

## 基于三焦点张量点转移的多摄像机协同\*

梁华<sup>1+</sup>, 刘云辉<sup>2</sup>, 蔡宣平<sup>1</sup>

<sup>1</sup>(国防科学技术大学 电子科学与工程学院, 湖南 长沙 410073)

<sup>2</sup>(香港中文大学 机械自动化工程学系, 香港)

### Multi-Camera Collaboration Based on Trifocal Tensor Transfer

LIANG Hua<sup>1+</sup>, LIU Yun-Hui<sup>2</sup>, CAI Xuan-Ping<sup>1</sup>

<sup>1</sup>(School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

<sup>2</sup>(Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China)

+ Corresponding author: E-mail: lh2078@163.com

Liang H, Liu YH, Cai XP. Multi-Camera collaboration based on trifocal tensor transfer. *Journal of Software*, 2009,20(9):2597-2606. <http://www.jos.org.cn/1000-9825/3571.htm>

**Abstract:** This paper presents a head detection and trifocal tensor pointer transfer based multi-camera collaboration. In this approach, people's head position is detected after background subtraction and tracked by Kalman and PDA. Trifocal tensor transfer is used to locate objects in the virtual top view by the corresponding head points in two camera views. Compared with existing approach, the approach doesn't need camera calibration and coplanar precondition, nor does the deterioration of epipolar transfer exist. The experimental comparison between this approach and other traditional ones proves its effectiveness and accuracy.

**Key words:** multi-camera collaboration; head detection; epipolar transformation; trifocal tensor; point transfer

**摘要:** 针对复杂遮挡环境下多摄像机协同的问题,提出一种基于三焦点张量点转移的多摄像机协同目标定位方法.该方法利用单摄像机视图中头部检测的结果进行对极匹配,然后利用虚拟顶视图和摄像机视图之间的三焦点张量关系进行点转移,确定行人在虚拟顶视图中的位置.该方法的优势在于:不需要标定摄像机参数,不需要假设物体在共同平面上运动,不存在利用对极关系进行点转移时出现的交点退化的问题.与传统方法的实验结果对比表明了该方法的有效性和准确性.

**关键词:** 多摄像机协同;头部检测;对极变换;三焦点张量;点转移

中图法分类号: TP391 文献标识码: A

多摄像机协同也称摄像机的传递.与单摄像机监控相比,多个摄像机协作可以扩大摄像机的观测范围、增加观测角度,在众多领域存在着广泛的应用前景,是近年来计算机视觉领域研究的热点之一<sup>[1]</sup>.多摄像机协同涉及的问题较多,根据摄像机视野(field of view)的覆盖关系可以分为视野有重叠(overlapping)<sup>[2-6]</sup>和视野无重叠(non-overlapping)<sup>[7-15]</sup>两类.在摄像机协同问题中可供利用的约束条件大致可以分为3类:(1)目标自身特征的一致性,即目标本身的特性在穿越多个摄像机视野时是不变的,如目标的颜色、形状等外观特征.但由于不同摄

\* Supported by the National Natural Science Foundation of China under Grant No.60675056 (国家自然科学基金)

Received 2008-09-15; Accepted 2009-01-20

像机的视角、场景光照的不同,这类特征在不同的摄像机视图中存在着较大的差异,寻找这些差异中的规律性,如不同视图间亮度转移关系<sup>[10]</sup>,仍是一个较难的问题.这类特征在目标匹配问题中应用较多,但准确度不高,算法耗时随着特征维数呈级数增加.(2) 摄影几何约束,如空间目标的 3D 方位、对极约束、共面约束等,这类约束比较严格,有良好的数学基础,通过对单个摄像机的标定或多个摄像机间关系的确定,能够准确地定位目标的空间位置<sup>[2-5]</sup>,这类约束多应用于视野有重叠的问题.(3) 另外,还有一类比较松的约束条件,如目标在不同摄像机之间运动的路径拓扑结构,目标进入或离开视图大致位置的分布,以及目标穿越摄像机视野的大致时间分布等,这类约束随机性强,一般需要通过大量观测数据进行统计学习来确定<sup>[7-9,11-15]</sup>,多用于无法获得严格摄影几何约束的视野无重叠问题.

基于射影几何约束的方法具有严格的数学基础,速度快、精度高,得到了广泛的研究和应用.这类方法根据代数几何学原理的不同大致可以归纳为以下几类:(1) 标定摄像机,利用摄像机矩阵,对观测点反向投影,利用最小二乘直接重构获得点的 3D 坐标,如线性三角形法(见附录(III)).(2) 假设目标在一个共同的平面内运动,在此假设下,目标点与摄像机视图中的观测点存在着单应(homography)的关系,而此单应矩阵可根据至少 4 组对应的点求出.利用此单应可将摄像机视图中的点或线映射至地平面或虚拟顶视图中获得目标的 2D 坐标<sup>[3-5]</sup>.(3) 对极转移的方法,利用两个视图中至少 7 组对应点可以计算出基础矩阵(fundamental matrix),根据基础矩阵可以计算出摄像机视图中的观测点在虚拟顶视图中的对极线,通过两个观测点的对极线的交点可以确定目标在虚拟顶视图中的 2D 坐标<sup>[2]</sup>.

在实际应用中直接射影重构的方法虽然可以准确地计算目标的 3D 位置,但由于需要对摄像机进行标定,特别是当摄像机观测范围较大、距离目标较远时,需要较大的标定物,标定过程比较繁琐,所以多局限于精确测量领域.假设目标在共同平面内运动,如人在地面上行走,可以近似地认为观测到的人的着地点在共同的地平面内,近似符合多数视频监控环境,而且单应关系计算简单,所以应用较广.但是在比较复杂的观测环境中,如目标不在一个平面内运动,或者由于存在较多的遮挡物,无法观测到人的着地点时,这种方法就不再适用.利用对极变换点转移的方法可以解决上述的不足,但是它也存在一个严重的缺陷使得它很难被实际计算采用,即当两条对极线在第 3 幅视图中接近重合时会产生对极线交点退化的问题,在本文中作详细讨论.

针对这种情况,本文提出一种基于三焦点张量点转移的多摄像机协同方法.该方法的步骤是:(1) 单摄像机视图中,在背景差<sup>[16]</sup>和阴影去除<sup>[17]</sup>的基础上检测行人头部位置并进行跟踪<sup>[18]</sup>;(2) 在摄像机视图间利用对极约束确定观测点间的匹配关系;(3) 利用摄像机视图与虚拟顶视图间的三交点张量关系进行点转移,确定目标在虚拟顶视图中的位置.

## 1 基于头部检测的对极匹配

利用不同摄像机视图中目标的观测值确定目标在统一坐标系下的位置,首先要解决的问题是如何确定不同摄像机视图中目标观测的对应关系,即确定哪些观测来自同一个目标.针对行人距离摄像机较远、身体部分存在严重遮挡的环境,我们以行人头部位置作为观测,利用两摄像机视图之间的对极关系建立匹配关系.

### 1.1 头部检测

头部检测常被用来确定行人在图像中的位置信息.在图像中进行头部检测的常见方法是建立头部特征模板,遍历图像与模板进行匹配,匹配较好的区域即符合头部特征的区域,但这类方法在图像遍历和模板匹配时所需时间较长,不适用于对速度要求比较高的视频应用.

视频中检测头部的的方法多在运动前景检测<sup>[16]</sup>结果的基础上进行.一种比较常见的方法是对运动前景区域进行垂直投影,通过聚类计算投影峰值的数量和位置.聚类的数量即头部的数量,类心即头部在横轴上的位置.头部的垂直位置可以通过在纵轴上搜索符合头部特征的区域获得.该方法的问题在于当两个运动区域水平距离较近时,会聚类成一个目标,很难将目标分离开.

本文提出一种基于肤色和发色特征点聚类的方法.首先对检测出的运动前景进行直方图均衡,然后提取其中符合头发和皮肤颜色特征的像素.发色一般比较暗(RGB 值均较小),经过直方图均衡后,发色的 RGB 值延展至

0 附近,便于设定检测阈值.肤色一般接近黄色,由于环境光线、摄像机成像等原因,造成饱和度和亮度不同而呈现不同的颜色,但色调一般保持稳定,所以在 HSV 空间中,通过 H 色调设定阈值可以较好地地区分.将提取的符合上述特征的点进行基于行程编码的连通区域标识,形成块,去掉面积较小的噪声块后,选择靠近整个运动目标上部的目标块作为头部区域.该方法的优点在于当行人重合时仍可以检测出头部位置.

### 1.2 对极匹配

根据对极几何,当两个摄像机观测的目标为同一目标时,目标在图像中所处的位置应该落在相应的对极线上.如图 1 所示,目标 1 在摄像机视图 1 中的观测值  $x$  和视图 2 中的观测值  $x'$ ,它们分别落在各自在另一幅视图图中对应的极线  $l_{21}$  和  $l_{12}$  上.由于观测误差,对应的观测点一般不会准确地落在极线上,但是一般情况下都在对极线的附近.理论上应该根据对极线的包络<sup>[19]</sup>进行寻找,但是由于极线包络计算的不稳定性和复杂性,一般较少采用.在本文中,我们假设观测点与对极线的距离满足正态分布  $N(0, \sigma)$ ,其中  $\sigma^2$  与目标的面积成正比(因为目标面积越大时,头部检测的误差越大).定义观测值间的匹配权值  $w$  为观测值  $x$  和  $x'$  源自同一个空间目标  $X$  的概率,即  $X$  的投影出现在  $x$  和  $x'$  时的概率,假设  $X$  出现在  $x$  和  $x'$  时的概率相互独立,则有:

$$w = P(x, x' / X) = P(x / X)P(x' / X) \tag{1}$$

$$w \approx P(x / l_{21})P(x' / l_{12}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{d}{\sigma}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma'} \exp\left(-\frac{1}{2}\left(\frac{d'}{\sigma'}\right)^2\right) \tag{2}$$

简化后得到:

$$w \propto \frac{1}{\sigma\sigma'} \exp\left(-\frac{1}{2}\left[\left(\frac{d}{\sigma}\right)^2 + \left(\frac{d'}{\sigma'}\right)^2\right]\right) \tag{3}$$

其中,  $d = |x^T l_{21}| / \|l_{21}\|$ ,  $d' = |x'^T l_{12}| / \|l_{12}\|$  为观测点到对极线的距离,  $x, x'$  和  $l_{21}, l_{12}$  分别为观测点和相应的对极线的列矢量,  $l_{12} = F_{12}x$ ,  $l_{21} = F_{12}^T x'$ .  $\sigma^2$  和  $\sigma'^2$  分别设置为前景运动目标的面积.从物理意义上讲,当观测点  $x, x'$  源自同一目标点时,观测点准确落在对应的极线上,  $w$  匹配权值为 1, 随着与极线距离的增加,这种可能性减小,  $w$  地相应减小并趋近于 0.

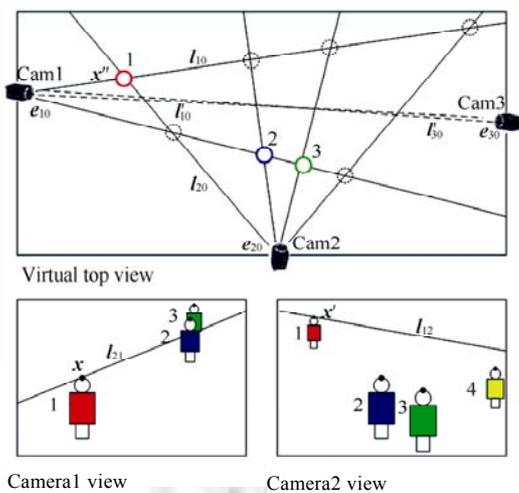


Fig.1 Epipolar view  
图 1 对极关系示意图

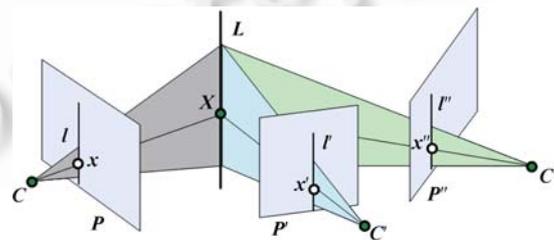


Fig.2 Trifocal view  
图 2 三视图关系示意图

当视图中存在多个观测点时,我们利用匹配权值  $w$  建立匹配矩阵  $W$ .如当视图 1 中有  $m$  个目标,视图 2 中有  $n$  个目标时,  $W$  为一个  $m \times n$  的矩阵.  $w_{ij}$  是观测值  $x_i$  和  $x'_j$  的匹配权值.这样,匹配的问题就转化为在  $W$  中,  $n!/(n-m)!$  (假设  $m < n$ ) 种匹配的组合中,寻找一组权值和最大的最优匹配组合.

在实际应用中,权值和最大的最优匹配组合常常极不稳定.如当  $c$  个目标距离较近时,会有  $c!$  个匹配组合的权值和比较接近.由于噪声的影响,最优匹配组合往往会出现错误.对于这种情况,我们采用了 Leave-One-Out 方法<sup>[7]</sup>剔除不稳定项.对于最优的匹配组合,我们逐一去掉组合中的一项,如果去掉该项后权值和发生较大的变化,则说明没有其他合理的匹配存在,就认为该项是稳定的,接受该项,否则反之.

**Table 1**  
**表 1**

	$x$	$y$	$z$
$a$	0.68	0.75	0.01
$b$	0.15	0.55	0.56
$c$	0.27	0.5	0.5

如表 1 所示,最优的匹配为  $\{a \leftrightarrow x, b \leftrightarrow z, c \leftrightarrow y\}$ , 权值和为 1.74. (注意到虽然对  $a$  而言  $a \leftrightarrow y$  是最优的匹配,但是整体而言并不是最优的). 如果去掉  $a \leftrightarrow x$ , 那么最优的组合为  $\{a \leftrightarrow y, b \leftrightarrow z, c \leftrightarrow x\}$ , 权值和为 1.58. 设置门限  $Th$  为 0.05, 那么有  $1.74 - 1.58 > Th$ , 所以保留  $a \leftrightarrow x$ , 因为没有其他合理的匹配存在. 如果去掉  $b \leftrightarrow z$ , 那么最优组合  $\{a \leftrightarrow x, b \leftrightarrow y, c \leftrightarrow z\}$  的权值和为 1.73, 而  $1.74 - 1.73 > Th$ , 所以剔除  $b \leftrightarrow z$ . 通过提高门限  $Th$  的值, 保留了更稳定的匹配项, 减少了错误, 但同时也会减少一些匹配的数量.

## 2 基于三焦点张量点转移的定位

给定一个目标在两幅视图中的一对匹配的观测点, 我们希望确定该目标在第 3 幅视图上的观测点, 即点的转移. 在本文中第 3 幅视图为虚拟顶视图. 当已知摄像机矩阵时, 可以采用直接射影重构的方法, 即通过观测点反向投影为空间直线, 利用最小二乘直接获得交点的 3D 坐标, 如线性三角形法(见附录(III)). 下面讨论对极点转移和三焦点张量点转移, 均针对摄像机矩阵未知的情况.

### 2.1 对极点转移

如图 1 所示, 已知摄像机 1、2 和虚拟顶视图间的基本矩阵  $F_{10}$  和  $F_{20}$ , 两个摄像机视图 1、2 中的匹配点为  $x$  和  $x'$ , 希望求出在目标在顶视图上的位置  $x''$ . 所求点  $x''$  与摄像机视图 1 的点  $x$  对应, 根据对极几何关系, 它必然在对应与  $x$  的对极线  $l_{10} = F_{10}x$  上, 同理,  $x''$  也在  $x'$  对应的对极线  $l_{20} = F_{20}x'$  上, 通过下式可求出两条极线的交点, 即

$$x'' = l_{10} \times l_{20} = (F_{10}x) \times (F_{20}x') \quad (4)$$

尽管对极变换可被用于点转移, 但它有一个严重的缺陷往往在实际中被忽略. 如当目标出现在摄像机 2、3 中间位置上, 那么当两条对极线(顶视图中虚线  $l_{10}$  和  $l'_{30}$  表示)在虚拟顶视图上接近重合时, 通过(4)计算得到交点的结果, 由于观测点的误差, 变得非常的不稳定. 即当虚拟顶视图中的目标点  $x''$  和摄像机 1、3 在虚拟顶视图中的对应极点  $e_{10}, e_{30}$  共线时, 对极变换将失效, 极线越接近病态就越严重, 即交点越不稳定. 可见, 当目标  $X$  在 3 个视图(两个摄像机视图和一个虚拟顶视图)中心确定的三焦点平面上时, 对极点转移将失效. 而这种问题可以利用三焦点张量点转移的方法解决.

### 2.2 三焦点张量点转移

三焦点张量在三视图中的作用类似于基本矩阵在两视图中的作用, 囊括了 3 幅视图间不依赖于景物结构的射影几何关系. 这里采用标准的矢量和矩阵符号代替张量符号表示, 以方便对几何关系计算的理解. 三焦点张量为 3 个  $3 \times 3$  矩阵的集合  $\{T_1, T_2, T_3\}$ , 记为矩阵的形式  $T = [T_1, T_2, T_3]$ . 如果 3 幅视图的摄像机矩阵的规范形式为:  $P = [I|0], P' = [A|a_4], P'' = [B|b_4]$ , 其中  $A$  与  $B$  是  $3 \times 3$  矩阵, 矢量  $a_i$  和  $b_i$  是矩阵的第  $i$  列,  $i=1, \dots, 4$ , 则

$$T_i = a_i b_4^T - a_4 b_i^T \quad (5)$$

如图 2 所示, 对于 3 个视图中的点-点-点对应,  $T$  满足关系:

$$x \leftrightarrow x' \leftrightarrow x'' : [x']_k \left( \sum_i x^i T_i \right) [x'']_k = \mathbf{0}_{3 \times 3} \quad (6)$$

其中,  $[x]_k$  是  $x$  的反对称矩阵,  $x^i$  是  $x$  的第  $i$  列,  $i=1, 2, 3$ .

可以通过三焦点张量  $T$  求解摄像机视图两两间的基本矩阵  $F$ , 过程如下:

首先, 可以通过三焦点张量计算对极点  $e'$  和  $e''$ ,  $e'$  和  $e''$  分别是第 1 幅视图的中心  $C$  在第 2 和第 3 幅视图中对应的极点. 令  $u_i$  和  $v_i$  分别为  $T_i$  的左零矢量和右零矢量, 即  $U_i^T T_i = \mathbf{0}^T, T_i V_i = \mathbf{0}$ , 则对极点可以通过求下面的  $3 \times 3$

矩阵的零矢量得到:

$$e'^T[u_1, u_2, u_3] = 0^T, e''^T[v_1, v_2, v_3] = 0^T \tag{7}$$

然后,利用  $e'$  和  $e''$ ,可以求解第 1 幅视图与第 2 幅和第 3 幅视图间的基本矩阵  $F_{21}$  和  $F_{31}$ :

$$F_{21} = [e']_x [T_1^T, T_2^T, T_3^T] e'', F_{31} = [e'']_x [T_1^T, T_2^T, T_3^T] e' \tag{8}$$

当已知两个摄像机视图 1、2 中的匹配点  $x$  和  $x'$ ,求解第 3 个视图中的对应点  $x''$  时,一个比较自然的想法是利用三焦点张量的点-点-点性质进行求解,即求方程组(6)的最小二乘解,这种方法计算量较大.另一种比较简便的方法是利用点-线-点性质进行求解,如图 2 所示,通过第 2 幅视图中的直线  $l'$  反向投影可以得到一个空间平面,而该平面确定了一个从视图 1 中点  $x$  到视图 3 中点  $x''$  的单应关系,这个从第 1 幅视图到第 3 幅视图的单应为

$$H_{13}(l') = [T_1^T, T_2^T, T_3^T] l' \tag{9}$$

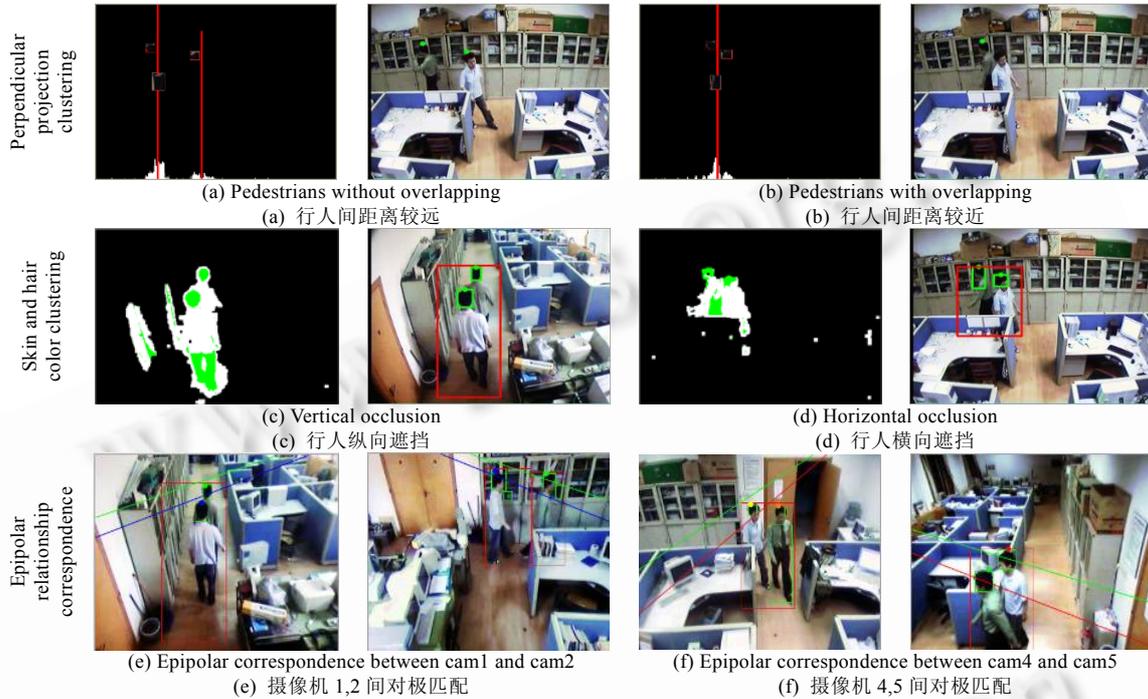


Fig.3 Head detection and epipolar correspondence results

图 3 头部检测与对极匹配实验结果

显而易见,该转移对于在三焦点平面上的一般点  $X$  是不退化的.但值得注意的是,需要保证  $l'$  不是  $x$  在视图 2 中的对极线.因为在这种情况下( $x$  反向投影的直线在  $l'$  反向投影平面内), $[T_1^T, T_2^T, T_3^T] l' x = 0, x''$  没有定义.所以在选择  $l'$  的时候应该避免选到  $x$  的对极线,一个比较好的方法是选择与  $x$  的对极线垂直的直线.下面给出已知 3 幅视图中的若干对应点组情况下,利用三焦点张量进行点转移的步骤:

- (1) 根据(附录(II))方法计算三焦点张量  $T=[T_1, T_2, T_3]$ .
- (2) 利用式(7)、式(8),通过  $T$  求出  $F_{21}$ .
- (3) 对于摄像机视图 1、2 中的对应点  $x = (x_1, x_2, 1)^T$  和  $x'$ ,计算过点  $x'$  并垂直于  $x$  的对极线  $l'_e = F_{21} x$  的直线  $l'$ .如果令  $l'_e = (l_1, l_2, l_3)$  和  $x' = (x_1, x_2, 1)^T$ ,那么  $l' = (l_2, -l_1, -x_1 l_2 + x_2 l_1)^T$ .
- (4) 利用  $T$  和  $l'$  计算摄像机视图 1 和虚拟顶视图间的单应:  $H_{13}(l') = [T_1^T, T_2^T, T_3^T] l'$ .
- (5) 确定在第 3 幅视图中  $X$  的投影  $x''$ :  $x'' = H_{13}(l') x$ .

### 3 实验

我们建立了由 5 路摄像机组成的室内视频监控系统,在观测空间中设置了标志点,经测量后绘制了室内虚拟顶视图,实验系统界面如图 4 所示,下方为 5 路摄像机视频,上方为虚拟顶视图.摄像机与观测目标间距离大于 3 米,而且由于存在大量隔板,行人长时间被遮挡,无法检测到完整的行人,唯一可行的方法是检测行人的头部位置.而由于行人高度和姿态不同导致头部并不处于同一平面内,因此基于单应关系的方法在此环境中不适用.

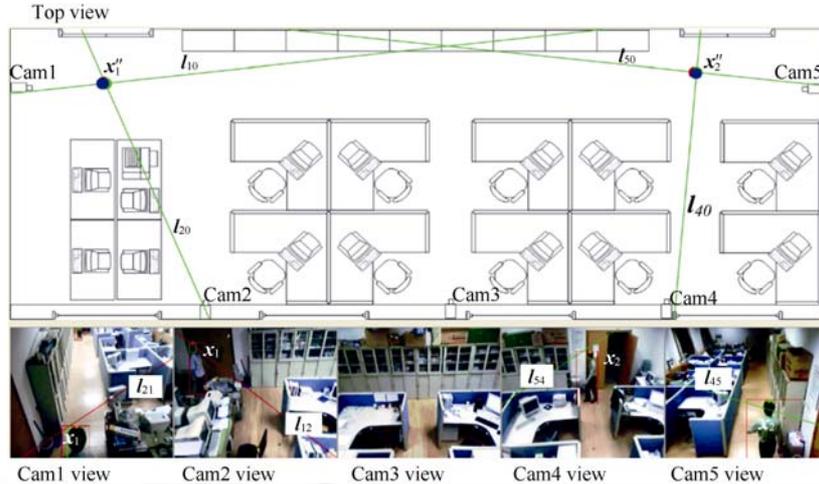


Fig.4 Virtual top view location result 1 without point transfer deterioration

图 4 虚拟顶视图定位结果 1(无对极线交点退化的情况)

我们采集了室内环境下的 13 组视频(352×288),共 65 000 多帧视频数据,包括下列不同类型:不同数量(单个人、多个人)、不同速度(快速、中速和慢速行走)、不同光照(白天日光、白天灯光、夜晚灯光)、不同颜色衣着(白色、黑色、蓝色、红色、绿色、灰色等)进行头部检测、匹配和点转移的实验.实验程序运行于 Pentium4 1.6GHz Windows 平台,采用 C++和 Intel OpenCV 开发,以下实验数据均针对上述视频和实验环境.

头部检测实验:在垂直投影聚类方法中,采用 K-means 算法对投影点进行聚类,初始聚类数量为 8,进行 10 次迭代,在迭代过程中将距离较近的类合并,并重新计算类心.在基于肤色聚类检测的方法中,肤色 RGB 分量门限设置为 50,而肤色 H 通道阈值设置为  $0^{\circ}\sim 30^{\circ}$ ,检测出的像素进行基于行程编码的连通区域标识,得到符合头部特征的块,选择处于运动区域上部(1/5 高度区域)的块.实验结果如图 3 所示,图 3(a),图 3(b)中白色区域为运动区域垂直投影产生的直方图,灰色垂线标识出聚类的类心,灰色矩形框为沿类心垂直方向上搜索到的头部区域;图 3(c),图 3(d)中白色区域为运动前景,灰色区域为检测为发色和肤色的部分,矩形框标识出运动前景和头部位置;图 3(e),图 3(f)中对极线颜色相同表示对应关系)垂直投影聚类方法中,当两个行人距离较近时,聚类方法只能检测出一个类心,无法作出正确判断;发色肤色检测的方法在行人纵向和横向遮挡时都可以得到较好的检测效果.实验表明,在检测速度相当的情况下(均为 0.05s~0.07s/帧),目标距离较近时,后者正确检测率比前者高出 80%.

接下来,对头部检测的结果进行跟踪滤波,数据关联和跟踪不是本文的重点,可参见文献[18].实验中采用了基于 PDA 数据关联的 Kalman 滤波进行跟踪,采用了概率似然比检测判断目标轨迹的起始与中止.对于多目标的情况,采用椭圆形跟踪门来判决新目标的产生.对于视频跟踪中遮挡问题的处理,可参见文献[20].通过数据关联和滤波,去除了大多数误检数据,并填补了一些漏检的数据.

对滤波后的数据进行对极匹配实验,图 3 展示了对极匹配的实验结果.实验结果表明,对极匹配方法在头部检测基本正确的情况下正确率可以达到 95%,且对头部检测的噪声有一定的鲁棒性.但是当多个行人的头部在不同摄像机中的位置都趋于重合时,仍会出现错误或不稳定的匹配,这时通过 Leave-One-Out 方法,对这种错误

起到了一定程度的抑制作用。

在点转移定位实验中,我们以直接射影重构的结果为标准(见附录(III)),比较三焦点张量点转移和对极点转移的结果.准备工作如下:(1) 通过室内设置的标志点,依据文献[21]中的摄像机模型和标定方法对摄像机进行标定,计算出摄像机矩阵  $P$ 、摄像机的内外参数以及径向扭曲系数(实验中视图经过了径向反扭曲,标定过程忽略了切向扭曲).(2) 人工选择摄像机视图和虚拟顶视图中的多组对应点,测量出图像坐标,然后通过基于 RANSAC 的方法计算出视图之间的基础矩阵  $F$ (见附录(I)).(3) 人工选择两两摄像机视图与虚拟顶视图中的多组对应点,测量出图像坐标,利用最小化代数误差算法(见附录(II))计算出三焦点张量  $T$ .

如图 4 所示,两个目标  $X_1$  和  $X_2$  分别投影于摄像机 1,2 和 4,5 的视图中,观测点为  $x_1, x'_1, x_2, x'_2$ , 顶视图中直线  $l_{10}, l_{20}, l_{40}, l_{50}$  是 4 幅视图中的观测点在虚拟顶视图中的对极线,点转移的结果为  $x''_1$  和  $x''_2$ . 3 种方法的点转移结果分别用不同灰度标出,其中灰色为直接射影重构的结果(参考标准),浅灰色为对极点转移结果,即对极线(浅灰色直线)交点,黑色为三焦点张量点转移结果.可见,在一般情况下,后两种方法可得到与直接射影重构相近的结果,误差主要取决于  $F$  和  $T$  计算时的误差.实验中,当精确提取摄像机图像特征点(采用 SIFT 特征点),控制对应特征点误差在  $\pm 3$  个像素以内时,转移后的虚拟顶视图中的误差不大于  $\pm 10$  个像素(3 种颜色的点基本重合).

如图 5 所示,一个目标出现在 1,2,5 视图中.摄像机 1,2 视图中的观测点为  $x_1, x'_1$ ,通过视图 1,2 的对极点转移结果为  $x''_1$ , 3 种方法结果相近( $x''_1$  处 3 个点基本重合),但是对于视图 1,5 中的观测点  $x_2, x'_2$ ,点转移后结果为  $x''_2$  和  $x''_2$  (error),对极转移产生了较大的误差的  $x''_2$  (error),而通过三焦点张量点转移的结果  $x''_2$  仍然和直接射影重构结果相近(浅灰色和黑色点距离较远,转移误差超过 400 个像素,而灰色点与黑色点基本重合,误差小于 10 个像素).这正是由于对极点转移的退化问题所导致的,当目标的对极线  $l_{10}, l_{50}$  在顶视图中接近重合时,观测值上一个较小的误差就可以造成对极线交点一个较大的变化,从而导致点转移结果不够稳定。

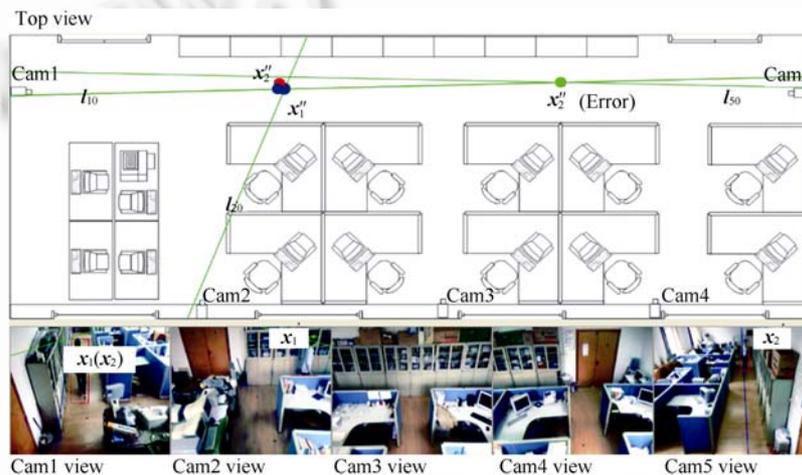


Fig.5 Virtual top view location result 2 with point transfer deterioration  
图 5 虚拟顶视图定位结果 2(对极线交点退化的情况)

#### 4 结 论

本文建立了一个室内遮挡环境下的多摄像机协同跟踪定位系统,提出了一种基于头部检测对极匹配和三焦点张量点转移定位的多摄像机协同的方法.该方法首先在单摄像机视图中检测行人头部位置并进行跟踪,然后利用摄像机视图间的对极约束建立观测点间的匹配关系,最后通过摄像机视图与虚拟顶视图间的三焦点张量关系进行点转移、确定目标在虚拟顶视图中的位置.该方法的优点在于在复杂遮挡环境中,在不进行摄像机标定和点转移无退化的情况下,确定行人的位置。

在头部检测实验中,我们将本文方法与传统的垂直投影方法进行了对比,在点转移实验中,我们以直接射影

重构法为标准,将本文三焦点张量点转移的方法与对极点转移的方法进行了对比,验证了本文方法的准确性和鲁棒性.

#### References:

- [1] Yang JQ, Liang H, Wang CY. A review of video surveillance techniques. *Modern Electronics Technique*, 2006,(21):84–88, 91 (in Chinese with English abstract).
- [2] Mittal A, Davis LS. M2Tracker a multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo. In: *Proc. of the 7th European Conf. on Computer Vision (ECCV)*. 2002. <http://citeseer.ist.psu.edu/584689.html>
- [3] Kim K, Davis L. Multi-Camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. In: *Proc. of the 9th European Conf. on Computer Vision 2006*. 2006. 98–109. doi=10.1.1.85.3402
- [4] Hu W, Hu M, Zhou X, Tan T, Lou J, Maybank S. Principal axis-based correspondence between multiple cameras for people tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006,28(4):663–671.
- [5] Black J, Ellis T, Rosin P. Multi view image surveillance and tracking. In: *Proc. of the IEEE Workshop on Motion and Video Computing*. 2002. doi=10.1.1.14.8531
- [6] Khan S, Javed O, Rasheed Z, Shah M. Human tracking in multiple Cameras. In: *Proc. of the 8th Int'l Conf. on Computer Vision. Vancouver, British Columbia*. 2001. doi=10.1.1.21.160
- [7] Huang T, Russell S. Object identification in a bayesian context. In: *Proc. of the 15th Int'l Joint Conf. on Artificial Intelligence*. 1997. 1276–1283. doi=10.1.1.51.9216
- [8] Kettmaker V, Zabih R. Bayesian multi-camera surveillance. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 1999. 253–259. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?isnumber=17024&arnumber=784638&type=ref](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?isnumber=17024&arnumber=784638&type=ref)
- [9] Pasula H, Russell S, Ostland M, Ritov Y. Tracking many objects with many sensors. In: *Proc. of the 16th Int'l Joint Conf. on Artificial Intelligence*. 1999. doi=10.1.1.34.7997
- [10] Javed O, Shafique K, Shah M. Appearance modeling for tracking in multiple non-overlapping cameras. In: *Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2005. doi=10.1.1.76.2620
- [11] Makris D, Ellis T, Black J. Bridging the gaps between cameras. In: *Proc. of the Computer Vision and Pattern Recognition*. 2004. 205–210. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1315165](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1315165)
- [12] Black J, Makris D, Ellis T. Validation of blind region learning and tracking. In: *Proc. of the Joint IEEE Int'l Workshop on Visual Surveillance and Performance Evaluation (VS-PETS2005)*. 2005. doi=10.1.1.139.3656
- [13] Javed O, Rasheed Z, Shafique K, *et al.* Tracking across cameras with disjoint views. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. Nice: ACM Press, 2003. 165–171. doi=10.1.1.57.9874
- [14] Stauffer C. Learning to track objects through unobserved regions. In: *Proc. of the IEEE Workshop on Motion*. 2005. 96–102. doi=10.1.1.58.9537
- [15] Tieu K, Dalley G, Grimson WE L. Inference of non-overlapping camera network topology by measuring statistical dependence. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. Beijing, China 2005, 1842–1849. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?tp=&arnumber=1544940&isnumber=32976](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1544940&isnumber=32976)
- [16] Liang H, Liu YH. Adaptive and efficient background subtraction using multi-models. *Journal of Image and Graphics*, 2008,13(2):345–350 (in Chinese with English abstract).
- [17] Wu L, Zhou DX, Liang H, Cai XP. A shadow detection algorithm based on phong lighting and radiosity model. *Modern Electronics Technique*, 2008,31(5):124–127 (in Chinese with English abstract).
- [18] Zhou HR, Wang PD. *Maneuvering Target Tracking*. Beijing: National Defense Industry Press, 1991 (in Chinese).
- [19] Hartley R, Zisserman A, Wrote; Wei S, Yang SJ, Zhang QB, Hu ML, *Trans. Multiple View Geometry in Computer Vision*. Hefei: Anhui University Press, 2002 (in Chinese).
- [20] Liang H, Liu YH. Video objects tracking under occlusions. *Application Research of Computers*, 2008,25(3):950–956 (in Chinese with English abstract).

- [21] Heikkila J, Silven NO. A four-step camera calibration procedure with implicit image correction. In: Proc. of the 1997 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'97). San Juan: IEEE Computer Society Press, 1997. 1106-1112.
- [22] Forsyth DA, Ponce J, Wrote; Lin XY, Wang H, Trans. Computer Vision: A Modern Approach. Beijing: House of Electronics Industry, 2004 (in Chinese).

#### 附中文参考文献:

- [1] 杨建全,梁华,王成友.视频监控技术的发展与现状.现代电子技术,2006,(21):84-88,91.
- [16] 梁华,刘云辉.自适应多模快速背景差算法.中国图象图形学报,2008,13(2):345-350.
- [17] 吴亮,周东翔,梁华,蔡宣平.一种基于 Phong 物体光照模型的阴影检测算法.现代电子技术,2008,31(5):124-127.
- [18] 周宏仁,王培德.机动目标跟踪.北京:国防工业出版社,1991.
- [19] Hartley R, Zisserman A, 著;韦穗,杨尚骏,章权兵,胡茂林,译.计算机视觉中的多视图几何.合肥:安徽大学出版社,2002.
- [20] 梁华,刘云辉.视频中遮挡情况下目标的跟踪.计算机应用研究,2008,25(3):950-956.
- [22] Forsyth DA, Ponce J, 著;林学闾,王宏,译.计算机视觉——一种现代方法.北京:电子工业出版社,2004.

## 附录

### (I) 基础矩阵 $F$ 的计算:随机采样一致算法 RANSAC

RANSAC 算法在计算机视觉中有着广泛的应用,将该算法应用于基本矩阵的计算,可以提高在匹配点存在野点时算法的鲁棒性.步骤如下:

- (1) 从数据中均匀随机的抽取 7 组对应点.
- (2) 利用这 7 组对应点求基础矩阵  $F$  的最小配置解,7 点算法可见文献[22].
- (3) 用未抽取到的每一组对应点来验证  $F$ ,如果距离足够小,那么这个对应是一致的.
- (4) 如果对应的一致足够多,那么就断定这个拟合是好的,然后用所有的点重新拟合基础矩阵  $F$ .反之,则回到步骤(1).

### (II) 三焦点张量 $T$ 的计算:最小化代数误差算法

简要介绍如何通过已知 3 幅视图中 7 组或者更多的对应点计算三焦点张量  $T$  的归一化最小化代数误差算法.详细内容可见文献[19].这里将算法步骤整理如下:

- (1) 将对应点组进行归一化,使质心在图像中心,点到质心的平均距离为  $\sqrt{2}$ .
- (2) 利用式(6)和对应点建立形如  $A\mathbf{t}=\mathbf{0}$  的齐次方程组.
- (3) 利用 SVD 求解  $T$  的初始估计.
- (4) 根据式(7)从  $T$  求两个对极点  $\mathbf{e}'$  和  $\mathbf{e}''$ ,使对极点  $\mathbf{e}'(\mathbf{e}'')$  同时垂直于 3 个  $T_i$  的左(右)零矢量.注意到,对于摄像机矩阵的规范形  $P=[I|0]$ ,有:

$$\mathbf{e}' = P'C = [A | \mathbf{a}_4](0, 0, 0, 1)^T = \mathbf{a}_4, \mathbf{e}'' = P''C = [B | \mathbf{b}_4](0, 0, 0, 1)^T = \mathbf{b}_4.$$

- (5) 根据式(5),构造使得  $\mathbf{t}=\mathbf{E}\mathbf{a}$  的  $27 \times 18$  矩阵  $\mathbf{E}$ ,其中  $\mathbf{t}$  是  $T$  的元素组成的矢量, $\mathbf{a}$  表示  $\mathbf{a}_i$  和  $\mathbf{b}_i$  的元素的矢量,而  $\mathbf{E}$  表示线性关系  $T_i = \mathbf{a}_i \mathbf{e}''^T - \mathbf{e}' \mathbf{b}_i^T$ .

- (6) 求解最小化问题:在满足  $\|\mathbf{E}\mathbf{a}\|=1$  的条件下最小化  $\|\mathbf{A}\mathbf{E}\mathbf{a}\|$ ,误差矢量  $\mathbf{e}=\mathbf{A}\mathbf{E}\mathbf{a}$ .这样,  $\mathbf{E}\mathbf{a}$  就是一个满足所有约束并且在给定对极点下最小化代数误差的三焦点张量.

- (7) 将  $T$  进行反向归一化.

### (III) 射影重构:非齐次线性三角形法

简要介绍已知 3D 点  $\mathbf{X}$  在两幅视图上的投影  $\mathbf{x}, \mathbf{x}'$  和摄像机矩阵  $P, P'$ , 计算该点空间位置的非齐次线性三角

形法.详细内容可见文献[19].

令  $X$  在两幅视图上的投影  $x=(x,y), x'=(x',y')$  和摄像机矩阵  $P, P'$ , 有  $x=PX, x'=P'X$ , 那么有  $x \times (PX)=0, x' \times (P'X)=0$ , 将其展开后得到:

$$\left. \begin{aligned} x(P^{3T}X) - (P^{1T}X) &= 0, & x'(P'^{3T}X) - (P'^{1T}X) &= 0 \\ y(P^{3T}X) - (P^{2T}X) &= 0, & y'(P'^{3T}X) - (P'^{2T}X) &= 0 \\ x(P^{2T}X) - y(P^{1T}X) &= 0, & x'(P'^{2T}X) - y'(P'^{1T}X) &= 0 \end{aligned} \right\} \quad (10)$$

其中,  $P^i$  和  $P'^i$  是  $P$  和  $P'$  的行. 该方程关于  $X$  的分量是线性的, 将式(10)中前两行组成形如  $AX=0$  方程, 即

$$\begin{bmatrix} xP^{3T} - P^{1T} \\ yP^{3T} - P^{2T} \\ x'P'^{3T} - P'^{1T} \\ y'P'^{3T} - P'^{2T} \end{bmatrix} X = 0 \quad (11)$$

令  $X=(X,Y,Z)^T$ , 这个齐次方程组可化为 3 个未知量的 4 个方程的非齐次方程组, 然后通过 SVD 求最小二乘解.



梁华(1977-), 男, 陕西宝鸡人, 博士生, 主要研究领域为计算机视觉, 视频监控.



刘云辉(1965-), 男, 博士, 教授, 博士生导师, 主要研究领域为人工智能, 自动控制, 计算机视觉.



蔡宣平(1962-), 男, 博士, 教授, 主要研究领域为图形图像, 计算机视觉.