基于混合特征的人体动作识别改进算法*

郭 利[1], 姬晓飞[2], 李 平[1], 曹江涛[1]

(1. 辽宁石油化工大学 信息与控制工程学院, 辽宁 抚顺 113001; 2. 沈阳航空航天大学 自动化学院, 沈阳 110136)

摘 要: 运动特征的选择直接影响人体动作识别方法的识别效果。单一特征往往受到人体外观、环境、摄像机设置等因素的影响不同,其适用范围不同,识别效果也是有限的。在研究人体动作的表征与识别的基础上,充分考虑不同特征的优缺点,提出一种结合全局的剪影特征和局部的光流特征的混合特征,并用于人体动作识别。实验结果表明,该算法得到了理想的识别结果,对于 Weizmann 数据库中的动作可以达到 100% 的正确识别率。

关键词: 动作识别; 剪影特征; 光流特征; 留一法

中图分类号: TP391.41 **文献标志码**: A **文章编号**: 1001-3695(2013)02-0601-04

doi:10.3969/j.issn.1001-3695.2013.02.079

Mixed features based improved human action recognition algorithm

GUO Li[1], JI Xiao-fei[2], LI Ping[1], CAO Jiang-tao[1]

(1. *School of Information & Control Engineering, Liaoning Shihua University, Fushun Liaoning* 113001, *China*; 2. *School of Automation, Shenyang Aerospace University, Shenyang* 110136, *China*)

Abstract: The choice of the motion features affects the result of the human action recognition method directly. Many factors often influence the single feature differently, such as appearance of human body, environment and video camera. So the accuracy of action recognition is limited. On the basis of studying the representation and recognition of human actions, and giving full consideration to the advantages and disadvantages of different features, this paper proposed a mixed feature which combined global silhouette feature and local optical flow feature. This combined representation was used for human action recognition. The experimental results demonstrate that this algorithm can recognize human actions and achieve high recognition rates. This algorithm achieves 100% correct recognition rate for the human actions in the Weizmann database.

Key words: action recognition; silhouette; optical flows; leave one out

0 引言

基于视频的人体动作识别是视频分析的重要研究方向之一,属于计算机视觉的中低层分析。它在视频监控、视频检索、人机交互、虚拟现实等领域有着广阔的应用前景。

视频序列中的动作信息通常可以用光流、剪影、兴趣点等特征来描述。文献[1]提出用光流直方图来描述动作的运动信息,然后用支持向量机作分类器识别运动员的击球动作。基于光流的表示法能在没有背景区域任何先验知识的条件下,实现对运动目标的检测和跟踪;但缺点是受噪声、光照强度变化的影响较大,并且计算方法复杂、计算量大。文献[2]基于剪影图像通过提取质心到边界点的距离特征来将运动目标分为车辆、人体、人群。基于剪影的描述方法易于实现,受光照条件影响小,但是它主要依赖于形状的边界信息,不能获得形状的内部结构且容易受到背景变化的影响。文献[3]将缩放后的标准人体姿态剪影图像分成 $m \times n$ 个子块,通过计算每个子块中人体的像素数占所有子块的像素数最大值的比例来分析人体行为。这种方法描述剪影内部信息,计算也简单,但是对视角变化和行为主体变化很敏感。文献[4]构建人体姿态的点模型采用 13 个特征点来表示人体姿态,每种动作可以表示为这些点的运动轨迹。该算法建立的模型较复杂,计算相对复杂,实时性较差。文献[5]提出一种时空兴趣点的检测方法并用兴趣点构成的点集来表示动作。类似的文献[6]提出一种基于 Gabor 滤波器的时空兴趣点检测算法并用于动作识别。这类方法对于低分辨率视频以及摄像头运动造成的干扰均具有一定的鲁棒性,但是兴趣点检测算法中的参数需要根据不同的使用条件进行调整,否则不能保证检测点的准确性。

在充分考虑不同特征表示的优缺点及适用范围的基础上,提出一种新的将全局形状描述的静态特征和局部光流描述的动态特征有机结合的混合特征。首先利用背景减除法确定出运动的大致区域,得到人体剪影,并利用剪影轮廓向量表示人体外观的整体信息;然后在运动区域内提取光流,并利用分区域的局部光流信息来表示人体运动的局部特征,以此来提高光流的抗噪能力;最后将全局的剪影特征与局部的光流特征结合作为混合特征。实验结果表明,混合特征的鲁棒性及识别性能比单一特征好。

收稿日期: 2012-06-15; **修回日期**: 2012-07-26 **基金项目**: 国家自然科学基金青年基金资助项目(61103123)

作者简介: 郭利(1987-),女,硕士研究生,主要研究方向为模式识别、视频监控(guoli3377@gmail.com);姬晓飞(1978-),女,博士,主要研究方向为视频处理及模式识别理论;李平(1964-),男,教授,博士,主要研究方向为工业过程控制与优化;曹江涛(1978-),男,教授,博士,主要研究方向为智能监控、智能优化与控制.

1 特征表示与提取

1.1 图像预处理

图像预处理可以减少研究的图像面积及处理数据,减少计算量。通常利用背景减除法确定出运动的大致区域及人体剪影,如图 1(b)所示。假定所有的动作是在静态背景前执行的,根据剪影信息,可确定兴趣区域,如图 1(c)所示,白色矩形框内的为兴趣区域。兴趣区域确定后,可以只处理兴趣区域内的信息。

(a)视频中原始帧图像 (b)人体剪影 (c)被标记原始帧图像

图 1 背景减除法

1.2 剪影特征表示与提取

单帧图像中人体剪影可以用来描述人体运动整体形状变化信息。选用剪影特征有以下几个优点:a)剪影特征可简单直观地描述运动人体的形状信息;b)剪影特征易于提取;c)二值剪影图对前景图像的纹理和颜色不敏感。这一步旨在将原始视频的运动信息转换成与之关联的形态特征序列,这些形态序列反映出运动过程变化。

设有一段运动视频 V 有 T 帧图像 I,如式 $V=[I_1,I_2,\cdots,I_T]$,相对应的运动剪影序列为 $S=[s_1,s_2,\cdots,s_T]$,s 在图像预处理时已获得。为了简单起见,运用轮廓向量[7]的方法描述人体剪影的整体形状信息。具体过程如下:

a)用 Canny 算子求得每帧剪影的边缘轮廓如图 2(a)所示,并求取此边缘轮廓的坐标表示,如图 2(b)所示。这样人体轮廓可用 n_t 个点表示,即$\{(x_1,y_1),(x_2,y_2),\cdots,(x_{n_t},y_{n_t})\}$,$t=1,2,3,\cdots,T$。

b)人体轮廓的质心用式(1)求得。

$$(x_c,y_c)=\left(\frac{1}{n_t}\sum_{i=1}^{n_t}x_i,\frac{1}{n_t}\sum_{i=1}^{n_t}y_i\right) \tag{1}$$

其中,(x_c,y_c)为质心,(x_i,y_i)为轮廓边缘点,n_t 为第 t 幅图像中边缘点数。

c)质心到边缘点的距离可以用式(2)求得。

$$d_i=\sqrt{(x_i-x_c)^2+(y_i-y_c)^2}\quad i=1,2,\cdots,n_t \tag{2}$$

其中:d_i 为第 i 个边缘点对应的质心到边缘点的距离。计算时从轮廓图的最左边最上边的点开始,顺时针方向依次计算,这样就从单帧图像 I_t 的二维轮廓图得到对应一维的轮廓向量 D_t。

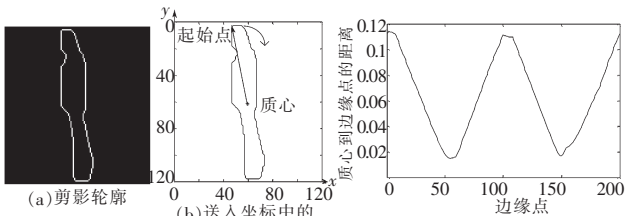
$$D_t=[d_1,d_2,d_3,\cdots,d_i,\cdots,d_{n_t}]\quad i=1,2,\cdots,n_t \tag{3}$$

d)为了消除空间尺度和距离长度的影响,使用 2-范数对轮廓向量作归一化处理。由于每帧图像的边缘点 n_t 大小不等,此处对归一化处理后的轮廓向量进行等间隔重采样,获得固定点数 N。本文对不同数值 N 进行实验,得出 $N=200$ 计算量相对较少,同时能完整表达运动信息,单一特征及混合特征均能达到最高的识别率。当采样点 $N=200$ 时,轮廓向量结果如图 2(c)所示。

1.3 光流特征表示与提取

剪影图像提取不准确可能导致轮廓向量特征信息不能准确地表达动作特征。此时,光流特征可以有效、准确地表示视频序列中的动作信息。在运动区域内提取光流,并利用分区域的局部光流信息来表示人体运动的局部特征,以此来提高光流的抗噪能力。光流的提取与表示具体过程如下:

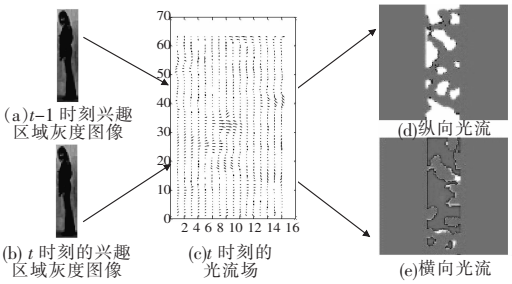
a)确定当前帧图像 I_t 对应的兴趣区域位置,剪切出此兴趣区域位置对应的当前帧和前一帧图像对应的灰度图像区域,如图 3(a)(b)所示。然后利用 Lucas-Kanade 的方法在当前帧和前一帧的兴趣区域内作光流检测,得到的光流场如图 3(b)所示,并将光流分解成纵向和横向的两个分量,即纵向光流、横向光流。



(a)剪影轮廓 (b)送入坐标中的剪影轮廓 (c)最终得到的轮廓向量特征

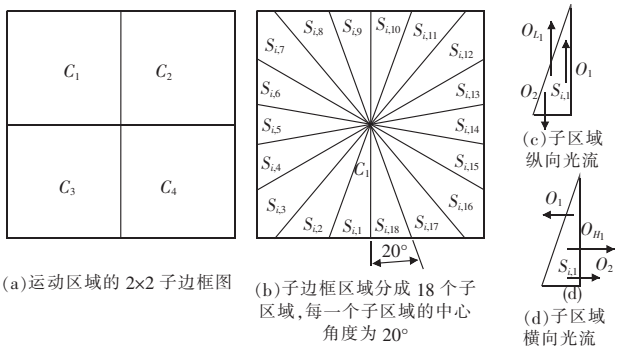
图 2 轮廓向量结果

b)为了降低光流信息维数,找到有辨识能力的数据表示,运用分区域径向直方图方法来统计光流特征[8]。首先采用按照长边缩放的前提下,将得到的兴趣区域光流图像标准化为 120×120 维的统一大小光流图,如图 3(d)(e)所示。将标准化后的光流图分成 2×2 的子边框 S_1、S_2、S_3、S_4,子边框为 60×60,其中心点分别为 C_1、C_2、C_3、C_4,如图 4(a)所示。然后以子边框的中心点为中心将子边框分成 18 个子区域分别为 $S_{i,1}$,$S_{i,2},\cdots,S_{i,18}$($i=1,2,3,4$),每个中心角占 20°,这样就形成了 72 个子区域,如图 4(b)所示。



(a)$t-1$ 时刻兴趣区域灰度图像 (b)t 时刻的兴趣区域灰度图像 (c)t 时刻的光流场 (d)纵向光流 (e)横向光流

图 3 对应灰度图



(a)运动区域的 2×2 子边框图 (b)子边框区域分成 18 个子区域,每一个子区域的中心角度为 20° (c)子区域纵向光流 (d)子区域横向光流

图 4 光流特征提取

c)在子区域 S_{ij} 中有 k 个纵向光流(或者横向光流),将所有的纵向光流(或是横向光流)相加就得到子区域 S_{ij} 的纵向光流之和 $O_{L_{i,j}}$(横向光流之和 $O_{H_{i,j}}$)。式(4)(5)为分别计算子区域的纵向光流之和和横向光流之和。

$$O_{L_{i,j}}=\sum_{m=1}^{k}O_{Lm}\quad(x_{O_{Lm}},y_{O_{Lm}})\in S_{i,j} \tag{4}$$

$$O_{H_{i,j}}=\sum_{m=1}^{k}O_{Hm}\quad(x_{O_{Hm}},y_{O_{Hm}})\in S_{i,j} \tag{5}$$

d)整帧图像 I_t 的光流信息就可以由 72 个子区域的纵向光流之和 O_L、横向光流之和 O_H 来表示,如式(6)~(8)所示。

$$O_L=[O_{L_{1,1}},\cdots,O_{L_{1,18}},\cdots,O_{L_{4,1}},\cdots,O_{L_{4,11}}] \tag{6}$$

$$O_H=[O_{H_{1,1}},\cdots,O_{H_{1,18}},\cdots,O_{H_{4,1}},\cdots,O_{H_{4,11}}] \tag{7}$$

$$O_t=[O_L,O_H] \tag{8}$$

其中:式(6)为 72 个子区域的纵向光流组成的纵向局部光流向量 O_L;式(7)为 72 个子区域的横向光流组成的横向局部光流向量 O_H;式(8)中 O_t 为局部光流向量。光流特征提取时,参数设置参考文献[8]。

e)使用 2-范数对 O_t 归一化处理就得到了当前帧图像 I_t 的局部光流向量的径向直方图表示的特征。

1.4　混合特征表示

1.2、1.3 节的特征是从视频单帧图像中提取的。为了改善动作识别的准确性,将轮廓向量、局部光流向量组合在一起,形成混合特征向量,如式(9)所示。

$$F_t=[O_t,D_t] \tag{9}$$

其中:F_t、O_t、D_t 分别为单帧图像 I_t 的混合特征向量、局部光流向量、轮廓向量。

2　识别方法介绍

有很多解决统计分类问题的方法,本文主要测试提出特征的辨识能力,所以此处选用最简单的最近邻分类器[9]。具体算法如下:

a)找到测试序列每一帧的最近邻。设测试样本序列第 t 帧的特征向量为 $M_Q^t(t=1,2,\cdots,T)$,训练样本所对应的第 n 帧特征向量为 M_T^n。用欧几里德距离来测试 M_Q^t、M_T^n的相似性,与 M_Q^t距离最小的训练样本帧就是测试样本序列第 t 帧的最近邻,如式(10)所示。

$$s_q=\min\|M_Q^t-M_T^n\| \quad n=1,2,\cdots,N \tag{10}$$

b)将测试帧对应的最近邻的训练帧所属动作的标号赋给当前的测试帧,这样测试序列的每一测试帧都将得到一个动作的标号。

c)将测试序列每一帧的动作标号进行统计,测试序列类别对应为票数最多的标号对应的动作。例如,diaria_bend 序列中有 83 帧,每帧都用最近邻动作标号标记,统计结果为[69,3,0,5,2,0,1,0,2,1],即有 69 帧被标记为 1 号动作,3 帧被标记为 2 号动作,依次类推。票数最多为 69,其对应的动作标号为 1,则此序列将被识别为 1 号动作。

3　实验分析(算法验证)

为了验证本文算法的有效性,在公开的 Weizmann 视频数据库和 KTH 数据库上作了大量的对比实验。

3.1　数据库介绍

本实验是在 MATLAB2010a 中运行实现的。Weizmann 动作视频库中有十种动作,分别为 bend、jack、jump、pjump、run、side、skip、walk、wave1、wave2,每种动作由 9 个人完成。视频背景和视角均不变,每帧图像的分辨率为 144×180,帧速率为 25 fps,示例如图 5 所示。KTH 数据库中有六种动作,分别为 boxing、handclapping、handwaving、jogging、running、walking,每种动作由 25 个不同的人在四个场景下完成,一共有 599 段视频;背景相对静止,除了镜头的拉近/拉远,摄像机的运动相对轻微,如图 6 所示。

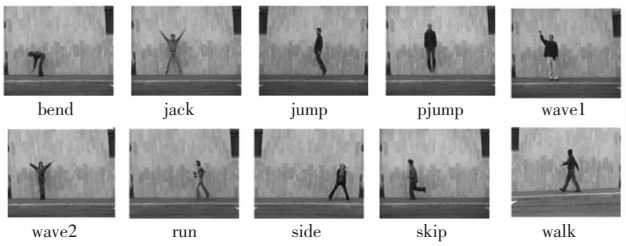


图 5　Weizmann 数据库十种动作

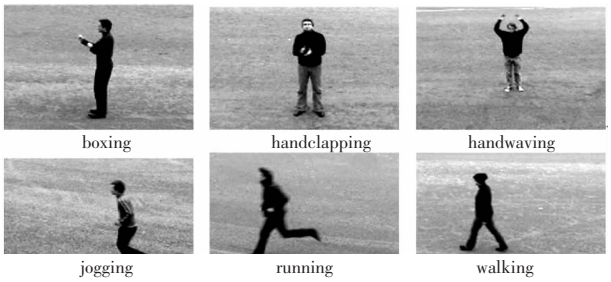


图 6　KTH 数据库六种动作示意图

3.2　实验及结果分析

本文采用第 1 章的方法分别提取人体动作轮廓向量特征、光流特征以及混合特征来表征动作。为了获取无偏估计准确性,采用留一法(leave one out)来验证实验效果,即每次实验选择数据库中的一个人所有动作为测试样本集,而余下的作为训练样本集。然后循环,每个人的动作都将作为测试样本进行测试,并统计识别结果。利用最近邻的方法进行分类识别。光流特征、轮廓距离特征及两者结合的混合特征识别结果如表 1 所示。混淆矩阵如图 7 所示,其中第一列为 Weizmann 数据库各特征混淆矩阵框图,第二列为 KTH 数据库各特征混淆矩阵框图。

表 1　选择不同特征对应的识别率

数据库	轮廓向量特征/%	光流特征/%	混合特征/%
Weizmann	88.89	98.89	100
KTH	83.33	93.33	95.00

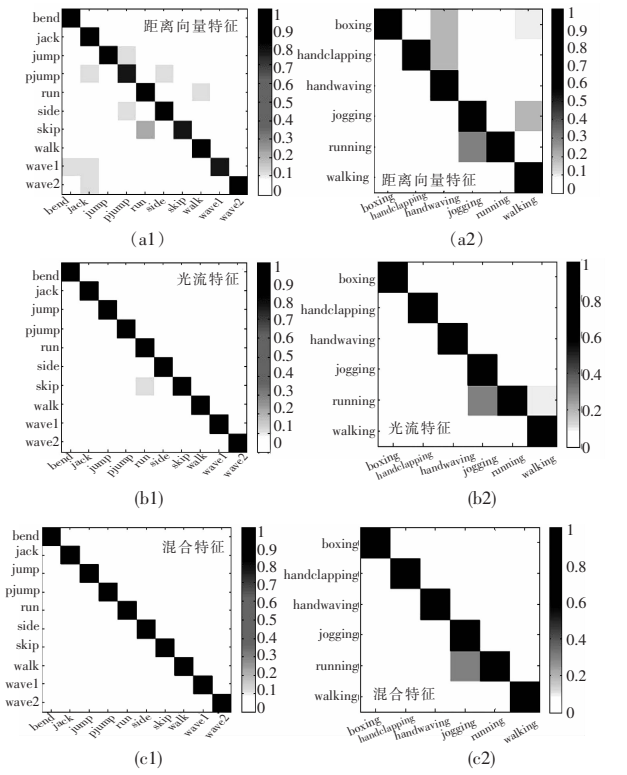


图 7　混淆矩阵

由表 1 可以看出,混合特征的识别率要比单个特征识别率

高,对于 Weizmann 数据库可得 100% 的正确识别率,对于 KTH 数据库可得高达 95% 的识别率。

从图 7(a1)(a2)可以看出,识别结果中有很多的错误识别,如(a1)中 skip 误判为 run,jump 误判为 pjump 等;完全正确识别的动作只有三个,分别为 bend、jack、walk。其主要原因是拍摄视频的距离相对较远、分辨率较低,造成提取人体剪影轮廓图时四肢模糊不易分辨。从图 7(b1)(b2)可以看出,光流特征的识别要比图 7(a1)(a2)的结果好很多,Weizmann 数据库只有动作 skip 被误判为 run,而 KTH 数据库只有一种动作 run 被误判,说明光流特征可以对相对远距离、分辨率低的视频进行很好的识别。将光流特征和轮廓向量特征相结合,得到图 7(c1)(c2)的识别结果,可以明显地看出识别结果非常好,充分验证了提出特征的有效性。

本文方法与近期的相关方法基于 Weizmann 数据库识别性能比较如表 2 所示。从表 2 可以看出,虽然选择的特征相近,但本文提出的特征及其算法的识别性能要优于其他算法。此外,所提出的特征易于提取和表征、具有较高的可靠性,避免了基于人体模型的特征提取等方法的复杂运算。

表 2 不同特征结合对应的识别率

方法	所用特征	识别率/%
Ahmad 等人[10]	光流 + 形状流	88.29
Sawant 等人[11]	剪影 + 局部光流	97.8
Tran 等人[8]	局部剪影 + 局部光流	96.7
本文方法	整体轮廓 + 局部光流	100

4 结束语

本文提出一种基于混合特征的人体动作识别的算法,将轮廓质心到边缘点的轮廓向量距离特征和光流特征进行结合,组成混合特征进行动作识别。从第 3 章的识别结果可以看出,这种混合特征在 Weizmann 及 KTH 数据库上得到了 95% 以上的正确识别率,充分证明了该算法的有效性及可行性。下一步工作将尝试对特征进行降维处理,进一步提高算法的计算效率,以期实现算法的实时在线应用。

参考文献:

[1] ZHU Guang-yu, XU Chang-sheng, HUANG Qing-ming. Action recognition in broadcast tennis video [C]//Proc of the 18th International Conference on Pattern Recognition. [S. l.]: IEEE Press, 2006: 251-254.

[2] WANG Liang, SUTER D. Recognizing human activities from silhouettes: motion subspace and factorial discriminative graphical model [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2007: 1-8.

[3] DEDEOLU Y. A silhouette-based method for object classification and human action recognition in video[C]//Proc of European Conference on Computer Vision. [S. l.]: IEEE Press, 2006: 64-77.

[4] GRITAI A, SHEIKH Y, SHAH M. On the use of anthropometry in the invariant analysis of human actions[C]//Proc of the 17th International IEEE Conference on Pattern Recognition. [S. l.]: IEEE Press, 2004: 923-926.

[5] LAPTEV I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2-3): 107-123.

[6] DOLLAR P, RABAUDV, COTTRELL G. Behavior recognition via sparse spatio-temporal features[C]//Proc of the 2nd Joint IEEE International Workshop on VS-PETS. [S. l.]: IEEE Press, 2005: 65-72.

[7] RONALD P. A survey on vision-based human action recognition [J]. Image and Vision Computing, 2010, 28(6): 976-990.

[8] TRAN D, SOROKIN A. Activity recognition with metric learning [C]//Proc of European Conference on Compute Vision. [S. l.]: IEEE Press, 2008: 61-66.

[9] WANG Liang, GENG Xin, LECKIE C, *et al*. Moving shape dynamics: a signal processing perspective[C]//Proc of IEEE International Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2008: 1649-1656.

[10] AHMAD M, LEE S. Human action recognition using shape and CLG-motion flow from multi-view image sequences[J]. Pattern Recognition, 2008, 41(7): 2237-2252.

[11] SAWANT N, BISWAS K. Human action recognition based on spatio-temporal features [C]//Proc of the 3rd International Conference on Pattern Recognition and Machine Intelligence. Berlin: Springer-Verlag, 2009: 357-362.

(上接第 594 页)

参考文献:

[1] AZUMA R. Recent advances in augmented reality[J]. IEEE Computer Graphics and Application, 2001, 21(6): 340-347.

[2] DAVID G L. Distinctive image features from scale-invariant key-points [J]. International Journal of Computer Vision, 2004, 20(2): 91-110.

[3] BAY H, TUYTELAARS T, GOOL L V. Speeded-up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[4] CALONDER M, LEPETIT V, STRECHA C. BRIEF: binary robust independent elementary features[C]//Proc of the 11th European Conference on Computer Vision. Berlin: Springer-Verlag, 2010: 778-792.

[5] ETHAN R, VINCENT R, KURT K, *et al*. ORB: an efficient alternative to SIFT or SURF[C]//Proc of International Conference on Computer Vision. 2011: 2564-2571.

[6] ROSTEN E, DRUMMOND T. Fusing points and lines for high performance tracking[C]//Proc of the 10th International Conference on Computer Vision. 2005: 1508-1515.

[7] KLEIN G, MURRARY D. Improving the agility of keyframe-based SLAM[C]//Proc of the 10th European Conference on Computer Vision. 2008: 802-815.

[8] ASSAD M, CARMICHAEL D J, CUTTING D. AR phone: accessible augmented reality in the intelligent environment[C]//Proc of OZCHI. 2003: 232-237.

[9] MIKA H, CHARLES W, MARK B. Augmented assembly using a mobile phone[C]//Proc of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. 2008: 167-168.

[10] ROSTEN E, PORTER R, DRUMMOND T. Faster and better: a machine learning approach to corner detection[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32(1): 105-119.

[11] MICHAEL R, BEAT G. Using camera-equipped mobile phones for interacting with real-world objects[C]//Advances in Pervasive Computing. [S. l.]: Austrian Computer Society, 2004: 265-271.

[12] ANDERS H, MARK B, MARK O. Face to face collaborative AR on mobile phones[C]//Proc of the 4th IEEE International Symposium on Mixed and Augmented Reality. 2005: 80-89.

[13] ARTH C, WAGNER D, KLOPSCHITZ M, *et al*. Wide area localization on mobile phones[C]//Proc of the 8th IEEE International Symposium on Mixed and Augmented Reality. 2009: 73-82.

[14] WAGNER D, MULLONI A, LANGLOTZ T. Real-time panoramic mapping and tracking on mobile phones [C]//Proc of IEEE Virtual Reality Conference. 2010: 211-218.