

# 基于视觉心理学的物体大小恒常性计算

吴爱民<sup>1,2</sup>, 须 德<sup>1</sup>, 王海霞<sup>1</sup>, 吴 静<sup>1</sup>

(1. 北京交通大学计算机研究所, 北京 100044; 2. 山东省东营职业学院, 山东东营 257091)

**摘 要:** 知觉恒常性是人类感知世界最重要、最突出的方面, 它为解决计算机图像理解和物体识别等经典计算机视觉难题提供了新的思路. 大小恒常性是最重要的知觉恒常性之一. 正确的图像物体大小恒常性感知的关键在于准确计算物体在图像中的感知深度. 本文总结了人眼使用的各种图像深度线索, 提出了这些线索融合与冲突的解决方案, 然后用数学方法建立了图像物体大小恒常性计算模型. 实验结果表明该模型是有效的. 本文是应用视觉心理学来解决计算机视觉问题的一次成功而有益的探索.

**关键词:** 大小恒常性; 视觉心理学; 图像理解; 图像深度线索

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2006) 06-1096-08

## Object Size Constancy Computation Based on Visual Psychology

WU Ai-min<sup>1,2</sup>, XU De<sup>1</sup>, WANG Hai-xia<sup>1</sup>, WU Jing<sup>1</sup>

(1. Institute of Computer Science, Beijing Jiaotong University, Beijing 100044, China;

2. Dongying Vocational College, Dongying, Shandong 257091, China)

**Abstract:** Perceptual constancy is the most salient aspect of the way the world looks to us. It may enlighten some new ideas for many open issues of computer vision, such as automatic image understanding and image object recognition. Size constancy is one of the most important perceptual constancy. The key to correct size constancy perception of image object is to correctly estimate the depth of objects in the image. This paper introduces most of image depth cues used by human vision, and presents a method to resolve the fusion and confliction of these cues. Then we propose a model to implement size constancy computation of image objects. Experimental results suggest that our model can work reasonably well. In addition, this paper sets a good example for applying visual psychology to resolving computer vision problems.

**Key words:** size constancy; visual psychology; image understanding; image depth cue

## 1 引言

根据几何光学知识, 物体在视网膜的映像轮廓不同于物体的轮廓, 会随着人和环境不断变化, 而且几乎每时每刻都在发生变化. 但是对我们而言, 外界的物体看上去都是一样的, 有着标准的形状、大小、颜色、明度和位置关系. 例如, 随着观察者与桌子的相对运动或照明的变化, 桌子的视网膜映像发生了很大的变化, 但我们对它的感知却基本上没有变化. 这种现象称作知觉恒常性 (Perception constancy), 它是人类感知世界最重要、最突出的方面. 知觉恒常性使人类视觉系统能超越不完全的、易于失真的、模糊的、二维视网膜映像, 而建立起丰富的、稳定的、通常正确的、三维的客观世界表象<sup>[1]</sup>.

从光学成像的角度, 人眼和照相机有着几乎完全相同的成像机制 (初始状态相同). 根据 Marr 的观点, 计算机视觉问题与人类视觉问题几乎完全相同, 两者都是从图像中

发现客观世界有什么事物, 这些事物在什么地方 (目的状态相同). 不同的是, 人类视觉能轻易地完成计算机视觉中的许多经典难题, 如边缘检测、图像分割、物体识别等等. 目前, 大多数研究者认为, 无论在生物社区还是在机器人社区, 人类视觉系统都是最好的、最通用的, 而且绝大多数计算机视觉算法的参考标准要靠人眼标定. 不难想象, 在计算机求解视觉问题的中间状态时, 如果能完全共享人类视觉的计算理论与算法, 从理论上讲计算机视觉的性能就有可能与人类视觉系统相媲美<sup>[2,3]</sup>. 知觉恒常性是人类感知世界的基本方式, 所以计算机视觉也应充分利用人类知觉恒常性的有关理论与算法.

恒常性理论对图像物体识别有着特别重要的意义. 因为随着成像视点的变化, 客观世界中的任一物体都可以产生无限多个二维图像投影, 所以从二维图像出发, 识别出对应的客观世界物体是一对多的数学问题, 也是计算机视觉中的经典难题. 恒常性理论最吸引人的地方是: 面对连

续变化的刺激特征,物体能被稳定、唯一地感知.所以恒常性理论特别有助于解决物体识别中的视点不变(View-point invariant)难题<sup>[4]</sup>.恒常性主要包含如下种类:大小恒常性、形状恒常性、明度恒常性及颜色恒常性等.文献[5]指出:除了有了颜色恒常性的报告外,没有发现计算机视觉学者对其他恒常性的研究文献.我们通过对国内外的中、英文文献资料的检索,也得出了同样的结论.故本文将对大小恒常性进行计算研究.大小是标识物体的一个重要属性.例如,在日常生活中,矮个子被感知为小孩的概率较大,高个子被感知为成年人的概率较大.而且,正确感知物体的大小具有重要的生物学意义.对许多食肉动物而言,小老虎是它们可能的美餐,而大老虎则是它们的杀手.故自动计算图像物体的正常大小对于图像物体识别无疑是十分重要的,这也正是图像物体大小恒常性计算的意义及应用所在.

## 2 大小恒常性心理学基础

视觉心理学的研究表明:尽管物体视网膜映像的大小在变,但看上去它的大小基本不变.心理学家称这一现象为大小恒常性(Size constancy).图1显示了一些大小恒常性的例子.

视觉心理学已经揭示了大小恒常性的计算理论<sup>[1]</sup>,用公式表示为:

$$S = B * A * D \quad (1)$$

S 为物体的感知大小, A 为物体的成像视角, D 为物体的感知深度(也称感知距离),即人类视觉系统感知到的图像

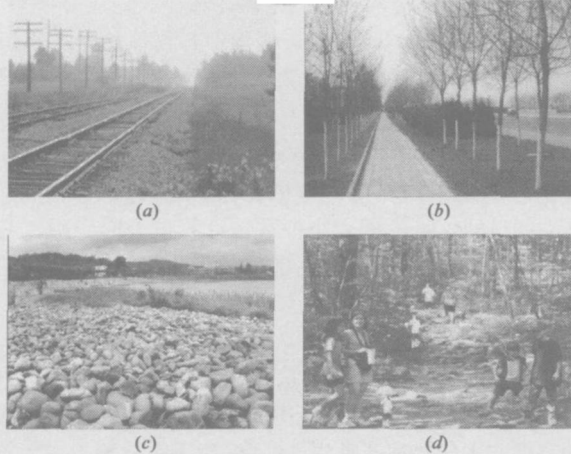


图1 大小恒常性的例子.(a)中的电线杆在图像中的高度不同,但对它们的感知高度却基本相同.(b)中的树在图像中的高度不同,但对它们的感知高度却基本相同.(c)中的鹅卵石在图像中的大小不同,但对它们的感知大小却基本相同.(d)中的上身穿白衣服的三个人在图像中的高度完全相同,但是对他们的感知高度却差别很大;而最上面穿白色上衣的男人,中间穿彩色上衣的男人、及下方背黑包个子最高的女人在图像中的高度差别虽然很大,但对他们的感知高度却基本相同.

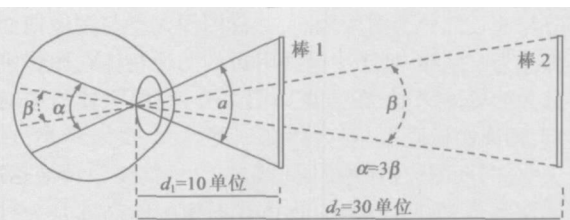


图2 大小恒常性计算理论的例子.同样高度的两根棒放在离观察者不同的距离上供观察,它们在视网膜上的像高是不同的.如果能正确地感知物体的实际深度(距离),应用式(1)就能实现大小恒常性计算.

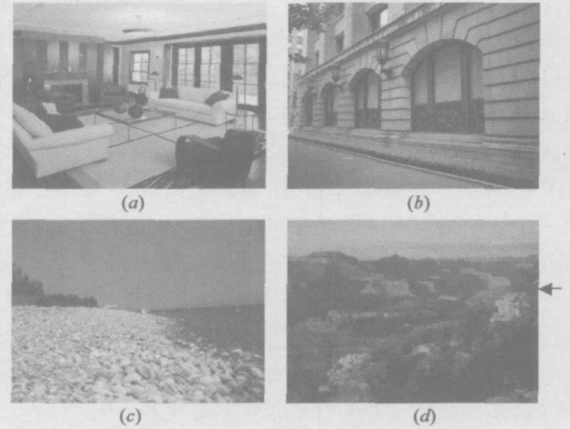


图3 图像深度线索的例子.(a)室内图像:地板(地面)部分越靠近图像底端,感知深度越小;天花板(天空)部分越靠近图像顶端,感知深度越小.(b)线性透视中的会聚线指示灭点与深度变化方向.(c)纹理梯度指示图像深度的变化方向:纹理越粗糙,分辨率越大,感知深度越小,反之亦然.(d)空气透视的例子:呈蓝色、模糊的山脉感觉较远.

上物体在成像时离照相机的距离, B 为与眼睛(相机)有关的成像缩放系数(对于同一次成像, B 值对所有物体都是相同的).物体的成像视角 A 可用物体在图像中的一维大小来表示.图2演示了一个实例.棒1和棒2分别放在离观察者10个单位、30个单位远的位置上,即  $D_{棒1} = 1/3 D_{棒2}$ .根据几何光学知识,物体成像视角的大小与物体离眼睛的距离成反比,故两棒成像视角的关系为  $A_{棒1} = 3A_{棒2}$ .如果能正确感知两棒的实际深度(距离)  $D_{棒1}$ ,  $D_{棒2}$ , 应用式(1)就能实现大小恒常性,即  $S_{棒1} = S_{棒2}$ .

视觉心理学还揭示了人眼使用的各种图像深度线索(Image depth cues),主要包括:物体在图像中的高度(Elevation of object in the image)、线性透视(Line perspective)、纹理梯度(Texture gradient)、大气透视(Aerial perspective)等.下面对它们分别介绍,主要参考资料见文献[1, 6, 7].

(1) 物体在图像中的高度 室外深度图像一般同时包含低处的地面部分与高处的天空部分,如图1(a)~(c)、图3(c)~(d).室内深度图像一般也同时包含低处地板部分与高处的天花板部分,如图3(a).我们分别统称室外图像的天空部分与室内图像的天花板部分为图像天空,地面部分与地板部分为图像地面,并称图像天空与图像地面的分界线为中间

线. 处于中间线附近的图像物体在图像中具有最大的感知深度. 在图像地面部分, 图像物体离中间线越近, 即离图像底端越远, 感知深度越大, 反之越小; 在图像天空部分, 图像物体离中间线越近, 即离图像顶端越远, 感知深度越大, 反之越小. 物体在图像中的高度是一种最重要的图像深度线索. 图像有时也没有中间线, 此时只有地面部分, 如图 1(d)、图 3(b).

(2) 线性透视 客观世界中向远处延伸的平行线, 在图像平面中将靠得越来越远, 甚至会聚 (Converge). 这样一组线称为会聚线 (Converging line), 它们的会聚点称为灭点 (Vanishing point). 在图像中, 平行线指示平坦的表面, 会聚线指示向远处延伸的表面. 对于室外图像, 线性透视效果一般仅出现在图像地面部分, 如图 1(a)、(b)、图 3(b). 但对于室内图像, 同时作用于地面部分与天空部分, 如图 3(a). 线性透视的深度感知规律是: 图像中的物体离灭点越近, 感知深度越大, 反之越小. 同时, 会聚线的中心线也能给出图像感知深度变化最快的方向. 如图 1(a) 两铁轨的中心线 (与图像底端边线夹角约  $45^\circ$ ) 指示感知深度变化最快的方向; 而图 1(b) 小径的中心线 (与图像底端边线夹角约  $90^\circ$ ) 为感知深度变化最快的方向.

(3) 纹理梯度 (纹理密度) 许多表面如墙面、路面及田野里的花朵都有纹理. 当这些纹理表面向远处延伸时, 表面离观察者越远, 分辨率越小、纹理也变得越小. 所以, 对于图像中的同质纹理区域, 分辨率越小, 纹理越小, 感知深度越大; 纹理越大, 分辨率越大, 感知深度越小. 纹理梯度方向可用来指示感知深度变化最快的方向, 如图 1(c)、图 3(c).

(4) 大气透视 大气散射来自各个方向的光线. 来自远处物体的光线也应被大气散射. 因蓝光更容易被散射, 所以物体距观察者越远, 物体附着的蓝色越深. 又因并不是所有的光都以直线进入观察者的眼睛, 所以物体距观察者越远, 物体看上去越模糊. 这种自然现象称大气透视, 它也是一种图像深度线索: 图像中的物体越模糊, 附着的蓝色越深, 感知深度越大, 反之越小. 图 3(d) 是一张大峡谷的照片. 照片中间靠近右侧 (如箭头所示) 的部分清晰、带蓝色较浅, 感知深度较小; 而同一高度靠近左侧的部分模糊, 带蓝色较深, 感知深度较大.

上述所有线索都是相对深度线索. 第一种深度线索 (物体在图像中的高度) 既适用于自然场景图像又适用于人造场景图像, 是最为重要的深度线索. 第二种深度线索 (线性透视) 一般仅适用人造场景图像. 前两种线索都能对整幅图像产生强烈的、难以抗拒的深度感知. 纹理梯度要求有大范围的规则纹理表面. 大气透视仅适用于室外、天气晴朗且实际距离较远的物体. 后两种线索适用范围有限. 一般情况, 各种线索是互补的, 但有时也会产生冲突. 冲突如何解决, 心理学家并没有给出可参考的结论. 本文将在这方面做一些有益的尝试. 其他深

度线索如物体重叠、阴影及对物体的熟悉程度等, 我们在此不作介绍. 原因是: 一对它们的表示与计算比较困难; 二通常它们只局部影响深度感知, 在成像场景景深较大时, 对深度感知的影响很小.

### 3 大小恒常性计算模型

尽管视觉心理学早已揭示了人类视觉系统大小恒常性的计算理论, 但是多年来, 计算机学者没有应用此项成果来解决计算机视觉问题, 所以计算机也就一直没能获得图像物体大小恒常性感知的能力. 应用式 (1) 作为计算理论, 本文提出了图像物体大小恒常性的计算模型, 试图使计算机像人一样, 对单幅二维图像中的各物体能实现相对大小恒常性感知. 实验结果表明此模型的效果确实如此.

该模型的输入是单幅二维直立图像; 输出是图像中的各物体在一维维度上和指定方向上 (一般是垂直或水平方向) 的相对感知大小; 中间过程是按式 (1) 进行的相对大小恒常性计算; 相机模型为针孔成像模型. 直立图像是指: 图像天空位于图像中间线的上面, 图像地面位于图像中间线的下面. 相对大小恒常性是指: 在客观世界中, 人与人之间、人与各正常物体之间及各正常物体之间的大小比例关系是基本恒定不变的, 如成人与儿童的相对高度、成人与房子的相对高度、餐桌与旁边椅子的相对高度等等. 这是人类视觉系统具有大小恒常性感知的客观物质基础, 但是在针孔成像过程中, 这种恒定的大小比例关系被破坏了. 所以必须自动恢复图像中各物体间的大小比例关系, 这个过程即本文要研究的大小恒常性计算.

由式 (1) 可知, 要实现对图像中各物体相对大小恒常性感知, 需要正确计算图像物体的成像视角  $A$  和相对感知深度  $D$ . 成像视角  $A$  可用物体在图像中的一维大小表示, 即可用它在图像中沿某一方向覆盖的像素点数量表示. 对给定轮廓的图像物体, 计算机能轻易完成这项计算任务. 本文的目的仅是论证大小恒常性计算模型的正确性与有效性, 故假定图像物体的轮廓都是人工给定的, 其余的工作都是由计算机自动完成.

估计相对感知深度  $D$  一直是计算机视觉中的难题. 本文从上述心理学的结论出发, 提出了一种简单、新颖、有效的求解方法, 其计算原理见图 4. 首先, 利用物体在图像中的高度与大气透视两种深度线索, 用天空检测技术计算出中间线  $L_1$ , 把图像地面部分从整幅图像中分离出来. 其次, 在图像地面部分, 利用线性透视与纹理透视两种深度线索, 可算出从图像底端边线到中间线的深度变化最快方向直线  $L_2$ .  $L_2$  与  $L_1$  的交点  $V(V_x, V_y)$  为图像中的感知深度最大的点, 即灭点.  $L_2$  与图像地面底端边线的交点  $U(U_x, U_y)$ , 为图像中的感知深度最小的点, 我们称之为近点. 根据针孔成像模型, 在一定的范围内, 图像深度是线性

变化的.故从近点  $U$  向图像中间线, 沿着深度变化最快方向直线  $L_2$ , 图像深度值线性递增, 直至灭点  $V$  达到最大. 最后, 计算图像地面相对感知深度图. 与  $L_2$  相垂直的直线上的所有点具有相同的深度. 如直线  $L_3$  过点  $P(m, n)$  且与  $L_2$  垂直, 则  $L_3$  上的所有点的感知深度与点  $P$  相同. 故点  $P$  的相对感知深度可用近点  $U$  到  $L_3$  的距离  $D_{U-L_3}$  表示. 这样就可以自动计算出图像地面各点的相对感知深度, 进而形成稠密相对感知深度图.

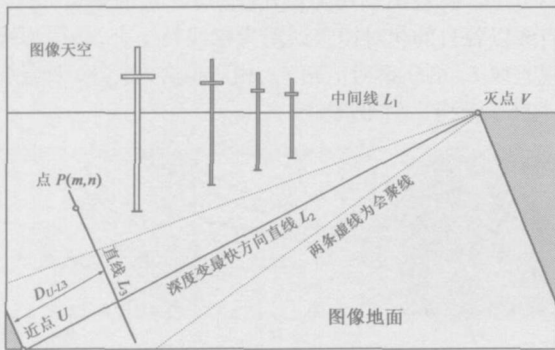


图 4 相对感知深度  $D$  计算原理示意图. 共分三步: (1) 计算中间线  $L_1$ , 将图像地面部分从整幅图像中分离出来; (2) 用线性透视与纹理梯度线索计算地面深度变化最快方向直线  $L_2$ ; (3) 计算图像地面上任一点  $P(m, n)$  处的感知深度, 形成稠密相对感知深度图.

得到了各物体的成像视角  $A$  和相对感知深度  $D$ , 应用式 (1), 计算机就能实现图像物体相对大小恒常性计算. 因为我们只计算相对感知大小, 故可将式 (1) 的中  $B$  值设为 1. 计算时, 用物体中点处的感知深度代表整个物体的感知深度. 其原因将在本文实验部分中讨论.

#### 4 感知深度算法

如上所述, 大小恒常性计算的关键是计算各图像物体的感知深度. 因为我们仅想实现相对大小恒常性, 故仅需要计算各物体间的相对感知深度, 即物体间感知深度的比例关系. 人类对位于地面上的物体能较好地实现大小恒常性感知, 而对处于天空中物体的感知能力较差 (如月亮错觉). 这是因为相对于天空而言, 地面的深度线索比较多, 比较有效, 所以本文也仅计算图像地面各点的相对深度.

已有很多方法可用来估计图像深度, 按成像设备与输入图像数量的不同可分成三类: 一是多相机多图像, 即基于立体成像的双目匹配 (Stereo based binocular matching), 精度较高, 但仅能产生稀疏的深度图<sup>[8]</sup>; 二是单相机多图像, 即基于变焦的单目图像 (Defocus based single image), 精度差, 但能产生稠密深度图<sup>[9, 10]</sup>; 三是单相机单图像, 精度差, 但适用范围广<sup>[11, 12]</sup>. 这些方法的共同缺点是: 一仅从相机成像模型及其相应的数学变换关系出发来计算图像深度, 数学运算复杂, 而且需要先验知识约束 (如

相机参数、对应点等); 二很少有效利用心理学已经给出的各种图像深度线索, 从而忽视了人的主观感受. 本文从心理学结论出发, 提出了一种简单、新颖的单相机单图像深度计算方法, 它无需先验知识约束, 可产生较稠密的相对深度图. 大小恒常性的实验结果表明这种方法是有效的. 感知深度算法由以下三步组成.

##### 4.1 计算中间线 $L_1$

我们使用天空检测技术来计算图像中间线, 把图像地面从整幅图像中分离出来. 目前天空分离算法主要有基于天空物理模型的方法<sup>[13]</sup>与二维多项式模型的方法<sup>[14]</sup>. 大气透视线索用来区分天空与类似天空颜色的表面, 如水域、墙面、玩具、衣服等. 这些方法计算复杂, 而且需要先进行分类训练. 我们的目的仅是想验证图像物体大小恒常性计算模型的可行性, 所以提出了一种简单的天空分离方法, 详述如下:

图像天空部分 (包括天花板) 的颜色一致性较好, 布局较简单. 利用这个特性, 使用图像分割技术能将天空分离出来. 因  $H$  分量对彩色描述能力相对来说与人的视觉最接近, 故先把 RGB 空间转换成 HSI 空间. 因待处理图像都是直立的, 故在图像的上半部分必定存在天空区域, 故仅对图像的上半部分进行统计, 计算  $I-D$  颜色直方图. 具有最大值的直方条 (Bin) 对应的  $H$  值就是天空的  $H$  值, 记作  $H_{sky}$ . 为了提高计算速度和避免地面部分的单点被误判为天空, 图像被分成  $2 \times 2$  的小块, 它的  $H$  值为 4 个像素的平均值. 设  $W$  为任一图像小块, 它的  $H$  值记为  $H_w$ , 如果  $|H_{sky} - H_w| \leq T_1 * H_{sky}$ , 则块  $W$  属于天空.  $T_1$  为相似性阈值, 实验取值为 0.05. 天空的计算在整幅图像上进行. 若计算出天空的面积小于图像的 5%, 我们便认为此图像中不包含天空. 图像每列中位于最下方的天空点形成天地分界线. 用最小二乘法把天地分界线拟合成水平直线, 这条水平直线就是中间线  $L_1$ .

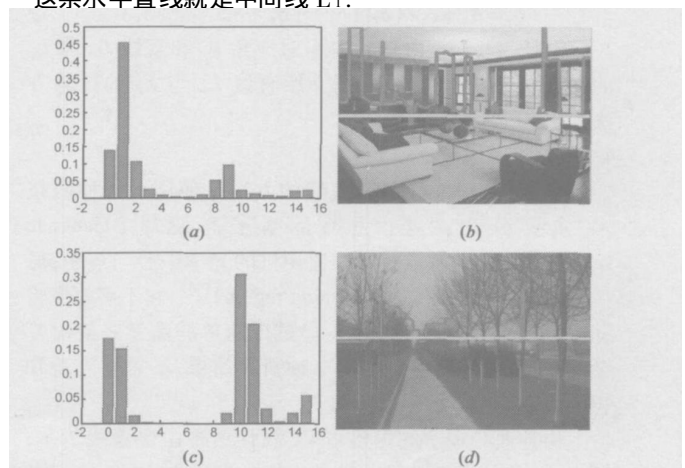


图 5 中间线计算的例子. (b) 与 (d) 中的折线是根据  $H$  分量直方图计算出的天地分界点的连线, 水平线是用天地分界点拟合的中间线. (a)、(c) 分别是 (b)、(d) 上半部分的  $H$  分量直方图.

当图像中不包含天空时,中间线一般退缩到图像的顶端边线,如图 1(d)所示;有时也会出现在图像的两侧边,如图 3(b)所示,在图像的左侧边.因为所有的图像都是直立的,中间线是出现在图像的底端边线.此时,中间线的位置由灭点位置与深度变化最快方向直线  $L_2$  决定.当图像中包含天空时,图像地面是由中间线、底端边线及两侧边所形成的区域;当图像中不包含天空时,图像地面是整幅图像.图 5 是中间线计算的例子,可以看出这种简单的方法基本能满足我们的要求.

#### 4.2 计算地面深度变化最快方向直线 $L_2$

由前面心理学的内容可知,线性透视与纹理梯度两种深度线索可用来指示地面深度变化最快方向.这两种线索仅在图像地面部分有效,所以计算直线  $L_2$  的图像支持范围仅是图像地面部分.单独利用线性透视线索,可算出一条从图像底端边线到中间线的深度变化最快方向直线.我们称此直线为线性透视直线  $L_P$ .单独利用纹理梯度线索,也可算出一条从图像底端边线到中间线的深度变化最快方向直线.我们称此直线为纹理梯度直线  $L_T$ .  $L_P$  与  $L_T$  的计算方法稍后介绍.现在假设这两条直线已经求出.一般情况下,这两条直线不会重合,所以在共同指示地面深度变化最快方向时,不可避免地会产生冲突.因这两条直线都是用最小二乘法拟合产生的,故可以认为,直线的相对拟合误差越大,它所指示的深度变化最快方向越不准确.一种冲突的解决办法是:两直线以各自的相对拟合误差为权,线性组合来求解地面深度变化最快方向直线  $L_2$ ,相对拟合误差越大,对应直线的组合权值越小,具体方法如下:

设深度变化最快方向直线  $L_2$ 、线性透视直线  $L_P$ 、纹理梯度直线  $L_T$  的相对拟合误差分别为  $\delta_2$ 、 $\delta_P$ 、 $\delta_T$ ,它们斜率对应的角分别为  $\theta_2$ 、 $\theta_P$ 、 $\theta_T$ ,所有  $\theta$  的取值范围为  $[-\pi/2, \pi/2]$ ,则有

$$\theta_2 = \theta_P * \delta_T / (\delta_P + \delta_T) + \theta_T * \delta_P / (\delta_P + \delta_T) \quad (2)$$

$$\delta_2 = \delta_P * \delta_T / (\delta_P + \delta_T) + \delta_T * \delta_P / (\delta_P + \delta_T) \quad (3)$$

于是,直线  $L_2$  由它的斜率对应角  $\theta_2$  和直线  $L_P$  与  $L_T$  的交点唯一确定.下面分别介绍直线  $L_P$  与  $L_T$  的计算方法.

##### 4.2.1 求解线性透视直线 $L_P$

目前,求解线性透视直线的方法是使用灭点检测技术,主要分为两类:一类是基于高斯球(Gaussian sphere)<sup>[15]</sup>,它的缺点是要知道相机的光学中心与焦距;另一类是基于灭点区域(Vanishing region)<sup>[16]</sup>,它不需要先验知识,但数学模型复杂.本文尝试用简单的数学来解决复杂的视觉问题,所以提出了一种新而简单、不需要先验知识的线性透视直线求解方法.

由前面心理学的内容可知,线性透视由会聚线指示,所以我们先提取直线.目前,提取直线方法主要有两种<sup>[17]</sup>:一是基于图像亮度梯度幅度(Gradient magnitude),一般使用 Hough 变换技术实现;另一种是基于图像亮度梯度方向(Gradient or

rientation),一般使用相位编组技术实现.这两种方法是互补的.因为第一种方法简单易行,所以我们在实验中使用它.因提取的直线越长,越能体现线性透视效果. Hough 变换技术使用投票机制,其原理就是直线越长,越先被提取.这是我们使用第一种方法的另一个原因.对每幅图像,先使用 Hough 变换技术找出最长的 10 条直线分别对应的图像点集,然后用最小二乘法把这些点集分别拟合直线,并得到每条直线的方程、斜率对应角  $\theta$  及相对拟合误差  $\delta$ . 利用类似式(2)、式(3)的思想,由这 10 条直线以各自的相对拟合误差为权线性组合,易得到线性透视直线  $L_P$  的斜率对应角  $\theta_P$ 、相对拟合误差  $\delta_P$  及直线方程.图 6 显示了一个例子.

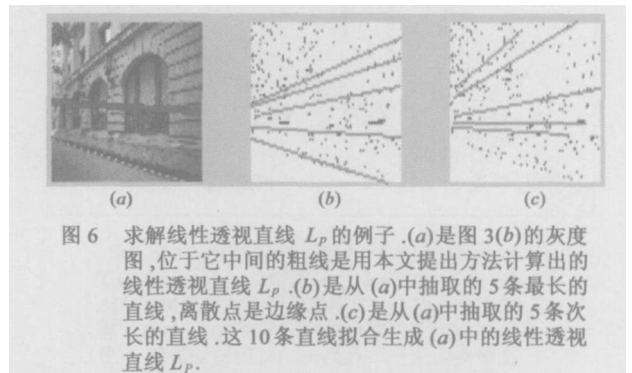


图 6 求解线性透视直线  $L_P$  的例子.(a)是图 3(b)的灰度图,位于它中间的粗线是用本文提出方法计算出的线性透视直线  $L_P$ . (b)是从(a)中抽取的 5 条最长的直线,离散点是边缘点.(c)是从(a)中抽取的 5 条最长的直线,这 10 条直线拟合生成(a)中的线性透视直线  $L_P$ .

##### 4.2.2 求解纹理梯度直线 $L_T$

纹理梯度已被广泛地应用于计算机视觉各领域中,如从纹理求形状<sup>[18]</sup>,基于纹理梯度图像分割<sup>[19]</sup>等等.在这些应用中,纹理梯度的表示与计算都比较复杂,故本文提出了一种简单的纹理梯度表示方法.

由前面心理学的内容可知:表面离观察者越远,纹理变得越小.其原因是:离视点越近,相同面积的视网膜(成像平面)区域包含的同质物体越少,即图像分辨率越大,纹理元素的尺寸越大.在物体内部区域,像素亮度的差异小,故物体一般被感知为同质区域.这也就意味着:从统计意义上讲,离视点越近,相同图像区域内的像素亮度差异之和应越小.为此,我们用各像素点的亮度差异度作为它的纹理梯度,并用它进一步求解纹理梯度直线  $L_T$ ,具体的计算过程如下:

(a) 设  $I(m, n)$  为图像地面任一像素点处的亮度  $I = (R + G + B) / 3$ ,按下式计算该点处的亮度差异度  $I_{diff}(m, n)$ .

$$I_{diff}(m, n) = \frac{\left[ \sum_{i=-Z_1}^{Z_1} \sum_{j=-Z_1}^{Z_1} |I(m, n) - I(m+i, n+j)| \right]}{(2 * Z_1 + 1)^2} \quad (4)$$

$Z_1$  决定各像素亮度差异的计算范围,实验表明取 1, 2, 3 中的某个值为宜.

(b) 将图像地面部分均匀地分成  $Z_2 * Z_2$  小块,设水平方向(行)与垂直方向(列)的块数分别为  $S$ 、 $T$ .每块的亮度差异度  $M_{diff}$  为块中所有点的像素亮度差异度  $I_{diff}$  之和,找

出每行(水平方向)中具有最小亮度差异度  $M_{diff}$  的块, 分别记作  $R_1, R_2, \dots, R_{T-1}, R_T$ . 从统计意义上讲, 块  $R_1, R_2, \dots, R_{T-1}, R_T$  代表各行中离视点最近的区域. 实验表明  $Z_2$  的值不宜过大, 取 5 左右为宜.

(c) 用最小二乘法对块  $R_1, R_2, \dots, R_{T-1}, R_T$  的中心点坐标进行拟合, 便可计算出纹理梯度直线  $L_T$  的斜率对应角  $\theta_T$ 、相对拟合误差  $\delta_r$  及直线方程. 例子见图 7.

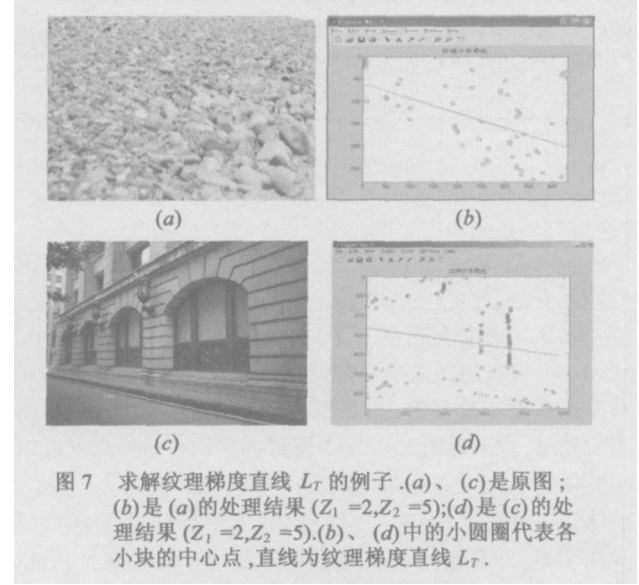


图 7 求解纹理梯度直线  $L_T$  的例子.(a)、(c)是原图;(b)是(a)的处理结果( $Z_1=2, Z_2=5$ );(d)是(c)的处理结果( $Z_1=2, Z_2=5$ ).(b)、(d)中的小圆圈代表各小块的中心点, 直线为纹理梯度直线  $L_T$ .

### 4.3 计算地面感知深度图

由图 4 可知, 近点  $U$  的感知深度最小, 设为  $D_U$ , 它的值等于相机离客观世界场景最近成像点的距离除以相机缩放系数  $B$ . 由本文第三节的内容可知, 在图像地面部分各点感知深度的变化规律是: 从近点  $U$  到图像中间线, 沿着深度变化最快方向直线  $L_2$ , 图像深度值线性递增, 直至灭点  $V$  达到最大; 与深度变化最快方向直线  $L_2$  相垂直的直线上的所有点具有相同的深度(等深线). 设  $P(m, n)$  为图像地面部分任一坐标为  $m, n$  的像素点, 求解点  $P(m, n)$  处的相对感知深度  $D_P$  的方法如下:

设深度变化最快方向直线  $L_2$  的斜率为  $K_2$ , 直线  $L_3$  过点  $P(m, n)$  且垂直于直线  $L_2$ , 故直线  $L_3$  的斜率  $K_3 = -1/K_2$ , 则直线  $L_3$  的方程为:

$$X + K_2 Y - mK_2 - n = 0 \quad (5)$$

表 1 选用图 1(a) 中的 6 根杆子进行计算大小恒常性实验, 从左到右分别记为杆 1~杆 6. 图像中的高度为各杆沿垂直方向在图像中覆盖像素点的数量除以 10. 脚点感知深度是用式(7)计算出的杆子脚点(杆在图像中的最低点)处感知深度. 脚点感知大小= 图像中的高度\* 脚点感知深度, 它是以脚点处的感知深度代替整个杆子的感知深度时, 利用式(1)计算出的杆子的感知高度. 中点感知深度与中点感知大小亦同理计算. 相对分离程度=  $|X_i - \text{平均}(X)| / \text{平均}(X)$ , 例如杆 1 第 3 列相对分离程度=  $|5.70 - 4.08| / 4.08 = 0.40$ , 其值越小, 表示所有杆的平均高度差异越小.

序号	图像中的高度	相对分离程度	脚点感知深度	脚点感知大小	相对分离程度	中点感知深度	中点感知大小	相对分离程度
杆 1	5.70	0.40	5.60	31.92	0.15	7.10	40.47	0.03
杆 2	5.00	0.23	7.80	39.00	0.04	8.90	44.50	0.07
杆 3	4.80	0.18	9.70	46.56	0.24	10.90	52.32	0.26
杆 4	4.10	0.00	11.00	45.10	0.20	11.80	48.38	0.16
杆 5	2.80	0.31	12.50	35.00	0.07	12.90	36.12	0.13
杆 6	2.10	0.49	13.00	27.30	0.27	13.40	28.14	0.32
平均	4.08	0.27	9.93	37.48	0.16	10.83	41.66	0.16

设近点  $U$  到直线  $L_3$  的距离为  $D_{U-L_3}$ , 则有:

$$D_{U-L_3} = |U_x + K_2 U_y - mK_2 - n| / (1 + K_2^2)^{1/2} \quad (6)$$

故点  $P(m, n)$  处的感知深度  $D_P$  为:

$$D_P = D_U + D_{U-L_3} \quad (7)$$

一般情况, 近点的感知深度  $D_U$  难以估计, 考虑它与  $D_{U-L_3}$  相比小得多, 故在后面的实验中不予考虑, 被设为 0. 本文在实验部分将讨论  $D_U$  被设为 0 对计算结果的影响.

按照这种方法, 依次计算图像地面各点的相对感知深度, 便可形成稠密相对深度图. 但是本方法有一个缺陷, 图 4 中两个三角形阴影区域处于所有的等深线之外, 位于其中像素点的感知深度无法计算, 故无法产生完整的图像感知深度图. 根据我们对多幅图像的观察, 在这两个区域中经常存在着冲突的深度线索, 人眼的感觉也经常出错. 图像物体大小恒常性现象很少发生在这些区域, 所以这个缺陷不会妨碍本文的研究工作. 当地面深度变化最快方向直线恰好与图像的底端边线垂直时, 则不存在这两个三角形区域.

### 5 实验结果与分析

本文提出的方法在 MATLAB 的环境下实现, 微机基本配置为: CPU P4-2 666G, 内存 128M\* 2. 本文所列的试验最长运行时间为 2.5min, 主要用于 Hough 变换. 目前, 我们没有发现图像物体大小恒常性的计算模型与算法, 所以没有进行算法比较, 仅是用下面的例子来验证了我们所提出方法的正确性与有效性, 并对算法的关键问题进行了分析. 为了便于验证文中提出模型的正确性, 各组试验中选用的图像物体都具有同一属性: 它们在图像中的大小不同, 但感知大小却相同.

表 1 是对图 1(a) 中的 6 根杆子大小恒常性的处理结果. 表 2 是对图 1(d) 中的 3 个人的大小恒常性的处理结果. 从表 1、表 2 可以看出, 无论以脚点还是以中点的感知深度代表整个物体感知深度, 经大小恒常性处理后, 各物体间的平均相对分离程度都显著减少. 物体间的平均相对分离程度越小, 说明对物体的这种测量结果越趋于相同. 如表 1 所示, 各杆子高度平均相对分离程度由图像的 27% 减少到 16%(脚点), 16%(中点); 如表 2 所示, 3 人的高度



表 2 选用图 1(d) 中的 3 个人, 分别是最上面穿白色上衣的男人、中间穿彩色上衣的男人、及下方背着黑包个子最高的女人, 并依次以人 1、人 2、人 3 标记。表中其余各项含义与表 1 中的相同。

序号	图像中的高度	相对分离程度	脚点感知深度	脚点感知大小	相对分离程度	中点感知深度	中点感知大小	相对分离程度
人 1	9.20	0.87	0.4	3.68	0.79	4.80	44.16	0.31
人 2	3.00	0.39	8.5	25.5	0.43	10.00	30.00	0.11
人 3	2.60	0.47	9.35	24.31	0.36	10.50	27.30	0.19
平均	4.93	0.58	6.08	17.83	0.53	8.43	33.82	0.20

平均相对分离程度由图像的 58% 减少到 53% (脚点), 20% (中点)。这些结果都说明本文提出的方法能有效实现类似人眼的图像物体大小恒常性感知。在表 1 中, 脚点与中点的大小恒常性调整的效果相同。但在表 2 中, 两者的效果差别很大: 脚点的平均相对分离程度为 53% 与图像的 58% 相差无几, 而中点的为 20%, 与图像的 58% 相差很大。故以中点代表整个物体进行大小恒常性计算, 效果更令人满意。所以在计算感知深度时, 一般以物体的中点代表整个物体, 其原因后面分析。

表 3、表 4 分别是对室外建筑物图 3(b) 与室内建筑物图 3(a) 的计算结果。表 3 的大小恒常性计算效果较好, 平均相对分离程度从 31% 减少到 8%, 而表 4 则只有 2% 的减少。这是因为表 3 对应的室外图像场景景深较大, 图像物体的失真也较大, 深度线索较多而且强烈, 故调节效果较好。而表 4 对应的室内场景正好相反, 故调整效果较差。

表 3 选用图 3(b) 中的 4 个门, 从右到左分别记为门 1、门 2、门 3、门 4, 其余各项含义与表 1 中的相同。

序号	图像中的高度	相对分离程度	中点感知深度	中点感知大小	相对分离程度
门 1	8.10	0.59	8.20	66.42	0.05
门 2	5.30	0.04	15.10	80.03	0.15
门 3	3.90	0.24	18.40	71.76	0.03
门 4	3.10	0.39	20.40	63.24	0.09
平均	5.10	0.31	15.53	70.36	0.08

表 4 选用图 3(a) 中的 3 个窗户, 从右到左分别记为窗 1、窗 2、窗 3, 其余各项含义与表 1 中的相同

序号	图像中的高度	相对分离程度	中点感知深度	中点感知大小	相对分离程度
窗 1	3.40	0.12	8.75	29.75	0.10
窗 2	3.00	0.01	9.00	27.00	0.08
窗 3	2.70	0.11	9.15	24.71	0.01
平均	3.03	0.08	8.97	27.15	0.06

表 5 选用图 1(b) 右侧 5 棵树, 从右到左分别记作树 1~树 5 树的高度以计算树干上涂了防腐剂的白色部分。其余各项含义与表 1 相同。

序号	图像中的高度	相对分离程度	中点感知深度	中点感知大小	相对分离程度
树 1	4.90	0.47	0.60	2.94	0.66
树 2	3.80	0.14	2.30	8.74	0.00
树 3	3.16	0.05	3.34	10.55	0.21
树 4	2.56	0.23	4.15	10.62	0.22
树 5	2.26	0.32	4.70	10.62	0.22
平均	3.34	0.24	3.02	8.70	0.26

表 6 为表 5 中不计算树 1 的结果, 表中其余各项的含义与表 1 相同

序号	图像中的高度	相对分离程度	中点感知深度	中点感知大小	相对分离程度
树 1					
树 2	3.80	0.29	2.30	8.74	0.14
树 3	3.16	0.07	3.34	10.55	0.04
树 4	2.56	0.13	4.15	10.62	0.05
树 5	2.26	0.23	4.70	10.62	0.05
平均	2.95	0.18	3.62	10.14	0.07

表 5、表 6 都是对图 1(b) 右侧的 5 棵树的处理结果。从表 5 可以看出经大小恒常性计算后, 这 5 棵树的平均相对分离程度不降反升, 由 24% 增大至 26%, 效果很差。而表 6 显示, 若不计树 1, 其余 4 棵树的平均相对分离程度由 18% 降到 7%, 效果很好。经仔细分析, 与其他树相比, 树 1 的感知深度很小, 由式 (7) 可推测, 人为忽略近点  $D_U$  的感知深度  $D_U$  可能造成树 1 的感知深度计算误差很大, 所以出现表 5 中反常的计算结果。

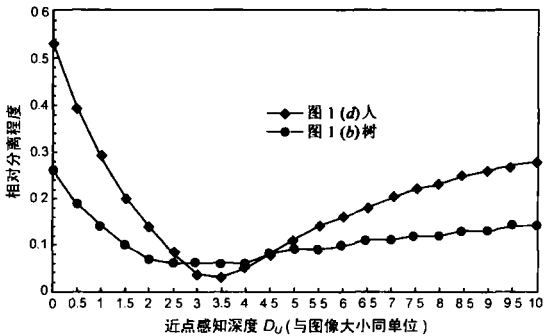


图 8 近点感知深度  $D_U$  不同取值对大小恒常性计算的影响

图 8 显示了近点感知深度  $D_U$  不同取值对大小恒常性计算的影响。圆点曲线表示  $D_U$  不同取值对图 1(b) 中右侧的 5 棵树大小恒常性计算的影响 (表 5 为  $D_U = 0$  的情况)。可以看出, 当  $D_U$  为 2.5 个单位时, 平均相对分离程度达到最小为 6%, 此时的大小恒常性计算效果最好。按式 (7), 此时树 1 的真实中点感知深度应为  $D_P = D_U + D_{U-L3} = 2.5 + 0.6 = 3.1$  (单位), 而表 5 中 ( $D_U = 0$ ) 的中点感知深度才 0.6 单位, 相差甚远。这是表 5 中大小恒常性计算结果反常的真正原因。方点曲线表示  $D_U$  不同取值对图 1(d) 中 3 个人的大小恒常性计算的影响 (以脚点代表整个物体, 表 2 为  $D_U = 0$  的情况)。可以看出, 当  $D_U$  为 3.5 个单位时, 平均相对分离程度由最初的 53% 减少到 3%。所以, 如果我们能正确估计图像的  $D_U$  值, 脚点也能够很好地代表整个物体。

心理学的研究结论也表明: 人眼在测定物体与我们的距离时, 一般以物体的脚为参考点。但是图像的  $D_U$  值一般很难计算, 当人为的把它设为 0 时, 物体脚点比中点受到的影响大, 这是因为图像物体的脚点比它的中点离近点  $U$  要近一些。故以中点代表整个物体也是对人为设  $D_U = 0$  的一种补偿。即使补偿过度, 比不补偿造成的误差要小。这可从图 8 中曲线的走势看出: 在到达最佳的平均相对分离程度之后, 曲线走势趋于平缓, 每单位  $D_U$  对相对分离程度的影响变得越来越小。

## 6 结论

本文从视觉心理学的结论出发, 提出了图像物体大小恒常性计算模型, 实验结果说明该模型是有效的: 对图像中大小不同的物体, 如果人眼感觉它们的大小是基本相同的, 本文提出的模型就能实现恒常性感知。视觉心理学表明: 人类视觉系统用简单的数学方法解决了许多复杂的问题, 所以本文也力图使用简单的数学(平面解析几何)来建立复杂的恒常性计算模型。在计算大小恒常性的同时, 本文也提出了一种简单而新颖的单幅图像相对深度计算方法。实验表明, 这种深度计算方法是有效的。

忽略近点  $U$  处的感知深度, 当成像场景景深很小时, 大小恒常性的计算结果不好, 可增加先验知识(如相机参数)或用机器学习的方法来估计近点  $U$  处的感知深度, 从而弥补本文的不足。本文只实现了相对大小恒常性感知, 而人类以自身高度或以某一熟悉物体尺寸为绝对深度参考点, 对周围物体具有绝对大小恒常性感知的能力。计算机要想获得这种能力, 必须在图像中寻找绝对深度参考点。这些是我们下一步的研究目标。

## 参考文献:

- [1] I Rock. Perception[M]. New York: Scientific American Books, Inc, 1984. 1- 80.
- [2] Wu Aimin, Xu De, et al. Cognition theory based performance characterization in computer vision[J]. Lecture Notes in Computer Science, 2005, 3708(1): 210- 221.
- [3] 郑南宁. 认知与计算机视觉[EB/OL]. <http://www.nsf.gov.cn/nsfc/cen/10yy/zy/004.htm>. 2004-10-19/2005-10-18.
- [4] Wu Aimin, Xu De, et al. Generic solution for image object recognition based on vision cognition theory[J]. Lecture Notes in Computer Science, 2005, 3614(2): 1265- 1275.
- [5] D Forsyth, J Ponce. Computer Vision: A Modern Approach (影印版)[M]. 北京: 清华大学出版社, 2004. 129- 130.
- [6] S E Palmer. Vision science[M]. London: Cambridge, Mass: MIT Press, 1999. 229- 236.

- [7] John H Krantz. Vision and art[EB/OL]. <http://psych.harvard.edu/Krantz/art/index.html>, 2003-11-04/2005-10-18.
- [8] 马颂德, 张正友. 计算机视觉- 计算理论与算法基础[M]. 北京: 科学出版社, 1998.
- [9] P Favaro, S Soatto. A geometric approach to shape from defocus[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2005, 27(3): 406- 417.
- [10] A N Rajagopalan, S Chaudhuri, et al. Depth estimation and image restoration using defocused stereo pairs[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2004, 26(11): 1521- 1525.
- [11] Lai Shang-Hong, Fu Chang-Wu, Chang Shyang. A generalized depth estimation algorithm with a single image[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1992, 14(4): 405- 411.
- [12] A Torralba, A Oliva. Depth estimation from image structure[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(9): 1226- 1238.
- [13] Jiebo Luo, S P Etz. A physical model-based approach to detecting sky in photographic images[J]. IEEE Transaction on Image Processing, 2002, 11(3): 201- 212.
- [14] A C Gallagher, J Luo, W Hao. Improved blue sky detection using polynomial model fit[A]. Meng-Hwa ER. Proc of ICIP-2004[C]. New York: IEEE Press, 2004. 2367- 2370.
- [15] J A Shufelt. Performance evaluation and analysis of vanishing point detection techniques[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1999, 21(3): 282- 288.
- [16] A Almansa, A Esolneux, S Vamech. Vanishing point detection without any a priori information[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2003, 25(4): 502- 507.
- [17] Jin Xiaoying, Curt H Davis. An integrated system for automatic road mapping from high-resolution multispectral satellite imagery by information fusion[J]. Information Fusion, 2005, 6(4): 257- 273.
- [18] M Clerc, S Mallat. Texture gradient equation for recovering shape from texture[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(4): 536- 549.
- [19] P R Hill, C N Canagarajah, D R Bull. Image segmentation using a texture gradient based watershed transform[J]. IEEE Transactions on Image Processing, 2003, 12(12): 1618- 1633.

## 作者简介:

吴爱民 男, 1972 年生于江西上饶, 博士研究生, 高级工程师, 主要研究方向为计算机视觉及心理学在视觉信息处理中的应用。

E-mail: wuaimin@sohu.com