

基于遮挡区域建模和目标运动估计的动态遮挡规避方法

张世辉^{1,2} 何琦¹ 董利健¹ 杜雪哲¹

摘要 对于运动视觉目标, 如何对遮挡区域进行规避是视觉领域一个具有挑战性的问题. 本文提出了一种新颖的基于运动视觉目标深度图像利用遮挡信息实现动态遮挡规避的方法. 该方法主要利用遮挡区域最佳观测方位模型和视觉目标运动估计方程, 通过合理规划摄像机的观测方位逐渐完成对遮挡区域的观测. 主要贡献在于: 1) 提出了深度图像遮挡边界中关键点的概念, 利用其构建关键线段对遮挡区域进行快速建模; 2) 基于关键线段和遮挡区域建模结果, 提出了一种构建遮挡区域最佳观测方位模型的方法; 3) 提出一种混合曲率特征, 通过计算深度图像对应的混合曲率矩阵, 增加了图像匹配过程中提取特征点的数量, 有利于准确估计视觉目标的运动. 实验结果验证了所提方法的可行性和有效性.

关键词 深度图像, 遮挡信息, 遮挡区域建模, 下一最佳观测方位, 运动估计

引用格式 张世辉, 何琦, 董利健, 杜雪哲, 基于遮挡区域建模和目标运动估计的动态遮挡规避方法. 自动化学报, 2019, 45(4): 771–786

DOI 10.16383/j.aas.2017.c170221

Dynamic Occlusion Avoidance Approach by Means of Occlusion Region Model and Object Motion Estimation

ZHANG Shi-Hui^{1,2} HE Qi¹ DONG Li-Jian¹ DU Xue-Zhe¹

Abstract How to avoid the occlusion region of a moving visual object is a challenging problem in the visual field. Based on the occlusion information in the depth image of a moving visual object, this paper proposes a novel dynamic occlusion avoidance approach which plans the camera position by utilizing the best view model of occlusion area and the visual target motion estimation equation. This work has three contributions. The first one is the concept of key point, which constitutes the key line segment to construct the model of occlusion region. The second one is the approach for constructing the best view model of occlusion region based on the key line segment and the occlusion region model. The third one is the feature of mixed curvature. The number of feature points extracted in the process of image matching is increased by calculating the mixed curvature matrix corresponding to the depth image, which is conducive to estimating motion of visual object accurately. Experimental results demonstrate the feasibility and effectiveness of the proposed approach.

Key words Depth image, occlusion information, occlusion region modeling, next best view, motion estimation

Citation Zhang Shi-Hui, He Qi, Dong Li-Jian, Du Xue-Zhe. Dynamic occlusion avoidance approach by means of occlusion region model and object motion estimation. *Acta Automatica Sinica*, 2019, 45(4): 771–786

遮挡规避是目标跟踪、三维重建、物体识别、场景绘制、运动估计等视觉领域一个重要且困难的问题, 受到了学者们的广泛关注^[1–3]. 遮挡规避是指当视觉目标在某一时刻发生遮挡时, 基于检测出的遮

挡信息及摄像机当前观测方位信息, 通过改变摄像机的观测位置和方向等参数对遮挡区域进行观测以获取遮挡区域信息的过程.

根据视觉目标的状态是静止的还是运动的, 可将遮挡规避分为静态视觉目标遮挡规避和动态视觉目标遮挡规避两大类. 目前, 国内外学者在静态遮挡规避问题的研究方面已经取得了一定的成果. Maver 等^[4] 通过对遮挡区域外接表面进行多边形建模, 获取摄像机的下一最佳观测方位. Li 等^[5] 提出了一种通过 B 样条构建物体模型计算最大信息熵的方法, 将摄像机能够获取模型最大信息熵的方位作为下一最佳观测方位. Banta 等^[6] 提出一种组合方法来确定下一最佳观测方位. Wu 等^[7] 利用基于密度的分层轮廓拟合算法来确定下一最佳观测方位. Pito^[8] 利用 PS (Positional space) 算法从大量的观测方位

收稿日期 2017-05-02 录用日期 2017-11-22
Manuscript received May 2, 2017; accepted November 22, 2017
国家自然科学基金 (61379065), 河北省自然科学基金 (F201420319) 资助
Supported by National Natural Science Foundation of China (61379065) and Natural Science Foundation of Hebei Province (F2014203119)
本文责任编辑 王亮
Recommended by Associate Editor WANG Liang
1. 燕山大学信息科学与工程学院 秦皇岛 066004 2. 河北省计算机虚拟技术与系统集成重点实验室 秦皇岛 066004
1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004 2. Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004

中确定下一最佳观测方位. Potthast 等^[9]提出了一种基于遮挡环境下信息熵的不同确定下一最佳观测方位的方法. 张世辉等^[10]利用投影降维思想通过对遮挡区域外接表面构建小平面集合从而确定下一最佳观测方位. Doumanoglou 等^[11]利用稀疏自编码器对深度不变 RGB-D 块提取特征训练霍夫森林, 并基于训练好的霍夫森林通过计算最小化期望熵来规划下一最佳观测方位. Bissmarck 等^[12]提出面向边界和体积的分层射线追踪算法, 利用边界信息和空间信息去除冗余射线确定下一最佳观测方位. 同静态遮挡规避相比, 动态遮挡规避问题的研究相对较少, 主要是因为遮挡区域会随着视觉目标运动而发生改变, 如何合理规划摄像机观测方位对视觉目标遮挡区域进行观测是动态遮挡规避过程中的难点. 为了解决动态遮挡规避问题, 需要对视觉目标进行运动估计以确定运动后遮挡区域变化情况, 进而移动摄像机到遮挡区域最佳观测方位下对遮挡区域进行观测. 因此, 动态遮挡规避的最终目标是获得摄像机一系列的下一最佳观测方位. 目前可查到的动态遮挡规避方法是 Zhang 等^[13]通过对遮挡区域建模以及计算视觉目标运动方程, 从而规划摄像机的运动方式, 对遮挡区域进行动态规避, 但该方法存在建模速度较慢、运动估计误差较大等不足.

鉴于目前动态遮挡规避研究较少且已有方法存在的问题, 本文提出一种基于遮挡区域建模和目标运动估计的动态遮挡规避方法. 该方法主要目的是观测运动视觉目标的遮挡区域, 因为在诸多研究和应用如运动目标三维重建、主动跟踪、动态目标识别、机器人自主作业等领域, 视觉目标发生遮挡现象是一种普遍存在的情况, 此时遮挡区域可能包含更多的有用信息. 如果视觉系统能最大限度地观测到遮挡区域, 将会有助于视觉任务的顺利完成. 选取深度图像是考虑到深度图像更利于获取场景的三维信

息, 以便更加方便地确定摄像机的观测方位. 由于动态遮挡规避过程需要兼顾实时性和准确性, 因此在方法设计过程中, 主要解决了两个问题: 1) 如何快速构建动态视觉目标遮挡区域的模型并确定遮挡区域最佳观测方位; 2) 如何准确地对深度图像中的视觉目标进行运动估计. 针对第一个问题, 提出了关键点概念, 通过提取深度图像遮挡边界中的关键点, 构建关键线段实现对遮挡区域快速而准确的建模, 在此基础上, 利用建模结果和关键线段构建遮挡区域最佳观测方位模型, 以确定遮挡区域最佳观测方位. 针对第二个问题, 为了提高视觉目标运动估计的准确性, 提出了混合曲率的概念, 通过计算视觉目标深度图像对应的混合曲率矩阵, 增加了图像匹配过程中提取特征点的数量, 并依据筛选后的匹配点对视觉目标进行运动估计. 实验结果验证了本文所提方法的可行性和有效性.

本文第 1 节概述所提方法的基本思想; 第 2 节论述如何依据视觉目标深度图像遮挡信息实现动态遮挡规避过程; 第 3 节给出实验结果及分析; 第 4 节总结全文.

1 方法概述

1.1 动态遮挡规避问题分析

动态遮挡规避是指运动视觉目标在某一时刻发生遮挡时, 以该时刻视觉目标遮挡区域为规避对象, 同时对视觉目标进行运动估计, 通过合理规划摄像机的运动方式, 多次移动摄像机以观测到更多的遮挡区域. 图 1 展示了动态遮挡规避过程. 图 1(a) 为动态视觉目标在某一时刻观测方位下发生遮挡现象, 阴影区域即为发生遮挡区域, 为了获取到遮挡区域信息, 需要规划摄像机的运动方式, 使其能够跟随视觉目标并向着遮挡区域最佳观测方位运动. 图 1(b)

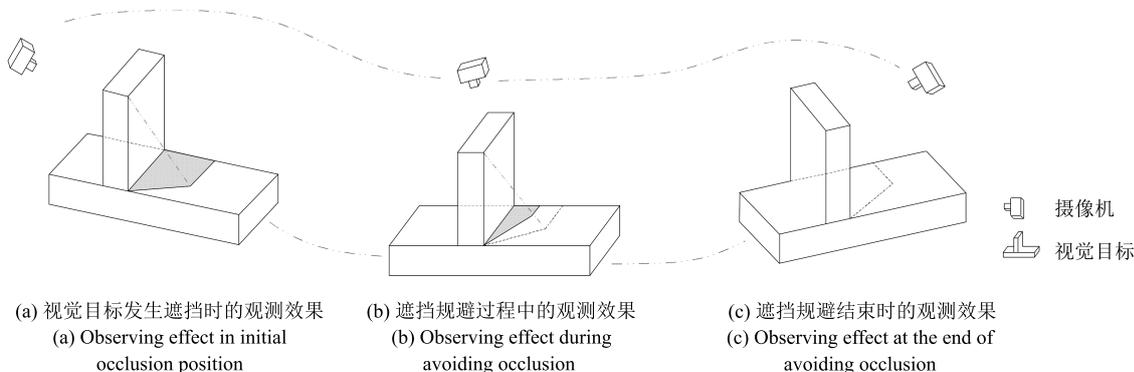


图 1 动态遮挡规避过程示意图

Fig. 1 Sketch map of dynamic occlusion avoidance

为规划摄像机跟随视觉目标运动并观测遮挡区域的中间过程, 随着摄像机的运动, 更多的遮挡区域被观测到, 遮挡区域面积也随之减少. 图 1(c) 为摄像机到达遮挡区域最佳观测方位时的观测效果, 此时遮挡区域被全部观测到, 摄像机完成了动态遮挡规避的全过程.

1.2 动态遮挡规避总体思想

基于上述对动态遮挡规避问题的分析, 本文提出一种解决动态遮挡规避问题的方法. 1) 在摄像机当前观测方位下采集视觉目标运动前后两幅深度图像, 将第一幅深度图像中的遮挡区域作为要规避的区域, 同时提取遮挡边界关键点以及与其相对应的下邻接点构建关键线段, 并利用关键线段对遮挡区域建模, 在此基础上, 利用遮挡区域建模结果和关键线段构建遮挡区域最佳观测方位模型. 2) 对两幅深度图像分别利用反投影变换计算图像中各像素点对应的三维坐标, 并利用三维坐标计算两幅深度图像的混合曲率矩阵. 3) 对两幅深度图像对应的混合曲率矩阵提取特征点进行匹配, 并对匹配特征点进行筛选, 根据筛选后匹配点的三维信息对视觉目标进行运动估计. 4) 利用遮挡区域最佳观测方位模型和视觉目标运动估计方程, 合理规划摄像机的下一最佳观测方位. 同时, 判断摄像机在相邻两次观测方位下获取视觉目标遮挡区域面积的差值是否小于给定阈值. 若不小于给定阈值则移动摄像机到下一最佳观测方位下继续采集视觉目标深度图像, 并利用遮挡区域最佳观测方位模型和新的运动估计方程重新规划摄像机下一最佳观测方位, 否则动态遮挡规避过程结束. 图 2 展示了动态遮挡规避过程的总体流程.

2 动态遮挡规避

2.1 构建遮挡区域最佳观测方位模型

2.1.1 提取遮挡边界关键点构建关键线段对遮挡区域建模

采集视觉目标发生遮挡时的第 1 幅深度图像,

利用文献 [14–15] 中的方法检测深度图像中的遮挡边界以及遮挡边界的端点. 为了对遮挡区域进行快速建模, 通过分析深度图像中遮挡边界点之间的位置关系, 本文提出了关键点概念. 其定义主要包括两部分: 1) 定义遮挡边界上的端点为关键点; 2) 如果任一遮挡边界点不为端点并且其八邻域内不存在其他关键点, 则将该遮挡边界点与八邻域内全部遮挡边界点组成线段, 若任意相邻线段所对应的斜率不同, 那么定义该遮挡边界点为关键点.

图 3 展示了视觉目标 Bunny 在当前观测方位下的遮挡边界和遮挡边界的关键点. 其中灰色点代表遮挡边界点, 黑色点代表遮挡边界的关键点.

获取完遮挡边界关键点之后, 即可对遮挡区域进行建模. 图 4 展示了构建关键线段并利用关键线段对遮挡区域建模的过程. 本文采用如下方式对遮挡区域进行建模: 依次选取遮挡边界上的关键点 m_i , 通过计算其八邻域内深度差值最大的点作为其下邻接点并记为 m'_i , 如图 4(a) 所示, 其中黑色点代表遮挡边界上的关键点, 白色点代表与关键点相对应的下邻接点. 然后在三维空间中依次将遮挡边界关键点 m_i 和与其相对应的下邻接点 m'_i 相连构成关键线段 l_i , 并将相邻关键线段之间所围成的空间四边形 $m_i m_{i+1} m'_{i+1} m'_i$ 记为 patch_i , 如图 4(b), 其中 i 为 1 到 $N-1$ 的整数, N 为关键线段个数. 所有 $N-1$ 个 patch 即为对遮挡区域的建模结果.

2.1.2 计算 patch 的面积、法向量和关键线段观测方向、中点

为了构建遮挡区域最佳观测方位模型, 需要计算 patch_i 的面积和法向量以确定关键线段的观测方向. 利用四边形面积公式计算 patch_i 面积, 计算公式为

$$S_i = \frac{1}{2} |m_i m'_{i+1}| |m'_i m_{i+1}| \sin \alpha$$

$$\text{s.t. } 1 \leq i \leq N-1 \quad (1)$$

其中, α 为 patch_i 中对角线 $m_i m'_{i+1}$ 和 $m'_i m_{i+1}$ 构成的夹角 (图 4(b)), S_i 为 patch_i 的面积.

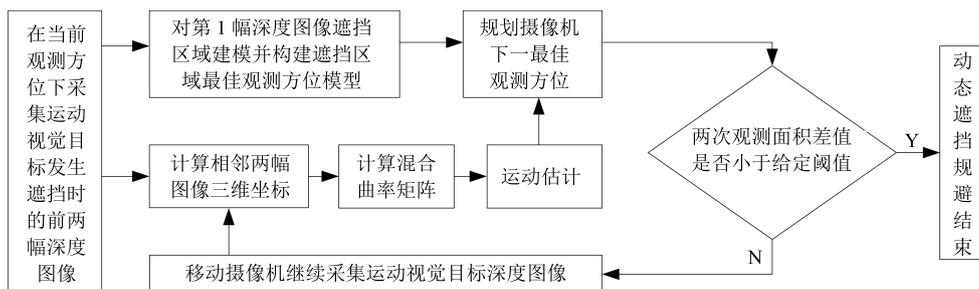


图 2 动态遮挡规避过程的总体流程

Fig. 2 The overall flow of the proposed dynamic occlusion avoidance

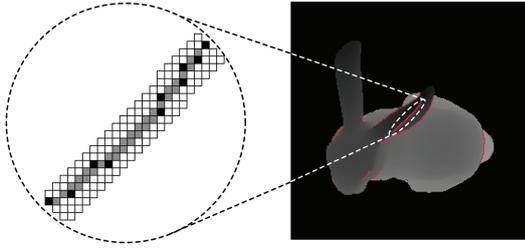


图3 Bunny的遮挡边界及其关键点

Fig. 3 The occlusion boundaries and key points of Bunny

由于 patch_i 的法向量可能指向视觉目标内部或外部, 为了保证计算出的 patch_i 实际可观测面积的有效性, 选取指向视觉目标外部的法向量作为 patch_i 的法向量. 在不考虑法向量方向的前提下, patch_i 法向量的 $\mathbf{V}_{\text{patch}_i}$ 计算公式为

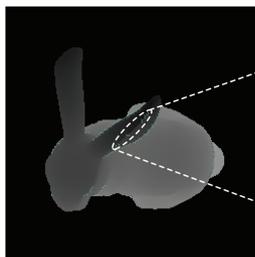
$$\mathbf{V}_{\text{patch}_i} = \boldsymbol{\mu}_i \times \boldsymbol{\gamma}_i \quad (2)$$

其中, $\boldsymbol{\mu}_i$ 为三维空间中从 m_i 指向 m'_{i+1} 的向量, $\boldsymbol{\gamma}_i$ 为三维空间中从 m'_i 指向 m_{i+1} 的向量. 为了保证 patch_i 的法向量方向指向视觉目标外部, 通过对关键点和下邻接点分析可知, 当关键点位于下邻接点左(上)侧, 则 patch_i 的法向量计算公式应为 $\boldsymbol{\mu}_i \times \boldsymbol{\gamma}_i$; 当关键点位于下邻接点右(下)侧, 则 patch_i 的法向量计算公式应为 $\boldsymbol{\gamma}_i \times \boldsymbol{\mu}_i$.

确定出 patch_i 的法向量之后, 即可在此基础上进一步计算关键线段观测方向. 考虑到组成关键线段的点存在端点的情况, 本文分两种情况对关键线段观测方向进行定义: 1) 若与关键线段 l_i 对应的点 m_i 为端点, 则定义 l_i 与其相邻关键线段围成 patch_i 的法向量为该关键线段的观测方向; 2) 若与关键线段 l_i 对应的点 m_i 不为端点, 则定义关键线段观测方向为

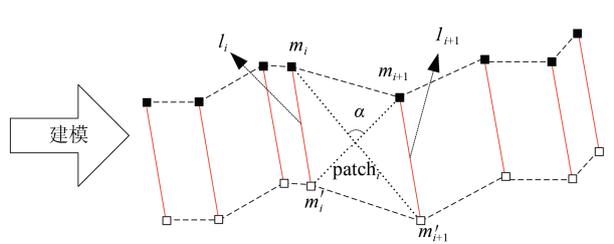
$$\mathbf{V}_{l_i} = \frac{w_i}{w} \mathbf{V}_{\text{patch}_i} + \frac{w_{i+1}}{w} \mathbf{V}_{\text{patch}_{i+1}} \quad (3)$$

s.t. $2 \leq i \leq N - E$



(a) 关键点及其对应的下邻接点

(a) The key points and their corresponding nether adjacent points



(b) 遮挡区域模型

(b) The model of occlusion region

图4 遮挡区域建模示意图

Fig. 4 The sketch map of modeling occlusion region

其中, \mathbf{V}_{l_i} 为关键线段 l_i 的观测方向, w_i 为 patch_i 的面积值, $\mathbf{V}_{\text{patch}_i}$ 为 patch_i 的法向量, w_{i+1} 为 patch_{i+1} 的面积值, $\mathbf{V}_{\text{patch}_{i+1}}$ 为 patch_{i+1} 的法向量, w 为 w_i 与 w_{i+1} 的和, E 为端点个数. 同时, 为了尽可能多地观测到关键线段, 定义关键线段的中点 c_i 作为观测点, 计算公式为

$$c_i = \frac{1}{2}(m_i + m'_i) \quad (4)$$

2.1.3 构建遮挡区域最佳观测方位模型

依据上述方法完成对遮挡区域建模和获取关键线段观测方向后, 下一步需要构造遮挡区域最佳观测方位模型. 由于摄像机对遮挡区域内未知表面的观测可近似等价于对相邻关键线段之间围成四边形区域的观测, 因此能够最大量观测到相邻关键线段之间围成的四边形面积信息的观测方位即为遮挡区域最佳观测方位. 基于此思想, 本文定义遮挡区域最佳观测方位模型为

$$(x_p, y_p, z_p) = \arg \max_{(x_p, y_p, z_p)} \sum_{i=1}^N \frac{\sin(\theta_i) S_i}{1 + \exp(-\beta \cos(\varphi_i))} \quad (5)$$

其中, (x_p, y_p, z_p) 为摄像机的观测位置. β 为常数, β 值越大函数越精确. φ_i 为将关键线段 l_i 的端点 m'_i 平移到当前摄像机观测中心点时, 关键线段 l_i 的观测方向 \mathbf{V}_{l_i} 与其中点 c_i 到摄像机当前观测位置 (x_p, y_p, z_p) 构成向量的夹角, θ_i 为将关键线段 l_i 的端点 m'_i 平移到当前摄像机观测中心点时, 摄像机观测方向与关键线段 l_i 所成的夹角, 如图 5 所示. $1/(1 + \exp(-\beta \cos(\varphi_i)))$ 是一个激活函数, 当 φ_i 夹角在 $0^\circ \sim 90^\circ$ 时, $1/(1 + \exp(-\beta \cos(\varphi_i)))$ 约等于 1, 即关键线段能被摄像机观测到才参与计算; 当 φ_i 在 $90^\circ \sim 180^\circ$ 时, $1/(1 + \exp(-\beta \cos(\varphi_i)))$ 约等于 0, 表示关键线段不能被摄像机观测到, 此时关键线段不参与计算. $\sin(\theta_i)$ 为第 i 条关键线段的投影长度

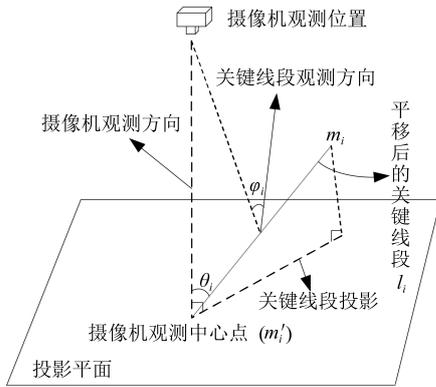


图5 当前摄像机观测中心点与关键线段的位置关系

Fig.5 The position relation between the observation center point of current camera and the key line segments

与关键线段长度的比值. S_i 为关键线段 l_i 与其相邻关键线段围成 patch 的面积, 当关键线段 l_i 对应的点 m_i 为端点时, S_i 为 l_i 与其相邻的关键线段围成一个 patch 的面积; 当关键线段 l_i 对应的点 m_i 不为端点时, S_i 为 l_i 与其相邻的关键线段围成两个 patch 的面积之和.

2.2 视觉目标运动估计

2.2.1 获取深度图像混合曲率矩阵

鉴于视觉目标在运动过程中其物体表面的曲率信息并不会发生改变. 因此, 对两幅深度图像分别利用反投影变换计算图像中各像素点对应的三维坐标, 并利用三维坐标计算两幅深度图像的混合曲率矩阵. 在此基础上, 通过对视觉目标运动前后两幅深度图像对应的混合曲率矩阵提取特征点进行匹配, 依据匹配信息计算视觉目标运动方程. 混合曲率计算公式为

$$\text{pixel}_i = \sqrt{\text{MC}_i^2 + k\text{GC}_i} \quad (6)$$

其中, pixel_i 表示深度图像中第 i 个像素点的混合曲率, 本文采用 Besl 等^[16] 提出的方法计算第 i 个像素点的平均曲率 MC_i 和高斯曲率 GC_i , k 为常数, k 值越大, 图像表面信息越丰富. 图 6 展示了视觉目标 Bunny 的深度图像和与其深度图像对应的混合曲率矩阵图像.

2.2.2 采用 SURF 算法匹配图像计算运动方程

依次获取视觉目标运动前后两幅深度图像对应的混合曲率矩阵之后, 下一步工作是对提取的混合曲率矩阵进行匹配, 考虑到动态遮挡规避过程的实时性与准确性要求, 本文采用 SURF 算法^[17] 对两幅深度图像对应的混合曲率矩阵进行匹配. SURF 算法优点是采用 Hessian 矩阵以及积分图像的概念, 加快了算法的运行时间与效率, 并且对旋转、仿射变

换、噪声都有一定的鲁棒性. 具体匹配过程如下: 1) 对视觉目标运动前后两幅深度图像对应的混合曲率矩阵提取特征点. 2) 计算与特征点对应的特征描述符. 3) 对两幅深度图像对应混合曲率矩阵提取出的特征点进行匹配. 由于视觉目标运动过程中产生的误差可能导致算法得到的匹配点不准确并且部分特征点在视觉目标外部, 因此采用三角形法则^[18] 对匹配结果进行筛选并去除视觉目标外部的特征点. 图 7 展示了视觉目标 Bunny 运动前后深度图像所对应混合曲率矩阵提取出的特征点. 图 8 展示了筛选之后的匹配效果.

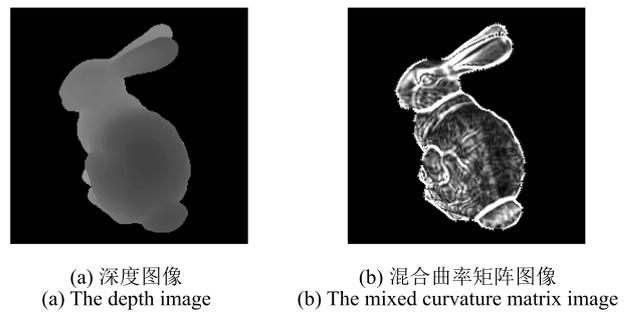


图6 Bunny 的深度图像及其对应的混合曲率矩阵图像
Fig.6 The depth image of Bunny and its corresponding mixed curvature matrix image

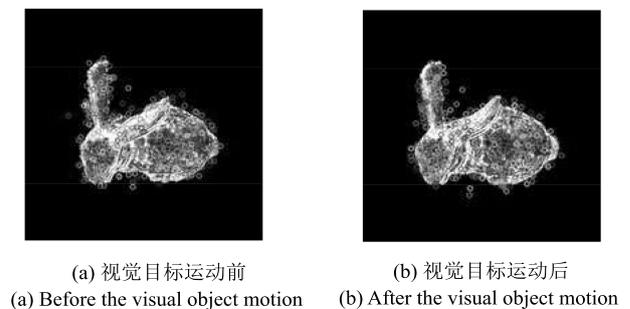


图7 视觉目标 Bunny 运动前后的特征点
Fig.7 The feature points of Bunny before and after motion

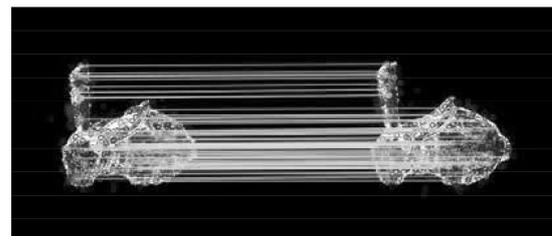


图8 筛选之后的匹配效果
Fig.8 The matching results after filtering

依据筛选后的匹配点, 可以对视觉目标进行运动估计. 假设视觉目标运动前后匹配点的对应关系为

$$X'_o = RX_o + T \quad (7)$$

其中, X'_o 为视觉目标运动后匹配点的三维坐标, R 为单位正交旋转矩阵, X_o 为视觉目标运动前匹配点的三维坐标, T 为平移矩阵.

从式 (7) 可以看出, 运动估计的结果就是求得一组 R 、 T , 使之满足全部运动前后匹配点的三维坐标. 考虑到动态遮挡规避过程的实时性, 本文采用奇异值分解法^[19] 计算式 (7) 中的 R 和 T . 定义一个矩阵 A

$$A = \sum_{o=1}^n X'_l X_l^T \quad (8)$$

其中,

$$X'_l = X'_o - \frac{1}{n} \sum_{k=1}^n X'_k, \quad X_l = X_o - \frac{1}{n} \sum_{k=1}^n X_k$$

n 为筛选后的匹配点个数.

对矩阵 A 进行奇异值分解可以得到 $A = U\Sigma V^T$, 其中 $R = VU^T$, 将其代入式 (7) 可得到 T . 至此完成了视觉目标的运动估计.

2.3 规划摄像机下一最佳观测方位

获取遮挡区域最佳观测方位模型与运动估计方程之后, 即可在此基础上规划摄像机的下一最佳观测方位. 通过分析动态遮挡规避过程可知, 若把摄像机观测位置代入运动估计方程, 便可实现摄像机随动于视觉目标, 但为了获取遮挡区域信息, 需要在摄像机随动于视觉目标的同时引入摄像机为规避遮挡区域所需运动的偏移量. 式 (5) 并非一个凸模型, 故很难直接求得最优解去计算摄像机所需运动的偏移量. 考虑到动态遮挡规避的最终结果是获取摄像机一系列观测方位, 而最优化方法在求解非凸模型过程中计算出的搜索方向可以为摄像机运动偏移量的确定提供依据. 目前可行的最优化方法有最速下降法、牛顿下降法、共轭梯度法、信赖域法等. 由于该模型较易计算导数, 且牛顿下降法相比于其他方法具有迭代次数少、收敛快的特点, 从而有利于减少摄像机为规避遮挡区域所需运动的次数. 故综合考虑方法的可行性和时间消耗, 本文采用牛顿下降法求解模型的搜索方向来计算摄像机的运动偏移量, 将摄像机下一最佳观测方向 V_{next} 与观测中心点 P_{next} 的计算公式分别定义为

$$V_{\text{next}} = RV_{\text{position}} + T + \lambda \frac{f'(V_{\text{position}})}{f''(V_{\text{position}})} \quad (9)$$

$$P_{\text{next}} = RP_{\text{current}} + T \quad (10)$$

其中, V_{position} 代表摄像机当前观测位置, λ 为步长因子, λ 值越大摄像机移动距离越远, 但匹配点数量

会相应减少从而影响运动估计的准确性. P_{current} 代表摄像机当前观测中心点.

依据计算出的下一最佳观测方向 V_{next} 和观测中心点 P_{next} 便可计算摄像机的下一最佳观测位置 P_{nc} , 计算公式定义为

$$P_{\text{nc}} = P_{\text{next}} - \frac{d}{\|V_{\text{next}}\|} V_{\text{next}} \quad (11)$$

其中, d 为当前摄像机位置到观测目标中心点的固定距离. 此时, 将已经得到的下一最佳观测方位 ($P_{\text{next}}, P_{\text{nc}}$) 作为当前观测方位 ($P_{\text{current}}, V_{\text{position}}$), 在新的观测方位 ($P_{\text{current}}, V_{\text{position}}$) 下继续采集视觉目标的深度图像, 并计算所观测到的遮挡区域面积 S , 计算公式定义为

$$S = \sum_{i=1}^n \cos \theta_i S_i, \quad i = 1, \dots, n \quad (12)$$

其中, θ_i 为第 i 个 patch 的法向量与 patch 的中心点 (patch 四个点坐标的算术平均值) 到当前摄像机观测方位构成向量的夹角, S_i 为第 i 个 patch 的面积. 基于式 (12) 的求解, 判断相邻两次观测方位下观测到的视觉目标遮挡区域面积差值是否小于给定阈值, 计算公式定义为

$$S_{\text{next}} - S_{\text{current}} \leq \varepsilon \quad (13)$$

其中, S_{next} 为下一观测方位下观测到的视觉目标遮挡区域面积, S_{current} 为当前观测方位下观测到的视觉目标遮挡区域面积.

重复式 (6)~(13) 的过程直到式 (13) 成立为止, 此时获得了一系列摄像机下一最佳观测方位 ($P_{\text{next}}, P_{\text{nc}}$), 动态遮挡规避过程结束.

2.4 动态遮挡规避算法

输入. 运动视觉目标深度图像, 摄像机内外参数.

输出. 动态遮挡规避过程中摄像机的下一最佳观测方位集合.

步骤 1. 提取视觉目标深度图像遮挡边界的关键点以及与其相对应的下邻接点构建关键线段, 并利用关键线段对遮挡区域建模.

步骤 2. 依据式 (1)~(5) 构建遮挡区域最佳观测方位模型.

步骤 3. 根据式 (6) 计算相邻两幅深度图像的混合曲率矩阵; 采用 SURF 算法进行匹配, 并依据三角形法则对匹配点进行筛选; 根据式 (7) 和式 (8) 对视觉目标进行运动估计.

步骤 4. 根据式 (9)~(11) 规划摄像机下一最佳观测方位; 依据式 (12) 计算摄像机能观测到的遮挡区域面积; 若相邻两次面积差值大于阈值 ε , 跳转

至步骤 3 继续进行遮挡规避; 直到动态遮挡规避过程结束.

步骤 5. 输出摄像机的下一最佳观测方位集合.

3 实验及分析

3.1 实验方案及数据集

为了全面合理地验证所提方法的可行性和有效性, 本文做了 4 组实验. 1) 构建遮挡区域最佳观测方位模型过程中各个阶段可视化实验; 2) 针对不同曲率矩阵提取特征点的对比实验; 3) 动态遮挡规避实验; 4) 与已有动态遮挡规避方法的实验比较. 实验硬件环境为 Intel(R) Core(TM) i7-3770 CPU@3.40 GHz, 内存 8.00 GB. 实验中的 3 维模型来自深度图像领域著名的 Stuttgart Range Image Database^[20] 和 Kinect 采集的真实视觉目标. 实验采用 C++ 编程实现. 实验过程中采用 OpenGL 实现视觉目标运动, 设置 OpenGL 投影矩阵的参数为 (60, 1, 200, 600), 视窗大小为 400 像素 \times 400 像素, 深度图像的分辨率为 640 像素 \times 480 像素. 摄像机到观测目标中心点的固定距离 $d = 300$ mm, 常数 β

$= 200$, 常数 $k = 3$, 步长因子 $\lambda = 2.5$ (常数 β 、 k 和步长因子 λ 均为经验值), 考虑到动态遮挡规避对准准确性和实时性的要求, 以及为了与文献 [13] 进行对比实验等因素, 本文选取阈值 $\varepsilon = 30$ mm² (后文中, 世界坐标系的单位均为 mm).

3.2 实验结果及分析

3.2.1 构建遮挡区域最佳观测方位模型过程中各个阶段可视化实验结果和分析

为了更加形象地说明本文方法构建的遮挡区域最佳观测方位模型. 图 9 给出了构建遮挡区域最佳观测方位模型过程中各个阶段的可视化结果. 图中最左侧的文字表示视觉目标, 图 9(a) 是初始观测方位下深度图像, 图 9(b) 是遮挡边界、关键点及下邻接点, 图 9(c) 是关键线段的观测方向, 图 9(d) 是遮挡区域最佳观测方位下的可见关键线段, 图 9(e) 是遮挡区域最佳观测方位下的深度图像.

从图 9 可以看出, 对于视觉目标 Duck, 由于在当前观测方位下遮挡情况不太显著并且各个关键线段之间围成的 patch 面积差值较小, 故在遮挡区域最佳观测方位下观测到的遮挡区域面积较小; 对于

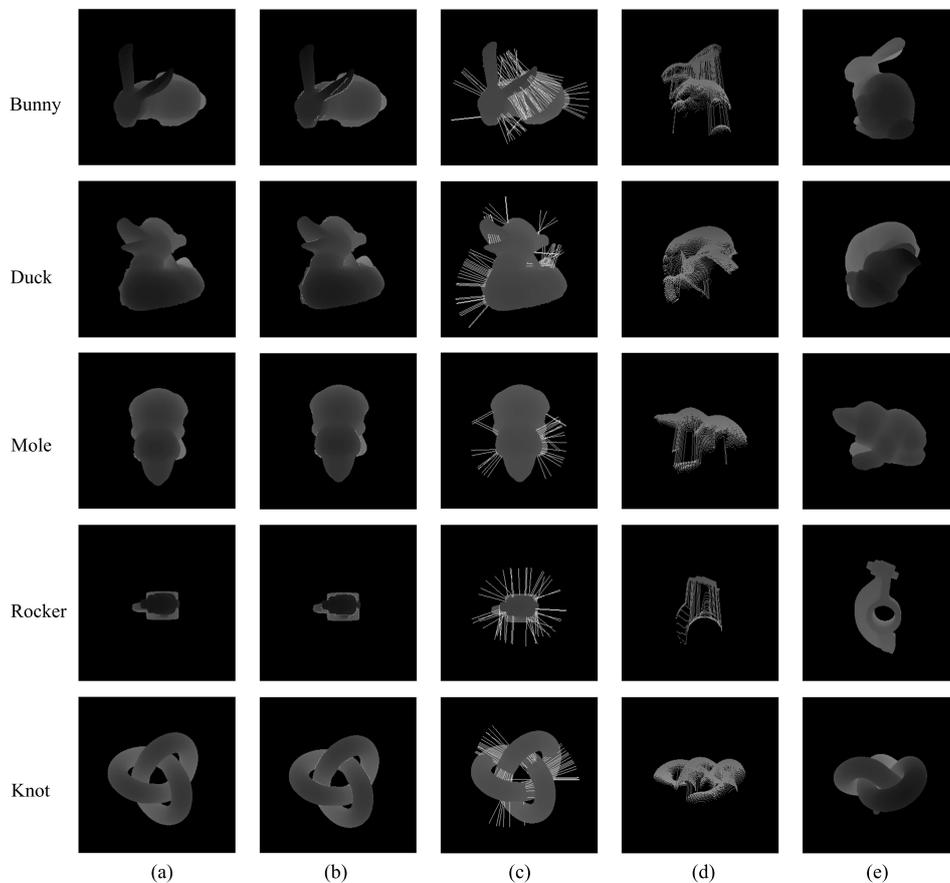


图 9 各个阶段的可视化结果

Fig. 9 The visualization results of each phase

视觉目标 Bunny, Mole, Rocker 和 Knot, 由于在当前观测方位下遮挡情况相对显著并且各个关键线段之间围成的 patch 面积差值较大, 故在遮挡区域最佳观测方位下观测到的遮挡信息较多. 由此可见, 当视觉目标遮挡现象越显著并且各个关键线段之间围成的 patch 面积差值越大, 本文方法构建的遮挡区域最佳观测方位模型越好, 进一步验证了所提关键点方案的可行性.

3.2.2 针对不同曲率矩阵提取特征点的对比实验

视觉目标深度图像所对应曲率矩阵将对最终提取的特征点产生较大的影响, 进而影响到视觉目标运动估计的准确性. 为了验证本文所提混合曲率特征的有效性, 分别对基于高斯曲率、平均曲率和混合曲率三种不同曲率计算出的曲率矩阵的实验结果进行了比较. 图 10 展示了针对不同视觉目标运用不同曲率获取与深度图像相对应的曲率矩阵的可视化结果. 图中最左侧的文字表示视觉目标, 图 10 (a) 是视觉目标深度图像, 图 10 (b) 是高斯曲率矩阵, 图 10 (c) 是平均曲率矩阵, 图 10 (d) 是混合曲率矩阵.

分析图 10 结果可知, 高斯曲率能较好地表达视觉目标内部表面的纹理信息, 平均曲率能较好地表达视觉目标边缘的轮廓信息, 本文所提的混合曲率则能够更好地表达视觉目标边缘轮廓与内部表面纹理信息.

为了更好地说明本文所提混合曲率特征的效果, 表 1 给出了与视觉目标深度图像对应不同曲率矩阵提取特征点的量化结果, 实验结果为 30 幅深度图像对应的曲率矩阵提取特征点的平均数 (取整). 其中 N_C 表示采用平均曲率矩阵提取特征点的个数, N_G 表示采用高斯曲率矩阵提取特征点的个数, N_{GC} 表示采用混合曲率矩阵提取特征点的个数. 从表 1 可以看出, 采用本文方法获取视觉目标特征点的结果明显优于其他两种方法获取视觉目标特征点的结果, 进而验证了本文所提混合曲率特征的有效性.

3.2.3 动态遮挡规避实验的结果和分析

为了验证本文所提动态遮挡规避方法的可行性和有效性, 分别进行了仿真实验和真实实验.

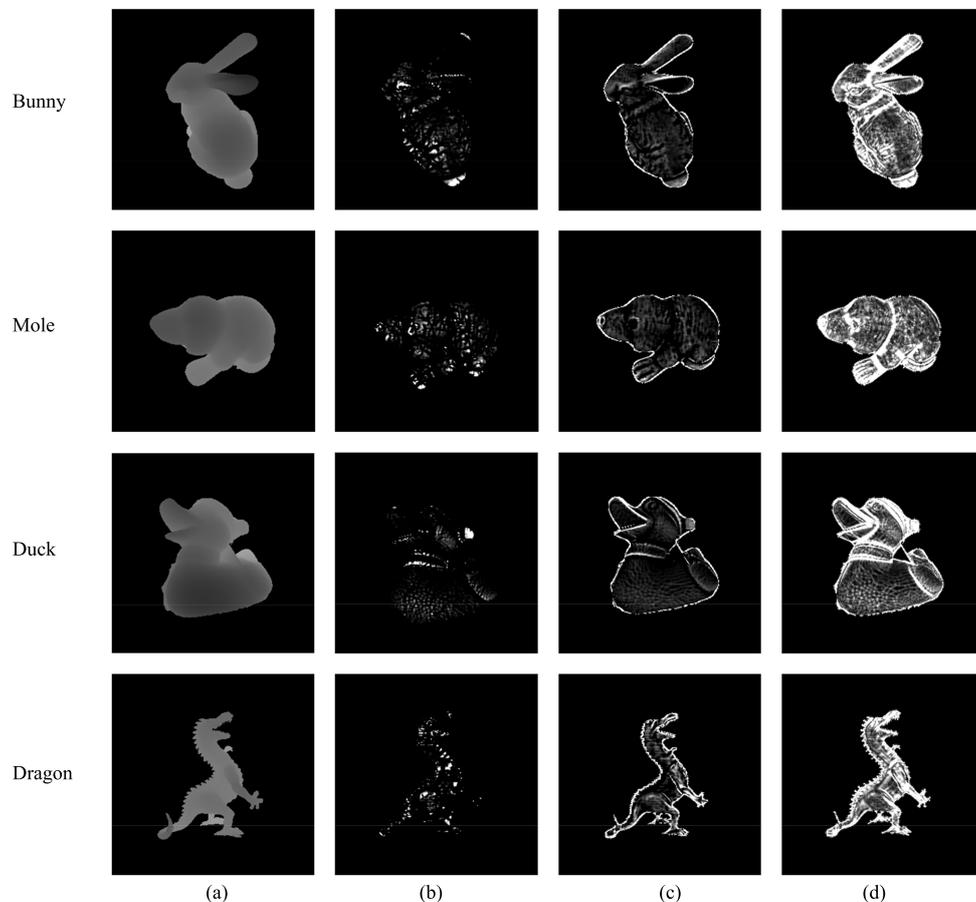


图 10 不同曲率矩阵的可视化结果

Fig. 10 The result of visualization for different curvature matrices

表 1 不同曲率矩阵提取特征点结果的量化评估

Table 1 Quantitative evaluation of the result of feature points extracted by different curvature matrices

视觉目标名称	N_C	N_G	N_{GC}
Bunny	210	109	387
Duck	201	95	378
Mole	144	110	277
Rocker	105	46	148
Knot	175	87	298

在仿真实验中, 选取 Stuttgart Range Image Database 中的 3 维视觉目标 Dragon 和 Bunny 作为实验对象, 使其在不同的运动方式下进行动态遮挡规避. 表 2 是视觉目标 Dragon 在初始观测方位

$[190, -200, 118]^T$ 下以 2 mm/s 的速度沿向量 $[-2, 0, 0]^T$ 做平移运动的动态遮挡规避过程. 表 3 是视觉目标 Bunny 在初始观测方位 $[-196, 170, 150]^T$ 下以 $5^\circ/\text{s}$ 速度绕向量 $[-1, 1, 2]^T$ 做旋转运动的动态遮挡规避过程. 表 4 是视觉目标 Bunny 在初始观测方位 $[190, -200, 118]^T$ 下以 3 mm/s 的速度沿向量 $[1, 0, 0]^T$ 平移的同时以 $2^\circ/\text{s}$ 速度绕向量 $[2, 1, 0]^T$ 做旋转运动的动态遮挡规避过程. 由于摄像机在动态遮挡规避过程中移动次数较多, 仅展示了摄像机在部分观测方位下的观测结果.

在真实实验中, 基于视觉目标 Printer (打印机), Kettle (水壶) 和 Plant (海葱) 的不同运动方式进行动态遮挡规避实验, 并利用 Kinect 采集它们的深度图像. 图 11 是视觉目标 Printer, Kettle 和 Plant 的 RGB 图像、深度图像、深度图像中的遮挡边界及其关键点. 分析图 11 (c) 中不同视觉目标的深度图像可

表 2 视觉目标 Dragon 做平移运动时的动态遮挡规避过程

Table 2 The dynamic occlusion avoidance process of visual object Dragon with translation motion

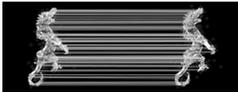
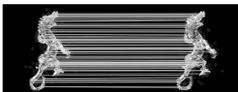
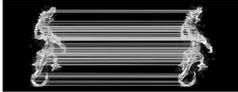
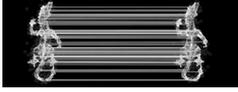
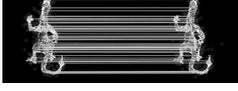
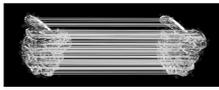
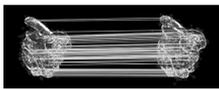
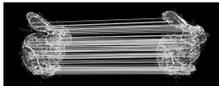
	深度图像	筛选后特征点匹配图像	筛选后匹配点数	观测方向	观测位置 (mm)	观测面积 (mm^2)
第 1 幅				$[-190.00,$ $200.00,$ $-118.00]$	$[190.00,$ $-200.00,$ $-118.00]$	
第 2 幅			125	$[-190.00,$ $200.00,$ $-118.00]$	$[190.00,$ $-200.00,$ $-118.00]$	79.85
第 3 幅			130	$[-177.41,$ $212.58,$ $-114.91]$	$[178.12,$ $-212.42,$ $114.78]$	142.64
⋮						
第 6 幅			131	$[-138.89,$ $245.81,$ $-101.09]$	$[139.54,$ $-245.54,$ $101.32]$	343.87
⋮						
第 10 幅			119	$[-77.45,$ $297.12,$ $-78.71]$	$[77.94,$ $-278.79,$ $78.89]$	653.20
⋮						
第 15 幅			121	$[-1.99,$ $296.93,$ $-46.55]$	$[2.96,$ $-296.31,$ $47.01]$	963.37
⋮						
第 16 幅				$[5.47,$ $297.18,$ $-43.25]$	$[-4.78,$ $-296.82,$ $43.57]$	990.57

表 3 视觉目标 Bunny 做旋转运动时的动态遮挡规避过程

Table 3 The dynamic occlusion avoidance process of visual object Bunny with rotation motion

	深度图像	筛选后特征点匹配图像	筛选后匹配点数	观测方向	观测位置 (mm)	观测面积 (mm ²)
第 1 幅				[196.00, -170.00, -150.00]	[-196.00, 170.00, 150.00]	
第 2 幅			212	[196.00, -170.00, -150.00]	[-196.00, 170.00, 150.00]	476.11
第 3 幅			143	[209.64, -197.55, -81.44]	[-208.74, 195.66, 89.22]	1 139.89
⋮						
第 5 幅			167	[211.73, -216.22, -10.29]	[-210.13, 212.87, 18.59]	2 165.26
⋮						
第 9 幅			181	[198.26, -220.89, -72.18]	[-195.49, 217.06, -66.95]	3 131.86
⋮						
第 14 幅			208	[149.48, -192.54, 178.01]	[-148.50, 191.19, -176.66]	3 902.76
第 15 幅				[145.25, -188.99, 184.46]	[-144.46, 188.14, -183.16]	3 928.31

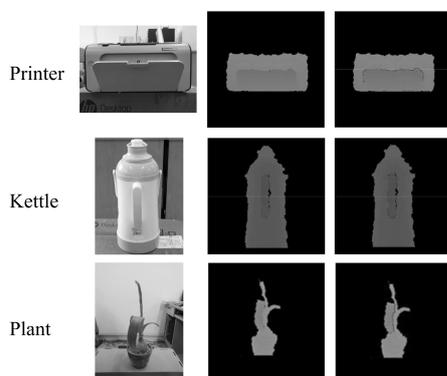


图 11 真实实验结果

Fig. 11 The result of real experiment

知, 在图像采集过程中, 由于 Kinect 自身误差、遮挡及柔性物体例如 Plant 易发生形变、尺度变化等

现象, 从而导致获取的深度图像存在一定的边界不连续、信息缺失或噪声. 但通过分析图像 11 (d) 的结果可知, 对于存在上述情形的遮挡边界, 所提方法依然能够较准确地提取到遮挡边界上的关键点, 从而验证了本文所提关键点筛选方法具有一定的鲁棒性.

在真实实验中, 由于目前实验硬件条件的限制, 无法使摄像机随动于视觉目标的同时对视觉目标进行拍摄. 考虑到实验的完整性, 通过对 Kinect 采集到的不同视觉目标的深度图像进行三维重建, 进而利用 OpenGL 模拟了不同视觉目标在不同运动方式下的动态遮挡规避过程. 表 5 是视觉目标 Printer 在初始观测方位 $[0, 1, 1200]^T$ 下以 2 mm/s 的速度沿着向量 $[0, 1, 0]^T$ 做平移运动的动态遮挡规避过程. 表 6 是视觉目标 Kettle 在初始观测方位 $[0, -1, 600]^T$ 下以 $2^\circ/\text{s}$ 的速度沿着向量 $[-2, 1, 0]^T$ 做旋转

表 4 视觉目标 Bunny 做平移和旋转运动时的动态遮挡规避过程

Table 4 The dynamic occlusion avoidance process of visual object Bunny with both translation and rotation motion

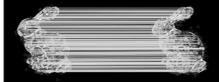
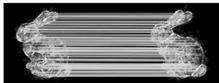
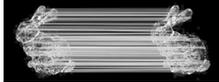
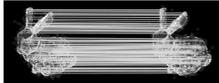
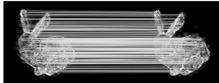
	深度图像	筛选后特征点匹配图像	筛选后匹配点数	观测方向	观测位置 (mm)	观测面积 (mm ²)
第 1 幅				[−190.00, 200.00, −118.00]	[190.00, −200.00, −118.00]	
第 2 幅			224	[−190.00, 200.00, −118.00]	[190.00, −200.00, −118.00]	192.57
第 3 幅			193	[−161.66, 212.61, −137.59]	[164.35, −211.53, 135.16]	301.96
⋮						
第 7 幅			204	[−93.59, 227.26, −173.45]	[95.53, −226.39, 172.17]	815.42
⋮						
第 15 幅			201	[18.87, 211.37, −213.50]	[−17.23, −201.95, 212.67]	1 476.87
⋮						
第 20 幅			202	[63.67, 191.36, −223.15]	[−62.42, −191.18, 222.66]	1 678.43
⋮						
第 21 幅				[71.12, 187.29, −224.59]	[−69.81, −187.05, 223.98]	1 707.26

表 5 视觉目标 Printer 做平移运动时的动态遮挡规避过程

Table 5 The dynamic occlusion avoidance process of visual object Printer with translation motion

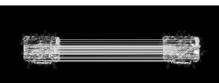
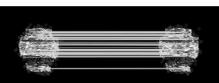
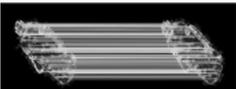
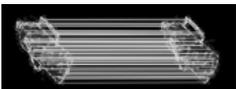
	深度图像	筛选后特征点匹配图像	筛选后匹配点数	观测方向	观测位置 (mm)	观测面积 (mm ²)
第 1 幅				[0.00, −1.00, −1 200.00]	[0.00, 1.00, 1 200.00]	
第 2 幅			73	[0.00, −1.00, −1 200.00]	[0.00, 1.00, 1 200.00]	321.47
⋮						
第 4 幅			85	[4.45, −126.90, −917.06]	[−3.12, 127.21, 917.43]	587.21
⋮						
第 7 幅			133	[−14.13, −350.78, −682.74]	[−13.78, 351.64, 682.95]	1 102.06
⋮						
第 18 幅			187	[49.91, −1 089.01, −151.39]	[−49.85, 1 088.96, 151.51]	4 537.99
⋮						
第 19 幅				[53.07, −1 191.18, −129.07]	[−52.93, 1 191.83, 129.37]	4 566.16

表6 视觉目标 Kettle 做旋转运动时的动态遮挡规避过程

Table 6 The dynamic occlusion avoidance process of visual object Kettle with rotation motion

	深度图像	筛选后特征点匹配图像	筛选后匹配点数	观测方向	观测位置 (mm)	观测面积 (mm ²)
第 1 幅				[0.00, 1.00, -600.00]	[0.00, -1.00, 600.00]	
第 2 幅			92	[0.00, 1.00, -600.00]	[0.00, -1.00, 600.00]	65.31
第 3 幅			160	[25.11, 12.97, -583.63]	[-24.37, -11.56, 581.70]	112.74
⋮						
第 7 幅			182	[368.26, 124.46, -398.30]	[-367.81, -123.11, 397.05]	581.09
⋮						
第 11 幅			192	[601.97, 274.61, -200.72]	[-601.32, -274.01, 200.08]	844.53
第 12 幅				[599.19, 272.53, -201.67]	[-598.77, -271.21, 201.16]	871.52

运动的动态遮挡规避过程. 表 7 是视觉目标 Plant 在初始观测方位 $[0, -1, 650]^T$ 下以 $2^\circ/\text{s}$ 的速度沿着向量 $[-2, 1, 0]^T$ 做旋转的同时以 2 mm/s 的速度沿着向量 $[0, 1, 0]^T$ 做平移运动的动态遮挡规避过程.

表 2~7 中, 第 1 列是摄像机在动态遮挡规避过程中所拍摄的视觉目标深度图像编号, 当遮挡规避过程结束时, 由不同表的第 1 列图像编号不同可知, 当视觉目标以不同的方式进行运动时, 摄像机的运动次数也是不同的. 第 2 列是摄像机在一系列下一最佳观测方位下拍摄到的视觉目标深度图像, 分析相关图像可知, 本文方法在动态遮挡规避过程中求得的一系列下一最佳观测方位都与人类视觉的观测习惯相符. 第 3、4 列分别是筛选匹配点之后的匹配图像和匹配点数, 可以看出筛选后的匹配点都更加准确, 从而有利于后续估计视觉目标运动以及规划摄像机的下一最佳观测方位, 以便摄像机获取更多的视觉目标遮挡信息. 第 5~7 列分别是计算出的摄像机下一最佳观测方向、观测位置和相应观测方位下求得的遮挡区域面积. 从这些数据及其变化趋势可以看出, 随着动态遮挡规避过程的进行, 越来越多的遮挡区域被观测到.

3.2.4 与已有动态遮挡规避方法的实验比较及分析

虽然目前动态遮挡规避方面的文献较少, 但为

了更全面合理地评估本文方法的效果, 将本文方法与同样基于深度图像且考虑遮挡信息的代表动态遮挡领域最新进展的文献 [13] 中的方法进行了三组对比实验. 1) 不同建模方法以及下一最佳观测方位的对比实验; 2) 运动估计的对比实验; 3) 动态遮挡规避最终结果的对比实验. 考虑到本文真实实验过程并不完整, 即动态遮挡规避部分仍然采用重建后的 3 维模型进行实验, 这与基于 Stuttgart Range Image Database 中的 3 维模型进行实验并无本质区别, 并且为了保持与文献 [13] 实验对比部分的一致性, 故此部分并未增加真实实验的对比数据. 下面分别给出三组对比实验的结果及分析.

表 8 是本文方法和文献 [13] 方法在相同初始观测方位下对遮挡区域建模的时间消耗对比结果. 从表 8 可以看出, 对于不同视觉目标遮挡区域, 本文建模方法在时间消耗上均低于文献 [13] 中的方法, 进一步验证本文方法在建模速度上更满足动态遮挡规避对实时性的要求.

为了更好地衡量本文所提建模方法, 将两种建模方法计算出的下一最佳观测方位进行实验对比. 图 12 是两种方法在所求下一最佳观测方位下获得的视觉目标深度图像. 其中, 第 1 行为当前观测方位下的视觉目标深度图像, 第 2 行为文献 [13] 方法计算出的下一最佳观测方位下的视觉目标深度图像, 第

表 7 视觉目标 Plant 同时做旋转和平移运动时的动态遮挡规避过程

Table 7 The dynamic occlusion avoidance process of visual object Plant with both translation and rotation motion

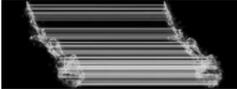
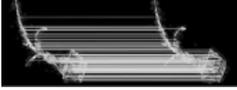
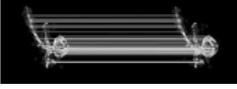
	深度图像	筛选后特征点匹配图像	筛选后匹配点数	观测方向	观测位置 (mm)	观测面积 (mm ²)
第 1 幅				[0.00, 1.00, -650.00]	[0.00, -1.00, 650.00]	
第 2 幅			167	[0.00, 1.00, -650.00]	[0.00, -1.00, 650.00]	89.47
⋮						
第 5 幅			209	[151.02, -126.37, -471.21]	[-149.56, 127.21, 471.44]	237.01
⋮						
第 7 幅			257	[218.62, -251.09, -367.44]	[-217.38, 251.64, 369.18]	437.91
⋮						
第 15 幅			119	[434.17, -648.52, -32.81]	[-436.55, 648.77, 33.81]	1 123.24
第 16 幅				[435.89, -648.97, -30.88]	[-437.68, 649.10, 31.14]	1 152.32

表 8 两种建模方法的时间消耗 (ms)

Table 8 Time consumption of two modeling methods (ms)

方法名称	时间消耗					平均时间消耗
	Bunny	Duck	Mole	Rocker	Dragon	
文献 [13]	69.51	40.03	56.04	42.87	83.03	58.29
本文	6.83	4.19	6.96	4.77	8.28	6.21

3 行为本文方法计算出的下一最佳观测方位下的视觉目标深度图像. 表 9 是两种不同建模方法求解出的下一最佳观测方位下获取的视觉目标表面点的量化结果. 其中 N_{nbv} 代表摄像机在下一最佳观测方位下获取到的视觉目标表面点的个数, N 代表被重复观测到的视觉目标表面点的个数, $N_{new} = N_{nbv} - N$. 代表实际新增点的个数, R 代表重合率, R_{new} 代表新增率.

结合图 12 和表 9 的分析可知, 由于本文方法与文献 [13] 均是通过遮挡区域建模, 进而计算下一最佳观测方位, 因此, 对于遮挡情况不太显著的视觉目标 Duck 和 Mole, 由于在初始观测方位下遮挡区域相对较少, 影响下一最佳观测方位计算的准确性, 从而导致在下一最佳观测方位下计算出的新增率相对偏低; 对于遮挡情况相对显著的视觉目标 Bunny,

Rocker 和 Dragon 而言, 由于下一最佳观测方位计算相对准确, 所以新增率相对较高. 与本文方法相比, 由于文献 [13] 方法在遮挡区域建模过程中, 对遮挡面积不同的区域并没有做任何区分, 而本文方法在遮挡区域建模过程中, 更注重对遮挡面积较大的区域进行观测, 根据实验结果来看, 在新增点和新增率上, 本文方法都比较好. 综合来看, 与文献 [13] 相比, 本文方法的适用性更好, 更符合人类的观测习惯.

为了衡量本文运动估计的准确性, 针对做不同运动的视觉目标 Bunny, 分别采用文献 [13] 和本文中的方法进行运动估计, 并将运动估计得到的结果对初始观测方位 $[0, -1, 300]^T$ 进行变换, 最后分析变换结果和时间消耗. 表 10 是两种运动估计方法的对比结果.

表 9 两种方法的下一最佳观测方位实验结果量化评估

Table 9 The quantitative evaluation of experimental results in next best view for two methods

视觉目标名称	文献 [13] 方法					本文方法				
	N_{nbv}	N_o	N_{new}	R_o	R_{new}	N_{nbv}	N_o	N_{new}	R_o	R_{new}
Bunny	16 071	1 835	14 236	11.42	88.58	16 087	1 023	15 064	6.36	93.64
Duck	17 622	5 187	12 444	29.38	70.62	17 418	4 286	13 132	24.61	75.39
Mole	12 354	1 930	10 424	15.62	84.38	15 996	1 942	14 054	12.14	87.86
Rocker	10 291	346	9 945	3.36	96.64	9 880	309	9 571	3.13	96.87
Dragon	9 090	416	8 674	4.58	95.42	9 358	317	9 041	3.39	96.61

表 10 两种运动估计方法的对比结果

Table 10 The comparison results of two motion estimation methods

运动方式	方法	结果 (mm)	消耗时间 (ms)
以 3 mm/s 速度沿向量 $[1, 0, 0]^T$ 平移	Ground truth	$[0.3, -1, 300]^T$	
	文献 [13]	$[0.398, -1.158, 300.114]^T$	434.14
	本文方法	$[0.341, -1.026, 300.015]^T$	119.55
以 $2^\circ/s$ 速度绕向量 $[2, 1, 0]^T$ 旋转	Ground truth	$[4.682, -10.364, 299.789]^T$	
	文献 [13]	$[3.541, -7.003, 299.551]^T$	573.09
	本文方法	$[4.531, -9.158, 299.865]^T$	212.95
以 3 mm/s 速度沿向量 $[1, 0, 0]^T$ 平移的同时 以 $2^\circ/s$ 速度绕向量 $[2, 1, 0]^T$ 旋转	Ground truth	$[4.982, -10.364, 299.789]^T$	
	文献 [13]	$[3.671, -7.119, 299.518]^T$	473.66
	本文方法	$[4.768, -9.690, 299.606]^T$	238.62

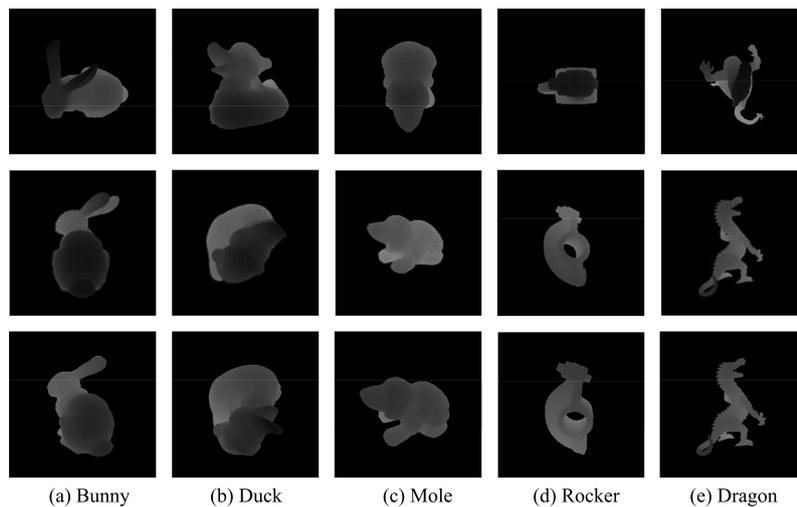


图 12 两种方法在下一最佳观测方位下获取的视觉目标深度图像

Fig. 12 The depth images of visual object in next best view for two methods

从表 10 可以看出, 本文方法的运动估计结果与 Ground truth 最为相近, 并且在时间消耗上明显低于文献 [13] 中的方法. 这是由于文献 [13] 采用 SIFT 算法匹配图像, 该算法需要计算每个关键点的 128 维特征向量, 从而导致运动估计的时间较长, 并且该方法没有对错误匹配点进行筛选, 致使最终运动估计结果与 Ground truth 相差相对较大. 而本文方法采用 SURF 算法匹配图像, 在相同条件下

SURF 算法的匹配速度是 SIFT 算法的 3 倍, 并且本文通过计算深度图像混合曲率矩阵, 增加了匹配点的数量, 同时, 对错误匹配点进行筛选, 以确保本文方法在耗时较少的情况下也能具有相对较高的准确性. 由此可见, 本文方法在运动估计及其误差方面具有一定的优势和改进.

此外, 在动态遮挡规避最终结果方面进行了对比实验. 表 11 是部分具有代表性的视觉目标在不同

表 11 两种动态遮挡规避方法的量化评估

Table 11 The quantitative evaluation of dynamic occlusion avoidance for two methods

视觉目标	实验组别	文献 [13] 方法					本文方法				
		$\sum_{j=1}^N S_j$ (mm ²)	S_{purpose} (mm ²)	S_{view} (mm ²)	η (%)	\bar{T} (s)	$\sum_{j=1}^N S_j$ (mm ²)	S_{purpose} (mm ²)	S_{view} (mm ²)	η (%)	\bar{T} (s)
Bunny	1	7289.26	2189.12	3025.58	94.87	0.93	7588.85	3729.86	3682.89	98.74	0.29
	2	7289.26	3189.12	2654.63	83.24	0.97	7588.85	3729.86	3467.21	92.96	0.35
	3	5028.92	3238.47	3017.48	93.18	0.91	4892.33	3556.86	3510.18	98.69	0.36
	4	4414.50	1758.18	1553.48	88.36	0.95	4575.69	1947.95	1837.64	94.34	0.37
Rocker	5	1905.09	1440.92	1363.45	94.62	0.93	2007.43	1567.96	1485.38	94.73	0.33
	6	3433.54	2707.33	1947.98	71.95	0.98	3364.32	2832.76	2396.03	84.58	0.40
Duck	7	1911.10	1025.27	733.58	71.55	1.02	1898.01	1071.93	830.93	77.52	0.39
	8	3307.46	2788.37	2748.35	98.56	0.93	3316.25	2928.41	2793.99	95.41	0.34
Mole	9	320.01	221.68	133.09	60.04	1.02	352.18	262.3	186.35	71.04	0.38
Dragon	10	5736.48	1948.61	1485.98	76.26	1.10	5831.1	2061.78	1710.62	82.87	0.36

观测方位和运动方式下对应的实验结果. 其中, $\sum_{j=1}^N S_j$ 代表视觉目标要规避的遮挡区域总面积, S_{purpose} 代表目标观测方位下观测到的遮挡区域面积, S_{view} 为动态遮挡规避过程结束时的观测方位下观测到的遮挡区域面积, η 表示有效规避率即 S_{view} 与 S_{purpose} 的比值, \bar{T} 代表动态遮挡规避过程中计算下一最佳观测方位所需时间 (平均值).

分析表 11 可知, 与文献 [13] 相比, 本文方法通过计算与视觉目标深度图像相对应的混合曲率矩阵, 增加提取特征点的数量并且对匹配点进行筛选, 为准确估计视觉目标运动奠定了基础. 因此, 无论视觉目标做简单运动 (第 1、3、5 组实验) 还是复杂运动 (第 2、4、6、7、9、10 组实验), 在最能反映实际动态遮挡规避效果的有效规避率 η 指标上, 实验结果都明显高于文献 [13] 方法的结果. 对于第 8 组实验, 由于在当前观测方位下相邻关键线段之间构成的 patch 面积差值不明显, 故本文方法的有效规避率低于文献 [13] 方法的结果. 对于第 6、7、9、10 组实验, 由于视觉目标在初始观测方位下遮挡情况不太显著, 进而影响下一最佳观测方位计算的准确性, 最终导致有效规避率相对偏低. 由此可见, 当遮挡情况越显著, 通过遮挡区域最佳观测方位模型计算的下一最佳观测方位越准确, 本文所提方法的有效规避率就相对越高. 对于每次获取下一最佳观测方位所需的时间 \bar{T} , 主要受运动估计以及计算摄像机下一最佳观测方位等因素的影响, 而本文方法在运动估计方面的时间消耗低于文献 [13] 中的方法, 并且采用牛顿下降法减少下一最佳观测方位计算过程中的迭代次数, 故本文方法获取下一次最佳观测方位

所需的时间明显少于文献 [13] 方法的结果. 综合来看, 与文献 [13] 相比, 本文方法在对遮挡区域建模、视觉目标运动估计以及规划摄像机下一最佳观测方位等方面, 更好地达到了动态遮挡规避问题的求解要求.

4 结论

本文提出了一种利用深度图像遮挡信息对运动视觉目标遮挡区域进行规避的方法. 所提方法的贡献主要体现在三个方面: 1) 提出了深度图像遮挡边界关键点的概念, 并利用关键点与其对应的下邻接点构建关键线段实现对视觉目标遮挡区域的快速建模; 2) 依据遮挡区域建模结果和关键线段构造遮挡区域最佳观测方位模型, 并结合最优化方法求解模型, 为摄像机下一最佳观测方位的确定提供了依据; 3) 提出了一种混合曲率特征, 通过计算深度图像对应的混合曲率矩阵, 使得 SURF 算法能够提取到更多的特征点, 增加了匹配点的数量, 为准确估计视觉目标运动奠定了基础. 实验结果表明, 所提方法能够实现运动视觉目标遮挡区域的动态规避过程, 且与人类的视觉观测习惯相符. 需要指出的是, 所提方法构建的遮挡区域最佳观测方位模型是一个非凸模型, 未来将在尝试构建更优的最佳观测方位模型等方面展开研究.

References

- 1 Koller D, Weber J, Malik J. Robust multiple car tracking with occlusion reasoning. In: Proceedings of the 3rd European Conference on Computer Vision. Stockholm, Sweden: Springer-Verlag, 1994. 189–196

- 2 Senior A, Hampapur A, Tian Y L, Brown L, Pankanti S, Bolle R. Appearance models for occlusion handling. *Image and Vision Computing*, 2006, **24**(11): 1233–1243
- 3 Chang T H, Gong S G, Ong E J. Tracking multiple people under occlusion using multiple cameras. In: Proceedings of the 2000 British Machine Vision Conference. Bristol, UK, Greece: BMVC, 2000. 566–575
- 4 Maver J, Bajcsy R. Occlusions as a guide for planning the next view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(5): 417–433
- 5 Li Y F, Liu Z G. Information entropy-based viewpoint planning for 3-D object reconstruction. *IEEE Transactions on Robotics*, 2005, **21**(3): 324–337
- 6 Banta J E, Wong L R, Dumont C, Abidi M A. A next-best-view system for autonomous 3-D object reconstruction. *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans*, 2000, **30**(5): 589–598
- 7 Wu B, Sun X, Wu Q C, Yan M L, Wang H Q, Fu K. Building reconstruction from high-resolution multiview aerial imagery. *IEEE Geoscience and Remote Sensing Letters*, 2015, **12**(4): 855–859
- 8 Pito R. A solution to the next best view problem for automated surface acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, **21**(10): 1016–1030
- 9 Potthast C, Sukhatme G S. A probabilistic framework for next best view estimation in a cluttered environment. *Journal of Visual Communication and Image Representation*, 2014, **25**(1): 148–164
- 10 Zhang Shi-Hui, Liu Jian-Xin, Kong Ling-Fu. Determining next best view based on occlusion information in depth image. *Chinese Journal of Computers*, 2015, **38**(12): 2450–2463
(张世辉, 刘建新, 孔令富. 基于深度图像利用遮挡信息确定下一最佳观测方位. *计算机学报*, 2015, **38**(12): 2450–2463)
- 11 Doumanoglou A, Kouskouridas R, Malassiotis S, Kim T K. Recovering 6D object pose and predicting next-best-view in the crowd. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 3583–3592
- 12 Bissmarck F, Svensson M, Tolt G. Efficient algorithms for Next Best View evaluation. In: Proceedings of the 2015 IEEE International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany: IEEE, 2015. 5876–5883
- 13 Zhang S H, He H, Zhang Y C, Li X, Sang Y. Dynamic self-occlusion avoidance approach based on the depth image sequence of moving visual object. *Mathematical Problems in Engineering*, 2016, **2016**: Article No. 4783794
- 14 Zhang S H, Liu J X. A self-occlusion detection approach based on depth image using SVM. *International Journal of Advanced Robotic Systems*, 2012, **9**(6): Article No. 230
- 15 Zhang Shi-Hui, Han De-Wei, He Huan. Determining next best view using occlusion and contour information of visual object. *Journal of Electronics and Information Technology*, 2015, **37**(12): 2921–2928
(张世辉, 韩德伟, 何欢. 利用视觉目标遮挡和轮廓信息确定下一最佳观测方位. *电子与信息学报*, 2015, **37**(12): 2921–2928)
- 16 Besl P J, Jain R C. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1988, **10**(2): 167–192
- 17 Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features. In: Proceedings of the 9th European Conference on Computer Vision and Image Understanding. Graz, Austria: ECCV, 2006, **110**: 404–417
- 18 Zhang Dong-Xing, Zhu Ming-Bo, Zou Jian-Wu, Li Xiang-Ping. Research on wrong match pairs elimination based on similar triangles in the SIFT algorithm. *Computer Engineering and Science*, 2012, **34**(4): 66–70
(张东兴, 祝明波, 邹建武, 李相平. 基于相似三角形的 SIFT 错误匹配点剔除算法研究. *计算机工程与科学*, 2012, **34**(4): 66–70)
- 19 Zhao Feng, Huang Qing-Ming, Gao Wen. An image matching algorithm based on singular value decomposition. *Journal of Computer Research and Development*, 2010, **47**(1): 23–32
(赵峰, 黄庆明, 高文. 一种基于奇异值分解的图像匹配算法. *计算机研究与发展*, 2010, **47**(1): 23–32)
- 20 Hetzel G, Leibe B, Levi P, Schiele B. 3D object recognition from range images using local feature histograms. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA: IEEE, 2001. 394–399



张世辉 燕山大学信息科学与工程学院教授. 主要研究方向为视觉信息处理, 模式识别, 智能机器人. 本文通信作者.

E-mail: sshhzz@ysu.edu.cn

(**ZHANG Shi-Hui** Professor at the College of Information Science and Engineering, Yanshan University. His research interest covers visual information processing, pattern recognition and intelligent robot. Corresponding author of this paper.)



何琦 燕山大学硕士研究生. 主要研究方向为计算机视觉和模式识别.

E-mail: heqi0109@126.com

(**HE Qi** Master student at Yanshan University. His research interest covers computer vision and pattern recognition.)



董利健 燕山大学硕士研究生. 主要研究方向为计算机视觉和模式识别.

E-mail: donglijian8087@163.com

(**DONG Li-Jian** Master student at Yanshan University. His research interest covers computer vision and pattern recognition.)



杜雪哲 燕山大学硕士研究生. 主要研究方向为计算机视觉和模式识别.

E-mail: duxuezhe@yeah.net

(**DU Xue-Zhe** Master student at Yanshan University. His research interest covers computer vision and pattern recognition.)