

DOI:10.11918/j.issn.0367-6234.201610081

无人机视频的分层表达与实时压缩

沈 秋, 李小凡, 严小乐, 孔繁铨

(南京航空航天大学 航天学院, 南京 210016)

摘 要: 为克服无人机视频应用中计算能力和传输带宽的限制,提出一种视频分层表达方法,实现快速、高质量、灵活的视频编码. 将无人机视频分为背景层和目标层,并分别采用基于全局运动估计和基于局部块匹配的方法进行压缩,压缩后的码流可以根据实际的网络情况 and 应用需求进行单独传输或组合传输. 实验结果表明:在极低码率如 50 kbit/s 时,H.264 的 PSNR 低于 28 dB,且图像主观质量超出可接受范围,而采用分层表达中的背景层压缩,可以在保证 PSNR 在 28 dB 以上,且图像细节清晰、主观质量较高;在码率较高时,采用背景层和目标层结合,在同等码率条件下,目标的细节比 H.264 更清晰;整个压缩过程消耗的时间只有 H.264 的 18%. 本文方法在不降低视频质量的同时提高了压缩效率,且码流具备灵活性,适用于低延时、低带宽、复杂的无人机应用.

关键词: 视频压缩;无人机视频;分层表达;全局运动估计;H.264

中图分类号: TN919.81 **文献标志码:** A **文章编号:** 0367-6234(2017)05-0068-05

Hierarchical description and real-time compression for UAV video

SHEN Qiu, LI Xiaofan, YAN Xiaole, KONG Fanqiang

(College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To overcome the limitations of compute capability and bandwidth in UAV (Unmanned Aerial Vehicle) video, this paper proposes a novel video compression algorithm based on hierarchical description. The video is split into background layer and object layer. Background layer is encoded with global motion estimation, while object layer is encoded with local block matching. The experimental results demonstrate that, when bandwidth is narrow e.g. 50 kbit/s, using background layer encoding can achieve higher PSNR than H.264, and the subjective quality is much more acceptable; in most cases, combining background and object layer can give more details than H.264 with the same bitrate; the consuming time is only 18% of H.264's. Additionally, alternative bitstream can be extracted for different bandwidth. Consequently, the proposed algorithm is suitable for low latency and complexity applications in UAV video due to its good performance, high efficiency and flexibility.

Keywords: video compression; UAV videos; hierarchical description; global motion estimation; H.264

视频作为无人机的眼睛在其任务执行中发挥着不可替代的作用,但却因其数据量巨大一直是无线传输的挑战,因此,视频压缩是无人机应用中非常重要的环节. H.264 作为当前最成熟、稳定的编码算法正广泛应用于各个行业,也是无人机视频压缩的主流算法. 文献[1-3]分别实现了基于 H.264 的无人机视频压缩的软硬件系统,文献[4-5]则针对无人机带宽条件分别提出了码率控制算法和抗干扰算法. 尽管如此,H.264 复杂的编码过程和无人机有限的计算、传输能力,使得无人机视频传输仍然存在质量不高、延时较大等问题. 为了克服这些不足近几年出现了一些非标准算法,如文献[6-8]采用的图

像拼接方法,文献[9-10]研究的全局运动估计算法,文献[11]提出的基于感兴趣区域的方法,都一定程度上提高了无人机视频压缩的效率和质量. 然而这些方法仍无法在压缩质量、可靠性、复杂度、灵活性之间取得最优的综合性能.

无人机视频往往具有一定的特殊性:飞行高度远大于地表的起伏程度,因此可以将视频背景简化为平面;视频中背景的运动主要来自于无人机的运动,具有一致性;视频中运动目标,则存在独立的运动特点. 结合以上特征,本文提出一种视频的分层表达方法,将独立运动物体从背景中分割出来,分别构建背景层和目标层,并根据每层各自的特点设计有针对性的压缩算法. 其中,背景层采用基于全局运动估计的方法,目标层则采用局部块匹配的方法. 这样不仅可以获得更快的计算速度和更高的压缩比,同时可以保留精确的运动信息和纹理信息,而且还可以根据实际应用场景的需求,灵活地选择不同

收稿日期: 2016-10-21
基金项目: 国家自然科学基金(61201365, 61401200);
江苏省普通高校研究生科研创新计划项目(SJLX15_0138);
中央高校基本科研业务费专项资金资助
作者简介: 沈 秋(1982—),女,博士,讲师
通信作者: 沈 秋,shenqiu@nuaa.edu.cn

的码流,提供更高效、可靠的视频传输方案.

1 基于全局运动估计的运动目标分割算法

由于无人机机载设备的限制,所用运动估计算法复杂度不能太高,所以本文对原有的全局运动估计算法进行改进,具体流程如图 1 所示. 将图像划分为固定数量的网格,并在网格中提取固定数量的特征点,这样不仅降低了传统特征点提取算法的复杂度,也避免了陷入局部最优. 为了适应旋转和尺度变化并达到实时性要求,本文选用文献[12]提出的 ORB(ORiented Brief)进行特征点提取与匹配,采用文献[13]提出的 RANSAC(RANdom Sample Consensus)去除误匹配,并使用最小二乘法得到视频全局运动估计的初步结果. 因无人机视频中独立运动目标的存在会对视频全局运动估计产生影响,接下来提取与全局运动不一致的区域,并进一步剔除其中的特征点,再重新进行全局运动参数的估算. 最后根据精确计算的全局运动参数,结合帧差法、形态学操作完成视频中运动目标的分割.

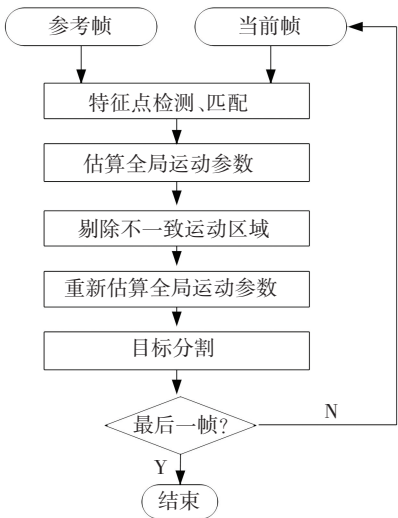


图 1 基于全局运动估计的运动目标分割流程图

Fig.1 Object segmentation based on global motion estimation

2 基于分层表达的无人机视频压缩

本文提出的分层表达主要包括两层(如图 2 所示):第一层为背景层,由静止的地面背景组成,在视频中的运动主要来自于相机运动,因此可以通过建立全局运动模型来刻画背景的运动;第二层为目标层,由实际运动的物体组成,在视频中体现出来的运动是由相机运动和物体运动共同组成,无法通过简单的运动模型来刻画. 因此,表达中的每一层应采用不同的方法进行压缩.

基于分层表达的无人机视频压缩算法主要由两部分组成(如图 3 所示),分别为目标分割和分层压

缩. 在目标分割部分,先对输入视频进行预处理以消除噪声对目标分割的影响,再采用第 1 节描述的基于全局运动估计的方法来进行分割;在分层压缩部分,根据各自的特征,背景层采用结合全局运动估计的方法,目标层则采用基于局部块匹配的方法,两者结合起来得到最终编码码流.

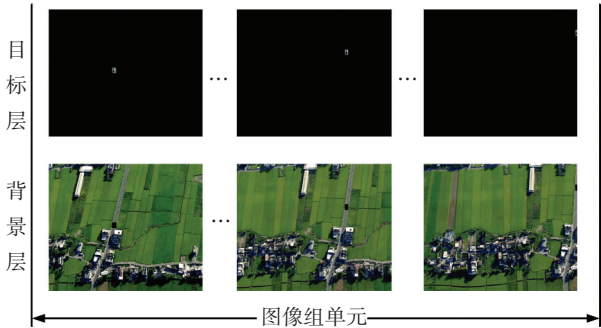


图 2 视频的分层表达结构

Fig.2 Hierarchical description of video

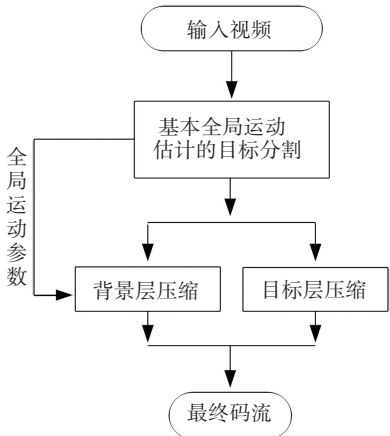


图 3 视频分层表达编码框架图

Fig.3 The frame of video coding based on hierarchical description

2.1 背景层压缩

背景层压缩算法是在 H.264 基础上实现的,通过在运动估计部分增加全局运动估计实现低复杂度的背景层压缩算法,如图 4 虚线框所示. 通过判断当前帧中每个宏块的位置,决定该宏块采用全局运动估计还是 H.264 的基于宏块的运动估计.

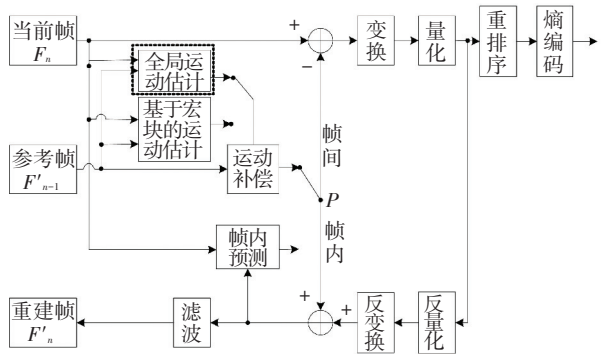


图 4 背景层压缩算法编码框图

Fig.4 Coding method of background layer

具体实现步骤如下(假设一个图像组由 N 帧组成):

1) 关键帧的选择. 关键帧是背景层中要编码的第一帧, 随后的帧根据全局运动参数映射到前一帧坐标系中, 如图 5 所示.

2) 利用全局运动估计的参数矩阵 H_{nn-1} , 将当前帧 F_n 映射到参考帧 F'_{n-1} 坐标系中, 计算每一个宏块 (x, y) 映射后的新坐标 (x', y') (注: 该坐标是指宏块左上角像素坐标).

3) 当映射后的宏块坐标超过参考帧的范围时, 被标识为未曾出现区域(图中斜线区域), 使用非 skip 模式编码; 而运动目标出现的区域(图中灰色区域)也使用非 skip 模式编码; 其余宏块(图中白色区域)使用 skip 模式编码. 映射参数将在补充增强信息(SEI)中编码.

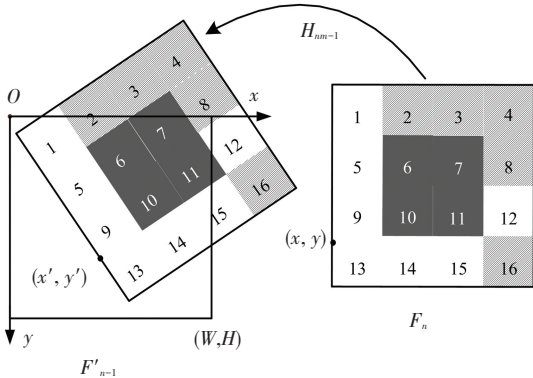


图 5 背景层编码模式判定

Fig.5 Determination of coding mode in background layer

2.2 目标层压缩

运动目标构成目标层, 采用基于局部块匹配的方式进行编码. 图 6 是目标层编码模式判定示意图, 背景黑色表示像素值为 0. 具体过程为:

1) 利用运动目标分割的结果, 得到运动目标整体在图像坐标系中位于左上角的坐标 $P_{tl}(x_{\min}, y_{\min})$

和右下角的坐标 $P_{br}(x_{\max}, y_{\max})$.

2) 运动目标所在区域为以宏块为单位的包括 $P_{tl}(x_{\min}, y_{\min})$ 和 $P_{br}(x_{\max}, y_{\max})$ 的最小矩形框.

3) 目标区域的宏块采用 H.264 帧间预测模式进行编码, 而非运动目标区域采用 H.264 的 skip 模式编码. 通过这种方式, 可以使编码速度大大增加, 同时编码性能也得以提高.

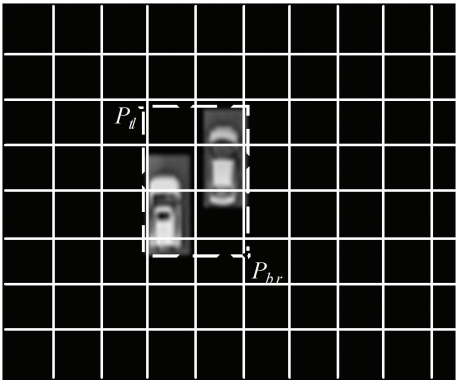


图 6 目标层编码模式判定

Fig.6 Determination of coding mode in object layer

3 实验结果与分析

实验分为仿真视频和实际航拍视频两组, 限于篇幅, 每组只选择一个视频进行展示. 实验的硬件平台为 HP Z820 图形工作站(主频 2.3 GHz、内存 14 G), 软件平台为 Win7 操作系统+ Microsoft Visual Studio 2010、OpenCV2.3.1 以及 H.264 压缩标准参考软件平台 JM12.1.

3.1 仿真视频实验

仿真视频是利用谷歌地球的飞行模拟器从一定海拔高度, 以一定的速度和航线对地面垂直俯拍所得, 视频信息如表 1 所示, 并且仿真视频中没有运动目标.

表 1 仿真视频参数信息

Tab.1 Parameters of artificial video

视频名称	视频分辨率/(pixel×pixel)	帧数	拍摄高度/m	运动目标个数	视频运动情况
仿真视频.yuv	720×480	30	650	0	平移和旋转

对该视频分别采用本文提出的背景层压缩算法和 H.264 进行编码, 得到的 R-D(Rate-Distortion, 率失真)曲线如图 7 所示. 当码率小于 75 kbit/s 时, 本文提出的算法不需要将运动矢量及模式选择信息写入到码流中, 因而可以以很低的码率得到较好的压缩质量. 但随着码率的增加, 本文算法的 PSNR 却没有像 H.264 一样较大幅度的提高, 主要原因是: 一方面, H.264 压缩算法的运动搜索采用 1/4 像素精度,

高于本文使用的基于特征匹配的全局运动估计; 另一方面, 以较少的全局运动参数代替所有宏块的运动参数一定存在微小的误差, 这些误差导致重建的图像与原始图像会有 1 个像素左右的位移, 这虽然不影响视频的主观质量, 但以像素为单位进行计算的 PSNR 值却会受到极大的影响. 图 8 给出了两种方法的在不同码率下重建图像的局部细节, 可见无论在低码率还是高码率的情况下, 本文算法的重建

图像的主观质量都高于 H.264.

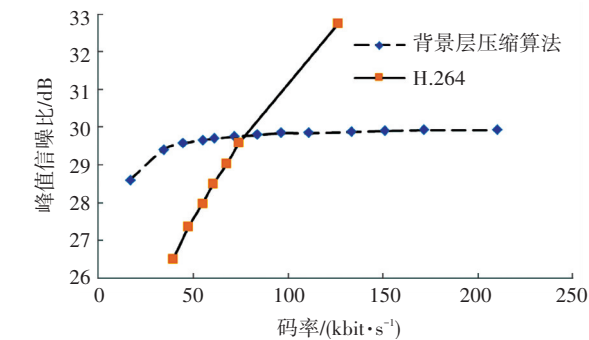


图 7 低码率区间编码率失真曲线对比

Fig.7 R-D curve comparison at low bitrate

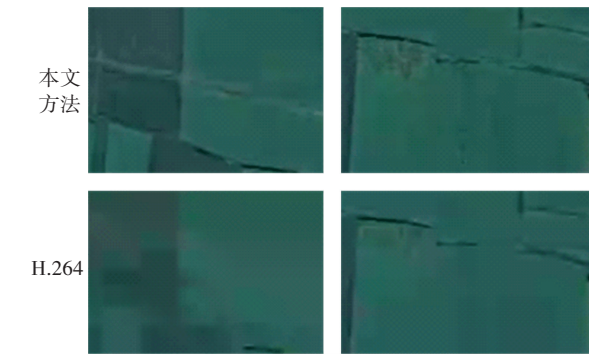


图 8 不同码率下的主观质量对比

Fig.8 Comparison of subjective quality at different bitrate

本文算法相比于 H.264 的最大优势在于计算效率,从表 2 可以看出,本文方法因采用全局运动估计代替了 H.264 中最耗时的基于宏块的运动估计和模式选择,大大减少了编码时间.

表 2 仿真视频运动估计时间及压缩编码时间比较
Tab.2 Comparison of time consuming of motion estimation and encoding

视频名称	压缩方式	运动估计 时间/s	压缩编码 时间/s	压缩编码 时间比
仿真视频.yuv	背景层压缩	0.33	7.43	1 : 16.50
	H.264	98.73	122.61	

3.2 实际航拍视频实验

该视频于 2015 年 6 月 7 日在南京市江宁区一处农田附近使用大疆精灵 3 无人机拍摄,视频中有 2 辆并行行驶的汽车以及行人等运动目标. 无人机对地俯拍,行人在视频影像中所占像素数量相对于汽车而言可以忽略不计,属于极微小的运动目标. 本文为简化处理,认为视频中独立运动目标数量为 2 个,分割出的运动目标也为 2 个.

获取的视频信息如表 3 所示. 由于背景层与原始视频相比缺少了运动目标,无法进行客观质量的评价,在相似码率下比较本文方法的背景层压缩和 H.264 压缩的主观质量,图 9 给出了背景部分的局部放大图. 可见本文算法在 rate = 134.3 kbit/s 时背

景部分边缘清晰、纹理详细,而 H.264 在 rate = 158.8 kbit/s 时压缩后的背景部分非常模糊.

表 3 实际航拍视频参数信息
Tab.3 Parameters of aerial video

视频分辨率/ pixel×pixel	帧数/ frame	拍摄高度/ m	运动目标 个数	视频运动 情况
1 920×1 080	30	55	2	平移、旋转、缩放

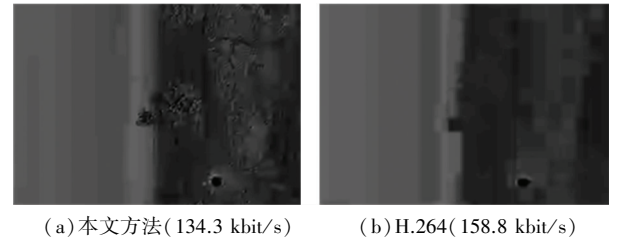


图 9 本文方法的背景层和 H.264 压缩主观质量对比

Fig.9 Subject quality comparison of background layer with H.264

目标层压缩实验用于比较本文提出的目标层压缩算法和 H.264 标准压缩算法压缩目标层的效率和性能. 如图 10 所示,两者 R-D 曲线非常接近,甚至目标层编码略高于 H.264. 而在压缩效率方面,由于目标层压缩是大部分的宏块直接采用 skip 模式编码,不需再进行复杂的模式选择,编码所需时间仅相当于 H.264 的 6% 左右(如表 4 所示).

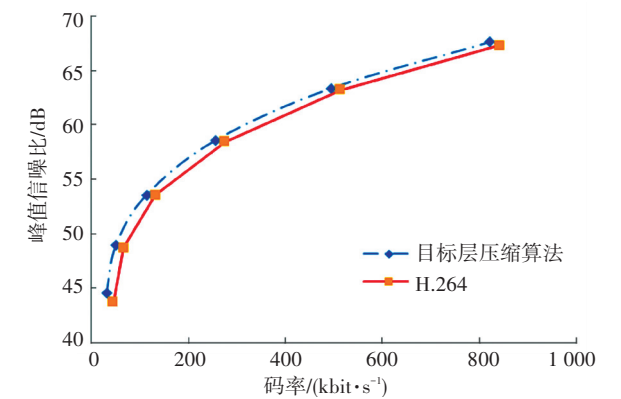


图 10 目标层压缩 R-D 曲线对比
Fig.10 R-D curve comparison of object layer

表 4 目标层压缩时间对比
Tab.4 Coding time comparison of object layer with H.264

压缩方式	压缩编码时间/s	压缩编码时间比
目标层压缩	43.27	1 : 16.98
H.264	734.75	

为了达到更好的压缩效果,本文提出的视频的分层表达可以实现对背景层和目标层采用不同压缩方法和压缩质量,从而适应不同的带宽条件和应用需求. 一方面,在极低码率时,无法提供运动目标的细节,则可以只传输背景层码流,使接收方对视频的整体内容有更好的体验,并且背景层码流还可以通过全局运动补偿实现快速的全景图拼接;另一方面,

在一般码率条件下,本文方法可以为背景层和目标层设置不同的量化参数,充分保证目标层的质量.图 11 为在 $\text{rate} = 203.89 \text{ kbit/s}$ 时,背景层采用量化参数 $\text{QP} = 42$,而目标层采用量化参数 $\text{QP} = 35$ 时编码重建帧的主观图.可见本文算法重建帧中背景和目标的清晰度不同,目标区域更加清晰,而背景区域相对目标而言较模糊,从而实现了对目标区域的感兴趣编码,并且在相似码率下,本文算法背景部分和目标部分都比 H.264 标准压缩算法重建帧更清晰.此外,从表 5 可知,本文算法的编码时间也远远小于 H.264.

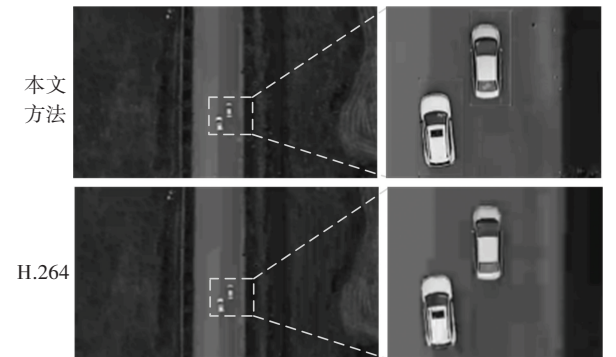


图 11 实际航拍视频压缩后主观质量对比

Fig.11 Subject quality comparison of compressed aerial video

表 5 实际航拍视频压缩编码时间比较

Tab.5 Coding time comparison of aerial video

压缩方式	压缩编码时间/s	压缩编码时间比
背景层+目标层压缩	134.11	1 : 5.71
H.264	766.21	

4 结 语

本文针对无人机视频的特征和特定需求,提出一种新的视频压缩编码方法,通过分层表达对视频进行分解和压缩,从而提供了更高效、快速、灵活的无人机视频传输方案.一方面,该算法在大部分码率条件下都可以得到高于 H.264 的压缩质量,同时需要的计算时间只有 H.264 的 18% 左右;另一方面,这种分层的表达形式可以根据实际的传输条件,选择不同的码流,具有更好的适应性.本文算法的不足在于:在高码率时 PSNR 值较低,体现为整体画面可能与实际存在微小的位置偏移;当无人机飞行高度低,背景不能被视为平面的情况时,压缩效果将降低.因此下一步的研究重点是进一步借鉴更新的编码算法,优化编码步骤,建立更精确的运动模型,提高算法的适用性和有效性.

参考文献

[1] 李枫. 基于 H.264 的航拍视频实时压缩系统[D].北京:北京理

工大学,2011.
LI Feng. System of real-time compression of aerial video based on H.264[D]. Beijing: Beijing Institute of Technology, 2011.
[2] 郭昕. 微型无人机视频压缩的研究[D].天津:天津大学,2012.
GUO Xin. Research of video compression on micro-UAV[D]. Tianjin: Tianjin University, 2012.
[3] 吴笑天,鲁剑锋,王宇庆,等. 基于 DM368 的无人机视频压缩系统的设计[J]. 液晶与显示,2014,29(6):1117-1123.
WU Xiaotian, LU Jianfeng, Wang Yuqing, et al. Design of UAV video encoding system based on DM368[J]. Chinese Journal of Liquid Crystals & Displays, 2014, 29(6):1117-1123.
[4] 赵春蕾,戴明,尹传历. 无人机机载侦察视频的高效视频码率控制[J]. 光学精密工程,2015,23(12):3500-3508.
ZHAO Chunlei, DAI Ming, YIN Chuanli. HEVC rate control for unmanned aerial vehicle airborne reconnaissance[J]. Optics & Precision Engineering, 2015, 23(12):3500-3508.
[5] 张露. 无人机低时延抗误码音视频编解码系统软件设计[D].杭州:浙江大学,2016.
ZHANG Lu. Software design of low-latency error-resilient audio-video codec system for UAV[D]. Hangzhou: Zhejiang University, 2016.
[6] KRUTZ A, GLANTZ A, SIKORA T. Recent advances in video coding using static background models[C]//Picture Coding Symposium (PCS). Nagoya: IEEE, 2010: 462-465.
[7] 谢清鹏. 无人机序列图像压缩方法研究[D].武汉:华中科技大学,2005.
XIE Qingpeng. Research on compression methods of unmanned aerial vehicle's image sequence[D]. Wuhan: Huazhong University of Science and Technology, 2005.
[8] KANG C B, DING W R, YUAN Y X, et al. Integrated compression coding method for UAV reconnaissance video and still image[M]. Foundations of Intelligent Systems. Berlin, Heidelberg: Springer, 2014: 1067-1075.
[9] 王云丽,张鑫,高超,等. 航拍视频拼图中基于特征匹配的全局运动估计方法[J]. 航空学报, 2008, 29(5): 1218-1225.
WANG Yunli, ZHANG Xin, GAO Chao, et al. Feature matching based global motion estimation in aerial video mosaicing[J]. Acta Aeronautica et Astronautica Sinica, 2008, 29(5): 022.
[10] 沈秋,李小凡,孔繁镔,等. 基于仿射模型的无人机视频实时压缩算法[J]. 电子与信息学报, 2014, 36(12): 2855-2860.
SHEN Qiu, LI Xiaofan, KONG Fanqiang, et al. A real-time video compression for UAV based on affine model[J]. Journal of Electronics & Information Technology, 2014, 36(12): 010.
[11] BHASKARANAND M, GIBSON J D. Low-complexity video encoding for UAV reconnaissance and surveillance[C]//Military Communications Conference, MILCOM 2011. Baltimore: IEEE, 2011: 1633-1638.
[12] RUBLEE E, RABAU V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//2011 IEEE International Conference on Computer Vision (ICCV). Barcelona: IEEE, 2011: 2564-2571.
[13] 张波,罗海勇,刘冀伟,等. 视觉传感器网络中基于 RANSAC 的稳健定位算法[J]. 通信学报, 2013, 34(8): 62-69.
ZHANG Bo, LUO Haiyong, LIU Jiwei, et al. RANSAC based robust localization algorithm for visual sensor network[J]. Journal on Communications, 2013, 34(8):62-69.

(编辑 王小唯, 苗秀芝)