一种基于 H.264/AVC 压缩域的 GOP 级视频场景转换检测算法

高 宇,卓 力,王素玉,沈兰荪
(北京工业大学信号与信息处理研究室,北京 100124)

摘 要: 本文提出了一种基于 H.264/AVC 压缩域的 GOP(Group of Pictures)级视频场景转换检测算法.该算法利用 H.264/AVC 基本档次码流中的帧内预测模式、运动矢量和宏块编码类型等可用信息,提出了基于子块的色度模式差异、累积运动值和累积帧内宏块数等三个判决准则,然后综合利用这三个判决准则,提出了一种 GOP 级的视频场景转换检测算法.实验结果表明,与现有的一个 GOP 级场景检测算法对比,本文提出的 GOP 级视频场景转换检测算法可以获得更好的检测性能.

关键词: 视频; 场景检测; H.264/AVC; 压缩域
中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2009) 02-0382-05

A GOP Level Video Scene Change Detection Algorithm in H.264/AVC Compression Domain

GAO Yu, ZHUO Li, WANG Su-yu, SHEN Lan-sun
(*Signal and Information Processing Lab*, *Beijing University of Technology*, *Beijing* 100124, *China*)

Abstract: A GOP level video scene change detection algorithm in H.264/AVC baseline profile compression domain is proposed. The algorithm is based on intra prediction modes, motion vectors and macroblock coding types which can be extracted from H.264/AVC baseline bitstream. Three new criteria, the difference of chroma prediction modes based on sub-block, the accumulative motion amount and the accumulative intra coding macroblock amount, are proposed. Experimental results show that compared with another typical GOP level method, the proposed algorithm can achieve better performance.

Key words: video; scene change detection; H.264/AVC; compression domain

1 引言

视频场景检测是视频内容分析与过滤、视频索引与检索等应用中的关键技术,其性能的好坏对后续的工作有着直接的影响.现有的视频场景检测方法可以分为基于像素域的检测和基于压缩域的检测两种.基于压缩域的场景检测,是指通过挖掘视频压缩码流中包含的信息,力争在不解码或部分解码的情况下检测出场景的转换.与基于像素域的场景检测方法相比,基于压缩域的检测具有处理的数据量少,速度快,实时性高,特征提取方便等优点.

根据应用需求的不同,视频场景的转换检测的精度可以不同,如帧级别和 GOP 级别.帧级别的视频场景检测可以确定具体哪一帧发生了场景的转换,而 GOP 级的视频场景检测可以检测出一个 GOP 中是否发生了场景的转换.但是无论是场景突变还是短渐变,都无法指出具体哪一帧发生了突变或哪几帧发生了短渐变.

许多学者对 MPEG 压缩域的视频场景检测技术做了大量的研究.MPEG 压缩视频码流中的可用信息包括 DCT 系数[1]、运动矢量[2]、宏块编码模式[3]以及编码比特数[4]等.近年来,随着 H.264/AVC 标准的不断普及,基于 H.264/AVC 压缩域的视频场景检测引起了人们的广泛关注.但是,与基于 MPEG 压缩域的检测技术相比,基于 H.264/AVC 标准的压缩域场景转换检测存在很多的技术难点.

首先,H.264/AVC 引入了帧内预测,这使得帧内宏块的 DCT 系数无法独立提取.预测残差的变换编码和熵编码使得帧内和帧间宏块 DCT 系数提取的计算量接近于完全解码.这样经典的 MPEG 压缩域 DC 图方法无法应用于 H.264/AVC 压缩码流.另外,H.264/AVC 采用

收稿日期:2008-10-20;修回日期:2008-12-17
基金项目:国家自然科学基金(No.60772069);国家 863 高技术研究发展计划(No.2008AA01A313);北京市自然科学基金(No.4102008)

了基于率失真最优的编码模式选择策略,纹理平缓的宏块往往采用帧内编码模式,这使得 H.264/AVC 的宏块编码类型统计信息与 MPEG 码流有所不同,因此统计宏块或运动矢量类型方法也不能直接应用于 H.264/AVC 码流.最后,当发生场景转换时,H.264/AVC 中 I 帧比特率的波动往往比相应的 MPEG 码流小,应用于 MPEG 码流中的比特率判决法如果直接用来检测 H.264/AVC 码流中的场景转换,其性能会变差.

可以看出,基于 H.264/AVC 压缩域的视频场景转换检测是一个极具挑战性的研究课题.近年来,一些学者开展了这方面的研究工作,主要利用 H.264/AVC 压缩码流中的帧内预测模式和帧间编码模式等信息作为判决的主要依据,来进行场景的检测[5~9].文献[5]提出了一种基于 H.264/AVC 压缩域的 GOP 级视频场景转换检测方法,该方法定义了基于子块的相似性度量准则作为场景检测的判决准则,通过区分 I 帧的平滑区域(16×16)和纹理区域(4×4)检测 GOP 中是否发生了场景转换.实验结果表明,该方法对于纹理信息非常丰富的视频,检测的效果会比较差.文献[6]则是通过统计 13 种亮度预测模式的数量来表示亮度值空间分布规律,得到帧间差异,然后根据帧间的差异来判断是否发生了场景的转换.文献[7]利用帧内预测模式的八个预测方向产生的边沿直方图来定位场景转换.另外,帧内预测模式还可以与其他方法相结合,如隐式马尔可夫模型[8]、直方图(亮度预测直方图[6]、边沿直方图[7])、子块划分[5~9]等,进行场景的转换检测.

还有一些学者用帧内编码宏块数作为辅助判决准则,进行突变和短渐变的检测.当突变发生时,P 帧的前向相关性变弱,大多数宏块采用帧内编码模式.当短渐变发生时,连续几帧中会存在大量的帧内编码宏块.因此,可以用经过高斯平滑滤波的帧内编码宏块数序列来检测场景转换[9].

本文基于 H.264/AVC 基本档次码流中的帧内预测模式、运动矢量和宏块编码类型等可用信息,分别提出了基于子块的色度模式差异、累积运动值、累积帧内宏块数等三种判别准则,并据此提出了一种新的基于 H.264/AVC 压缩域的 GOP 级视频场景检测算法.实验结果表明,与 Kim 等人提出的方法相比,本文方法可以获得更好的检测结果.

2　基于 H.264/AVC 压缩域的 GOP 级视频场景转换检测算法

为了增强检测算法的鲁棒性和有效性,本文首先利用 H.264/AVC 基本层码流中的色度预测模式、运动矢量和宏块预测类型等可用信息,提出了基于子块的色度模式差异、累积运动值和累积帧内宏块数等三种判决准则,然后综合使用这些判决准则,提出了一种新的 GOP 级视频场景转换检测算法.接下来本文将首先介绍这三种判决准则,然后介绍提出的 GOP 级视频场景转换检测算法.

2.1　基于子块的色度模式差异—D'_{GOP_CM}

对于视频场景检测,如果采用 13 种亮度模式预测信息(4×4 和 16×16)进行帧间差异的度量,会造成对同一场景中的变化过于敏感等问题,不利于场景检测.考虑到光照变化对亮度分量的影响比色度分量大,本文选用色度分量进行判决,以减少光照变化对场景转换检测所造成的影响.

为此,本文提出了一种基于子块的色度模式差异度量准则,来度量相邻 I 帧之间对应子块中四种色度预测模式在数量上的差异,用于检测 GOP 中发生的场景转换.基于子块的色度模式差异的具体计算方法为:首先将相邻的两个 I 帧划分为相对应的若干个子块 S_k,每个子块 S_k 中包含 $N_{MB}^{S_k}$ 个宏块;然后计算每个子块的色度模式 m 在分布上的差异,将这些差异求和就可以得到基于子块的色度模式差异 D_{GOP_CM},可以表示为:

$$D_{GOP\_CM}=\sum_{\forall k}\frac{1}{N_{MB}^{S_k}}\sum_{m=0}^{3}\left|NC_m^{i,S_k}-NC_m^{j,S_k}\right| \quad (1)$$

其中,i 为当前 I 帧,j 为前一个 I 帧.NC 表示某子块中,某一种帧内色度预测模式的总数.如果基于子块的色度模式差异较大,则可以初步判定 GOP 中发生了场景转换.

当 GOP 中有物体的快速运动或突然出现了新物体时,对于快速运动物体经过或新物体出现的区域,相应子块之间的色度预测模式数量之差会很大,这时就会导致 D_{GOP_CM} 的数值较大,从而造成误检.为了消除这些影响,本文采用排序算法找出最大的几个色度模式差异值,在计算时将这些值排除在外.而对于大多数视频来说,即使没有快速运动的物体和突然出现的新物体,排除几个最大值也不会对最后的判定产生特殊的影响.

为此,本文对 D_{GOP_CM} 进行了改进,得到判决准则 D'_{GOP_CM},具体定义为:

$$D'_{GOP\_CM}=\sum_{\forall k}\frac{1}{N_{MB}^{S_k}}\sum_{m=0}^{3}\left|NC_m^{i,S_k}-NC_m^{j,S_k}\right|,k\in RS \quad (2)$$

其中,RS 表示剩余子块序号的集合.D'_{GOP_CM} 准则可以排除大多数由于物体的快速运动或新物体的突然出现对检测结果造成的影响,提高算法的检测性能.

2.2　累积运动值—S_{GOP_MV}

仅仅采用上述的判决准则 D'_{GOP_CM} 还不足以获得一个理想的检测结果.经过研究发现,在同一个场景

中,如果视频中包含有较多的运动物体或者镜头在运动时,即使将色度模式差异值较大的子块排除在外,D'_{GOP_CM}的值仍会较大.但是,如果增加被排除子块的数量,就会对物体运动缓慢或没有镜头运动的视频的检测产生不良的影响.针对这个问题,本文提出了一个辅助判决准则,即累积运动值 S_{GOP_MV}. S_{GOP_MV}定义为两个 I 帧之间所有 P 帧 F_{GOP}的运动矢量平均幅度之和,可以表示为:

$$S_{GOP\_MV}=\sum_{F_{GOP}}\frac{1}{N_{mv}}\sum_{\forall k}\sqrt{(X^{i,MB_k})^2+(Y^{i,MB_k})^2} \quad (3)$$

其中,i 为当前的 P 帧;X、Y 分别表示运动矢量中的 X 和 Y 分量,MB_k 表示第 k 个宏块,N_{mv}为 GOP 中 P 帧的个数.若累积运动值 S_{GOP_MV}较小,D'_{GOP_CM}仍较大,就可以初步推测当前 GOP 中存在场景转换.累积运动值准则有助于进一步降低 D'_{GOP_CM}的门限,从而提高算法的检测性能.

2.3 累积帧内宏块数—S_{GOP_Intra}

H.264/AVC 压缩码流中还包含一种非常重要的信息,即 P 帧中宏块的编码类型(帧内或帧间).为了充分利用这一信息,提高场景检测的准确率,本文提出了第三个判决准则—累积帧内宏块数 S_{GOP_Intra}.提出该判决准则的出发点是:如果 P 帧中的帧内编码宏块的数量占到了总宏块数的 90% 以上,则这一帧就很有可能是场景突变帧.如果连续几个 P 帧的帧内编码宏块的数量都比较高(比如,占总宏块数的 50% 左右),则这帧图像就很有可能是短渐变.如果某一 P 帧的帧内编码宏块数占了一定的比例,但前后几帧却几乎不存在帧内编码的宏块,那么这个 P 帧就很有可能是由于较多区域比较平滑、运动情况较复杂或新物体较大所造成,而不是场景突变帧,也不是场景渐变中的一帧.

为此,本文引入"累积"的概念来消除这种干扰,并且在计算 S_{GOP_Intra}时采用门限 T_{NI}将这些干扰因素排除,也就是说,将一个 GOP 中大于 T_{NI}的帧内编码宏块数进行累加.累积帧内宏块数准则 S_{GOP_Intra}的定义如下:

$$S_{GOP\_Intra}=\sum_k NI_k,\ k\in\{F_{GOP}\mid NI_k>T_{NI}\} \quad (4)$$

其中,k 是符合条件的帧号,NI_k 就是第 k 帧的帧内编码宏块数.若 S_{GOP_Intra}值较大,就可以初步判定 GOP 中存在场景转换.

2.4 检测算法的流程

本文将上述的三个判决准则有效地结合在一起,提出了一种新的 GOP 级场景转换检测算法,该算法的流程图如图 1 所示.

在图 1 中,T_{CM1}、T_{CM2}、T_{CM3}、T_{CM4}分别表示 D'_{GOP_CM}的门限值($T_{CM1}>T_{CM2}>T_{CM3}>T_{CM4}$). T_{MV1}、T_{MV2}、T_{MV3}分别表示 S_{GOP_MV}的门限值($T_{MV1}>T_{MV2}>T_{MV3}$). T_{Intra1}、T_{Intra2}分别表示 S_{GOP_Intra}的门限值($T_{Intra1}>T_{Intra2}$).在本文中,它们的取值如表 1 所示.

表 1 本文提出的三种判决准则使用的门限值

D'_{GOP_CM}的门限值		S_{GOP_MV}的门限值		S_{GOP_Intra}的门限值	
T_{CM1}	3.3	T_{MV1}	350	T_{Intra1}	280
T_{CM2}	2.7	T_{MV2}	250	T_{Intra2}	230
T_{CM3}	2.3	T_{MV3}	150		
T_{CM4}	2.0				

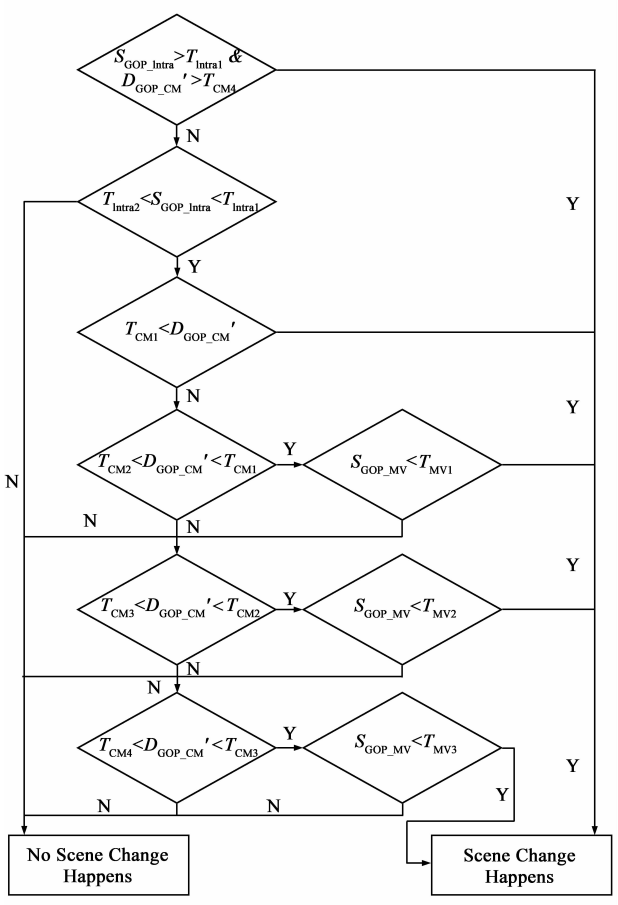

图1 GOP级视频场景检测算法流程图

GOP 级视频场景检测算法的基本流程如下:首先,如果 S_{GOP_Intra}很大且 D'_{GOP_CM}的值也大于最小的门限值 T_{CM4}(即本文认为的发生视频场景转换的最小色度模式差异值),就直接判定 GOP 中存在视频场景转换.如果 S_{GOP_Intra}大于本文认为的 GOP 中发生视频场景转换时的最小门限值 T_{Intra2},则进入下一步判决,否则判定 GOP 中不存在视频场景转换.接下来,将 D'_{GOP_CM}和 T_{CM1}、T_{CM2}、T_{CM3}、T_{CM4}四个门限分别进行比较,如果 D'_{GOP_CM}大于最大的门限 T_{CM1},则判定 GOP 中存在视频场景转换.如果 D'_{GOP_CM}小于最小的门限 T_{CM4},则判定 GOP 中

不存在视频场景转换.如果 D'_{GOP_CM}在 T_{CM1}和 T_{CM4}之间,则需要采用 S_{GOP_MV}来做辅助判别.在 D'_{GOP_CM}固定的情况下,S_{GOP_MV}的值越小,GOP 中发生视频场景转换的几率也就越大.

3　实验结果与分析

常用的视频场景转换检测的性能指标有查准率 *Precision* 和查全率 *Recall*.这两个指标分别是基于正确检出数 N_c、误检数 N_f 和漏检数 N_m 的比值给出的,二者的定义如下:

$$Precision = \frac{N_c}{N_c + N_f} \times 100 \ , \ Recall = \frac{N_c}{N_c + N_m} \times 100 \quad (5)$$

另外一种综合了查准率和查全率的评价指标是 Qi 等人提出的 *F*1,只有当查准率 *Precision* 和查全率 *Recall* 都较高的时候,*F*1 才会较高.*F*1 的定义如下[10]:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

目前,绝大部分压缩域视频场景检测的研究都采用自行选取的视频序列进行测试,没有一组标准的视频测试序列,这给算法之间的性能比较带来了一定的困难.为此,除了所选择的体育、动画、新闻、电影等四类视频外,本文还采用 Akiyo 等 9 个标准测试视频序列合成了一段长度为 10000 帧的视频序列,这个合成的视频测试序列包含了 200 处场景转换.所有测试视频的尺寸均为 320×240,编码参数为:基本档次,*I* 帧周期为 30 帧,帧率 30fps,编码码率为 300Kbps.

为了证明本文提出的算法的有效性,本文将该算法和 Kim 等人提出的 GOP 级视频场景检测算法进行了实验对比.表 2 和表 3 给出了分别采用 Kim 等提出的算法和本文提出的算法对测试序列进行测试得到的实验结果,其中两种算法采用的子块大小均为 5×3(单位:宏块).

表 2　采用 Kim 等人提出的算法时的场景转换检测结果(5×3)

测试视频	场景转换总数	N_c	N_m	N_f	*Precision*	*Recall*	*F*1
体育	15	13	2	11	54.2	86.7	66.7
动画	32	22	10	9	71.0	68.8	69.8
新闻	55	29	26	1	96.7	52.7	68.2
电影	54	17	37	20	45.9	31.5	37.4
合成序列	200	120	80	2	98.4	60.0	74.5

表 3　采用本文提出的算法时的场景转换检测结果(5×3)

测试视频	场景转换总数	N_c	N_m	N_f	*Precision*	*Recall*	*F*1
体育	15	14	1	12	53.8	93.3	68.3
动画	32	30	2	6	83.3	93.8	88.2
新闻	55	52	3	1	98.1	94.5	96.3
电影	54	53	1	11	82.8	98.1	89.8
合成序列	200	193	7	0	100.0	96.5	98.2

表 4 和表 5 给出的是当子块大小为 5×5(单位:宏块)时,分别采用 *Kim* 等提出的算法和本文提出的算法进行测试得到的实验结果.

表 4　采用 Kim 等人提出的算法时的场景转换检测结果(5×5)

测试视频	场景转换总数	N_c	N_m	N_f	*Precision*	*Recall*	*F*1
体育	15	12	3	16	42.9	80.0	55.8
动画	32	20	12	10	66.7	62.5	64.5
新闻	55	27	28	1	96.4	49.1	65.1
电影	54	18	36	20	47.4	33.3	39.1
合成序列	200	121	79	2	98.4	60.5	74.9

表 5　采用本文提出的算法时的场景转换检测结果(5×5)

测试视频	场景转换总数	N_c	N_m	N_f	*Precision*	*Recall*	*F*1
体育	15	10	5	2	83.3	66.7	74.1
动画	32	29	3	3	90.6	90.6	90.6
新闻	55	45	10	1	97.8	81.8	89.1
电影	54	49	5	8	86.0	90.7	88.3
合成序列	200	188	12	0	100.0	94.0	96.9

从表 2、3、4、5 中的实验结果可以看出,无论测试哪种视频,不管采用的是 5×5 的子块还是 5×3 的子块,本文提出算法的检测性能都要远远好于 Kim 等人提出的算法.这主要是因为本文提出的算法综合利用了帧内预测模式、运动矢量和宏块编码类型三种压缩域信息,而 Kim 等人提出的算法只利用了帧内预测模式的粒度信息(4×4 或 16×16),而没有进一步将包含方向信息的帧内预测模式值考虑进去,这使得该算法无法检测出连续两个纹理非常丰富的场景之间的转换.而且对于一个纹理逐渐丰富(或趋于平坦)的场景,Kim 等人的算法会将其判定为多个场景,从而发生误检.总之,Kim 等人对压缩域信息的挖掘还不充分,所以该算法的检测性能不如本文提出的算法.

另外,对比在不同子块划分情况下本文算法的检测结果可以看出,在门限值相同时,使用 5×3 的子块可以获得更高的 *Recall*,使用 5×5 的子块时则可以获得更高的 *Precision*.但是,也必须指出,无论采用哪种划分方式,它们的综合性能 *F*1 几乎不相上下.从实验结果也可以看出,体育类视频的检测结果要比其他类的差,这是因为体育视频中的运动物体较多,运动幅度较大.

需要指出的是,由于采用了 3 组门限值,当视频尺寸或帧率变化时,门限值的调整会较为复杂,这也是本文提出的算法存在的主要缺点.

4　结论

本文首先利用 H.264/AVC 压缩域的可用信息,提出了基于子块的色度模式差异、累积运动值、累积帧内宏块数等三个判决准则,并据此实现了一种新的基于 H.264/AVC 压缩域的 GOP 级视频场景转换检测算法.

实验结果表明,与 Kim 等人提出的算法相比,本文提出的算法可以获得更好的检测性能.本文提出的算法可以应用于关键帧提取、视频内容分析与过滤、视频索引与检索等多个不同的领域.

参考文献:

[1] B L Yeo, B Liu. Rapid scene analysis on compressed video [J]. IEEE Transactions on Circuits and System for Video Technology, 1995, 5(6): 533 – 544.

[2] Divakaran A, Sun H. A descriptor for spatial distribution of motion activity for compressed video[A]. Proc. SPIE Conference on Storage and Retrieval for Media Database[C]. San Jose, CA, USA: SPIE, 2000. 392 – 398.

[3] S C Pei, Y Z Chou. Efficient MPEG compressed video analysis using macroblock type information[J]. IEEE Trans Multimedia, 1999, 1(4): 321 – 333.

[4] Li H, Liu G, Zhang Z, Li Y. Adaptive scene – detection algorithm for VBR video stream[J]. IEEE Transactions on Multimedia, 2004, 6(4): 624 – 633.

[5] Sung Min Kim, Ju Wan Byun, Ghee Sun Won. A scene change detection in H.264/AVC compression domain[A]. Advances in Multimedia Information Processing[C]. Jeju Island, South Korea: Springer – Verlag, 2005. 1072 – 1082.

[6] Wei Zeng, Wen Gao. Shot change detection on H.264/AVC compressed video[A]. IEEE International Symposium on Circuits and Systems[C]. Kobe, Japan: IEEE, 2005. 3459 – 3462.

[7] Bohyun Hong, Minyong Eom, Yoonsik Choe. Scene change detection using edge direction based on intra prediction mode in H.264/AVC compression domain[A]. TENCON IEEE Region 10 Conference[C]. Hong Kong, China: IEEE, 2006. 4.

[8] Liu Yang, Wang Weiqiang, Gao Wen, Zeng Wei. A novel compressed domain shot segmentation algorithm on H.264/AVC [A]. International Conference on Image Processing[C]. Singapore: IEEE Computer Society, 2004. 2235 – 2238.

[9] De Bruyne S, De Neve W, De Wolf K, De Schrijver D, Verhoeve P Van de. Temporal video segmentation on H.264/AVC compressed bitstreams[A]. Advances in Multimedia Modeling, MMM2007[C]. Singapore: Springer, 2007. 1 – 12.

[10] Qi Y, Hauptmann A, Liu T. Supervised classification for video shot segmentation[A]. Proc. IEEE Conf. on Multimedia Expo (ICME)[C]. Baltimore, MD, USA: IEEE, 2003. 689 – 692.

作者简介:

高 宇 男,1983 年生于北京,北京工业大学硕士,主要研究方向为视频信号处理.
Email: yugaobob@hotmail.com

卓 力 女,1971 年生于江苏,北京工业大学教授、博士,主要研究方向为多媒体分析、视频编码与网络传输等.
Email: zhuoli@bjut.edu.cn

王素玉 女,1976 年生于河北,北京工业大学博士、讲师.主要研究方向为图像/视频超分辨率复原,智能视觉监控等.
Email: suyuwang@emails.bjut.edu.cn

沈兰荪 男,1938 年生于江苏,北京工业大学教授、博士生导师,主要研究方向是图像/视频信号处理、生物医学图像处理信息系统等.
Email: sls@bjut.edu.cn