

基于汉字标志的增强现实系统

董子龙 章国锋 邵元龙 华 炜

(浙江大学 CAD&CG 国家重点实验室, 杭州 310058)

摘要 增强现实是将虚实信息融合的技术,该系统目标是利用普通的网络摄像头,实时地在用户提供的包围在黑色方框中的汉字标志图像上附加相关的媒体信息,有效地扩展和丰富传统的看图识字。为了在复杂的光影下也能稳定地检测出标志,系统利用边缘提取优化了标志检测方法。为了稳定快速地求解摄像机参数,系统首先利用单应性矩阵快速求解出初值,并通过在优化函数中加入自适应的平滑性约束来抑制图像噪声的影响,使得求解的摄像机参数更加精确平滑。实验结果证明了该系统的稳定性和实用性。

关键词 增强现实 边缘提取 汉字标志 摄像机跟踪

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2009)07-1463-06

Chinese-character-marker Based Augmented Reality System

DONG Zi-long, ZHANG Guo-feng, SHAO Yuan-long, HUA Wei

(CAD&CG State Key Lab of Zhejiang University, Hangzhou 310058)

Abstract Augmented reality is a technique to fuse virtual and real worlds. The proposed augmented reality system aims to use an off-the-shelf web camera, and augments related virtual media information onto the Chinese-character-markers surrounded by black rectangles, which provides a new learning experience for Chinese character. The proposed method enhances the marker detection method with edge extraction, and performs well under complicated shading condition. To reliably solve the camera parameters, the proposed system first uses the homography to estimate the initial values, and then incorporates adaptive smoothness constraint to reduce the image noise. This constraint makes the solved camera parameters very accurate and stable. The experimental results demonstrate the effectiveness and robustness of the proposed system.

Keywords augmented reality, edge extraction, Chinese-character-marker, camera tracking

1 引言

增强现实 (augmented reality) 是将现实世界和计算机生成的信息融合在一起的技术,这些信息可以是图像、文字、3 维模型和光照。增强现实极大地丰富了现实世界的信息,在医学、教育、娱乐、军事、游戏上都有丰富的应用^[1-2]。为利用增强现实技术完成一个稳定的汉字学习系统,本文将计算机生成的多媒体信息与汉字图像结合起来,使学习手段更

加丰富有趣(如图 1 所示)。

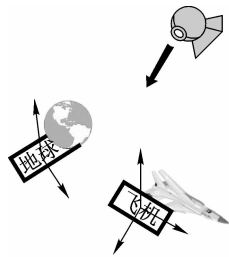


图 1 系统原型,在汉字标志上绘制相关媒体内容
Fig. 1 System prototype, augment related media information onto Chinese-character-markers

基金项目:国家高技术研究发展计划(863)基金项目(2007AA01Z326)

收稿日期:2008-09-01;改回日期:2008-10-24

第一作者简介:董子龙(1983 ~),男。浙江大学 CAD&CG 国家重点实验室计算机科学和工程专业博士研究生。主要研究领域为计算机视觉、增强现实。E-mail:zldong@cad.zju.edu.cn

对于增强现实系统来说,最重要的是准确地获取摄像机的方位参数,这样才能把虚拟信息和现实世界对齐,保证沉浸感。基于硬件感应器和基于视觉获得摄像机参数是常用的两个途径。基于硬件的方法通常成本高,装配复杂,但是比较稳定。反之,基于视觉的方法充分利用计算机视觉领域的成果,不需要额外的特殊装备,但是相对鲁棒性较弱。基于视觉的方法又包括自然特征和人工标志两类。对于本文的汉字学习系统而言,需要结合说明性的汉字图像,基于标志的视觉方法更为合适。

基于标志的方法^[3]特点是速度快而稳定,ARToolKit 和 ARTag 是其中应用最广泛的系统^[4-5],他们都是使用黑色方框标志。ARToolKit 的问题在于其二值化图像方法在复杂光影下很容易遗漏标志,而且摄像机求解不够稳定。ARTag 利用边缘提取和内建编码扩展了 ARToolKit 的标志检测方法,但是他们使用的标志是一些分散的方块,本身不具备用户可以感知到的信息。本文结合了两者的优势,提出一种新的标志系统,标志本身具有充分的信息,又能更稳定地被检测。

在系统中,标志是包围在黑色方框中的汉字,标志大小已知,每一个标志拥有独立的摄像机参数。由于标志作为求解输入的特征信息很有限,求解的结果充斥噪声和错误。利用标志的 4 个角点,优化投影误差来确定摄像机参数。这是一个非线性优化过程,为了得到优化结果,通常需要一个较好的初值。由于标志是一个平面,从单应性矩阵出发就可以得到一个满足要求的初值。最后,为了消除摄像机参数噪声导致的抖动^[6],利用加权的运动历史信息^[7],避免当前的解与之前的参数产生不合理的偏离,以保证参数的平滑。

系统的模块设计和流程如图 2 所示。

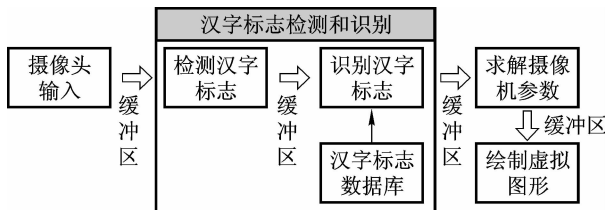


图 2 系统流程

Fig. 2 System flowchart

综上所述,本文完成 3 个方面的工作,(1)利用增强现实技术,实现了结合多媒体的汉字学习系统,

丰富了学习的趣味性。(2)优化了基于边缘提取的标志检测方法。(3)结合单应性矩阵,投影误差,和平滑性约束,求解的摄像机参数更加精确稳定。

2 相关工作

目前,有很多增强现实系统基于人工标志,因为相比自然特征,人工标志更容易跟踪和识别。Kato 和 Billingham 的 ARToolKit 是现今最常用的标志系统^[8],利用方框标志,他们实现了诸如网络会议,魔法书,画图板等应用。但是 ARToolKit 对光照变化的适应能力比较弱^[5],而且没有消除参数的抖动。ARTag 扩展了 ARToolKit,用更加鲁棒的边缘提取方法检测标志^[5],而且将标志的编码内嵌在标志图像里,不仅避免预处理的时间和存储,而且降低了识别的错误率。

还有一些特殊的标志,如 Ryuhei 等根据处理的背景物体设计对应的具有相似颜色的标志^[9],并将标志放置在物体的角落上,以降低其对应用的视觉影响。Keisuke 等人在大标志里嵌套小标志,使摄像机无论远近都能检测到适当的标志,增加了摄像机的活动范围^[10]。而 Shigeru 更是把标志设计成装饰壁纸,铺满房间的地面、墙面^[11],用户携带一个俯视的摄像机,可以在室内自由活动。

大多数系统都没有消除由于图像噪声而造成的参数抖动。Vacchetti 等人在当前帧的特征点和前一帧之间建立对应关系^[6],同时优化两帧的参数得到平滑的效果。Gordon 和 Lowe 则是直接利用前一帧的参数来平滑当前帧的参数^[7],但没有针对相机运动的旋转和平移分量的不同,对相应的平滑权重系数做自适应的调整。

3 摄像机模型

本文中的坐标为齐次形式,图像坐标表示为 $m = [u, v, 1]^T$,3 维坐标表示为 $M = [X, Y, Z, 1]^T$ 。我们遵循标准的针孔摄像机模型,3 维坐标在图像平面上的投影坐标如下计算,

$$m \sim PM = K[R | t]M \quad (1)$$

式中,~表示一个非零的尺度, R 和 t 摄像机的旋转和平移变换, P 是投影矩阵。 R 是 3×3 正交矩阵, t 是 3 维向量。 K 表示摄像机的内部参数,

$$K = \begin{pmatrix} f & s & c_x \\ 0 & \alpha f & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

其中, f 是焦距, α 是像素的纵宽比, (c_x, c_y) 是图像平面的中心, s 是倾斜系数。当前的工作假设 K 已知, 用现有的工具^[12] 确定。

4 汉字标志的检测和识别

本文提出的汉字标志与 ARToolkit 类似, 汉字置于黑色的包围框中, 而且其大小已知, 即其 3 维结构已知。假设标志所在的平面是 Z 平面, 如果其角点的图像坐标 $m = [u, v, 1]^T$, 则 3 维坐标 $M = [X, Y, Z, 1]^T = [u, v, 0, 1]^T$, 如图 3 所示。

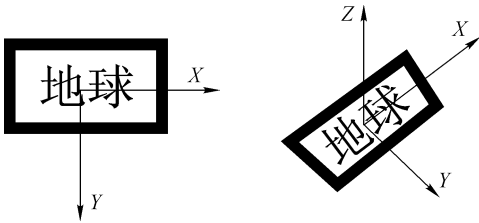


图 3 汉字标志图像及其默认的 3 维结构
Fig. 3 Chinese-character-marker image and default 3D structure

ARToolkit 将图像二值化来提取标志轮廓, 这样做的问题是很难设定一个统一的亮度阈值提取所有的轮廓。利用局部的对比度来确定轮廓是更加稳定的方法, 因此, 通过抽边来提取轮廓线。利用 Canny 算法^[13] 确定所有轮廓线上的像素之后 (如图 4(b) 所示), 需要拟合出包围框的 4 条直线。Hough 算法是拟合直线的通用方法, 但是由于离散采样, 不能完全得到精确的直线方程及其交点, 需要根据轮廓线上的点进一步拟合 (如图 4(c) 所示)。因此可使用类似多边形简化的更直接, 更高效的方法。具体步骤如下:

- (1) 确定轮廓上距离最远的两个顶点 v_0, v_2 , 然后分别确定到直线 v_0v_2 最远的 v_1, v_3 ;
- (2) 计算轮廓线上所有点到直线 $v_0v_1, v_1v_2, v_2v_3, v_3v_0$ 的距离 $\{d_i, i=0, 1, 2, 3\}$;
- (3) 将点分成 4 个集合 $\{c_i, i=0, 1, 2, 3\}$, 轮廓上任一点 $v \in c_i, t = \min_i \{d_i, i=0, 1, 2, 3\}$ (如图 4(b) 所示);
- (4) 分别对 $\{c_i, i=0, 1, 2, 3\}$ 中的点拟合直线, 如果得到的 4 条直线的残差都小于阈值, 实验中设为 1 个像素, 那么就认为找到一个矩形轮廓线。

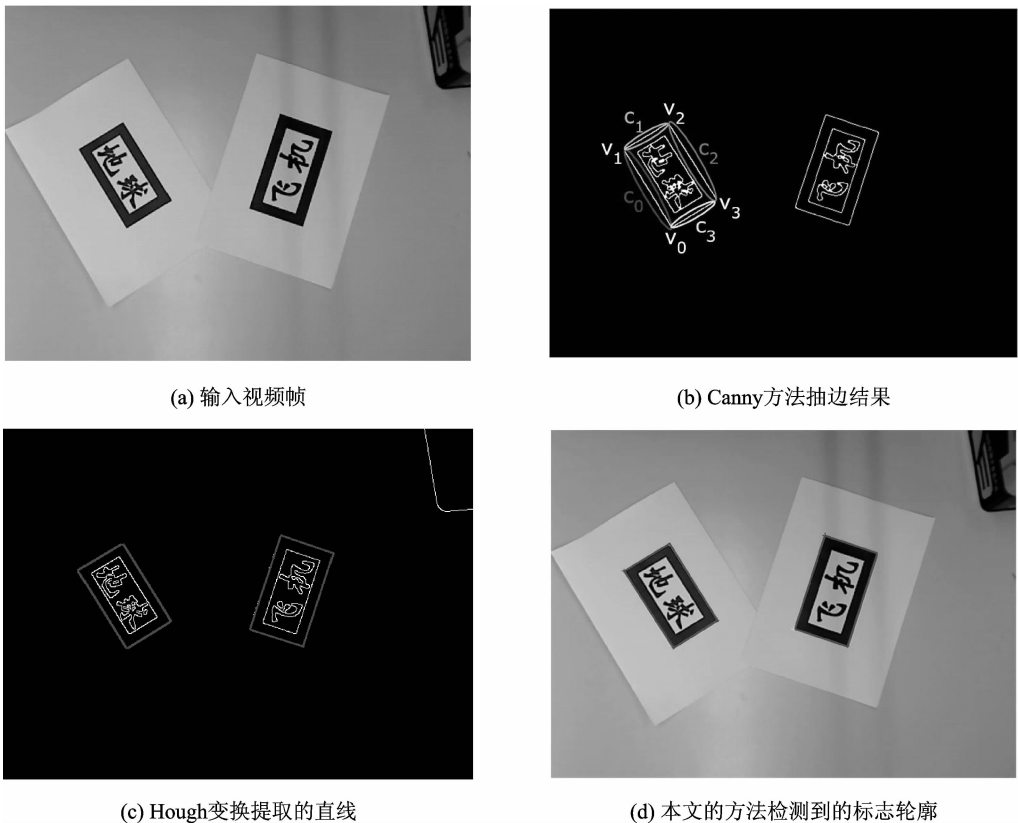


图 4 通过边缘提取来检测标志轮廓线
Fig. 4 Extract marker contours with edge detection

图 4(d) 是检测到的两个矩形轮廓线, 4 条直线的交点就是标志的 4 个角点。检测到标志区域之后, 通过模板匹配确定其包含的内容。在此之前, 利用标志的 4 个角点计算面与面之间的单应性矩阵, 将标志图像归一化到大小一致的图像, 再将其二值化。最后, 为了确定标志的旋转, 需要将归一化后的图像与 4 个方向的标志比较, 如图 5 所示。

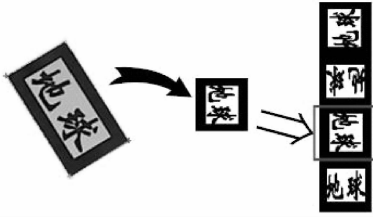


图 5 标志图像归一化及匹配

Fig. 5 Marker image normalization and match

5 摄像机参数的求解

系统中每一个标志有独立的坐标系, 可计算摄像机在每一个坐标系中的参数。如在图 1 中, 会获得两组摄像机参数。由于只有 4 个角点的信息, 而且处在同一个平面上, 优化很容易出现退化情况, 常常不能得到稳定的解。ARToolKit 通过标志包围框的 4 条边之间的关系求解摄像机参数, 可发现其对图像噪声导致的直线偏移并不足够鲁棒。因此选择从单应性矩阵推导出摄像机参数的初值, 并优化投影误差得到最终的参数。

5.1 摄像机参数的初始化

设投影矩阵 $\mathbf{P} = \mathbf{K}[\mathbf{R} | \mathbf{t}] = \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}]$, 根据式(1), Z 平面任意一点 $[X, Y, 0, 1]^T$ 在图像平面上的坐标

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (2)$$

从式(2)可以看出, $\mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}]$ 就是从 Z 平面到图像平面的一个单应性矩阵, 其中 \mathbf{K} 是摄像机的内参, \mathbf{r}_1 和 \mathbf{r}_2 分别是摄像机旋转矩阵的第 1 列和第 2 列, \mathbf{t} 是摄像机的平移向量。由于假定 \mathbf{K} 已知, 只要求得单应性矩阵 $\lambda \mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] = \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}]$, λ 是非零的尺度, 就可以通过下式算得摄像机的参数,

$\mathbf{r}_1 = \lambda \mathbf{K}^{-1} \mathbf{h}_1, \mathbf{r}_2 = \lambda \mathbf{K}^{-1} \mathbf{h}_2, \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, \mathbf{t} = \lambda \mathbf{K}^{-1} \mathbf{h}_3$, 式中, $\lambda = 1/\|\mathbf{K}^{-1} \mathbf{h}_1\| = 1/\|\mathbf{K}^{-1} \mathbf{h}_2\|$ 。于是, 只要求出标志的 4 个角点从 Z 平面到当前图像平面的单应性矩阵, 就可以估计出 \mathbf{R}, \mathbf{t} 的初值。

5.2 优化与去抖

通过单应性矩阵得到的摄像机外参只是一个估计值, 往往不够精确, 且很容易受噪声的影响。通过进一步优化, 抑制噪声带来的影响, 得到准确的外参, 从而使得合成的虚拟物体能够保持平稳而不发生漂移或抖动。首先是优化标志角点在图像平面上的投影误差,

$$\min_{\mathbf{R}^n, \mathbf{t}^n} \sum_i \|\mathbf{K}[\mathbf{R}^n | \mathbf{t}^n] \mathbf{M}_i - \mathbf{m}_i\|^2 \quad (3)$$

式中, n 表示第 n 帧图像, i 表示第 i 个角点。只优化目标函数(式(3))得到的摄像机参数, 比较容易受图像噪声的影响(因为图像噪声使得每一帧上提取的角点位置存在一定偏差)。即使当摄像机固定不动时, 求解得到的摄像机参数也会有明显的抖动现象, 所以需要进行进一步优化, 通过平滑性约束来消除图像噪声带来的影响。系统分别利用第 $n-1$ 帧获得的旋转和平移参数来平滑第 n 帧相应的参数, 特别是当摄像机的运动很缓慢时, 尽量使第 n 帧的解靠近第 $n-1$ 帧。因为摄像机的旋转运动和平移运动往往具有不一致的特点, 通常旋转的变化会更加剧烈, 所以旋转和平移的平滑项需要分别加权,

$$\min_{\mathbf{R}^n, \mathbf{t}^n} \left(\sum_i \|\mathbf{K}[\mathbf{R}^n | \mathbf{t}^n] \mathbf{M}_i - \mathbf{m}_i\|^2 + \alpha \|\mathbf{R}^n - \mathbf{R}^{n-1}\|^2 + \beta \|\mathbf{t}^n - \mathbf{t}^{n-1}\|^2 \right) \quad (4)$$

式中, α 和 β 分别用来控制旋转和平移运动的平滑度。首先, 根据式(3)计算 \mathbf{R}^n 和 \mathbf{t}^n , 这等价于式(4)中的 $\alpha = 0, \beta = 0$, 然后通过优化式(4)平滑参数, 每一次迭代, α 和 β 如下更新:

$$\alpha = \frac{e(\mathbf{R}_k^n, \mathbf{t}_k^n)}{\min\{e(\mathbf{R}^{n-1}, \mathbf{t}^{n-1}), e(\mathbf{R}_k^{n-1}, \mathbf{t}_k^{n-1})\}} \cdot \frac{e(\mathbf{R}_k^n, \mathbf{t}_k^n)}{\|\mathbf{R}_k^n - \mathbf{R}_k^{n-1}\|^2}$$

$$\beta = \frac{e(\mathbf{R}_k^n, \mathbf{t}_k^n)}{\min\{e(\mathbf{R}^{n-1}, \mathbf{t}^{n-1}), e(\mathbf{R}_k^{n-1}, \mathbf{t}_k^{n-1})\}} \cdot \frac{e(\mathbf{R}_k^n, \mathbf{t}_k^n)}{\|\mathbf{R}_k^n - \mathbf{R}_k^{n-1}\|^2}$$

其中, $e(\mathbf{R}, \mathbf{t}) = \sum_i \|\mathbf{K}[\mathbf{R} | \mathbf{t}] \mathbf{M}_i - \mathbf{m}_i\|^2$ 为投影误差计算公式。

根据摄像机的运动状态, 旋转和平移的平滑系数会做出自适应的调整: 保持上一帧的平移参数不变, 将当前帧得到的初始解的旋转参数代入上一帧的结果中, 如果得到的投影误差大, 则说明旋转参数变化大, 则减少旋转参数的平滑系数。反之, 则增加

平滑系数,以尽可能靠近上一帧的解,保持平滑性。旋转参数和平移参数的变化是独立的,因此系数分开计算,计算方法相同。 k 代表迭代的次数。在实验中,2~3次迭代已经足够。以上所有的非线性优化都使用 Levenberg-Marquardt 算法。

6 实验结果

本文使用的摄像头是罗技快看迷你高清版,Directshow 开发包实时捕获 640×480 图像。CPU 是酷睿 2 双核处理器 T7100,主频 1.8 G。绘制图形采用 OpenGL,因为 OpenGL 支持针孔摄像机投影模型,摄像机的内部参数用来设置 OpenGL 的投影矩阵,方位参数设置 OpenGL 的视图变换矩阵。只有将 OpenGL 这两个矩阵设置成与摄像机的现实参数一致,虚拟物体才可以和汉字标志配准。系统运行频率是 30 帧每秒,完全满足实时要求。图 6 显示部分帧的增强现实结果,第 1 行是摄像头输入的第 300,700 帧图像及检测到的汉字标志,第 2 行是增强现实的结果。系统支持图像、视频和 3 维模型等信息。



图 6 增强现实结果
Fig. 6 Augmented result

图 7 展示了标志检测方法在复杂光影变化时,仍然能在边缘提取的基础上稳定地检测到汉字标志。

图 8 和图 9 显示平滑约束使得摄像机参数更加平滑稳定。从图 8 可以看出,平滑参数对平移参数的影响很小,这是由于摄像机参数的抖动往往是由旋转引起的,从图 9 可以看出,在 X 和 Y 旋转方向

上,平滑约束有效地抑制了噪声。



图 7 基于边缘提取的标志检测方法
Fig. 7 Marker detection method with edge detection

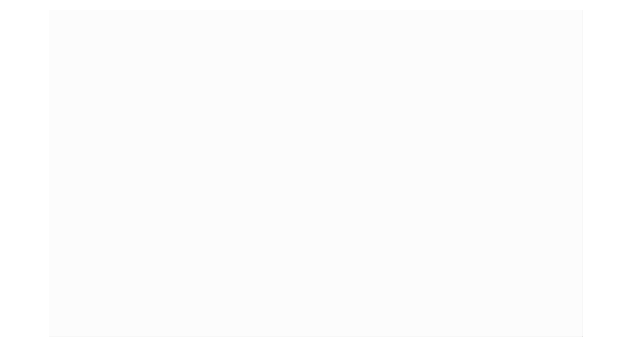


图 8 平滑约束对摄像机平移参数的影响
Fig. 8 Influence of the smoothness constraint on camera translation parameters



图 9 平滑约束对摄像机旋转参数的影响
Fig. 9 Influence of the smoothness constraint on camera rotation parameters

7 结 论

本文提出了一个基于汉字标志的增强现实系统。与之前的一些基于标志的增强现实系统相比,本系统将汉字学习与计算机生成的多媒体可视化信息结合起来,丰富了汉字学习的手段和乐趣。本文提出的边缘提取相比于传统的基于图像二值化的检测方法,对光照变化具有较好的容忍性,能够适应实际应用中复杂的光影变化。在摄像机参数的求解上,利用单应性矩阵可以快速求解出好的初值,从而有效避免了直接的非线性优化极易陷入局部最优解的问题。并且,摄像机参数求解加入了平滑性约束,并根据旋转和平移运动的不同特点,提出了一套自适应的权重系数调整,不仅有效抑制了图像噪声对求解的影响,而且改善了求解的精度和稳定性。实验结果显示,本系统能准确快速地识别出汉字标志,恢复的摄像机参数具有很高的精确性,证明了该算法的稳定高效,能够满足增强现实的应用需求。

参考文献 (References)

- 1 Ronald T. Azuma. A survey of augmented reality [J]. Presence: Teleoperators and Virtual Environments, 1997, 6(4): 355-385.
- 2 Ronald Azuma, Yohan Baillet, Reinhold Behringer, *et al.* Recent advances in augmented reality [J]. IEEE Computer Graphics and Applications, 2001, 21(6): 34-47.
- 3 Zhang Xiang, Stephan Franz, Nassir Navab. Visual marker detection and decoding in AR systems: a comparative study [A]. In: Proceedings of the First IEEE and ACM International Symposium on Mixed and Augmented Reality [C], Darmstadt, Germany, 2002: 97-106.
- 4 Artoolkit [EB/OL]. <http://www.hitl.washington.edu/artoolkit/>.
- 5 Mark Fiala. ARTag, a fiducial marker system using digital techniques [A]. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Washington, DC, USA, 2005: 590-596.
- 6 Luca Vacchetti, Vincent Lepetit, Pascal Fua. Stable real-time 3D tracking using online and offline information [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(10): 1385-1391.
- 7 Iryna Gordon, David G. Lowe. Scene modelling, recognition and tracking with invariant image features [A]. In: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality [C], Arlington, Virginia, 2004: 110-119.
- 8 Hirokazu Kato, Mark Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system [A]. In: Proceedings of the Second IEEE and ACM International Workshop on Augmented Reality [C], Washington, DC, USA, 1999: 85-94.
- 9 Ryuhei Tenmoku, Shinsuke Kaigawa, Fumihisa Shibata, *et al.* Visually elegant and robust semi-fiducials for geometric registration in mixed reality [A]. In: Proceedings of the Sixth International Symposium on Mixed and Augmented Reality [C], Nara, Japan, 2007: 13-16.
- 10 Keisuke Tateno, Itaru Kitahara, Yuichi Ohta. A nested marker for augmented reality [A]. In: Proceedings of the IEEE Virtual Reality Conference [C], Charlotte, North Carolina, USA, 2007: 259-262.
- 11 Shigeru Saito, Atsushi Hiyama, Tomohiro Tanikawa, *et al.* Indoor marker-based localization using coded seamless pattern for interior decoration [A]. In Proceedings of the IEEE Virtual Reality Conference [C], Charlotte, North Carolina, USA, 2007: 67-74.
- 12 Camera Calibration Toolbox for Matlab [EB/OL]. http://www.vision.ca.ltech.edu/bouguetj/calib_doc/.
- 13 John Canny. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.