

DOI:10.3969/j.issn.1673-4785.201403009

网络出版地址: <http://www.cnki.net/kcms/doi/10.3969/j.issn.1673-4785.201403009.html>

单摄像头下基于样本学习的人体深度估计

何磊¹, 苏松志^{1,2}, 李绍滋^{1,2}

(1. 厦门大学 信息科学与技术学院, 福建 厦门 361005; 2. 福建省仿脑智能系统重点实验室, 福建 厦门 361005)

摘要: 深度图像的研究是当前计算机视觉的研究热点。从图像中获取深度信息有2种方法: 1) 利用深度感应器, 该方法的缺点是成本高; 2) 基于一个场景的多幅图像或图像序列, 通过求取视差, 获得深度值, 该方法的缺点是需要摄像机参数, 专业知识要求较高。针对上述情况, 提出了一种简单有效的从单摄像头捕获的人体图像中估计出人体深度信息的方法, 利用深度摄像机建立人体的“表观深度”图像对, 然后对单摄像头获取的彩色图像进行人体表观特征提取, 根据该表观特征检索图像对数据库, 并对获得的人体深度进行估计和优化。最后, 在厦门大学的深度数据库上, 验证了该方法的有效性。

关键词: 深度图像; 单摄像头; 人体深度估计; 基于样本的学习; 特征提取; 特征匹配; 相似样本; 深度数据库

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1673-4785(2014)02-0161-07

中文引用格式: 何磊, 苏松志, 李绍滋. 单摄像头下基于样本学习的人体深度估计[J]. 智能系统学报, 2014, 9(2): 161-167.

英文引用格式: HE Lei, SU Songzhi, LI Shaozi. Human depth estimation on the basis of the example learning method under a single camera[J]. CAAI Transactions on Intelligent Systems, 2014, 9(2): 161-167.

Human depth estimation on the basis of the sample learning method under a single camera

HE Lei¹, SU Songzhi^{1,2}, LI Shaozi^{1,2}

(1. School of Information Science and Technology, Xiamen University, Xiamen 361005, China; 2. Fujian Key Laboratory of the Brain-Like Intelligent Systems, Xiamen 361005, China)

Abstract: Currently, the research on depth imaging is one of the hotspots concerning computer vision. There are two methods for acquiring depth information from images: 1) The utilization of depth sensors, with the disadvantage of this method being its considerable expense. 2) The utilization of multiple images or a sequence of images for the same scene by calculating the optical parallax for getting depth information, with the disadvantages of this method including the requirement of camera parameters and the need for a large amount of professional knowledge. In response to the circumstances mentioned above, this paper proposes a simple and efficient method that estimates human depth information from images captured by a single camera. The basic ideas of this method include establishing many pairs of human 'appearance depth' images by use of a depth camera, extracting human appearance features from colorful images captured by a monocular camera and then searching the image pairs database according to the appearance features, and estimating and optimizing human depth information obtained from the database of the pairs of images. Finally, simulation experimental results in the Xiamen University depth database established by ourselves were found to validate the effectiveness of the proposed method.

Keywords: depth image; a single camera; human depth estimation; example-based learning method; feature extraction; feature matching; similar samples; depth database

收稿日期: 2014-03-03. 网络出版日期: 2014-03-31.

基金项目: 国家自然科学基金资助项目(61202143); 福建省自然科学基金资助项目(2013J05100); 厦门市科技重点项目资助项目(3502Z20123017); 湖南省自然科学基金资助项目(12JJ2040).

通信作者: 苏松志. Email: ssz@xmu.edu.cn.

传统的机器视觉是把三维景物投影成二维图像, 然后通过建立起的图像数据与成像过程及景物特征的数学关系来恢复三维场景。成像的过程中损失了深度信息, 因而重构的三维场景并不是惟一的, 使机器视觉的发展和应用受到了限制。图像深度信息获

取的基本任务就是利用光学、几何学等方法从二维图像中估计出各物体之间以及同一物体不同部位之间与摄像头的距离,从而估计出图像的深度信息。深度图像获取的方法可归纳为主动式深度传感器的方法^[1]和被动式计算机立体视觉的方法^[2-3]。

主动式深度传感器的方法主要是利用物理学和光学等知识来获取深度图像。激光雷达深度成像^[4-5]的基本原理是每隔一定时间间隔向被测目标发射信号并检测回波,从而确定距离。莫尔条纹技术^[1-2]利用刻有高频等间距条纹的标尺光栅与指示光栅相重叠,并且二者之间有一个很小的夹角时相对运动形成低频莫尔条纹的原理。结构光法是近年来在激光逐点扫描法基础上发展起来的一种非接触测量方法。它用激光线光源,经柱面镜产生平面光照射在被测物上,在被测物上产生一条明亮的光带,通过 CCD 摄像机摄入,经数字信号处理可获得光带的数字图像,再经计算机处理即得物体在光切面上的二维轮廓信息,若进一步沿着第三维直线方向步进测量,就可获得目标的深度图像。国内外已有许多方法采用结构光实现三维测量。Rioux、Haggren、Lorenz 等发表了多种结构光单点测距系统。除了单点法,Shirai 和 Will 又采用了结构单线法。其他的主动式传感器的方法还有接触式测量仪、工业 CT、变焦距法、三角法和干涉法等。上述主动式的方法都有一个共同的缺点:需要借助特殊的设备,估计深度信息需要还原成像场景。无法通过普通的摄像机获得图像,对其估计深度信息。假如手头有一副图像,估计它的深度就需要通过特殊设备还原成像场景来对其估计深度。

相对于获取深度几何信息的主动式方法,采用传统计算机视觉的方法获得深度信息的方法称为被动式方法,又称计算机立体视觉。立体视觉方法按需要的图像数目可分为 3 类:1) 利用一副图像的图像理解方法^[6-7];2) 利用在 2 个不同的观察点获得的同一景物的 2 幅图像恢复三维立体信息的双目立体视觉^[8-10];3) 利用多个观察点获得多幅图像的多目立体视觉^[11]。其中双目立体视觉^[12]直接模仿了人和许多动物通过双眼获得景物的深度信息的方式,得到了更为深入的研究。Barnard^[13-14]将立体视觉技术划分为 6 个部分,分别为图像获取、摄像机定标、特征提取、图像匹配、深度确定、内插。与 Barnard 的划分方法不同,Dhond 和 Aggarward 将立体视觉技术分为 3 个主要步骤:预处理、匹配和深度信息恢复^[15-16]。计算机立体视觉虽然近年来应用越来越广泛,但是它们普遍存在的一个基本问题就是对一个场景景物的深度估计需要多幅图像,日常生活中往往不会对同一个场景拍摄不同角度的几幅图像,这就大大限制了它的应用范围。

从单幅图像中恢复深度信息是计算机视觉领域

的一个难点,需要考虑整幅图像的全局结构,也需要利用关于场景的先验知识。如何建立单幅图像彩色信息到深度信息的映射,具有广泛的理论意义和很好的应用价值。针对上述前 2 个问题,提出了一个基于样本学习的方法,通过单摄像头采集的一幅包含人体的图像,对其中的人体深度进行估计。该方法的基本思路是:建立人体深度数据库,学习人体特征,通过特征匹配从相似样本中对人体深度进行估计和优化。本文的方法不需要特殊的设备,也不需要对于一个场景采集多幅不同角度的图像,仅仅通过单摄像头采集的包含人体的图像,对人体深度进行估计,从而建立单幅图像彩色信息到深度信息的映射,同时克服上述 2 种方法各自的缺点。为估计单幅图像的人体深度信息提供了新的思路。

1 基于样本学习的人体深度估计

本文提出的方法主要理论基础是基于特征统计学习的方法,在建立的数据库中学习到所需要的重要信息,然后根据所学习到的知识,用机器学习的方法估计出新来的目标的深度信息。这种方法首先需要通过学习,对数据库的每个目标进行特征提取,这个特征可以是亮度、深度、纹理或者几何形状,然后对估计目标的亮度、深度、纹理、几何形状等特征各自建立概率函数,最后将测试目标与数据库中相似目标的相似程度表示为概率大小,取概率最大的目标深度作为估计的深度。

基于机器学习的方法有如下优点:只要数据库足够完备,任何和数据库目标一致的对象都能进行深度估计,并且估计质量和效率都很高,很少需要人工交互。这种重建技术最大的困难是建立完备的数据库。此外,如果能在匹配特征上有更好的选择,估计的深度信息会更加准确。本文的方法主要是通过单摄像头采集的单幅包含人体的图像,从已经建立的数据库中学习到的知识,估计出单幅图像人体的深度信息。其中需要解决的问题有:

1) 数据库如何建立,包含哪些内容,需要根据选择的匹配特征来决定;

2) 如何选择人体的部分特征作为学习和匹配的重要信息;

3) 如何根据选择的人体部分特征在数据库中找到需要的信息;

4) 找到了需要的信息后,如何根据这些信息,估计出单幅图像中人体的深度信息,并进行优化。

针对上述问题,本文设计了如图 1 所示的基于样例学习的人体深度信息估计算法基本流程,其核心思想是:相同姿态的人在以同一个角度面对镜头拍摄图像,他们的人体的各个部分有大致相同的深度分布。换言之,就是说具有相似轮廓的人体图像,他们的人体各部分的相对深度也是相似的。这

里所说的深度不是传统意义上的目标离摄像机光心的距离,这个相对深度的含义仅仅是指在人体上选取一个参考点,然后计算出的人体离参考点的距离。这个相对距离不受人体离摄像机距离的影响,把相同姿态的人体在离摄像机不同距离情况下所拍摄图

像的人体深度信息认为是相同的。当然,不是所有找到的样本都对目标图像估计准确,这也是为什么要找一些候选样本,然后用全局优化的方法对他们进行加权和内插的原因,而不是简单的利用这些样本的绝对深度信息。具体算法如算法 1 所示。

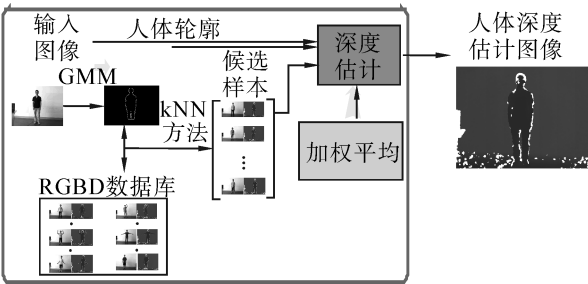


图 1 基于样例学习的人体深度信息估计算法基本流程

Fig.1 The basic flow of the human body depth information estimation algorithm based on sample learning

算法 1:基于样例学习的人体深度信息估计

输入:单幅 RGB 人体彩色图像;
输出:人体深度图像。

1) 给出一个目标图像,准确地分割出人体的轮廓;2) 在给定的 RGBD 数据库中,找到 k 个候选样本,这 k 个候选样本一定是与目标图像在人体轮廓上最相似的 k 个;3) 通过对 k 个候选样本的深度图像进行加权平均完成优化过程,得到估计的深度图像。论文的第 2、3 和 4 部分分别对算法中的每个步骤进行描述。

2 匹配特征

如何选择人体的部分特征作为匹配特征将直接关系到实验结果的好坏。可以选择亮度、深度、纹理、几何形状作为匹配特征。由于本文的方法是基于统计学的方法,认为单幅图像中人体各部位的深度最可能与数据库图像中人体姿态相似的那部分样本中人体各部位的深度相似。所以就选择相应的轮廓特征作为匹配的特征。

选择的轮廓特征是基于这样一个事实:在图像中的人体如果具有相似的姿态,那么它们的人体轮廓上各个点到它们重心的距离必然是相似的。那么基于最直观的角度,可以首先分割出人的轮廓,然后计算出人体的轮廓上各点到重心的距离,组成一个 N 维的向量,那么这个 N 维的向量就是所提取出的关于这个人体的轮廓信息。

假设人体的密度是均匀的,那么计算轮廓上各个点的平均横坐标和平均纵坐标就是这个人体轮廓重心,计算公式如式(1)、(2)所示:

$$\bar{x} = \sum_{i \in \theta} x_i / n \tag{1}$$

$$\bar{y} = \sum_{i \in \theta} y_i / n \tag{2}$$

式中: θ 是所有轮廓上点的集合, x_i 是轮廓上点的横

坐标, y_i 是轮廓上点的纵坐标, (\bar{x}, \bar{y}) 就是所计算的重心。得到重心以后,接下来计算出轮廓上各点到重心的距离,如式(3)所示:

$$a_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \tag{3}$$

式中: a_i 就是轮廓上点 i 到重心的距离,这样把轮廓上所有的点到重心的距离计算出来得到如下一组距离向量: $[a_1 \ a_2 \ a_3 \ \cdots \ a_T]$ 。

式中: T 是轮廓上的总点数。然后在这个不定维数的向量中均匀地选择 N 维,作为最后的特征向量: $A = [a'_1 \ a'_2 \ a'_3 \ \cdots \ a'_N]$ 。

这里会遇到一个问题,就是轮廓上的点数 T 一般不会刚好是 N 个点,有时候会多于 N 个,有时候会比 N 个少,这时候用线性插值的方法,均匀的在轮廓上取得 N 个点,计算出特征向量 A 。人体轮廓特征提取如图 2 所示。

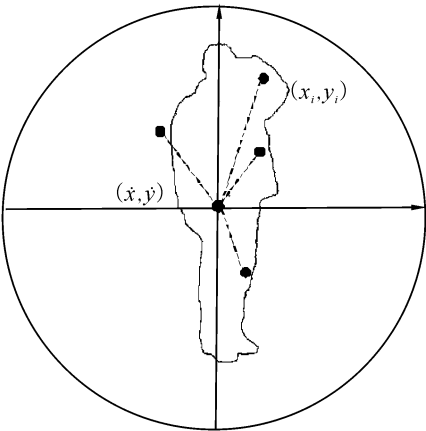


图 2 人体轮廓特征提取

Fig.2 The human body contour feature extraction

3 匹配最相似的样本

选择了轮廓特征作为匹配特征,通过线性内插法选择轮廓上的 N 个点计算到重心的距离得到 N

维向量 \mathbf{A} 。接下来就是如何利用这个 N 维向量 \mathbf{A} 在数据库中找到与这个 N 维向量最相似的一些样本。这是一个简单的问题,同样对数据库中的每个样本都计算出这样一个 N 维向量。然后在这些个 N 维向量中找到一些与测试样本中的 N 维向量距离最小的 N 维向量,与这些找到的 N 维向量对应的样本就是和测试样本最相似的那些样本。由于每个 N 维向量中的各维大小不仅与人的轮廓有关,而且与人离镜头的距离有密切相关,计算相似度的时候必须把这一因素考虑在外。所以必须对这些 N 维向量进行归一化,如式(4)所示:

$$\mathbf{A} = \frac{\mathbf{A} - \min(a'_1, a'_2, a'_3, \dots, a'_N)}{\max(a'_1, a'_2, a'_3, \dots, a'_N) - \min(a'_1, a'_2, a'_3, \dots, a'_N)} \quad (4)$$

这样对 N 维向量归一化的过程相当于把人体离摄像机距离的因素考虑在外, N 维向量的每个分量大小只包含了人体的轮廓信息。那么,现在需要做的就是找到与测试样本中人体轮廓最相似的一些样本。计算出测试样本对应 N 维向量与数据库中样本 N 维向量的距离来找到这些样本,如式(5):

$$d_i = \|\mathbf{A}_i - \mathbf{A}\| \quad (5)$$

式中: d_i 是测试样本对应 N 维向量与第 i 个 N 维向量之间的距离, \mathbf{A}_i 是第 i 个样本对应的 N 维向量, \mathbf{A} 是测试样本的 N 维向量。然后把 d_i 进行排序,找到最小的 k 个 d_i , 与之对应的样本就是最相似的样本。一般认为距离越小,那么样本中人体的轮廓与测试样本中人体的轮廓越相似。

4 深度估计与优化

通过提取测试样本中的轮廓信息,得出了一个关于轮廓信息的 N 维向量,基于这个 N 维向量,通过计算这个向量与数据库中各个样本的距离,找到与测试样本人体轮廓最相似的 k 个样本,接下来根据这 k 个样本的人体深度信息,来估计测试样本中人体的深度信息。

一般认为在摄像头面前拥有相似轮廓的人体,它们的姿态都是相似的,同时它们面对摄像机的角度也是相似的,那么人体上的 2 点之间的相对深度应该趋于一致,例如手相对于胸的距离在上述条件下应该趋于一致。

所以基本可以得出一个结论就是在摄像头面前人体轮廓越相似,那么它们人体各部分之间的相对深度就越相似,比如同样姿势的人体,手到头、头到胸、胸到脚之间的相对距离都是相似的。

到此为止可以给出一种最直观的估计测试样本人体深度的一个方法,如式(6)所示:

$$P_{\text{test}} = \frac{\sum_{i=1,2,\dots,K} P_i \frac{1}{d_i}}{\sum_{i=1,2,\dots,K} \frac{1}{d_i}} \quad (6)$$

式中: P_{test} 是所估计的测试样本人体深度, P_i 是找到的最相似的 k 个样本中的第 i 个样本对应的深度信息, d_i 是 k 个样本中第 i 个样本 N 维向量与测试样本 N 维向量的距离。这个公式说明了轮廓距离越小、越相似的样本,对应的深度信息在最后的估计结果中影响权值越大,反之影响权值越小。

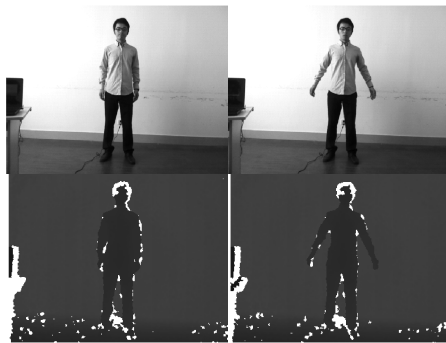
5 实验结果及分析

实验的目的是在建立的厦门大学深度数据库上,验证本文提出单摄像头下基于样本学习的人体深度估计方法的有效性。本实验分为 2 个部分:1) 选择多组测试样本进行实验,根据后文中提出的评价指标进行评价。2) 对同一组测试样本设定不同的 k 值,分别测得实验指标,说明 k 值对整个实验结果的影响。

5.1 数据库的建立和评价指标

由于本文所采用的方法比较新颖,所以无法使用国际上实验所用的标准数据库,必须自己建立所需要的数据库。

如上文所述,数据库中的样本至少由如下部分组成:一副 RGB 图像,与 RGB 对应的包含深度信息的图像,另一副对应的人体轮廓的图像。数据集的部分深度信息如图 3 所示。



(a) 图例 1



(b) 图例 2

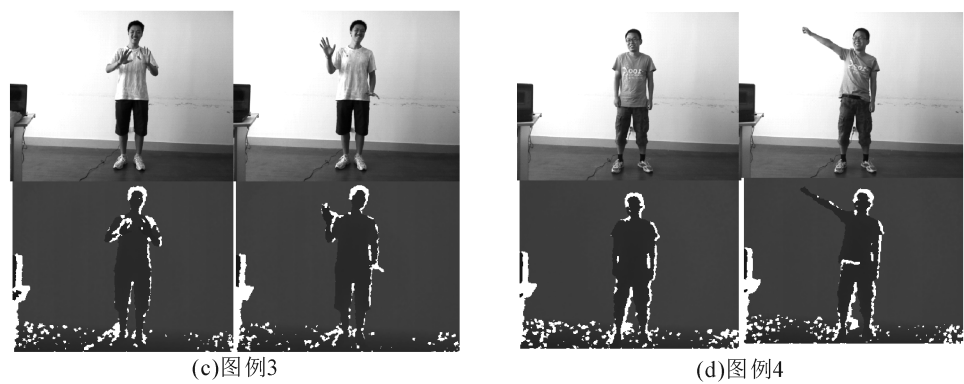


图 3 厦门大学深度数据库部分深度信息

Fig.3 Xiamen University depth section depth information database

因为需要深度信息,所以采集图像时需要用到 kinect,在采集一副图像的时候同时得到对应的深度信息。用图像分割的方法,分割出人体的部分,找到人体的轮廓,以便匹配人体轮廓。

最后是样本采集的环境问题。采集的样本全部是在室内的环境,由于室外环境复杂,kinect 采集的深度图像可能会出现很多噪声,导致深度信息不够准确影响实验结果。采集的样本必须包含人的全身,暂时不考虑只有部分人体的条件。样本数量必须足够,以便可以收集到人体在室内环境下的各种姿态,保证估计图像的人体轮廓在数据库中总能找到非常相似的那些样本,这对实验结果至关重要。

厦门大学深度库详细信息如表 1 所示。

实验必须在建立评价指标的基础上来说明本文方法的有效性。由于在计算机视觉领域对人体深度信息的估计还没有标准的、公开可用的基准实验指标,所以设计了自己的基准实验评价指标。采用文献[17]中的评价方案来比较每一幅估计的深度图

像 D_Q 和已知真值 D_{Q^*} 之间的误差:lg 误差, (lg) , $|lg(D_Q) - lg(D_{Q^*})|$, 均方根误差 (root mean square error, RMSE), $\sqrt{\sum_{i=1}^N (D_{Q_i} - D_{Q_i^*})^2 / N}$, 相对误差 (relative error, REL), $\frac{|D_Q - D_{Q^*}|}{D_{Q^*}}$ 。

5.2 结果及分析

第 1 部分实验是选择数据库里面的 11 组样本作为训练样本,然后剩下的 4 组作为测试样本。参数 k 为 5 固定不变。实验的输入为包含人体的彩色图像,先通过预处理得到人体轮廓。预处理方法可以用 GMM^[18-19]、Vibe^[20]、SOBs^[21]、Codebook^[22] 等背景减除的方法,这里采用 GMM。

然后通过本文提出的方法估计出人体深度信息。部分实验结果如图 4 所示。图 4 表示的是 4 组测试样本的部分实验结果,选择了几组包含不同人体姿态的测试样本进行说明。

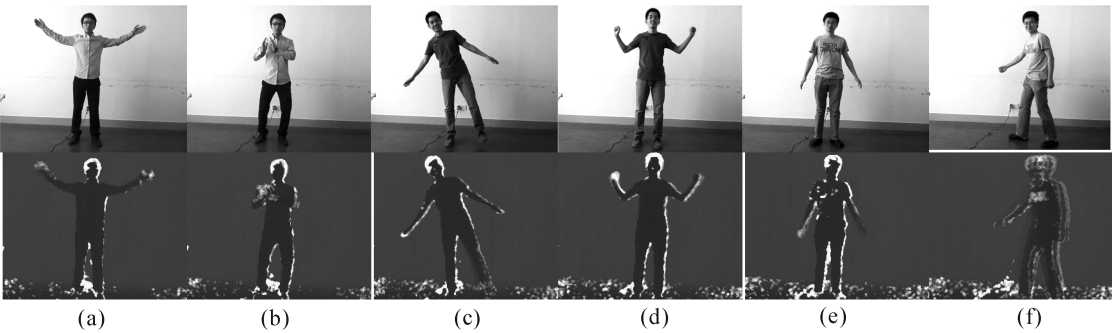


图 4 人体深度信息估计部分实验结果

Fig.4 The body part depth information estimation results

表 1 厦门大学表观深度图像数据集的统计信息

Table 1 Xiamen University statistics of apparent depth image data set

拍摄环境	拍摄距离	包含信息	人数	样本总数	训练样本	测试样本
室内	3~4 m	人体 RGB 彩色和深度图像对	15	18 805	4 288	14 517

图 4 中第 1 行的图像是实验测试图像,第 2 行图像是实验所得测试图像的人体深度估计信息。从图 4 中可以看出,由前 5 个测试样本实验所得到的人体深度信息估计中,除了人体边缘部分的深度信息估计比较模糊,存在误差,其他部分的深度信息估计的比较准确。因为虽然以人体轮廓作为匹配特征,但是由于相同姿态的人体轮廓特征总不可能完全相同,所以这就解释了在人体边缘部分估计的深度信息有所误差,而在其他剩余的部分深度信息比较准确的问题。图 4(f) 的实验结果不甚理想,主要原因是:这幅图像的人体姿态比较不常见,而在训练样本中没有类似人体姿态的样本能够匹配到,所以导致人体深度信息估计偏差较大,这也间接证明前文提出的一个观点:训练样本越多,包含的人体姿态越多,那么对人体深度信息的估计越准确,反之则相反。

4 组测试样本的估计结果分别计算的 $\lg 10$, RMSE 和 REL 误差如表 2 所示。4 组测试样本分别包含 865、1174、933、1316 幅各种人体姿态的 RGB 彩色图像。表 2 得到的是每组测试样本的评价误差。可以从 REL 相对误差这一项看出,本文的方法对人体深度信息的估计很准确\误差很小,这也符合从图 4 得出的主观感受。

表 2 4 组测试样本的估计误差

Table 2 Four groups of test sample estimation error

No.	lg	RMSE	REL
1	0.014 3	1 212.5	$9.131\ 0\times10^{-7}$
2	0.021 1	1 655.1	$1.320\ 8\times10^{-6}$
3	0.015 8	1 285.9	$9.935\ 6\times10^{-7}$
4	0.021 9	1 818.5	$1.382\ 7\times10^{-6}$

最后通过对同一组测试样本设定不同的 k 值来研究 k 值对实验结果的影响。分别计算测试样本在 k 值为 3、5、7、9 下的 \lg , RMSE 和 REL 误差,如表 3 所示。从表 3 中可以看出, k 值的变化对实验结果的影响并不大,不同的 k 值误差之间差别很小。但是可以看出随着 k 的增大误差在缓慢变大,这并不难理解。由于 k 值的增大,匹配到的相似样本数越多,那么匹配到错误样本的几率就逐渐增大,有可能相似样本中有一些和测试样本的人体姿态并不相同,但是也被选择为相似样本,这就在后面的优化过程中出现了误差。所以并不建议选择很大的 k 值,一般选择 5,在数据库样本充足的情况,可以适当的增大 k 值。

表 3 不同 k 值下的实验误差

Table 3 The experimental error under different k values

k	lg	RMSE	REL
3	0.010 6	945.7	$6.991\ 2\times10^{-7}$
5	0.014 3	1 212.5	$9.131\ 0\times10^{-7}$
7	0.016 6	1 364.4	$1.037\ 8\times10^{-6}$
9	0.018 3	1 469.0	$1.128\ 1\times10^{-6}$

6 结束语

文中提出了一种新的方法,利用单摄像头采集的单幅图像估计出人体的深度,突破了传统的通过特殊成像器材和多视图的方法来估计深度的范畴。从样本学习的角度出发,通过找到测试图像中人体与数据库中的相似信息,从数据库已有的信息来估计出人体的深度。通过实验也很好验证了该方法在简单的室内环境下的有效性。本文的实验在建立的厦门大学深度数据库上完成的,数据库中人体的姿态和面对摄像头的角度都是有限的,接下来的工作主要集中于如何在原有的数据库上扩充样本数量,使得数据库包含的人体姿态更多,人体面对摄像头的角度更加丰富,环境更加复杂普遍。如何从更加广泛的数据库上验证方法的可行性和有效性,讨论数据库的完备性和在庞大的数据下建立一种快速有效的检索匹配特征的方法,会是将来工作的重点和方向。

参考文献:

[1] NITZAN D, BRAIN A E, DUDA R O. The measurement and use of registered reflectance and range data in scene analysis[J]. Proceedings of the IEEE, 1977, 65(2): 206-220.

[2] LEWIS R A, Johnston A R. A scanning laser rangefinder for a robotic vehicle[C]//IJCAI. 1977: 762-768.

[3] 游素亚. 立体视觉研究的现状与进展[J]. 中国图象图形学报: A 辑, 1997, 2(1): 17-24.

YOU Suyu. The present situation and progress in the study of stereo vision[J]. Journal of Image and Graphics: A, 1977, 2(1): 17-24.

[4] HERSMAN M, GOODWIN F, KENYON S, et al. Coherent laser radar application to 3D vision and metrology[C]//Proc of Vision 87 Conf. London, 1987: 465-579.

[5] 赵远, 蔡喜平. 成像激光雷达技术概述[J]. 激光与红外, 2000, 30(6): 328-330.

ZHAO Yuan, CAI Xiping. Imaging laser radar overview[J]. Laser and Infrared, 2000, 30(6): 328-330.

[6] HORN B K P. Shape from shading: a method for obtaining

- the shape of a smooth opaque object from one view. AITR-232[R]. Cambridge, USA: MIT Artificial Intelligence Laboratory, 1970.
- [7] WOODHAM R J. Photometric method for determining surface orientation from multiple images[J]. Optical Engineering, 1980, 19(1): 139-144.
- [8] AKIMOTO T, SUENAGA Y, WALLACE R S. Automatic creation of 3D facial models[J]. Computer Graphics and Applications, 1993, 13(5): 16-22.
- [9] CHEN C L, TAI C L, LIO Y F. Virtual binocular vision systems to solid model reconstruction[J]. The International Journal of Advanced Manufacturing Technology, 2007, 35(3/4): 379-384.
- [10] 隋婧,金伟其.双目立体视觉技术的实现及其进展[J]. 电子技术应用, 2005, 30(10): 4-6.
SUI Jing, JIN Weiqi. The realization of the binocular stereo vision technology and its progress[J]. Application of Electronica Technology, 2005, 30(10): 4-6.
- [11] SEITZ S M, CURLESS B, DIEBEL J, et al. A comparison and evaluation of multi-view stereo reconstruction algorithms[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S.l.], 2006, 1: 519-528.
- [12] 吴立德.计算机视觉[M]. 上海:复旦大学出版社, 1993: 119-205.
- [13] BARNARD S T, FISCHLER M A. Computational stereo[J]. ACM Computing Surveys (CSUR), 1982, 14(4): 553-572.
- [14] LEMMENS M. A survey on stereo matching techniques[J]. International Archives of Photogrammetry and Remote Sensing, 1988, 27(B8): V11-V23.
- [15] DHOND U R, AGGARWAL J K. Structure from stereo—a review[J]. IEEE Transactions on Systems, Man and Cybernetics, 1989, 19(6): 1489-1510.
- [16] MAYHEW J E W, FRISBY J P. Psychophysical and computational studies towards a theory of human stereopsis[J]. Artificial Intelligence, 1981, 17(1): 349-385.
- [17] SAXENA A, SUN M, NG A Y. Make3d: learning 3d scene structure from a single still image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 824-840.
- [18] ZIVKOVIC Z. Improved adaptive Gaussian mixture model for background subtraction[C]//Proceedings of the 17th International Conference on Pattern Recognition. [S.l.], 2004, 2: 28-31.
- [19] 王亮,胡卫明,谭铁牛.人运动的视觉分析综述[J]. 计算机学报, 2002, 25(3): 225-237.
WANG Liang, HU Weiming, TanTieniu. People movement of the visual analysis overview[J]. Chinese Journal of Computer, 2002, 25(3): 225-237.
- [20] BARNICH O, VAN DROOGENBROECK M. ViBe: a universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724.
- [21] MADDALENA L, PETROSINO A. A self-organizing approach to background subtraction for visual surveillance applications[J]. IEEE Transactions on Image Processing, 2008, 17(7): 1168-1177.
- [22] KIM K, CHALIDABHONGSE T H, HARWOOD D, et al. Real-time foreground-background segmentation using codebook model[J]. Real-time Imaging, 2005, 11(3): 172-185.

作者简介:



何磊,男,1988年生,硕士研究生,主要研究方向为图像处理、模式识别、计算机视觉。



苏松志,1982年生,男,博士,助理教授,主要研究方向为人体行为分析与理解。主持国家青年基金一项,主持省青年基金一项,参与多项国家级基金项目,发表学术论文多篇,其中被SCI检索7篇。



李绍滋,1963年生,男,博士,教授,博士生导师,福建省仿脑智能系统重点实验室(厦门大学)副主任,福建省人工智能学会副理事长兼秘书长,主要研究方向为运动目标检测与识别、机器学习与计算机视觉、自然语言处理与多媒体信息检索等。发表学术论文160余篇,其中被SCI检索16篇、被EI检索142篇。