

增强现实中的虚实注册技术研究

明德烈^{1,2)} 柳健²⁾ 田金文²⁾

¹⁾(北京大学视觉与听觉信息处理国家重点实验室, 北京 100080)

²⁾(华中科技大学图象识别与人工智能研究所, 图象信息处理与智能控制教育部开放研究实验室, 武汉 430074)

摘要 增强现实是把计算机产生的虚拟物体或其他信息合成到用户看到的真实世界中的一种技术, 实现虚拟信息与真实景象的正确匹配或有机结合是每个增强现实实用系统都必须完成的一项重要任务, 且它的具体实现需依赖于计算机硬件和软件系统的相互配合及协调。为了有效地降低增强现实应用系统中的虚实注册误差, 在对基于位置跟踪和相机定标技术的传统虚实注册技术进行分析的基础上, 首先研究了仿射变换技术理论在增强现实中的应用, 并提出了在全局仿射坐标系下, 获取像素相对深度参数的一种策略; 然后在对 AR 中的虚实注册问题进行研究分析的基础上, 进一步讨论了图象分析技术在虚实注册问题中的应用问题, 并提出了一种基于数值背景表达的快速虚实注册策略, 以提高虚实注册过程中, 场景特征点提取的精度和速度; 最后还分别用该仿射理论进行了动态和静态的虚实注册实验, 实验结果表明, 该方法能有效地降低增强现实应用系统的虚实注册误差。

关键词 计算机仿真(520·6020) 增强现实 虚拟现实 虚实注册 仿射变换 相机定标

中图分类号: TP391.9 **文献标识码**: A **文章编号**: 1006-8961(2003)05-0557-05

A Study on the Registration Technology in Augmented Reality

MING De-lie^{1,2)}, LIU Jian²⁾, TIAN Jin-wen²⁾

¹⁾(National Laboratory on Machine Perception, Peking University, Beijing 100871)

²⁾(State Key Laboratory for Image Processing and Intelligent Control, Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, 430074)

Abstract Augmented reality is the merging of synthetic sensory information into a user's perception of a real environment. One of the most important tasks for every augmented reality application is to make the virtual scene and real scene merged correctly and become an organic whole. The implementation lies heavily on the cooperation between augmented reality hardware and software technology. Based on the analysis of traditional virtual-real registration technology which lies heavily on position tracking and camera calibration, we investigate the application of affine representations in augmented reality, and put forward a new method to get the relative depth of each rendered point. By studying the application of image analysis technology in AR technology, a numerical background expression method is put forward to facilitate the feature point extraction in the virtual-real registration process. We have performed both dynamic and static virtual-real registration experimentation using affine theory. The experiment results indicate that this method can effectively reduce virtual-real registration errors in applications of augmented reality.

Keywords Augmented reality (AR), Virtual reality (VR), Registration between virtual object and real scene, Affine representations, Camera calibration

0 引言

增强现实(Augmented Reality, 简称 AR)是一

个新的研究领域, 它是一种通过将计算机产生的图形、文字注释等虚拟信息有机地融合到使用者所看到的真实世界景象中, 用以对人的视觉系统进行景象增强或扩张的技术。Azuma 是这样定义增强现实

基金项目: 北京大学视觉与听觉信息处理国家重点实验室开放课题基金(2002-05)

收稿日期: 2002-06-07; 改回日期: 2002-09-26

的:虚实结合,实时交互,三维注册^[1].

增强现实系统必须能够实时地检测出观测者、相机相对于真实场景的位置和方向角,以及成像系统(摄像机)的内部参数(焦距和像元高宽比等),以便能够根据这些信息来实时确定所要添加的虚拟信息在真实空间坐标中的映射位置,并将这些信息实时显示在图象的正确位置,这就是AR中的虚实注册(Registration)所要完成的任务^[2].

无论是静态,还是动态的AR应用,都存在虚实注册问题.在静态AR应用中,放置在真实场景中的虚拟物体,在用户看来,无论从任何角度观察,都应该保持三维欧氏空间的几何一致性,除非用户或其他物体与它发生交互作用,其任何静态注册错误都将直接导致从不同角度观察到的虚拟物体,在真实场景中产生相对位置的漂移,并将使得后续动态虚实影像的注册更加困难;在动态AR应用中,则要求视频图象的刷新率起码要达到15Hz,理想的情况要达到30Hz,还要尽量减少系统的计算延时,因为如果虚拟物体的绘制速度跟不上用户的运动速度,那么不仅将直接导致人体视觉系统与运动系统的不匹配,还将给用户造成一种虚拟物体在三维空间漂浮的感觉.此外,虚实注册问题还与AR系统的具体实施方案紧密相关.

目前大多数的AR应用都是从虚实注册问题的自然属性出发,用一种直接的方式来解决AR中虚实影像的匹配问题.对于这些AR应用系统,虚实注册问题解决的关键就是要明确各坐标系之间的转换关系,因为只有明确了各坐标系之间的转换关系,才能将虚拟物体正确地融合到真实场景中,以生成增强场景图象.其实现过程大致可以分为位置跟踪(真实场景至摄像机)、相机定标(三维场景至二维成像平面)、测量定位(虚拟物体至真实场景)等几个步骤.

从虚实注册自身实现原理出发,上述传统的虚实注册方法必然存在其局限性,如位置跟踪和相机定标过程本身就极易出错,这都将直接导致虚实注册错误的发生^[3].

1 仿射变换在AR中的应用

在以上讨论的传统虚实注册问题解决方法中,人们都将虚实注册问题的解决归结在各坐标系之间变换方程的求解上,但却忽略了其他的一些可能解

决途径,其中很重要的一个方面就是对计算机图形学和计算机视觉技术的利用.随着近几年来,计算机图形学和计算机视觉技术的迅速发展,其中许多成熟的技术可用于解决AR中的虚实注册问题.

虚实影像正确注册的关键是要精确确定虚拟物体上的每一点在场景图象上的投影坐标.由上节讨论可知,在齐次坐标表示下,三维空间中,虚拟物体上的某一点 $[x_o, y_o, z_o, w_o]^T$ (下角o代表object)与其在增强场景平面上的投影点 $[u, v, h]^T$ 之间的转换关系,可以用下式表示^[4]

$$\begin{bmatrix} u \\ v \\ h \end{bmatrix} = P_{3 \times 4} C_{4 \times 4} O_{4 \times 4} \begin{bmatrix} x_o \\ y_o \\ z_o \\ w_o \end{bmatrix} \quad (1)$$

其中, $P_{3 \times 4}$ 、 $C_{4 \times 4}$ 和 $O_{4 \times 4}$ 分别对应于三维场景至二维成像平面、真实场景至摄像机和虚拟物体至真实场景的变换.由于这时虚拟物体、真实场景和相机被定义在各自不同的坐标系下,因此虚实注册问题的解决就依赖于上述坐标系之间转换关系的求解.这已经被证明是一个烦琐,且极易发生错误的过程.

1991年,Koenderink和Ullman等人得到这样一个结论:如果给定三维空间点集合中的至少4个非共面点,那么对于任何一个投影变换,三维空间点集合中的任意一个点的变换结果都可以用这4个点变换结果的一个线性组合来表示^[5].基于这个结论,不禁想到是否可以通过将物体坐标系、相机坐标系和场景坐标系合并,即建立一个全局仿射坐标系(非欧几里得坐标系)来将真实场景、相机和虚拟物体定义在同一坐标系下,以绕开不同坐标系之间转换关系的求解问题,从而使得虚实注册问题的解决不再依赖于相机内、外部参数的定标.这样在新的坐标系表示下,式(1)中的投影变换表达式就变为

$$\begin{bmatrix} u \\ v \\ h \end{bmatrix} = P_{3 \times 4} \begin{bmatrix} x_f \\ y_f \\ z_f \\ w_f \end{bmatrix} \quad (2)$$

其中, $[x_f, y_f, z_f, w_f]^T$ (下角f代表仿射坐标)为在新坐标系下虚拟物体的点坐标表示.这样,在定义了原点和3个坐标轴后,全局仿射坐标系就可以被确定,这时在增强场景中,虚实物体上的每一点都有各自固定的仿射坐标表示.虽然用来定义全局仿射坐标系的原点和坐标轴的4个点可以是3D空间中任意4个非共面的点,但在实际应用中,通常选用人工设

定的特征点,这样做是为了易于计算机图象特征的提取.这时,Koenderink 和 Ullman 等人的理论可以用图 1 来表达^[6].

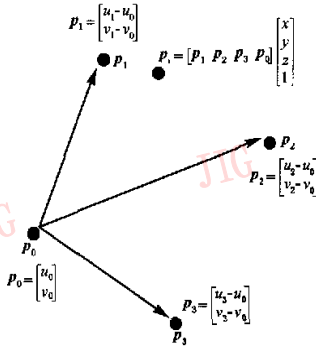


图 1 仿射坐标系中的点坐标关系示意图

其中, $p_i (i=0\sim 3)$ 为经过仿射坐标系下某一变换 $\Pi_{3\times 4}$ 后, 4 个基准点在二维投影平面上的对应投影点, p_i 表示仿射坐标系中的任一点经变换 $\Pi_{3\times 4}$ 后在二维平面上的投影点, $[x, y, z, 1]^T$ 为 p_i 的仿射坐标. 从图 1 中可以看出, 对于仿射坐标系下的某一变换 $\Pi_{3\times 4}$, 当通过图象分析, 已经确定 4 个基准点在二维投影平面上的坐标 $([u_i, v_i]^T, i=0\sim 3)$ 后, 则 3D 空间中任一点 $p [x_i, y_i, z_i, 1]^T$ 的二维投影坐标 $[u_p, v_p]^T$ 可以通过下式计算得到.

$$\begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} u_{p_1} - u_{p_0} & u_{p_2} - u_{p_0} & u_{p_3} - u_{p_0} & u_{p_0} \\ v_{p_1} - v_{p_0} & v_{p_2} - v_{p_0} & v_{p_3} - v_{p_0} & v_{p_0} \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\Pi_{3\times 4}} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (3)$$

从式(3)中可以看出, 只要确定了仿射坐标系下 4 个基准点在二维投影平面上的坐标, 就可以唯一确定变换矩阵 $\Pi_{3\times 4}$. 这时对于 3D 空间中的任一点, 就可以根据它的仿射坐标计算出它在二维投影平面上的对应平面坐标.

由此可见, 在全局仿射坐标系下, AR 虚实注册问题的解决就由烦琐的定标过程转化为二维投影平面上基准点的定位过程. 由于基准点通常易于计算机辨识的特征点加以标示, 所以后者的实现难度大大降低. 前面也多次提到, 为了获得 3D 空间中任一点的二维投影坐标, 很重要的一个前提就是要知道它的仿射坐标. 为了获取 3D 空间中任一点 P 的仿射坐标, 可以采用类似于双目立体视觉的方法. 当然, 这里不必受光轴平行等限制, 也不必知道相机焦

距和立视基线长度, 只需从任意不同的角度获取场景的两个二维投影平面, 再分别确定出 4 个基准点在两个投影平面上的坐标 $[u_i^{(j)}, v_i^{(j)}]^T (i=0\sim 3, j=1\sim 2)$ 以及点 P 在两个投影平面上的对应坐标 $[u_p^{(j)}, v_p^{(j)}]^T (j=1\sim 2)$ 即可. 这时点 P 的齐次仿射坐标 $[x_i, y_i, z_i, 1]^T$ 可以通过下式求得.

$$\begin{bmatrix} u_p^{(1)} \\ v_p^{(1)} \\ u_p^{(2)} \\ v_p^{(2)} \end{bmatrix} = \begin{bmatrix} u_{p_1}^{(1)} - u_{p_0}^{(1)} & u_{p_2}^{(1)} - u_{p_0}^{(1)} & u_{p_3}^{(1)} - u_{p_0}^{(1)} & u_{p_0}^{(1)} \\ v_{p_1}^{(1)} - v_{p_0}^{(1)} & v_{p_2}^{(1)} - v_{p_0}^{(1)} & v_{p_3}^{(1)} - v_{p_0}^{(1)} & v_{p_0}^{(1)} \\ u_{p_1}^{(2)} - u_{p_0}^{(2)} & u_{p_2}^{(2)} - u_{p_0}^{(2)} & u_{p_3}^{(2)} - u_{p_0}^{(2)} & u_{p_0}^{(2)} \\ v_{p_1}^{(2)} - v_{p_0}^{(2)} & v_{p_2}^{(2)} - v_{p_0}^{(2)} & v_{p_3}^{(2)} - v_{p_0}^{(2)} & v_{p_0}^{(2)} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (4)$$

式中, 下角 i 代表仿射坐标系.

2 全局仿射坐标系下投影点深度数据的计算

为了给用户足够的真实感, 必须保持增强场景中, 虚实物体间正确的遮挡关系. 这不仅需要确定 3D 空间中的任一点在投影平面上的二维坐标, 还需要知道每一投影点的深度信息, 即 3D 空间点相对于视点的距离, 以便通过计算机图形系统硬件的支持来实现场景中物体隐藏面的自动消隐.

在全局仿射坐标系下, 由于没有进行相机定标过程, 因此不可能知道具体的视点位置, 也就不可能恢复绝对意义上的投影点的深度数据. 这时, 可先通过 4 个基准点在投影平面上的坐标来定义一个三维平面 L , 然后计算仿射坐标系下每一空间点相对于平面 L 的距离, 并将其作为该点在投影平面上深度数据的一个近似. 实验证明, 由该近似得到的相对距离值可以正确地反映增强场景中, 虚实物体间的深度关系. 其中, 平面 L 定义如下(用平面法向量 φ 表示)

$$\varphi = \begin{bmatrix} u_{p_1} - u_{p_0} \\ u_{p_2} - u_{p_0} \\ u_{p_3} - u_{p_0} \end{bmatrix} \times \begin{bmatrix} v_{p_1} - v_{p_0} \\ v_{p_2} - v_{p_0} \\ v_{p_3} - v_{p_0} \end{bmatrix} \quad (5)$$

仿射坐标系下的任一空间点, 在投影平面上的深度数据的近似计算公式如下

$$\begin{bmatrix} u \\ v \\ h \\ 1 \end{bmatrix} = \begin{bmatrix} u_{p_1} - u_{p_0} & u_{p_2} - u_{p_0} & u_{p_3} - u_{p_0} & u_{p_0} \\ v_{p_1} - v_{p_0} & v_{p_2} - v_{p_0} & v_{p_3} - v_{p_0} & v_{p_0} \\ \varphi^T & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \quad (6)$$

公式中各参数的意义同上.

3 基于数值背景表达的基准点获取

前面曾经提到,由于只有定义了原点和 3 个坐标轴后,全局仿射坐标系才可以被唯一确定,因此,场景中用来定义原点和坐标轴的 4 个基准点的提取成为关键,并且需要实时地对场景中基准点进行定位,然后才能用这些基准点的定位信息来辅助虚实注册的全过程.这些基准点可能是人工在场景中设定的标志点,也可能就取场景中本身存在的结构特征点作为基准点.尽管用图象分析技术来从背景图象中提取基准点的方法很多,如基于彩色特征提取的方法、基于模版匹配的方法等等,且这些方法在一定条件下都能取得较好的效果,但在 AR 应用系统中,人们关心的不仅仅是基准点提取的精度,还必须考虑到基准点提取的速度,而且后者对于动态 AR 应用尤其显得重要,因为系统计算延迟是造成动态 AR 应用中虚实注册错误的一个主要原因.

本文以二值化后的梯度图象为对象,提出了一种进行背景数值表达的策略,以提高图象中基准点的提取速度.基于这种策略的基准点提取全过程可以用下面的流程图来表达:

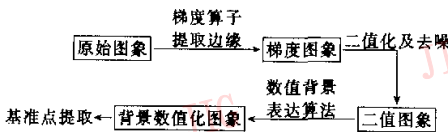


图 2 基于数值背景表达的基准点提取过程流程图

基准点提取时,首先用梯度算子(如:Roberts 算子、Prewitt 算子、Sobel 算子及形态学中的 Top-Hat 运算等)作用于原始图象来获得包含场景中物体轮廓信息的梯度图象;然后通过对该梯度图象进行二值化及去噪处理来获得场景中反映物体轮廓信息的二值图象.下面就是在这些轮廓信息中搜索基准点和其他一些可以用来辅助定位的轮廓特征点.这里,本文提出了一种背景数值表达的方法,其基本思想是通过二值图象中的每个像素按照一定规律进行编码,用来反映每个像素所处的背景信息.这种基于对像素所在的图象背景信息的数值描述以及对基准点所在位置的先验知识,可以利用句法模式识别的方法来获取基准点的准确位置.

数值背景表达就是对图象的背景点(非轮廓点)进行数值变换,以取得场景背景区域的结构特征.二

值化后的轮廓图象中,轮廓像素点的初始特征值为 1,背景点的初始特征值为 0,数值变换就是对图象进行扫描,同时对每一个背景点 P 进行变换.具体做法是从 P 点出发,向上、下、左、右 4 个方向引射线,如图 3 所示,若在方向 3 上遇到遮挡,则特征值加 1,若在方向 1 上遇到遮挡,则特征值加 2;若在方向 2 上遇到遮挡,则特征值加 4;若在方向 4 上遇到遮挡,则特征值加 8,如此形成背景点的特征值.

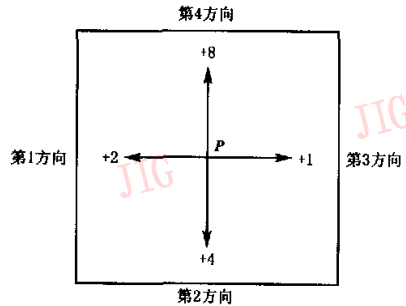


图 3 背景数值表达算法示意图

背景点经过这种变换后共形成 16 种特征值,这 16 种特征值分别代表该像素所处的不同背景环境,如:全封闭、上下左右均无遮挡、左边有遮挡等等.图 4 是一幅经过数值背景表达后的结果示意图.由于这种不同的背景结构特征具有显著差别,因此也是快速正确地场景轮廓中提取基准点和结构特征点的重要依据.在基准点提取过程中,通常要抽取典型区域的特征,经常用到的典型区域(见图 4)有 7 区(区域形状向上开口)、B 区(向下开口)、D 区(向左开口)、E 区(向右开口)和 F 区(全封闭区域),9 区(左边和下边开口)有时也会用到;经常用到的特征有:某种区域的个数及区域的大小、区域的顶部、区域的底部、区域的左边缘及右边缘以及区域最宽处的行坐标.这些特征清楚地表明了像素所在的背景结构特征,也给基准点的正确和快速提取带来明显的好处.



图 4 背景数值表达结果示意图

这些基准点的三维空间坐标被认为是已知的,它们可以取自场景中的自然特征(如告示牌、红绿灯

或其他醒目的特定场景特征),也可以是人工在场景中设置的标记点(如高饱和度、高亮度的彩色标记物体等),但无论是自然场景特征,还是人工添加的标记点,都应该以易于系统提取和识别为原则。因此,本文所选取的基准点与场景中的背景应该有很明显的特征差别,并很容易用本文所提出的数值背景表达方法,将基准点从背景区域中提取出来。

4 实验结果与分析

为了验证本文方法的效果,在实验室环境中,分别用本文提出的方法进行了动态和静态的虚实注册实验,实验结果如图5所示。

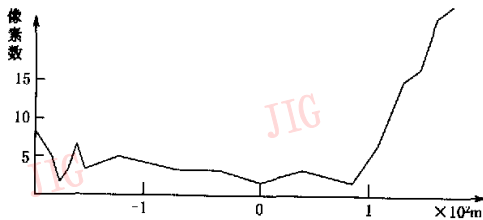


图5 虚实注册误差曲线

从图5中可以看出,在 512×512 大小的图象平面上,虚实注册的误差可以基本保持在15个像素的范围内,但随着相机相对于虚拟物体距离的增加,注册误差也会随之上升,因此,为了保持一定的虚实注册精度,应该将相机与虚拟物体的距离控制在一定范围之内。

本文探讨了图形学和计算机视觉中的仿射变换理论在AR中的应用问题,即通过引入全局仿射坐标系的定义,将虚拟物体、真实场景和相机置于同一个坐标系统下,从而绕开了传统虚实注册方法中的定标过程。虽然系统实现中,又带来一些新的要求和限制(例如,作为仿射坐标系基准的图象特征点的实时跟踪问题、深度数据的近似估计问题和虚拟物体在欧几里得坐标系与仿射坐标系之间的坐标转换问题等),但相对于传统方法中的烦琐定标问题而言,已经使得AR中的虚实注册问题大为简化。在对AR中的虚实注册问题进行研究分析的基础上,又进一步探讨了图象分析技术在虚实注册问题中应用,并提出了一种基于数值背景的表达策略,用以辅助场景图象中基准点的判别和提取。这种用数值来表征图象背景特征的方法还可以应用在文字识别、

工程图纸识别、机器人视觉等许多其他的研究领域。

参考文献

- 1 Azuma RT. A survey of augmented reality[J]. Tele-operators and Virtual Environments, 1997, 6(4): 355~385.
- 2 Kurulakos K N, Vallino J R. Calibration-free augmented reality [J]. IEEE Transactions on Visualization and Computer Graphics, 1998, 4(1): 1~20.
- 3 Durlach N I, Mavor A S. Virtual reality, Scientific and technological challengers[M]. Washington, DC USA: National Academy Press, 1995:234~256.
- 4 Ullman S, Basri R. Recognition by linear combinations of models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(10): 992~1006.
- 5 Koenderink J J, Van Doorn J J. Affine structure from motion [J]. Journal of the Optical Society of America A, 1991, 8(2): 377~385.
- 6 Barrett E B, Brill M H, Payton P M. Invariant linear methods in photogrammetry and model-matching [A]. In: Mundy J I., Zisserman A. Editors: Geometric Invariance in Computer Vision [M]. Cambridge Massachusetts, USA: The MIT Press, 1992: 277~292.



明德烈 1974年生,1999年获武汉理工大学计算机科学系计算机及应用硕士学位,2002年获华中科技大学图象识别与人工智能研究所模式识别与智能系统专业博士学位,现在华中科技大学电信系博士后流动站工作。研究方向为图象处理、模式识别、虚拟现实及增强现实。



柳健 1939年生,教授,博士生导师,现在华中科技大学电子与信息工程系,华中科技大学图象识别与人工智能研究所,图象信息处理与智能控制教育部重点实验室从事遥感图象处理、计算机视觉等研究。



田金文 1961年生,教授,现在华中科技大学图象识别与人工智能研究所,图象信息处理与智能控制教育部重点实验室从事小波理论、图象压缩、计算机视觉等研究。