

基于运动谱残差的视频显著性检测算法

刘宇光^{1,2}, 陈耀武^{1,2}

(1. 浙江大学数字技术及仪器研究所, 杭州 310027; 2. 浙江省网络多媒体技术研究重点实验室, 杭州 310027)

摘 要: 借鉴静态图像中通过计算像素灰度的谱残差检测显著性区域的方法, 提出基于运动谱残差的视频显著性检测算法。通过提取每一帧的运动矢量场, 对矢量场的水平和垂直分量分别计算运动谱残差, 融合 2 个方向的运动谱残差形成运动谱残差图, 将谱残差图进行顶帽变换形成视频显著图。实验结果表明, 该算法能准确地分割出各种运动场景的显著性区域, 在检测效果和抗噪能力方面优于现有视频显著性检测算法, 适用于运动微生物检测、行人车辆检测等领域。

关键词: 视频显著性; 谱残差; 光流法; 运动矢量; 顶帽变换; 图像增强

中文引用格式: 刘宇光, 陈耀武. 基于运动谱残差的视频显著性检测算法[J]. 计算机工程, 2014, 40(12): 247-250, 257.

英文引用格式: Liu Yuguang, Chen Yaowu. Video Saliency Detection Algorithm Based on Motion Spectral Residual[J]. Computer Engineering, 2014, 40(12): 247-250, 257.

Video Saliency Detection Algorithm Based on Motion Spectral Residual

LIU Yuguang^{1,2}, CHEN Yaowu^{1,2}

(1. Institute of Advanced Digital Technologies and Instrumentation, Zhejiang University, Hangzhou 310027, China;

2. Zhejiang Provincial Key Laboratory for Network Multimedia Technologies, Hangzhou 310027, China)

[Abstract] This paper references the idea of detecting salient area in still image through calculating spectral residual of the gray-level pixel value, so proposes a video saliency detection algorithm based on motion spectral residual. It extracts each frame's motion vector field, calculates motion spectral residual of the horizontal and vertical components of motion vector field, merges the two components' motion spectral residual to form a motion spectral residual image, and finally exerts a top-hat transformation on the motion spectral residual image to form video saliency image. Experimental results show that this algorithm can accurately segment saliency region out of various motion scenes, outperform existing video saliency detection algorithms in terms of detection accuracy and noise resistance. The algorithm can be applied to such field as moving microbe detection, pedestrian detection and vehicle detection.

[Key words] video saliency; spectral residual; optical flow method; motion vector; top-hat transformation; image enhancement

DOI: 10.3969/j.issn.1000-3428.2014.12.046

1 概述

视觉显著性和人类视觉神经系统紧密关联, 它使得视觉场景中的一个区域能从整个场景中突显出来, 从而捕捉到人类的注意力, 有助于对图像和视频语义的理解和感兴趣区域的提取, 因而视觉显著性成为目前图像处理 and 计算机视觉等领域的研究热点。

在过去十几年中, 视觉显著性研究目标主要集中在静态图像。文献[1]提出显著性模型, 计算灰

度、颜色对和方向这 3 个图像特征的多尺度中心-周边差异, 并将它们线性融合得到显著性图。文献[2]提出基于图的不相似度量的显著性计算方法。文献[3]利用图像的 FFT 谱残差得到了显著图。文献[4]提出基于直方图对比度和基于空间信息增强的区域对比度的方法。文献[5]提出基于对比度的显著性滤波器, 它计算图像的超像素成分的独特性和分布性, 并将两者按指数方式融合得到显著图。文献[6]提出基于图流形排序的显著性检测方法, 根据图像的超像素成分与先验背景的色彩差异进行排

基金项目: 国家自然科学基金资助项目(40927001)。

作者简介: 刘宇光(1988 -), 男, 硕士研究生, 主研方向: 图像处理, 视频分析; 陈耀武, 教授、博士。

收稿日期: 2013-12-19 **修回日期:** 2014-01-15 **E-mail:** yumkongleo@gmail.com

序得到显著图。

视频序列和静态图像在显著性检测方面有很大差别,这是由于在视频序列中一个快速移动的目标往往比静态背景更能吸引人的注意力。一个好的视频显著性检测方法应该充分利用时域特征和运动信息来捕捉最能吸引关注的区域。混合高斯模型^[7]通过高斯概率密度函数来建立背景模型并实时更新,具有显著运动特性的前景物体可以通过图像与背景相减得到。文献[8]通过时空线索检测视觉注意力,用 SIFT 点对来表达时域显著性,用颜色距离差表达空域显著性,将两者动态融合得到最终的显著图。文献[9]通过计算灰度、颜色对和运动信息的四元傅里叶变换的相位谱得到显著图。文献[10]通过将谱残差方法扩展到时域的方法来检测运动显著性。但以上方法只能检测出视频序列中显著目标的位置和大致轮廓,倾向于目标检测和跟踪应用,而在目标识别方面并不实用。本文提出基于运动谱残差的视频显著性检测算法,能够对视频序列中的显著目标边界进行准确检测,方便后续的特征提取和运动目标识别等工作。

2 运动谱残差

对于视频的运动显著性检测,关键是如何从种类繁多的背景中区分出前景物体有意义的运动,在静态图像中灰度、色彩的差异等特征是很好的显著性衡量方法,而动态视频流中运动矢量的差异往往更能吸引人的关注。受文献[3]的启发,对运动矢量做出了如下假设:视频序列相邻两帧所产生的运动矢量场 $MVF(frame)$ 可以分解为两部分:

$$MVF(frame) = MVF(innovation) + MVF(prior_knowledge) \quad (1)$$

其中, $MVF(innovation)$ 表示创新部分,该部分是运动显著性区域; $MVF(prior_knowledge)$ 表示先验知识部分,表示静止的背景部分或是场景整体的运动(例如摄像机的移动)。将若干视频序列的运动矢量场的水平分量进行叠加、归一化,然后求取它们的傅里叶谱幅值的对数(Logarithm of the Fourier Amplitude Spectrum, LFAS),得到的结果如图1、图2所示。

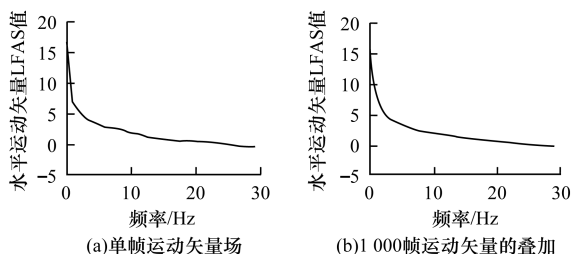


图1 运动矢量场水平分量的对数谱

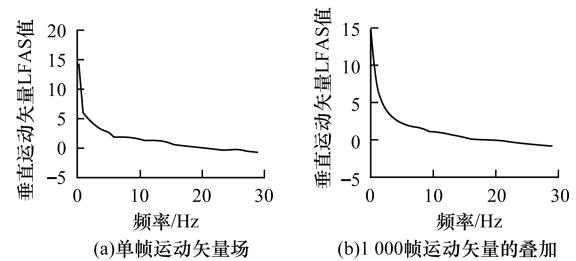


图2 运动矢量场垂直分量的对数谱

通过图1、图2可以看出,在单帧运动矢量场的水平和垂直分量在经过傅里叶变换后幅值的对数与频率的关系曲线是不平滑的,而多个运动矢量场的水平分量和垂直分量分别进行叠加后再进行同样运算所得的关系曲线是近似平滑的,这和文献[3]中对静态图像的灰度值的求对数谱得到的结果非常相似。因此,进一步进行运动矢量谱残差的计算和融合,得到的结果图体现了视频序列中运动反差强烈区域,即最能吸引人眼注意的显著区域。因此,证明了本文假设:运动矢量的谱残差可以表示视频的运动显著性。

3 视频显著性检测

视频显著性检测整体分为4个阶段:运动矢量场计算,运动矢量场分方向计算谱残差,不同方向的谱残差加权融合和运动显著图增强处理。整个算法流程如图3所示。

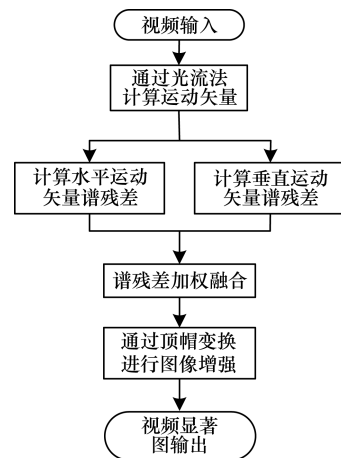


图3 视频显著性检测算法流程

3.1 运动矢量场计算

通过文献[11]中的光流计算方法(classic + nl-fast)类型的光流估计法来获得输入视频的运动矢量场信息。该方法得到与输入视频同样分辨率的运动矢量场 MVF ,再从 MVF 分别提取水平、垂直方向的运动矢量矩阵 MVF_{hor} 和 MVF_{ver} 。将运动矢量矩阵归一化,方便后续处理,计算公式为:

$$MVF_{hor_norm} = \frac{MVF_{hor} - \min(MVF_{hor})}{\max(MVF_{hor}) - \min(MVF_{hor})} \quad (2)$$

$$MV_{\text{ver_norm}} = \frac{MV_{\text{ver}} - \min(MV_{\text{ver}})}{\max(MV_{\text{ver}}) - \min(MV_{\text{ver}})} \quad (3)$$

3.2 运动矢量谱残差计算

水平方向和垂直方向的运动矢量的谱残差计算方法是相同的,这里以水平方向为例。首先对水平运动矢量矩阵进行二维傅里叶变换,获得变换后的幅值和相位。计算公式为:

$$A(f) = \text{Re}(\mathfrak{F}(MV_{\text{hor_norm}})) \quad (4)$$

$$P(f) = \text{Im}(\mathfrak{F}(MV_{\text{hor_norm}})) \quad (5)$$

其中, \mathfrak{F} 表示傅里叶变换; $A(f)$ 表示傅里叶谱的幅值(幅值谱); $P(f)$ 表示傅里叶谱的相位(相位谱)。

计算幅值谱的自然对数(对数谱),计算公式为:

$$L(f) = \ln(A(f)) \quad (6)$$

将对数谱通过与掩膜进行卷积的方式进行平滑滤波,计算公式为:

$$L'(f) = h_n(f) * L(f) \quad (7)$$

其中, $h_n(f)$ 为 $n \times n$ 的平滑掩膜,这里取 n 等于 5,即:

$$h_n(f) = \frac{1}{5^2} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (8)$$

然后计算对数谱与平滑滤波后的对数谱的差值:

$$R(f) = L'(f) - L(f) \quad (9)$$

将对数谱差值作为实部,原相位谱为虚部进行傅里叶反变换,并对结果进行平方,再通过高斯平滑滤波器进行平滑,得到显著图 $Salmap_{\text{hor}}$, 计算公式为:

$$Salmap_{\text{hor}} = g(x) * [\mathfrak{F}^{-1}(e^{R(f) + P(f)})]^2 \quad (10)$$

其中, $g(x)$ 表示高斯低通滤波器,这里选择标准差为 3; \mathfrak{F}^{-1} 表示傅里叶反变换。

垂直方向的运动显著图 $Salmap_{\text{ver}}$ 可以通过相同的计算步骤获得。

3.3 谱残差加权融合

在实验中发现简单将 3.2 节得到的 2 个方向的运动显著图进行相加并不会取得很好的效果,这是由于在特定视频中某一个方向的前景和背景的运动差异是处于主导地位的,相应的显著图可以很好地反映运动显著性,而另一个方向的前景背景运动差异是微弱的,更多地反映了运动噪声成分。因此,采用每个方向的运动矢量的标准差作为加权系数,将 2 个方向的运动显著图融合,计算公式为:

$$\lambda_{\text{hor}} = \text{std}(MV_{\text{hor_norm}}) \quad (11)$$

$$\lambda_{\text{ver}} = \text{std}(MV_{\text{ver_norm}}) \quad (12)$$

$$Salmap = \lambda_{\text{hor}} \times Salmap_{\text{hor}} + \lambda_{\text{ver}} \times Salmap_{\text{ver}} \quad (13)$$

3.4 显著图增强

由于视频序列中的图像光照不均且存在噪声,3.3 节得到的显著图会出现轮廓不连续的情况,利用顶帽变换的方法^[12]进行噪声消除和轮廓补全。

顶帽变换分为白顶帽变换和黑顶帽变换,分别定义如下:

$$T_{\text{white}}(f) = f - f \circ \text{mask} \quad (14)$$

$$T_{\text{black}}(f) = f \cdot \text{mask} - f \quad (15)$$

其中, \circ 表示形态学开运算; \cdot 表示形态学闭运算; mask 表示运算掩膜。原图像加白顶帽变换后的图像可以增加灰度值较大区域,减去黑顶帽变换可以抑制灰度值较小区域,因此利用式(16)获得最终显著图,其中,白顶帽变换和黑顶帽变换的掩膜均为直径为 10 个像素的圆形掩膜:

$$Salmap_{\text{final}} = Salmap + T_{\text{white}}(Salmap) - T_{\text{black}}(Salmap) \quad (16)$$

4 实验结果与分析

为验证上述算法的有效性,比较了本文提出的 MSR 方法和其他 4 种先进方法:混合高斯模型(GMM)^[7],基于图的抖动特征通道的显著性模型(GB)^[2],基于时空线索的显著性模型(STC)^[8]和基于谱残差的显著性模型(SR)^[3]。在这 4 种方法中,混合高斯模型(GMM)^[7]、基于图的抖动特征通道的显著性模型(GB)^[2]和基于时空线索的显著性模型(STC)^[8]是结合时空信息的显著性模型,基于谱残差的显著性模型(SR)^[3]是静态的空域信息的显著性模型。实验结果如图 4 所示。

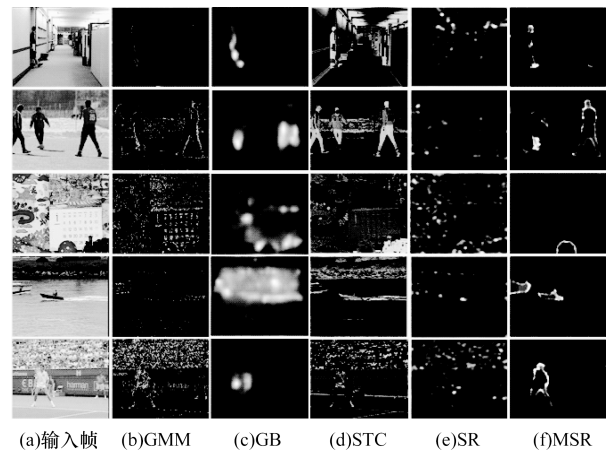


图4 视频显著性检测的对比实验结果

实验选取 5 类不同的视频序列进行测试:(1)有单个运动物体场景(图 4 中第 1 行 6 张图片);(2)有多个运动物体的场景(图 4 中第 2 行 6 张图片);(3)背景复杂的场景(图 4 中第 3 行 6 张图片);(4)相机在拍摄过程中抖动的场景(图 4 中第 4 行 6 张图片);(5)相机在拍摄过程中抖动且背景复杂

的场景(图4中第5行6张图片)。

由图4可以看出,本文方法(图4(f))在所有的5类视频场景中有突出的表现,可以清晰反映运动物体的轮廓,从而凸显运动显著性。具体来说,与图4(b)的GMM方法相比,MSR方法可以在所有视频场景中排除噪声干扰、清晰地表现出运动区域,而GMM对于复杂背景或者相机抖动的视频会产生较多噪声;对于图4(c),包含闪烁特征通道的GB方法只能反映出运动目标的大致位置,且在复杂背景或者相机抖动时也存在较多噪声;对于图4(d),STC方法存在着较多的背景信息,不利于对显著目标的精确提取;对于图4(e),它是基于静态图像的显著性检测,只能反映图像具有复杂纹理、强烈空域反差出的信息,并不能体现出运动显著性。总体来说,本文方法能在不同视频场景中工作良好是因为基于运动矢量场的方法可以有效地反映视频的运动信息,而且通过水平、垂直运动方向的谱残差运算和根据不同方向的运动强度信息的加权融合,可以有效地滤去噪声部分,体现出有效的运动信息。

为了客观评价本文方法的效果,按照文献[13]方法对各显著图进行分割,并与实际分割结果进行比较,得到精确率(Precision),查全率(Recall)和F度量值3个指标,定义分别如下:

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

$$Recall = \frac{TP}{TP + FN} \quad (18)$$

$$F = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (19)$$

其中,TP表示算法呈现的显著区域中本来就是显著部分的像素个数;FP表示算法呈现的显著区域中不应该是显著部分的像素个数;FN表示算法呈现的非显著区域中本来是显著部分的像素个数;参数 β^2 设为0.3。

图5显示了各算法对5种视频类型进行处理,得到的3个指标的平均值。

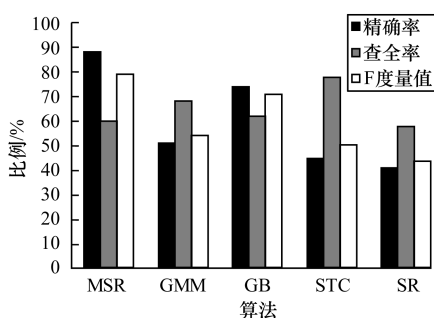


图5 视频精确率、查全率和F度量值的比较

从图5可看出,本文算法在整体指标上优于其他4种算法。

5 结束语

本文通过求运动矢量的谱残差获得视频序列的运动显著性检测算法。该算法通过运动矢量在水平、垂直2个方向的分量值分别进行谱残差计算,再进行线性融合和图像增强来获得轮廓分明、噪声很小的显著图。实验结果表明,与其他算法相比,本文算法无论在主观效果还是客观指标上都有较好的结果,但其生成的显著图并不一定是轮廓闭合的,这会导致在二值化分割时本来属于显著区域的部分被划分到背景区域中,查全率不高,而且使用光流法计算运动矢量较耗时。因此,如何合理地补全轮廓和提高运动矢量的计算效率将是后续研究的重点。

参考文献

- [1] Itti L, Koch C, Niebur E. A Model of Saliency-based Visual Attention for Rapid Scene Analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [2] Harel J, Koch C, Perona P. Graph-based Visual Saliency [C]// Proceedings of NIPS '06. [S. l.]: IEEE Press, 2006: 545-552.
- [3] Hou Xiaodi, Zhang Liqing. Saliency Detection: A Spectral Residual Approach [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE Press, 2007: 1-7.
- [4] Cheng Mingming, Zhang Guoxin, Mitra N J, et al. Global Contrast Based Salient Region Detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: Springer, 2011: 409-416.
- [5] Perazzi F, Krahenbuhl P, Pritch Y, et al. Saliency Filters: Contrast Based Filtering for Salient Region Detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE Press, 2012: 733-740.
- [6] Yang Chuan, Zhang Lihe, Lu Huchuan, et al. Saliency Detection via Graph-based Manifold Ranking [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE Press, 2012: 3166-3173.
- [7] Zivkovic Z. Improved Adaptive Gaussian Mixture Model for Background Subtraction [C]// Proceedings of IEEE Conference on Pattern Recognition. Cambridge, UK: IEEE Press, 2004: 28-31.
- [8] Zhai Yun, Shah M. Visual Attention Detection in Video Sequences Using Spatiotemporal Cues [C]// Proceedings of the 14th Annual ACM International Conference on Multimedia. New York, USA: ACM Press, 2006: 815-824.
- [9] Guo Chenlei, Ma Qi, Zhang Liming. Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2008: 1-8.

(下转第257页)

图7所示,采用本方法所述的自适应压缩算法传输实时图像,其延时与 $\omega = 40\%$ 情况进行对比。

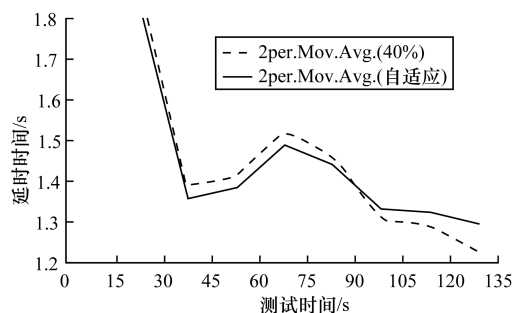


图7 使用 $\omega = 40\%$ 与使用自适应算法的延时比较

可以看到,使用本方法产生的延时并没有因为压缩质量因子在某些时刻的提高而加大,这是因为在网络传输能力强的前提下增加压缩质量,增加的数据量不会导致网络的瞬间堵塞;而当网络传输环境变差时,算法主动减小质量因子,减小网络负荷,保证传输的实时性。同样,在 $\omega = 60\%$ 和 $\omega = 80\%$ 时也可以得到相似结论。

5 结束语

本文通过引入贝叶斯分类器,实现了一种自适应的实时视频传输方法,并通过实验证明了该方法的有效性。下一步将针对分包和组包过程的延时较长问题,对本文方法进行改进,考虑自适应调整数据包的大小^[16],保证视频在高质量传输的同时^[17],进一步缩短传输延时,以提高系统的实用性。

参考文献

- [1] Liu Gaoping, Li Guosheng. Real-time Video Transmission Algorithm Based on Narrow and Variable Bandwidth[J]. Journal of Optoelectronics Laser, 2012, 55(1): 35-46.
- [2] Lee M L, Ceebin G. Classified Vector Quantisation with Variable Block-size DCT Models[J]. Browse Journals & Magazines, 1994, 141(1): 39-48.
- [3] 宋璿琪. 3D-DCT 自适应彩色图像压缩算法的研究[D]. 太原: 太原理工大学, 2012.

- [4] 黄战华, 蔡怀宇. 自适应量化表的 JPEG 压缩技术[J]. 光电子·激光, 2000, 11(5): 514-517.
- [5] 黄战华, 蔡怀宇. 自适应优化匹配量化表的 JPEG 压缩技术[J]. 光电子·激光, 2000, 11(6): 631-634.
- [6] 赵 嵩. 自适应数字图像压缩研究[D]. 大连: 大连海事大学, 2007.
- [7] Du Qinsheng, Jiang Baohua. Design and Implementation of the Embedded Based Web Camera System [J]. Journal of Software, 2012, 11(7): 2560-2566.
- [8] Hu Yuanyuan, Niu Xiamu. Image Hashing Algorithm Based on Robust Bits Extraction in JPEG Compression Domain [J]. Information Technology Journal, 2010, 9(1): 152-157.
- [9] 鲁业频, 李凤亭. 离散余弦变换编码的现状与发展研究[J]. 通信学报, 2004, 25(2): 106-118.
- [10] 叶冬芬, 董建平. H. 263 视频编解码器中量化过程的研究[J]. 浙江工业大学学报, 2006, 34(8): 429-432.
- [11] 李旭升, 郭耀煌. 基于多重判别分析的朴素贝叶斯分类器[J]. 信息与控制, 2005, 34(5): 581-584, 590.
- [12] Margineantu D D. Class Probability Estimation and Cost-sensitive Classification Decisions [C]//Proceedings of the 13th European Conference on Machine Learning. [S. l.]: IEEE Press, 2002: 270-281.
- [13] Mishra P, Singh D, Yamaguchi Y. Land Cover Classification of Palsar Images by Knowledge Based Decision Tree Classifier and Supervised Classifiers Based on SAR-observables [J]. Progress in Electromagnetics Research B, 2011, 30(1): 47-70.
- [14] Yi Mao Hsiao. H. 264 Video Transmissions over Wireless Networks: Challenges and Solutions [J]. Computer Communications, 2011, 34(14): 1661-1672.
- [15] Bradski G, Kaebler A. 学习 OpenCV [M]. 于仕琪, 刘瑞祯, 译. 北京: 清华大学出版社, 2013.
- [16] Goel S, Ismail Y, Bayoumi M. High-speed Motion Estimation Architecture for Real-time Video Transmission [J]. The Computer Journal, 2012, 39(1): 35-46.
- [17] Deng Chenwei, Lin Weisi. Performance Analysis, Parameter Selection and Extensions to H. 264/AVC FReXt for High Resolution Video Coding [J]. Journal of Visual Communication & Image Representation, 2011, 22(8): 749-759.

编辑 索书志

(上接第250页)

- [10] Cui Xinyi, Liu Qingshan, Metaxas D. Temporal Spectral Residual: Fast Motion Saliency Detection [C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2009: 617-620.
- [11] Sun Deqing, Roth S, Black M J. Secrets of Optical Flow Estimation and Their Principles [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: [s. n.], 2010: 2432-2439.

- [12] 张 雯, 葛玉荣. 基于形态学与不完全树形小波分解的图像纹理特征提取算法[J]. 计算机应用, 2011, 31(6): 1592-1594.
- [13] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned Salient Region Detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Press, 2009: 1597-1604.

编辑 陆燕菲