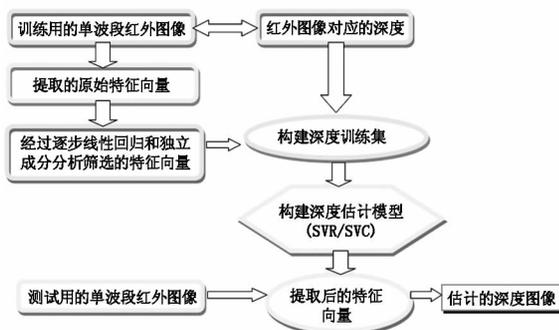


图1 算法流程图

Fig.1 algorithm flow chart

2 红外图像特征提取

2.1 特征粗提取

红外图像反映的是物体的表面温度分布,如图2所示,在红外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,仅对局部信息进行分析往往导致误差较大,因此本文沿用前人的基于“多尺度”和“空间上下文”^[6]信息提取图像特征的方法来对红外图像的某个区域或某个像素做出鉴别性的描述,下面简单介绍这种特征向量提取方法。

假设像素点 p_i 的基于“多尺度”和“空间上下文”信息的特征向量为 f_i ,该特征向量的构成如图2所示:将像素点 p_i 所属的图像块及该块的上、下、

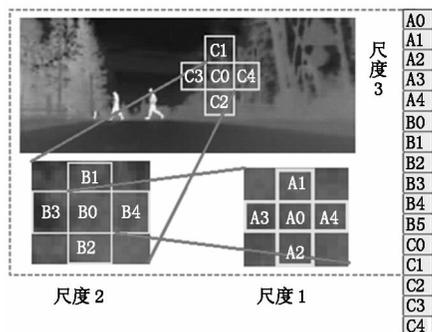


图2 像素点 p_i 的特征向量 f_i ,包含了像素点 p_i 及其周围四邻域从小到大的3个尺度上的信息

Fig.2 feature vectors f_i of pixel p_i , contains

information from this pixel and its four-neighbor-blocks in three scales

左、右相同大小的相邻图像块,共5个图像块视为其“特征区域”。为像素点 p_i 设置三个不同尺度上的“特征区域”,从小到大分别为尺度1、尺度2和尺度3,其中较大尺度上的图像块的中心块包含较小尺度下所有的图像块。

Claudia Hendrix 和 Woodrow Barfield 总结过能够用于计算机识别的深度线索类型,其中单目深度线索类型包括光信息、纹理信息、物体表面覆盖信息和运动信息^[7],以此为基础提取深度线索的特征向量,具体操作如下:

采用9个基本的二维“劳斯掩膜”记为 M_1, \dots, M_9 ,将红外图像 $I(x, y)$ 与每一个“劳斯掩膜”做卷积,得到 $T_n(x, y) = I(x, y) * M_n(x, y), n = 1, \dots, 9$ 。然后对于像素点 p_i 计算其尺度2和尺度3上每一个图像块 $N^i(m)$ 的纹理能量 E_n :

$$E_{nN^i(m)}^{\text{Scale}} = \sum_{x, y \in N^i(m)} |T_n(x, y)|, n = 1, \dots, 9, m = 0, \dots, 4, \text{Scale} = 2, 3 \quad (1)$$

对红外图像 $I(x, y)$ 的 x 方向和 y 方向分别求得梯度图,为 $I_{\text{gradx}}(x, y)$ 和 $I_{\text{grady}}(x, y)$,设角度 $\theta_i = i\pi/8, i = 0, \dots, 7$ 可以计算任一方向的梯度值 $G_i(x, y) = I_{\text{gradx}}(x, y) * \cos(\theta_i) + I_{\text{grady}}(x, y) * \sin(\theta_i)$,然后对像素点 p_i 计算其尺度2和尺度3上每一个图像块 $N^i(m)$ 的梯度能量 E_n^{grad} 、梯度方差 S_n 和梯度均值 M_n :

$$E_{nN^i(m)}^{\text{grad Scale}} = \sum_{x, y \in N^i(m)} |G(x, y)|, \\ M_{nN^i(m)}^{\text{Scale}} = \frac{1}{\text{size}_{\text{Scale}}} \sum_{x, y \in N^i(m)} (G(x, y)), \\ S_{nN^i(m)}^{\text{Scale}} = \frac{1}{\text{size}_{\text{Scale}}} \sum_{x, y \in N^i(m)} (G(x, y) - \bar{G}(x, y))^2, \\ n = 1, \dots, 8, m = 0, \dots, 4, \text{Scale} = 2, 3 \quad (2)$$

计算像素点 p_i 尺度2和尺度3的每个图像块 $N^i(m)$ 的清晰度 Sharp:

$$\text{Sharp}_{N^i(m)}^{\text{Scale}} = \frac{1}{\text{size}_{\text{Scale}}} \sum_{x, y \in N^i(m)} (I(x, y) - \bar{I}(x, y))^2 \quad (3)$$

最后再将像素点 p_i 及其在尺度1上的邻域像素点的灰度值也放在特征空间中。这样得出的特征向量 f_i 共有345维特征值。

2.2 特征筛选

由上可知我们已经得出了345维的特征向量,但该特征向量是通过前人对于彩色图像单目深度线索的研究得出的结果,它并不一定能够完全适用于红外图像,因此下一步将对特征进行筛选,寻找适合红外图像深度估计的深度线索。筛选的思路是通过

逐步线性回归保证筛选后的特征向量与深度之间的相关性,通过独立成分分析(ICA)保证特征向量之间的相互独立性^[8]。首先采用逐步线性回归法寻找与样本线性相关性最大的特征组。然后再通过快速 ICA 算法对已经过逐步线性回归分析的特征向量进行分析,使得各分量之间尽可能的独立。

我们将已知的红外图像对应的深度图作为样本,因为样本的数据量太大,所以每次随机抽取深度图的 5000 个像素点作为样本点,对每个像素点都求出其对应的 345 维特征向量,设定显著性水平为 0.05,并依次进行逐步线性回归和 ICA 分析。经过超过 30 次的分析后得出了 10 个特征符合 0.05 的显著性水平。这 10 个特征是尺度 1 的 5 个像素值和尺度 3 的 5 个图像块对应的清晰度值。

3 基于 SVM 学习的红外图像深度估计模型

支持向量机(SVM)建立在统计学习理论的 VC 维理论和结构风险最小原则基础上的^[9],其中支持向量分类(SVC)和支持向量回归(SVR)是 SVM 理论的重要应用。在此我们主要用到的是 SVR 模型。

用 depth 表示红外图像对应的深度图,提取一个有 1 组输入输出的深度训练样本集合 $\{f_i, \text{depth}_i\}$,其中 $i = 1, \dots, l$ 表示像素点 p_i , $f_i \in \chi$ 表示 p_i 的特征向量, depth_i 表示 p_i 对应的深度值。在支持向量回归中, f_i 首先被某一映射 $\Phi: \chi \rightarrow \mathcal{F}$ 映射到某一个新的特征空间 \mathcal{F} ,通常 \mathcal{F} 比 χ 具有更高的维数。若使用合适的映射 Φ ,那么在 χ 空间中不能被线性拟合的数据在映射到 \mathcal{F} 空间之后就很可能很好地被拟合。将深度训练样本 $\{f_i, \text{depth}_i\}$ 间的关系表示为如下形式:

$$\text{depth}_i = \mathbf{W}_d^T \phi(f_i) + b_d, f_i \in \chi, b_d \in \mathfrak{R}, i = 1, \dots, l \quad (4)$$

其中, $\mathbf{W}_d^T \in \mathcal{F}$ 是需要被求解的回归系数向量; b_d 是未知偏量。在支持向量理论中求解 \mathbf{W}_d^T 就是求解如下最优化问题:

$$\begin{aligned} \min_{\mathbf{W}_d^T, b_d, \xi, \xi^*} & \frac{1}{2} \mathbf{W}_d^T \mathbf{W}_d + C_d \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t. : } & \mathbf{W}_d^T \phi(f_i) + b_d - \text{depth}_i \leq \varepsilon + \xi_i, \\ & \text{depth}_i - \mathbf{W}_d^T \phi(f_i) - b_d \leq \varepsilon + \xi_i^*, \xi_i, \xi_i^* \geq 0, \\ & i = 1, \dots, l \end{aligned} \quad (5)$$

代入构造的拉格朗日函数后得出其对偶形式:

$$\begin{aligned} \min_{\alpha, \alpha^*} & \frac{1}{2} (\alpha - \alpha^*)^T \mathbf{Q} (\alpha - \alpha^*) + \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \\ & \sum_{i=1}^l \text{depth}_i (\alpha_i - \alpha_i^*), \end{aligned}$$

$$\begin{aligned} \text{s. t. : } & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C_d, \\ & i = 1, \dots, l \end{aligned} \quad (6)$$

其中,核函数 $Q_{ij} = K(f_i, f_j) = \phi(f_i)^T \phi(f_j)$,利用二次规划算法可以求最优解 α 和 α^* 。对于新提取的特征向量 f' ,由 Karush Kuhn Tucker 最优化条件(KKT 互补条件)相应的深度值 depth' 就可以被估计为:

$$\text{depth}' = \sum_{i=1}^l (-\alpha_i + \alpha_i^*) K(f_i, f') + b_d \quad (7)$$

4 实验

4.1 模型训练

训练样本是单幅红外图像,且该红外图像已有其对应深度图。为控制训练的计算量,并没有将所有图像中的像素点都参与训练,而是随机选取图像中的像素点共 t 个。对于每一个像素点都计算经过筛选得出的 10 个特征向量,深度值则选择该红外图像对应深度图的灰度值。这样就得到了训练样本集合 $\{f_i, \text{depth}_i\}, i = 1, \dots, t$ 。

在下面的实验中,红外深度估计模型的核函数选择的是径向基函数(RBF),RBF 中的参数 γ 和模型中的常数 C_d 由交叉验证法确定。每一次实验都随机选择红外图像中的 5000 个点作为训练样本。图 3 给出了源红外图像及其对应的深度图,样本集合便是在这两幅图像中取得的。



(a) 源红外图像



(b) 源红外图像对应的深度图

图 3 源红外图像及其对应的深度图

Fig.3 source infrared image and its depth map

4.2 结果分析

通过多次实验计算的参数及其对应的最小均方误差可知惩罚因子的结果都相对较大,所以拟合结果的损失也较大。选取其中一次实验取出前 500 个点的拟合结果进行查看,如图 4 所示在多个连续的峰值区间,拟合误差较大,其中 1 号线为原始深度值,2 号线为拟合的深度值。恢复的原始红外图像的深度图如图 5 所示,可以看出拟合的效果比较接

近原始深度。

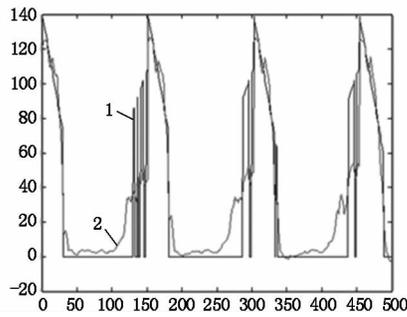


图4 实验1中对前500个点的拟合效果

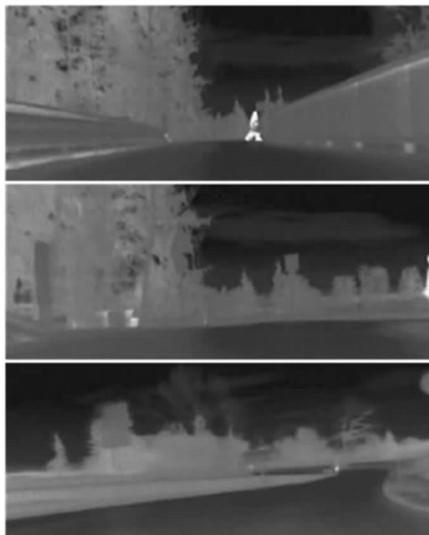
Fig.4 fitting results by selecting 500 data in the first experiment



图5 由学习模型估计的源红外图像的深度图

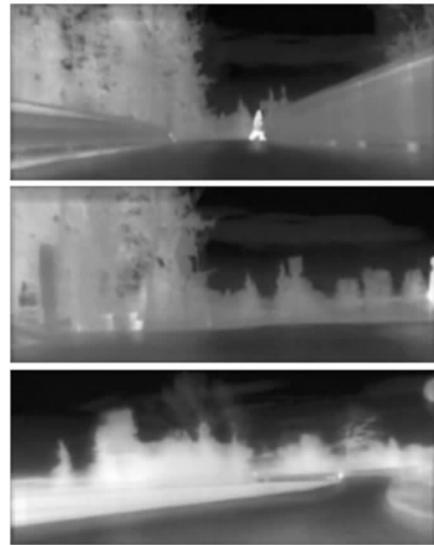
Fig.5 predicted depth map from source infrared image based on model

测试图像选取了与训练图像场景相近的图像,并分别计算了它们对应的深度图如图6所示。由于样本图像大小的限制,且源图像中路面的宽度较短,取到路面的样本点较少,因此对于之后的深度估计中路面的深度信息很少的被反映出来。另外训练用的对应于源红外图像的深度信息若不是很准确,那么训练出来的模型估计出的深度图像也会受到影响。但是从总体来说,除了道路深度信息的失真之外,其他场景在深度图中的反映还是比较准确的。



(a) 测试用的红外图像

(a) test infrared images



(b) 模型估计出的对应深度图像

(b) corresponding depth maps estimated using model

图6 测试用的红外图像及模型估计出的对应深度图像

Fig.6 test infrared images and corresponding depth maps estimated using model

5 结论

本文首次采用“监督学习”的方法通过非线性学习模型来估计单目红外图像的深度分布。该算法首先通过逐步线性回归和独立成分分析(ICA)寻找对于红外图像深度相关性较强的特征,然后在红外图像深度估计中采用“监督学习”的方法通过支持向量回归理论模型对筛选的特征进行训练,在训练过程中通过已知数据回归后的最小均方误差对模型参数进行修正,训练后的模型可对新的单目红外图像的深度分布进行估计。虽然现在的实验结果没有达到完全精确(对于路面的深度估计尚不精确),但综合来看该算法的实用性较好,能够利用已训练好的模型对场景的深度进行估计,将具有较好的应用前景。

参考文献:

- [1] Daniel Scharstein, Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal of Computer Vision, 2002, 47(1): 7-42.
- [2] Horn B K B. Obtaining shape from shading information, chapter 4 in the psychology of computer vision [M]. New York: McGraw Hill, 1975: 115-155.
- [3] Ashutosh Saxena, Andrew Ng, Sung Chung. Learning depth from single monocular images [C]. NIPS, 2005, 18: 44-58.

- [4] Derek Hoiem, Alexei A Efros, Martial Hebert. Automatic photo pop-up[J]. ACM Transactions on Graphics, 2005, 24(3):577-584.
- [5] Sun Xinde, Bo Shukui, Li Lingling. Study of infrared image clutter suppression based on background estimation [J]. Laser & Infrared, 2011, 41(5):586-590. (in Chinese)
孙新德, 薄树奎, 李玲玲. 基于背景估计的红外图像杂波抑制方法研究[J]. 激光与红外, 2011, 41(5):586-590.
- [6] Gu Xiaojing. Study on methods of natural color night vision based on image analysis[D]. Shanghai: DongHua University, 2011. (in Chinese)
古小婧. 基于图像分析的自然彩色夜视成像方法研究[D]. 上海: 东华大学, 2011.
- [7] Hendrix Claudia, Barfield Woodrow. Relationship between monocular and binocular depth cues for judgements of spatial information and spatial instrument design[J]. Displays, 1995, 16(3):103-113.
- [8] Wang Junfeng. Blind signal processing and its application based on PCA and ICA[D]. Wuhan: Huazhong University of Science & Technology, 2005. (in Chinese)
王峻峰. 基于主分量、独立分量分析的盲信号处理及应用研究[D]. 武汉: 华中科技大学, 2005.
- [9] Tan Ailing, Bi Weihong. Identification of honey by NIR spectroscopy technology based on KPCA and LSSVM[J]. Laser & Infrared, 2011, 41(12):1331-1336. (in Chinese)
谈爱玲, 毕卫红. 基于 KPCA 和 LSSVM 的蜂蜜近红外光谱鉴别分析[J]. 激光与红外, 2011, 41(12):1331-1336.