doi:10.11959/j.issn.1000-436x.2016033

视频检索中图像信息量度量

袁庆升[1,2]，张冬明[3,4]，靳国庆[3,4]，刘菲[3,4]，包秀国[1,2]
(1．中国科学院信息工程研究所，北京 100193；2．国家计算机网络应急技术处理协调中心，北京 100029；
3．中国科学院智能信息处理重点实验室，北京 100190；4．中国科学院计算技术研究所，北京 100190)

摘 要：综合考虑信息量度量的速度、性能要求，提出了相适应的显著图、多特征融合模型；基于区域划分融入空间关系，提出了分块信息熵的图像信息量度量方法（SEII）；构建了信息量度量的标注数据集，并设计了性能验证方法。实验结果表明该度量方法符合人眼视觉的评价结果。度量方法在实际视频检索系统中进行对比应用测试，测试表明 *mAP* 提高 4.4%，检索速度提高 1.5 倍。

关键词：视频检索；关键帧选择；图像信息量；显著区域；多特征融合
中图分类号：TP37 **文献标识码**：A

Image information measurement for video retrieval

YUAN Qing-sheng[1,2], ZHANG Dong-ming[3,4], JIN Guo-qing[3,4], LIU Fei[3,4], BAO Xiu-guo[1,2]
(1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100193,China;
2. National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China;
3. Key Lab of Intelligent Information Processing, Chinese Academy of Sciences, Beijing 100190, China;
4. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190,China)

Abstract: To meet the speed and performance requirements, Sub-region entropy based image information measurement (SEII) method was proposed, which integrates the salie region detection, region division and features fusion. And, performance evaluation method was designed and many experiments were carried out, proving SEII coordinates with human vision evaluation. Also, SEII is evaluated in a real video retrieval system, which shows increase about 4.4% of *mAP* with 1.5 times speedup.

Key words: video retrieval, key frame selection, image information, salient region, features fusion

1 引言

视频检索中，关键帧选择是影响检索性能的关键因素之一。现有关键帧选择算法按功能分为 2 类：1）在时间轴上按照一定模型抽取帧图像，去除视频时域冗余信息，减小视频特征数量，主要算法包括基于镜头的选择方法[1]、基于运动分析的选择方法[2]、基于聚类的选择方法[3]；2）从内容角度进行帧的二次筛选，去除不符合条件的视频帧。本文针对后者进行研究。

现有基于帧内容的筛选算法，主要从失真角度评价帧图像内容，目标是去除视频中模糊帧、切换帧，如文献[4]，由于模型单一，其不能准确度量图像内容。如图 1 所示的 6 幅图像来自于 6 段视频，在传统的失真评估模型中，由于图像的失真程度很低，所以图 1(a)~图 1(c)的质量评分很高；而图 1(d)~图 1(f)则由于模糊等原因，导致其质量评价较低。这恰好与人类的主观评估相悖，图 1(a)~图 1(c)相对于图 1(d)~图 1(f)，内容不够丰富，信息量较低。

收稿日期：2015-05-05；**修回日期**：2015-07-21
通信作者：张冬明，dmzhang@ict.ac.cn
基金项目：国家自然科学基金资助项目（No.61273247, No.61303159, No.61271428）；国家高科技研究发展计划（“863”计划）基金资助项目（No.2013AA013205）
Foundation Items: The National Natural Science Foundation of China (No.61273247, No.61303159, No.61271428), The National High Technology Research and Development Program of China (863 Program)(No.2013AA013205)

(a) 图像 1　(b) 图像 2　(c) 图像 3

(d) 图像 4　(e) 图像 5　(f) 图像 6

图 1　视频中图像帧信息量的对比

对于视频检索，选择图 1(d)~图 1(f)作为视频的关键帧比图 1(a)~图 1 (c)更具有区分性。因此，信息量是关键帧选择重要指标。高信息量的视频帧可增加视频数据特征的区分性；反之，低信息量视频帧，则可能导致检索性能下降：一方面它可能导致虚匹配，这类似于文本检索中，使用停止词检索会导致大量无关结果；另一方面它带来不必要的特征量增长，导致检索计算开销增加。但是，目前信息量度量尚没有在视频检索系统中取得有效应用，这主要是由于信息量度量方法尚存在不足。

信息量是信息多少的量度，熵可用于衡量信息量高低。应用于图像领域，熵值反映图像像素值的分布，与像素值大小无关。熵值越大，表示像素值越接近均匀分布。可通过融合局部空间信息将熵扩展至二维，若图像深度为 D，令 $P=2^D$，则像素值取值范围为$[0, P-1]$，则可根据式(1)来计算图像二维信息熵 $H(f)$，其中，i 是图像像素值，j 是其邻域像素均值，$p_{i,j}$ 是(i,j)像素值组合的分布概率。

$$H=-\sum_{i=0}^{P-1}\sum_{j=0}^{P-1}p_{i,j}\log(p_{i,j}) \tag{1}$$

但仅仅依靠信息熵并不足以实现对图像的信息量全面度量，主要表现为如下几方面不足：1）熵的计算方法具有对称性，即使图像的颜色直方图完全不同，但如果像素值具有相同的概率分布，那么信息熵相同；2）熵值易受噪声影响，并且在度量杂乱纹理图像时常常失效；3）全画面计算，没有突出图像中显著区域的重要性，不符合人眼视觉特点。

本文以图像信息熵为基础，融合人类视觉显著性模型，利用分块信息熵对图像的信息量进行度量，提出了多特征融合的图像分块信息熵度量（SEII, sub-region entropy based image information measurement）方法，从而实现对图像信息量的度量。该方法不仅结合了人眼的视觉特性，还为信息熵增加了空间信息，同时融合了颜色、纹理、显著性等特征，可以全面提升视频检索性能。

2　基于分块信息熵的信息量度量方法

与实际应用结合，理想的视频帧信息量度量方法应满足如下要求：1)度量结果与人眼视觉一致；2)计算简单，以满足各类实时性要求高的应用。

2.1　视觉显著性模型

认知心理学的研究表明，图像中有些区域能显著吸引人的注意，这些区域含有较大的信息量，这就是视觉显著区域，它是图像中最能引起用户兴趣、最能表现图像内容的区域。有许多数学模型可以用来模拟人的注意力机制，早期工作可以追溯到 Koch 和 Ullman 的基于生物视觉特性的计算模型[5]，Itti 等[6]在此模型的基础上进行了改进，通过结合多尺度的图像特征，在快速场景识别上取得了很好的效果。

对比度是像素与其邻域像素的差异程度，强对比度区域通常更能引起人们的注意，因而成为显著

性计算的重要特征之一。Ma 等[7]首次提出基于局部对比度的分析方法，局部对比度的方法倾向于给物体边缘赋予较大的显著值，而并非高亮整个显著区域。Zhai 等[8]提出了基于全局对比度的亮度通道显著模型，能够高亮整个区域。Cheng 等[9]提出了多颜色通道压缩表达的显著性模型，可进一步提升检测的准确率。频域分析是另一类显著性计算方法，文献[10]指出，这些基于频域分析的模型实际上等同于一个局部梯度算子并叠加一个高斯模糊，不适宜检测较大的显著区域。此外，还有一类基于学习的显著性模型，如文献[11]。这类方法在训练数据集上能取得不错的效果，但是通常十分耗时，在不同数据集下的表现也会存在较大的差异。

本文结合信息度量算法的要求，最终选择 2 类经典视觉特征——颜色和纹理的对比度来描述区域的显著性，以下分别描述其计算方法。

给定一个颜色集合 $Q=\{q_1,\L,q_B\}$，$h_{Dp}(b)=\{h_b\}$ 是图像区域 D_p 定义在组距 b（$b\in[1,B]$）上的颜色直方图，其中，h_b 是区域 D_p 中每个像素值 q_b（$q_b\in Q$）出现的概率。为了快速计算，并减小像素值微小变化带来的噪声，取 D_p 为 8×8 区块，对于深度为 8 的颜色集合，取 B=64。采用带归一化的巴氏距离[12]度量相邻区域的颜色直方图距离，如式(2)所示，其中，Q 是 D_p 的 $k\times k$ 邻域内的相邻区域，k 为奇数，k 控制着颜色对比度度量方法的敏感性，k 值越小，计算到的颜色对比度越敏感。

$$C(D_p)=\sum_{D_q\in Q}\sqrt{1-\sum_b\sqrt{\frac{h_{D_p}(b)h_{D_q}(b)}{\sum_b h_{D_p}(b)\sum_b h_{D_q}(b)}}} \tag{2}$$

纹理特征是图像的基本特征，也是经常被用来描述图像，能够有效度量图像亮度的变化情况。图像局部纹理的计算如下[6]

$$\hat{I}(i,j)=\frac{I(i,j)-m(i,j)}{s\ (i,j)+C} \tag{3}$$

其中，$I(i,\ j)$表示图像像素值，$\hat{I}(i,j)$ 是点$(i,\ j)$的去均值做除归一化后的系数，$i\in\{1,2,\L,M\}$ 和 $j\in(1,2,\L,N)$，M 和 N 分别为图像的长和宽，C=1 是一个常量，防止分母为零的情况发生。$m(i,j)$ 和 $s\ (i,j)$ 分别为

$$\begin{aligned} m(i,j)&=\sum_{k=-K}^{K}\sum_{l=-L}^{L}w_{k,l}I(i+k,j+l)\\ s\ (i,j)&=\sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L}w_{k,l}[I(i+k,j+l)-m(i,j)]^2} \end{aligned} \tag{4}$$

其中，$w=\{w_{k,l}\mid k=-K,\L,K,l=-L,\L,L\}$ 是经过采样并归一化后的二维圆对称高斯权重函数，K=L=5。

图 2 展示了图 1 中 2 个图像对应的显著图，可以看出，根据颜色对比度计算出的显著图对应于图像中的显著物体，通过纹理对比度计算出的显著区域与物体边界十分接近，边界部分在颜色和纹理上都具有较高的对比度值，上述模型能够很好地捕捉

(a) 原始图像 (b) 颜色显著图 (c) 纹理显著图

图 2 图像及其显著图对比示例

颜色和纹理信息。

2.2　显著图融合模型

形成最终视觉显著图，需要把上述颜色显著图和纹理显著图进行融合。现有融合算法可以分为 3 类：1）最直观的线性融合算法[13,14]，不同特征显著图给定不同权值，线性组合后得到最终的显著图；2）由于人脑的视觉系统是一个非线性处理系统，线性的融合算法具有很大的局限性，因此，一些研究者提出了非线性的融合算法[15,16]，为不同特征的显著图分别定义融合函数，得到相应的递推关系式，计算最终的融合结果；3）基于空间紧凑性和显著密度的方法[17]、采用遗传算法的融合方法[18]、根据上下文内容的融合方法[19]以及多种融合方式相结合的算法[20]等。

本文涉及到 2 个显著图的融合，在融合过程中，还考虑了其他显著图模型所忽略的邻域显著性。具体地，设计了式(5)对显著图进行融合。

$$F[C(D_p),T(D_p)]=\frac{W_c}{W_c+W_t}C(D_p)+\frac{W_c}{W_c+W_t}T(D_p) \tag{5}$$

其中，w_c 和 w_t 分别为区域 D_p 所对应的 $k\times k$ 邻域内的颜色显著值 $C(D_p)$和纹理显著值 $T(D_p)$。$T(D_p)$是 D_p 区域内所有像素点的纹理信息（按照式（3）计算）的均值。

得到融合的显著图后，根据像素的显著性进一步生成掩码图像，以屏蔽图中显著性较低的部分，文中以 7∶3 的比例将显著图划分为显著区域和非显著区域。从图 3(c)中可以看出，当图像信息量不丰富时，不显著区域占了图像的绝大多数，当图像信息量丰富时，显著的区域占了图像的绝大多数。

2.3　分块信息熵评估模型

利用融合显著图针对全画面计算，仅计算显著区域比例并不能精准地表达画面的信息量高低，而通过分块可以融合空间分布信息，显著提高信息量的表达能力。为此，本文引入分块信息熵概念。具体地，为避免复杂计算，保证信息度量的实时性，按下述步骤计算分块信息熵：首先按照横向 3 等分的方法将图像划分为 3 个区域；然后对图像的颜色空间进行简单变换，对变换后的图像分别计算水平分块后 3 个区域的水平方向信息熵 X_l、垂直方向信息熵 X_c 和整体的信息熵 X_a，为提高对像素值变化的顽健性，在信息熵计算中进行不同程度的量化；另外计算分块的均值 X_μ 和标准差 X_s 以及显著像素在每个区域中所占的比例 d 这 3 个特征值。图 4 给出特征提取的算法流程，最终形成 54 维特征向量。

为了得到图像的信息量度量值，需要利用训练图像集得到回归模型。在特征维数较高时，传统拟合方法通常采用增加高阶项的方法提高模型的适用性，容易造成模型过拟合问题。本文采用支持向量回归(support vector regression)模型[21]，该模型使用核函数代替线性方程中的高阶项，使原来的线性算法非线性化，实现非线性回归。引入核函数同时实现升维，低维空间中非线性问题投影到高维空间后可能变成线性问题，SVR 模型通过升维后在高维空间中构造线性决策函数实现回归。训练好 SVR 模型后，用待评估图像的分块信息熵特征作为模型参数，得到信息量度量值。

(a) 原始图像　　(b) 融合后显著图　　(c) 掩码图像

图 3　融合后的显著图及掩码图像

算法：图像信息量特征提取

for each image of $M\times N$ pixels **do**
计算显著图
横向 3 等分
for each region D **do**

$$r=\frac{R}{R+G+B}$$

$$g=\frac{G}{R+G+B}$$

$$T=R+G+B$$

for each feature $F\in\{r,g,T\}$ **do**

sl=vector of the sum of F value for each pixel of each row of region D

hl=histogram of sl on $\sqrt{N}$ bins

X_l=entropy(hl)

sc=vector of the sum of F value for each pixel of each column of region D

hc=histogram of sc $\sqrt{\frac{M}{3}}$ bins

X_c=entropy(hc)

h=histogram of F on $\sqrt{N\frac{M}{3}}$ bins

X_a=entropy(h)

X_μ=mean of F

X_s=std of F

d =proportion of saliency pixels

end for
end for
end for

图 4 图像信息量特征提取过程

2.4 算法流程

基于分块信息熵的图像信息量度量方法流程如图 5 所示。待检测图像首先经过颜色和纹理特征的提取，形成视觉显著性模型，然后对经过视觉显著性模型处理后的图像进行基于分块信息熵的特征提取，再由训练集图像得到 SVR 模型进行距离度量，最后得到信息量的估计值。

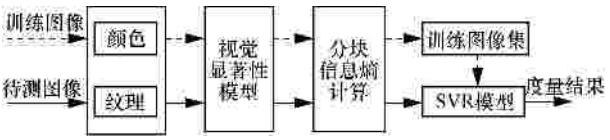


图 5 基于分块信息熵的图像信息量度量方法流程
（虚线表示训练流程，实线表示检测流程）

3 实验结果与分析

从信息量度量与主观测试一致性角度对度量方法的性能进行测试，进一步验证该度量方法在实际视频检索中的应用效果，以检验其实际应用价值。

3.1 度量一致性验证

目前针对图像的信息量度量方法还没有公认的评测数据集，本文参考图像质量评估模型的评价方法，采用主观评价法获得数据集构建了首个图像信息量度量标注测试集。构建方法如下：1）从 50 段网络视频中有选择地截取 500 幅图像；2）主观评价方法得到的结果会受到观测者自身条件和客观环境因素的影响，为减少这类因素带来的评价误差，选择中国科学院计算技术研究所的 50 名视力或矫正视力正常的研究生作为观测者，在安静、仅有观测者的房间中完成评测；3）50 名观测者对 500 幅图像按照表 1 所示的评分标准对图像的信息量进行打分，参考相关实验标准[22]，评价标准包括绝对尺度和相对尺度，用观测者给出的平均分数计算评价结果；4）根据打分结果，去除评价争议较大的图像，最终形成 469 幅图像的标注数据集。

表 1 主观图像信息量度量标准

级别	绝对尺度	相对尺度	得分
1	很好	该组最好	80~100
2	较好	高于该组的平均水平	60~79
3	一般	该组平均水平	40~59
4	较差	低于该组平均水平	20~39
5	很差	该组中最差	0~19

图 6 给出所建立的数据集中的部分图像，以及相应的主观评价的信息量得分。相对于客观评价方法，主观评价方法能够从人类视觉角度给出图像的信息量，结果比较准确。

在该标注数据集基础上，对本文的信息量度量方法与主观评价的一致性进行测试。与大多数训练模型相似，本文选取上述数据集中的 80%作为训练集，剩下的 20%作为测试数据[23]。为了减少实验对数据集的依赖性，本文对训练集和测试集进行了 10 次随机分割，分别进行实验，取平均值作为最终的实验结果。选择 SROCC[24]测量视频帧的信息量度量模型与主观评价值之间的相关性。本文还与 Peng 等[13]静态显著模型的算法进行对比实验，实验中用该算法替换 2.1 节所提显著图模型，并在本文所提的数据集上进行测试。表 2 给出 10 次随机分割实验的结果以及 10 次结果的平均值和标准差。可以看出，本文提出的基于分块信息熵的信息量度量方法与主观评价结果有较高的相关性，并且多次实验的评价结果变化不大，证明该方法能够有效地度量图像的信息量。使用文献[13]的方法替换本文中的

(a)33.6　(b) 49.7　(c) 53.6

(d)62.3　(e)87.5　(f)5.0

图 6　数据集中部分图像及其主观评价得分

显著图模型，对应 SROCC 均值降低、标准差增大，表明本文基于颜色和纹理的模型更适用于帧信息量度量。

表 2　本文方法与文献[13]方法的图像信息量度量结果

编号	训练图片数量	测试图片数量	SROCC 评价值
1	372	100	0.807/0.797
2	377	95	0.895/0.887
3	377	95	0.859/0.862
4	382	90	0.901/0.900
5	382	90	0.856/0.832
6	382	90	0.863/0.815
7	382	90	0.862/0.816
8	377	95	0.901/0.903
9	377	95	0.887/0.853
10	372	100	0.821/0.831
	10 次结果的均值		0.865/0.850
	10 次结果的标准差		0.032/0.038

3.2　应用效果验证

本文将基于底层特征查询的视频检索系统作为“基准”，与融合了关键帧选择检索系统的实验结果进行了比较。在“基准”中，关键帧选择利用颜色直方图的变化选取关键帧。特征提取相似度度量与本文的方法一致。查询结果都采用计算最长公共关键帧序列的方法获得。2 个系统采用的检索特征相同，都是基于直方图投影和密度分布的混合特征。

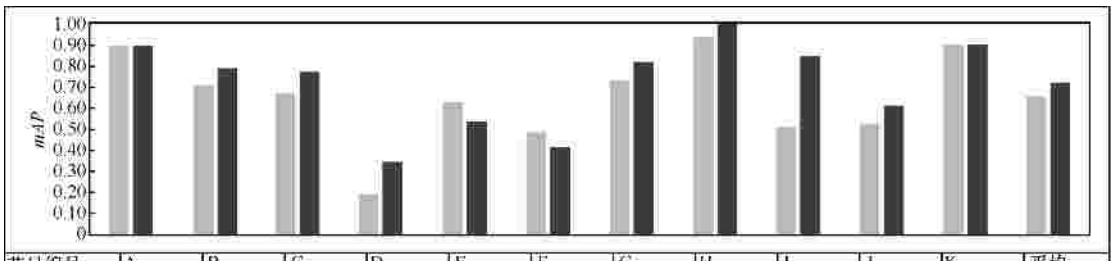
本文就 2 个视频检索系统在平均准确率（MAP, mean average precision）和时间性能上进行了比较。测试平台参数：CPU，Intel Core i5，3.1 GHz；内存，4 GB，操作系统，Windows7。

测试数据集由 TRECVID2013 的评测数据和互联网上收集到的视频数据组成，共 10 000 个视频片段。从中选择了几类典型节目作为查询片段，具体如表 3 所示。

表 3　查询节目对照

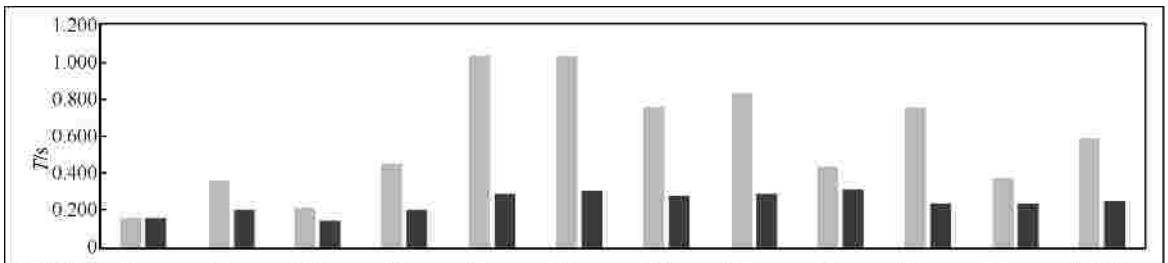
编号	节目名称	编号	节目名称
A	体育“高尔夫”	G	纪录片“我是普京”
B	人物访谈“刘欢”	H	综艺“快乐大本营”
C	人物访谈“刘嘉玲”	I	综艺“今夜有戏”
D	新闻节目“VOA 奥运新闻”	J	综艺“中国达人秀”
E	新闻节目“北京您早”	K	动画“猫和老鼠”
F	新闻节目“Student CNN”	—	—

图 7 给出 baseline 与本文系统的检索结果，其中，准确率的计算方法取前检索结果的前 15 位返回给用户。可以看出 *mAP* 平均提高了 4.4%，且节目 A 的查全率和准确率相对较高，这是由于这类节目的内容与其他节目存在明显区别。访谈类节目之间存在相互影响，但并不严重。节目 D 由于其内容上的丰富性，导致受其他类节目影响比较严重，对于该类节目，在信息量度量基础上应适当提高关键帧提取的密度。节点 E 和节目 F 检索结果相对较好。节目 G 涵盖的内容比较丰富，视频内容变化较大，因此检索效果比其他类型稍差。综艺类节目的检索结果存在很大差异，这是由于其内容的多样性造成的。节目 H 和节目 J 都在场景和人物上保持着很强



节目编号	A	B	C	D	E	F	G	H	I	J	K	平均
基准	0.90	0.71	0.67	0.19	0.63	0.48	0.73	0.94	0.50	0.53	0.90	0.65
本文	0.90	0.79	0.78	0.34	0.54	0.42	0.82	1.00	0.85	0.61	0.90	0.72

图 7 检索准确率 *mAP* 对比



节目编号	A	B	C	D	E	F	G	H	I	J	K	平均
基准	0.154	0.364	0.211	0.447	1.039	1.032	0.755	0.839	0.425	0.751	0.379	0.581
本文	0.154	0.196	0.148	0.198	1.282	0.309	0.275	0.289	0.306	0.227	0.225	0.237

图 8 检索时间对比

相关性，检索结果较好。节目 I 的不同期节目内容存在较大差异，检索结果较差。图 8 给出 2 个系统的检索时间，可以看出，本文提出的关键帧选择方法可以有效减少系统在检索时的时间消耗，检索速度平均提高 1.5 倍。2 个检索系统采用相同的特征索引方法，检索速度的提升主要是基于信息量度量结果的帧筛选：信息量过低的帧直接跳过，去除了特征提取和匹配的时间；低信息量帧不进入特征库，减少了匹配时间。

4 结束语

信息量度量对视频检索中关键帧选择具有重要作用，本文提出了基于分块信息熵的信息量度量方法，其融合了多个视觉显著性模型，采用颜色对比度和纹理对比度相融合的方法获得其显著图，模拟视觉显著性模型的工作原理，并根据像素的显著性对图像进行处理。该信息量度量方法以图像的信息熵为基础，提取基于图像分块的信息量特征，通过基于显著性分析的图像分块，不仅包含熵、均值、标准差等信息，还融合了方向和位置信息，实验表明其度量结果与人眼视觉保持较好一致性。

度量方法涉及计算量小，适用于大规模视频检索系统，在实际视频检索系统中的应用结果表明，其有助于减少视频特征量，提高检索准确率和速度。

参考文献：

[1] SHAHRARAY B, GIBBON D C. Automatic generation of pictorial transcripts of video programs[C]//Proc SPIE. c1995: 512-518.

[2] WOLF W. Key frame selection by motion analysis[C]//Acoustics, Speech, and Signal Processing, International Conference, c1996: 1228-1231.

[3] **章毓晋**. **基于内容的视觉信息检索**[M]. **北京**: **科学出版社**, 2003. ZANG Y J, Content based video information retrieval[M]. Beijing: Science Press, 2003

[4] SAAD M A, BOVIK A C, CHARRIER C. Blind image quality assessment: a natural scene statistics approach in the DCT domain[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3339-3352.

[5] KOCH C, ULLMAN S. Shifts in selective visual attention: towards the underlying neural circuitry[C]//Matters of Intelligence. c1987: 115-141.

[6] 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.

[7] MA Y F, ZHANG H J. Contrast-based image attention analysis by using fuzzy growing[C]//The eleventh ACM international conference on Multimedia. c2003:374-381.

[8] ZHAI Y, SHAH M. Visual attention detection in video sequences using spatiotemporal cues[C]//14th annual ACM international conference on Multimedia. c2006:815-824.

[9] CHENG M M, ZHANG G X, MITRA N J, *et al*. Global contrast based salient region detection[C]//Computer Vision and Pattern Recognition(CVPR). c2011:409-416.

[10] HOU X, HAREL J, KOCH C. Image signature: highlighting sparse salient regions[J]. IEEE Transactions on, Pattern Analysis and Ma-

chine Intelligence, 2012, 34(1):194-201.

[11] MAI L, NIU Y, LIU F. Saliency aggregation: a data-driven approach[C]//IEEE Conference. Computer Vision and Pattern Recognition(CVPR). c2013:1131-1138.

[12] BHATTACHARYYA B A. On a measure of divergence between two statistical populations defined by probability distributions[J]. Bulletin of the Calcutta Mathematical Society, 1943, 35:99-110,

[13] PENG J, QING X L. Keyframe-based video summary using visual attention clues[J]. IEEE MultiMedia, 2010, 17(2): 64-73.

[14] LAI J L, YI Y. Key frame extraction based on visual attention model[J]. Journal of Visual Communication and Image Representation, 2012, 23(1): 114-125.

[15] HUA X S, ZHANG H J. An attention-based decision fusion scheme for multimedia information retrieval[C]//Advances in Multimedia Information Processing-PCM 2004. Springer Berlin Heidelberg, c2005: 1001-1010.

[16] MA Y F, HUA X S, LU L, et al. A generic framework of user attention model and its application in video summarization [J]. EE Transactions on Multimedia, 2005, 7(5): 907-919.

[17] HU Y, XIE X, MA W Y, et al. Salient region detection using weighted feature maps based on the human visual attention model[C]//Advances in Multimedia Information Processing-PCM 2004. Springer Berlin Heidelberg, c2005: 993-1000.

[18] ARMANFARD Z, BAHMANI H, NASRABADI A M. A novel feature fusion technique in saliency-based visual attention[C]//Advances in Computational Tools for Engineering Applications, c2009:230-233.

[19] LAI J L, YI Y. Key frame extraction based on visual attention model[J]. Journal of Visual Communication and Image Representation, 2012, 23(1): 114-125.

[20] EJAZ N, MEHMOOD I, WOOK B S. Efficient visual attention based framework for extracting key frames from videos [J]. Signal Processing: Image Communication, 2013, 28(1): 34-44.

[21] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3): 199-222.

[22] PARK J S, CHEN M S, YU P S. Using a hash-based method with transaction trimming for mining association rules[J]. IEEE Transactions on Knowledge and Data Engineering, 1997, 9(5): 813-825.

[23] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transact ns on Image Processing, 2012, 21(12): 4695-4708.

[24] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444.

作者简介：

袁庆升（1980-），男，山东济南人，中国科学院信息工程研究所博士生，国家计算机网络应急技术处理协调中心副高级工程师，主要研究方向为多媒体大数据处理、网络与信息安全。

张冬明（1977-），男，江苏盐城人，中国科学院计算技术研究所副研究员、硕士生导师，主要研究方向为多媒体内容检索、模式识别、视频编码等。

靳国庆（1988-），男，山东单县人，主要研究方向为多媒体内容检索、模式识别等。

刘菲（1989-），女，河北唐山人，中国科学院计算技术研究所硕士生，主要研究方向为多媒体内容检索。

包秀国（1963-），男，江苏如皋人，国家计算机网络应急技术处理协调中心教授级高级工程师、博士生导师，主要研究方向为网络与信息安全。