

# 基于图像质量和注意力的宫腔镜视频 关键帧提取

苗强强\*

武汉理工大学信息工程学院, 湖北 武汉 430070

**摘要** 为解决传统方法中由注意力曲线等长分段导致的精确率偏低的问题,提出了基于图像质量曲线提取关键视频段的方法,从各段内选取注意力值最高的帧作为关键帧。针对本地数据库进行视频关键帧提取,精确率为 52.94%, $F$  值为 62.77%,分别比 Muhammad 的方法提高了 5.23%和 2.65%。

**关键词** 图像处理;宫腔镜视频摘要;注意力模型;图像质量评估;图像锐利度

中图分类号 TP399 文献标识码 A

doi: 10.3788/LOP56.061004

## Key Frame Extraction of Hysteroscopy Videos Based on Image Quality and Attention

Miao Qiangqiang\*

College of Information Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, China

**Abstract** In order to solve the problem of low accuracy rate caused by the equal length segmentation of the attention curve in the traditional methods, a scheme is proposed, in which the key video segments are extracted based on the image quality curve. The frame with the highest attention value is selected from each segment as the key frame. On the local database, the precision and  $F$ -measure are 52.94% and 62.77%, 5.23% and 2.65% higher than those by the Muhammad method, respectively.

**Key words** image processing; hysteroscopy video summarization; attention model; image quality assessment; image sharpness

**OCIS codes** 100.2000; 170.2150; 100.5010

## 1 引言

在医院或诊所里,妇科医生每天要为大量的病人做宫腔镜视频检查,相应的视频被保存在数据库中,用于病例对比、病史追踪、医学研究等。这些视频都是连续的、未经编辑的帧序列,其中只有少部分帧有助于诊断病情,所以宫腔镜视频中存在着大量的冗余数据。专家在查阅这些视频资料时,为避免错过关键信息,只能依次浏览所有图像帧,这一过程非常浪费时间,降低了医生的工作效率<sup>[1]</sup>。为此,相关的研究主要致力于从宫腔镜视频中提取关键帧(摘要),以供快速浏览和检索。具体实现时,采用的

技术主要有 K-means 聚类<sup>[2]</sup>、最小生成树<sup>[3]</sup>、支持向量机<sup>[4]</sup>、稀疏重构<sup>[5-7]</sup>等。此外,文献[8-10]中对多个低级特征(如颜色直方图、颜色矩、惯性矩、边缘等)进行加权融合,用于筛选关键帧。但这些方法都没有使用高级的语义特征,而且都是用于提取普通视频摘要的方法,所以用于宫腔镜视频提取时效果不佳。近年来,Ejaz 等<sup>[11-13]</sup>开始使用视觉注意力模型来解决这一问题。其中文献[11]中在评估注意力值时使用了运动显著性、多尺度颜色对比度和纹理显著性。运动显著性属于动态注意力,它反映了镜头的运动量,从而间接地反映了医生对当前帧的兴趣程度。文献[12]中在 Ejaz 等的基础上增加了

收稿日期: 2018-09-03; 修回日期: 2018-10-05; 录用日期: 2018-10-17

基金项目: 国家自然科学基金(61303028)

\* E-mail: amiao@whut.edu.cn

曲率特征,以强调方向不变性。随后,Muhammad等<sup>[13]</sup>又在此基础上做出改进:1)使用COC(color opponent color)颜色空间,更符合人类视觉系统<sup>[13]</sup>;2)使用积分图像,减少了重复计算。总之,基于注意力模型的方法融合了高级语义信息,因而更符合人类视觉感知,实验效果也较好。但是文献[11-13]都是将注意力曲线等分为 $N_{KF}$ 段,从各段内取注意力最高者为关键帧,而宫腔镜视频中的关键帧并不是均匀分布的,所以人为地强制等分注意力曲线缺乏合理性,也容易选出更多的错误帧,同时其实验效果比较依赖于 $N_{KF}$ 的取值。为此,本文提出了基于图像质量对注意力曲线进行自适应分段的方法。实现时,考虑了锐利度、亮度两个指标。此外,文献[11-13]在计算视觉注意力时均采用的是Lucas Kanade光流法<sup>[14]</sup>,而该算法处理时间较长,为此提出了一种基于SURF(speeded up robust features)特征点匹配的方法,用于计算注意力值。

## 2 基本原理

### 2.1 视觉注意力评估

医生在给患者做宫腔镜检查时,并不是以恒定速度移动摄像设备,其中大部分时间都是用于搜索感兴趣区域,一旦找到相关区域,就会以缓慢的速度移动镜头,甚至保持相对静止,以仔细观察该病理区域,这一区域的视觉注意力值较高。相反,当医生遇到非兴趣区域时,快速移动宫腔镜,相应的注意力值较低。这就是宫腔镜视频中隐含的视觉注意力模型。因此,可以由镜头内物体的位移量来估计镜头位移量(运动显著性),进而计算注意力值。本文的处理思路是从相邻帧中寻找匹配的特征点,然后根据匹配点的位置变化来求取物体的位移量。

本文采用了文献[15]中提出的SURF特征,这是一种局部图像特征,它对缩放、旋转、平移等具有较好的稳健性<sup>[16]</sup>。但是实验发现,直接对宫腔镜视频中的帧提取SURF特征时,无法获取足够多的稳定的特征点。文献[17]中指出,SURF特征点容易集中在图像中纹理特征丰富的区域,据此,先对原始帧求取对应的LBP(local binary patterns)图像,然后在此基础上提取SURF特征点。其中,LBP是一种纹理特征,具有较高的区分性、较低的计算量<sup>[18-19]</sup>。文献[20]中先对原图求取LBP图,然后从中提取SURF特征点,用于服装分类;而文献[21]中首先提取图像的颜色不变量边缘特征<sup>[22]</sup>和CS-LBP(central symmetry-local binary patterns)<sup>[23]</sup>纹

理特征形成融合特征灰度图,并从该图中提取SURF特征点。

对于某一宫腔镜视频,设其采样后共有 $l$ 帧,记为 $\mathbf{F}_{\text{frame}}^{(i)}$ ,其中 $i \in [1, l]$ 。在计算第 $i$ 帧 $\mathbf{F}_{\text{frame}}^{(i)}$ 的镜头位移量(运动显著性)时,具体算法如下:

1) 分别求取 $\mathbf{F}_{\text{frame}}^{(i)}$ 和 $\mathbf{F}_{\text{frame}}^{(i+1)}$ 的LBP图像 $\mathbf{B}_i$ 、 $\mathbf{B}_{i+1}$ 。

2) 从 $\mathbf{B}_i$ 、 $\mathbf{B}_{i+1}$ 中提取SURF特征点,得到对应的点集 $P_i$ 、 $P_{i+1}$ 。

3) 使用RANSAC(random sample consensus)<sup>[24]</sup>算法求取 $P_i$ 与 $P_{i+1}$ 的匹配点对<sup>[25]</sup>,得到 $n$ 个匹配点对的集合: $\{V_1, V_2, \dots, V_j, \dots, V_n\}$ ,其中 $V_j = \{c_1^{(j)}, c_2^{(j)} \mid c_1^{(j)} \in P_i, c_2^{(j)} \in P_{i+1}\}$ ,且 $c_1^{(j)}$ 、 $c_2^{(j)}$ 的像素位置分别为 $(x_1^{(j)}, y_1^{(j)})$ 、 $(x_2^{(j)}, y_2^{(j)})$ 。

4) 计算镜头位移量(运动显著性):

$$M_S^{(i)} = \left[ \sum_{j=1}^n \sqrt{(x_1^{(j)} - x_2^{(j)})^2 + (y_1^{(j)} - y_2^{(j)})^2} \right] / n. \quad (1)$$

按照步骤1)~2)依次计算各帧的运动显著性,得到集合 $\{M_S^{(i)} \mid i=1, 2, \dots, l\}$ 。然后对其进行归一化处理:

$$\tilde{M}_S^{(i)} = M_S^{(i)} / \max\{M_S^{(i)} \mid i=1, 2, \dots, l\}. \quad (2)$$

并依次求各帧的注意力值:

$$A_{\text{IF}}^{(i)} = 1 - \tilde{M}_S^{(i)}. \quad (3)$$

### 2.2 基于离散余弦变换(DCT)系数的图像锐利度评估

文献[26]的研究发现:对于散焦图像,其归一化DCT系数趋于低频加强;对于聚焦准确的图像,其归一化DCT系数趋于均匀分布,所以可以通过评估图像DCT系数分布的均匀程度,来计算图像的锐利度<sup>[26-27]</sup>。图像的DCT系数越趋于均匀分布,则其锐利度越高,而衡量一个分布的均匀程度可以使用熵的方法。贝叶斯熵<sup>[28]</sup>的定义为

$$H^2 = 2 \left( 1 - \sum_{e \in T} p_e^2 \right), \quad (4)$$

式中: $e$ 为域 $T$ 内的点; $p_e$ 是定义在域 $T$ 内的分布函数; $H^2$ 是熵的值。由(4)式可知,分布越均匀,则贝叶斯熵越大。所以可以借鉴贝叶斯熵的形式来定义锐利度的计算方法<sup>[26]</sup>。在本文的应用背景下, $p_e$ 是图像的归一化DCT变换。对于某一帧图像 $\mathbf{F}_{\text{frame}}^{(i)}$ ,其中 $i \in [1, l]$ ,可以将其分为 $M \times N$ 个不重叠的图像块(用 $f_{ab}$ 表示), $a \in [1, M]$ , $b \in [1, N]$ 。然后,计算每个图像块的锐利度(局部锐利度),锐利

度可表示为

$$M_B(f_{ab}) = 1 - \frac{\sum_{\omega=0}^{r-1} \sum_{\nu=0}^{c-1} \tilde{U}_C(\omega, \nu)^2}{\left[ \sum_{\omega=0}^{r-1} \sum_{\nu=0}^{c-1} |U_C(\omega, \nu)| \right]^2}, \quad (5)$$

式中:  $M_B(f_{ab})$  是  $f_{ab}$  的锐利度;  $\tilde{U}_C(\omega, \nu)$  是  $f_{ab}$  的归一化离散余弦变换;  $U_C(\omega, \nu)$  是  $f_{ab}$  的离散余弦变换;  $\omega, \nu$  是频率;  $r, c$  分别是  $f_{ab}$  的高度和宽度, 此处取  $r=c=8$  pixel。

图像  $\mathbf{F}_{\text{frame}}^{(i)}$  的整体锐利度可表示为

$$S_P[\mathbf{F}_{\text{frame}}^{(i)}] = \frac{\sum_{a=1}^M \sum_{b=1}^N M_B(f_{ab})}{M \times N}. \quad (6)$$

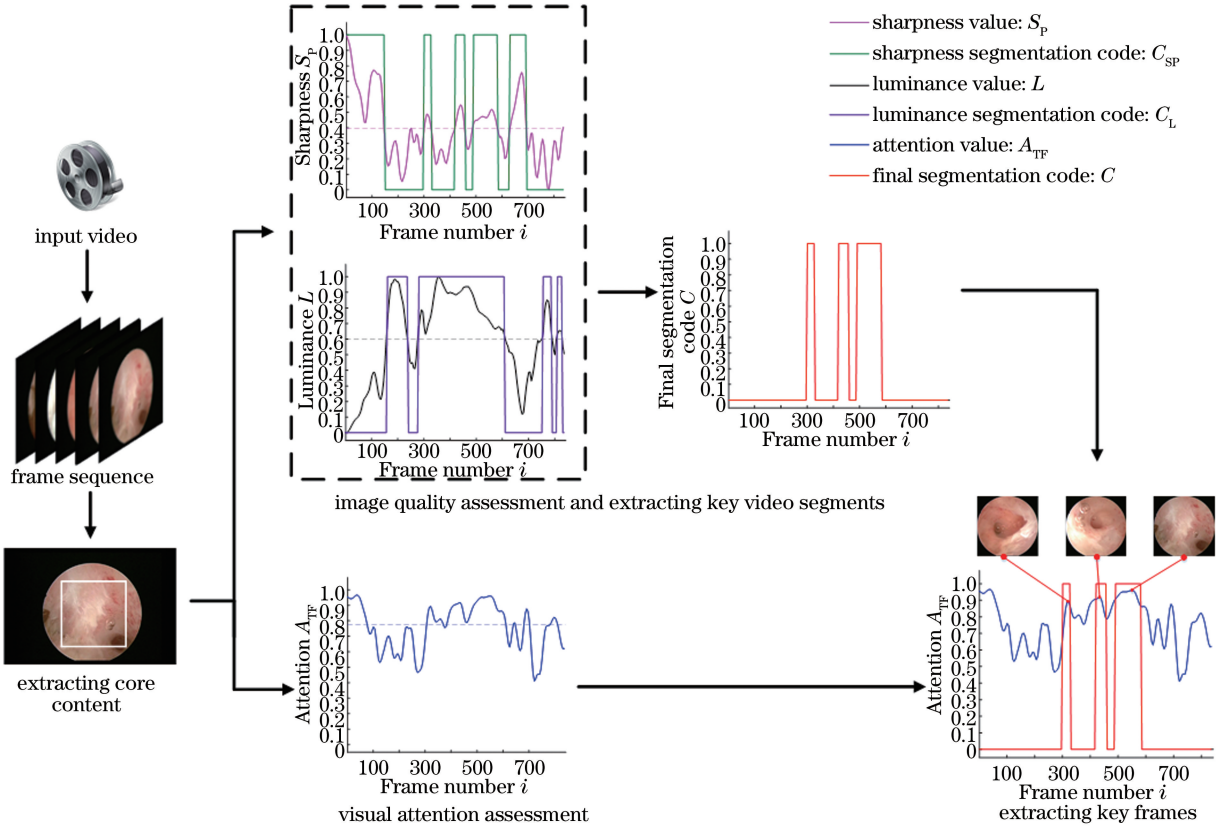


图1 所提方法的整体框架

Fig. 1 Overall framework of proposed method

### 3.1 粗筛选

对于给定的某一视频(采样后,共有  $l$  帧),依次计算其中各帧的图像质量值(锐利度、亮度),得到相应的序列:  $\{S_P(\mathbf{F}_{\text{frame}}^{(i)}) | i \in [1, l]\}$ 、 $\{L(\mathbf{F}_{\text{frame}}^{(i)}) | i \in [1, l]\}$ 。其中,  $S_P(\mathbf{F}_{\text{frame}}^{(i)})$  对应于锐利度,  $\delta_L(\mathbf{F}_{\text{frame}}^{(i)})$  对应于亮度(用灰度均值衡量)。分别计算这两种图像质量序列的均值,得到  $\delta_{S_P}$ 、 $\delta_L$ 。

## 3 基于图像质量和注意力的关键帧提取

所提方法的整体框架如图1所示,主要包括提取核心内容、视觉注意力评估、图像质量评估、提取关键视频段、提取关键帧几个部分。其中在提取核心内容时,先对图像进行二值化处理,通过水平投影、垂直投影来确定视镜有效区域,并截取相应的子图,用于后续的处理。评估视觉注意力的方法如2.1节所述。此外,人体内腔室的复杂性,使得操作者不能很好地控制宫腔镜,导致一些视频帧质量较差,这些帧无助于诊断病情,反而增加了视频冗余度。因此,在图像质量评估方面,综合考虑镜头亮度和图像锐利度,2.1节是用灰度均值来衡量的,2.2节是基于DCT系数来计算的。

以此作为阈值,对相应的图像质量曲线进行分割,高于阈值的视频段为关键视频段,相应的编码为1;反之为无意义视频段,编码为0。于是分别得到锐利度、亮度对应的编码序列  $C_{S_P}$ 、 $C_L$ ,图1展示了相关实例。

两种图像质量评分标准各有侧重,二者互为补充。因此,本文约定对于视频中某个片段,只有当所

有图像质量指标都达标时,才能成为关键视频段。相应地,本文将两种图像质量的编码序列进行与操作,即

$$C = C_{sp} \& C_L, \quad (7)$$

式中:符号“&”表示“与操作”; $C$  为最终的编码序列(图 1 展示了相关实例)。

而那些长度太短的关键视频段,可以被视为噪声、干扰。因此,对长度低于  $T_{\min}$  帧的关键片段予以滤除,这有利于提高视频摘要提取的精确率。其中,取  $T_{\min} = 25$ 。

通过图像质量评估、提取关键视频段,去除大部分图像质量较差的无意义帧,缩小了筛选关键帧的范围,这一部分被视为关键帧的粗筛选。

### 3.2 提取关键帧

对于给定的宫腔镜视频,按照 3.1 节的算法,可以将其分为数个关键片段、无意义片段(二者相间分布)。其物理意义是,每一个关键视频段都对应于一个观察目标(或一个观察区域),而无意义视频段则是其间的过渡片段。

但是,关键视频段内的帧并非都是关键帧。所以需要使用视觉注意力做进一步的筛选。当宫腔镜移动到感兴趣的病理区域时,医生会缓慢地移动镜头,甚至保持相对静止,以仔细观察相应的目标(此时注意力值较高)。根据这一理论基础,本文在各关键视频段中选择注意力值最高的帧作为关键帧(如图 1 所示)。通过此步骤,可以去除那些图像质量较好、但动态视觉注意力较低的帧。

## 4 实验及分析

### 4.1 实验设置

实验环境为 Intel i5-3230M 处理器,12 GB 内存,采用 Matlab R2016a 编程。实验所用数据集包含了 30 个宫腔镜视频,其视频长度从 0.5 min 到 2 min 不等,帧速为 25 frame/s,帧尺寸为 768 pixel $\times$ 576 pixel。另外,由两名妇科专家负责标记出关键视频片段(ground truth),这一点与文献

[13]一致。最终,专家标注出的片段共有 48 个,即  $N_{GT} = 48$ 。相应地,将专家标记的结果转换为编码  $C_{GT}$ ,其中关键片段的编码为 1,否则为 0。

此外,由于所用的视频数据都是实际的宫腔镜检查视频,情况比较复杂,在视频首尾都各有一个无意义片段(准备开始阶段、准备结束阶段),时长均约为 6 s,这属于该数据集本身的特性。所以在实验中,对所有方法都增加如下步骤:遍历所有选出的关键帧,若相应的帧被包括在首尾无意义片段内,则予以去除。

### 4.2 评价机制

对于某宫腔镜视频,假设专家标出  $N_{GT}$  个视频段,记为  $\{K_{seg}^{(g)} \mid g = 1, 2, \dots, N_{GT}\}$ ,某种方法从该视频中提取出了  $N_M$  个关键帧,记为  $\{K_{frame}^{(h)} \mid h = 1, 2, \dots, N_M\}$ 。且在集合  $\{K_{frame}^{(h)} \mid h = 1, 2, \dots, N_M\}$  中,有  $\xi$  个元素分别落在各自对应的片段  $K_{seg}^{(g)}$  内,其中片段序号  $g \in [1, N_{GT}]$ (可能会有多个  $K_{frame}^{(h)}$  指向同一个  $K_{seg}^{(g)}$ )。此时,在集合  $\{K_{seg}^{(g)} \mid g = 1, 2, \dots, N_{GT}\}$  中,有  $\mu$  个元素与集合  $\{K_{frame}^{(h)} \mid h = 1, 2, \dots, N_M\}$  中的元素存在映射关系,即意味着该方法识别出了  $\mu$  个视频段。

在以上约定下计算召回率、精确率和  $F$  值。其中,召回率表征样本中的正例被正确预测的比例;精确率表征在预测为正例的样本中,真正的正样本所占的比例; $F$  值是一种兼顾召回率和精确率的综合性评价指标,是在评价各个方法时的主要参考标准。相应地,计算召回率、精确率和  $F$  值的方法分别为

$$R = \mu / N_{GT}, \quad (8)$$

$$P = \xi / N_M, \quad (9)$$

$$F = \frac{2 \times R \times P}{R + P}. \quad (10)$$

### 4.3 实验结果及分析

表 1 展示了各个方法的实验结果。实验所用视频共 30 个,专家标注出的关键片段共有 48 个( $N_{GT} = 48$ )。

表 1 各方法的实验结果对比

Table 1 Comparison of experimental results by each method

Method	$N_M$	$\xi$	$\mu$	$R / \%$	$P / \%$	$F / \%$	Running time / h
Method in Ref. [1]	121	56	35	72.92	46.28	56.62	15.56
Method in Ref. [2]	192	58	22	45.83	30.21	36.42	<b>0.13</b>
Method in Ref. [11]	110	50	38	79.17	45.45	57.75	12.36
Method in Ref. [13]	109	52	39	<b>81.25</b>	47.71	60.12	12.21
Proposed	102	54	37	77.08	<b>52.94</b>	<b>62.77</b>	1.64

由表 1 可知,在  $F$  值方面,所提方法最佳,为 62.77%;在精确率方面,所提方法也优于其他方法,为 52.94%;但是在召回率方面,文献[11,13]中的方法均比所提方法好,其中,文献[13]的方法最高,为 81.25%。在计算时间方面,文献[13]中使用了积分图像,因此其计算效率高于文献[11]的方法;本文在计算视觉注意力时,没有使用 Lucas Kanade 光流法,而是提出了基于匹配 SURF 特征点的方法(如 2.1 节所述),相应的运行时间明显小于文献[1,11,13]的方法。此外,文献[2]的方法在精确率、召回率、在  $F$  值方面均明显低于其他方法,所以这种通用的视频摘要方法并不适用于宫腔镜视频。

文献[11,13]都是将视频等分为  $N_{KF}$  段,在每个段内选择注意力值最高者作为关键帧。但是关键帧并非一定是均匀分布在整个视频中的,不同的视频情况不同。所以这种方法会选出更多错误的帧,相应地,其精确率偏低。而所提方法考虑到了宫腔镜视频内容的连续性、渐变性,所以识别并去除一系列连续的无意义帧,这样可以避免遗漏。相比于那些将视频序列等分为  $N_{KF}$  段的方法,本文基于视频帧的图像质量变化趋势对自适应视频片段进行分割,分离出关键视频段,再从这些片段中提取关键帧,因而更符合镜头内容的实际变化情况。这就是所提方法的精确率和  $F$  值比其他方法高的原因。

以视频 4 为例,专家标记出了一个关键片段。图 2 给出文献[13]的方法对视频 4 的处理结果,该方法将视频等分为 5 段,然后从各段内选择注意力值最高者作为关键帧,其中,从第 1、5 个片段内选出的两帧都在首尾无意义片段内(首尾无意义视频段对应的注意力曲线显示为绿色),故被滤除。最终该方法选出了 3 个关键帧,但有一个是错误帧(不在专

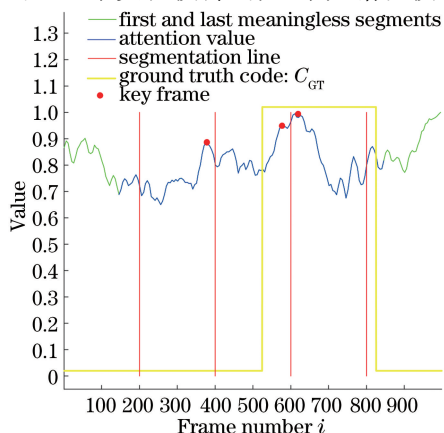


图 2 文献[13]中的方法对视频 4 的处理结果

Fig. 2 Processing results of video 4 by method in Ref. [13]

家标注的关键片段内)。图 3、图 4 给出所提方法对视频 4 的处理结果,所提方法提取出了 3 个关键帧,且全部都在专家标记的关键片段内。这进一步证明了本文所提出的关键帧粗筛选方法的有效性。

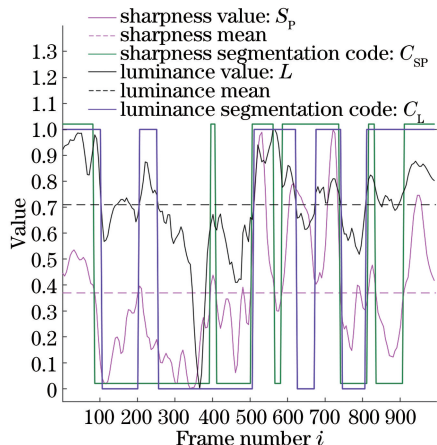


图 3 视频 4 的图像质量曲线及其相应的视频分段结果

Fig. 3 Image quality curve of video 4 and corresponding video segmentation results

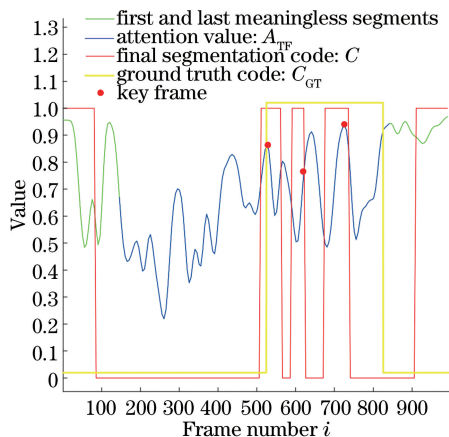


图 4 视频 4 的最终分段结果及所提方法选择出的关键帧

Fig. 4 Final segmentation results of video 4 and key frames selected by proposed method

图 5 展示了各个方法从视频 8 中提取出的关键帧,以及专家标注的真图。其中文献 [1-2] 的方法筛选出的错误帧偏多;文献[13]的方法只选出了一个正确的关键帧;而文献 [11] 的方法以及所提方法的处理结果与专家标注结果最为接近,不过二者都包含了一个错误帧。

## 5 结 论

通过图像质量评估获得相应的质量曲线,并以此进行自适应的视频分段,选出关键片段,从而实现关键帧的粗筛选,这一过程滤除了大部分无意义帧,有效地提高了精确率。此外,取相邻的两帧,求得其

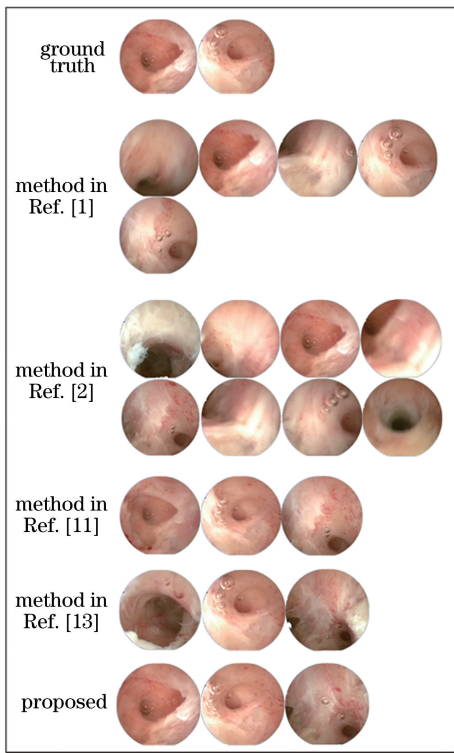


图 5 各个方法在视频 8 中提取的关键帧的对比  
Fig. 5 Comparison of key frames extracted from video 8 by each method

LBP 图像,然后从中提取 SURF 特征点,并根据匹配特征点的像素位置计算运动量,进而评估当前帧的注意力值。在评估视觉注意力时,与 Lucas Kanade 光流法相比,所提方法的计算速度更快,且其精确率、 $F$  值也更高。

### 参 考 文 献

- [1] Gavião W, Scharcanski J, Frahm J M, *et al.* Hysteroscopy video summarization and browsing by estimating the physician's attention on video segments[J]. *Medical Image Analysis*, 2012, 16(1): 160-176.
- [2] de Avila S E F, Lopes A P B, da Luz A, Jr, *et al.* VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method[J]. *Pattern Recognition Letters*, 2011, 32(1): 56-68.
- [3] dos Santos Belo L, Caetano C A, Jr, do Patrocínio Z K G, Jr, *et al.* Summarizing video sequence using a graph-based hierarchical approach [J]. *Neurocomputing*, 2016, 173: 1001-1016.
- [4] Chen J, Zou Y X, Wang Y. Wireless capsule endoscopy video summarization: A learning approach based on Siamese neural network and support vector machine[C] // *International Conference on Pattern*

Recognition, December 4-8, 2016, Cancún Center, Cancún, México. New York: IEEE, 2016: 1303-1308.

- [5] Meng J J, Wang H X, Yuan J S, *et al.* From keyframes to key objects: video summarization by representative object proposal selection [J]. *Proceedings of the IEEE*, 2016: 1039-1048.
- [6] Li J T, Yao T, Ling Q, *et al.* Detecting shot boundary with sparse coding for video summarization [J]. *Neurocomputing*, 2017, 266: 66-78.
- [7] Ma M Y, Met S, Hon J, *et al.* Nonlinear kernel sparse dictionary selection for video summarization [C] // *IEEE International Conference on Multimedia and Expo*, July 10-14, 2017, Hong Kong, China. New York: IEEE, 2017: 637-642.
- [8] Ioannidis A, Chasanis V, Likas A. Weighted multi-view key-frame extraction [J]. *Pattern Recognition Letters*, 2016, 72: 52-61.
- [9] Chen L, Wang Y H. Automatic key frame extraction in continuous videos from construction monitoring by using color, texture, and gradient features [J]. *Automation in Construction*, 2017, 81: 355-368.
- [10] Hamza R, Muhammad K, Lü Z, *et al.* Secure video summarization framework for personalized wireless capsule endoscopy [J]. *Pervasive and Mobile Computing*, 2017, 41: 436-450.
- [11] Ejaz N, Mehmood I, Baik S W. MRT letter: Visual attention driven framework for hysteroscopy video abstraction [J]. *Microscopy Research and Technique*, 2013, 76(6): 559-563.
- [12] Muhammad K, Ahmad J, Sajjad M, *et al.* Visual saliency models for summarization of diagnostic hysteroscopy videos in healthcare systems [J]. *SpringerPlus*, 2016, 5(1): 1495.
- [13] Muhammad K, Sajjad M, Lee M Y, *et al.* Efficient visual attention driven framework for key frames extraction from hysteroscopy videos [J]. *Biomedical Signal Processing and Control*, 2017, 33: 161-168.
- [14] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [C] // *International Joint Conference on Artificial Intelligence*, August 24-28, 1981, Vancouver, British Columbia. [S. l. : s. n.], 1981: 674-679.
- [15] Bay H, Tuytelaars T, van Gool L. SURF: Speeded up robust features [C] // *European Conference on Computer Vision*. Berlin, Heidelberg: Springer, 2006: 404-417.

- [16] Wang M, Li Z Y, Wang C, *et al.* Key frame extraction algorithm of sign language based on compressed sensing and SURF features[J]. *Laser & Optoelectronics Progress*, 2018, 55(5): 051013.  
王民, 李泽洋, 王纯, 等. 基于压缩感知与 SURF 特征的手语关键帧提取算法[J]. *激光与光电子学进展*, 2018, 55(5): 051013.
- [17] Han T Q, Zhao Y D, Liu S L, *et al.* Spatially constrained SURF feature point matching for UAV images[J]. *Journal of Image and Graphics*, 2013, 18(6): 669-676.  
韩天庆, 赵银娣, 刘善磊, 等. 空间约束的无人机影像 SURF 特征点匹配[J]. *中国图象图形学报*, 2013, 18(6): 669-676.
- [18] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. *Pattern Recognition*, 1996, 29(1): 51-59.
- [19] Yang H X, Chen Y, Zhang F, *et al.* Face recognition based on improved gradient local binary pattern [J]. *Laser & Optoelectronics Progress*, 2018, 55(6): 061004.  
杨恢先, 陈永, 张翡, 等. 基于改进梯度局部二值模式的人脸识别[J]. *激光与光电子学进展*, 2018, 55(6): 061004.
- [20] Surakarin W, Chongstitvatana P. Classification of clothing with weighted SURF and local binary patterns[C] // *International Computer Science and Engineering Conference*, Nov. 23-26, 2015, Chiang Mai, Thailand. New York: IEEE, 2015: 1-4.
- [21] Luo T J, Liu B H. Fast SURF key-points image registration algorithm by fusion features[J]. *Journal of Image and Graphics*, 2015, 20(1): 95-103.  
罗天健, 刘秉瀚. 融合特征的快速 SURF 配准算法[J]. *中国图象图形学报*, 2015, 20(1): 95-103.
- [22] Geusebroek J M, van den Boomgaard R, Smeulders A W M, *et al.* Color invariance [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(12): 1338-1350.
- [23] Heikkilä M, Pietikäinen M, Schmid C. Description of interest regions with center-symmetric local binary patterns[M]. *Computer Vision, Graphics and Image Processing*. Berlin, Heidelberg: Springer, 2006: 58-69.
- [24] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [25] Jin J J, Lu W L, Guo X T, *et al.* Position registration method of simultaneous phase-shifting interferograms based on SURF and RANSAC algorithms[J]. *Acta Optica Sinica*, 2017, 37(10): 1012002.  
靳京京, 卢文龙, 郭小庭, 等. 基于 SURF 和 RANSAC 算法的同步相移干涉图位置配准方法[J]. *光学学报*, 2017, 37(10): 1012002.
- [26] Kristan M, Perš J, Perše M, *et al.* A Bayesian-spectral-entropy-based measure of camera focus using a discrete cosine transform[J]. *Pattern Recognition Letters*, 2006, 27(13): 1431-1439.
- [27] Wang Z M. Review of no-reference image quality assessment[J]. *Acta Automatica Sinica*, 2015, 41(6): 1062-1079.  
王志明. 无参考图像质量评价综述[J]. *自动化学报*, 2015, 41(6): 1062-1079.
- [28] Devijver P A. On a new class of bounds on Bayes risk in multihypothesis pattern recognition [J]. *IEEE Transactions on Computers*, 1974, C-23(1): 70-80.