DOI:10.3969/j.issn.1673-4785.201203003
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20120712.1041.005.html

视觉关注转移的事件检测算法

张丽坤[1], 孙建德[1], 李静[2]
(1.山东大学 信息科学与工程学院, 山东 济南 250100; 2.山东省工会管理干部学院 信息工程学院, 山东 济南 250100)

摘 要:智能监控系统已广泛应用于银行、超市、公交车等公共场合,监控视频的事件检测已经成为智能监控中的关键技术.提出了一种基于视觉关注转移的事件检测方法,该方法首先分别通过对视频帧进行动态和静态受关注模型的提取得到视觉关注显著图,然后根据视觉关注显著图的时域特性形成视觉关注节奏曲线,根据视觉关注节奏的变化强度选取关键帧,以关键帧形式表示受关注事件的发生.实验结果表明,算法提取的关键帧可以准确地标示监控视频中特征事件的发生,并且可以做到实时地检测事件.

关键词:智能监控;事件检测;事件发生;视觉关注模型;视觉关注节奏;关键帧提取
中图分类号:TP18;TN911.73 **文献标志码**:A **文章编号**:1673-4785(2012)04-0333-06

Event detection based on visual attention shift

ZHANG Likun[1], SUN Jiande[1], LI Jing[2]
(1. School of Information Science and Engineering, Shandong University, Ji'nan 250100, China; 2. School of Information Engineering, Shandong Institute of Trade Unions' Administration Cadres, Ji'nan 250100, China)

Abstract: Intelligent surveillance systems have been widely used in banks, supermarkets, buses, and other public places. Event detection in a surveillance video is a key technology for this field. In this paper, a visual attention shift-based event detection algorithm was proposed for intelligent surveillance in which the dynamic and static visual attention regions were detected to obtain the visual saliency map. After that, the visual attention rhythm was derived from the visual saliency map temporally. According to the visual attention rhythm, the key frames were selected to label the occurrence of the events. Experimental results demonstrate that the proposed algorithm can label the occurrence of the events with the extracted key frames correctly, and that the event detection is performed in real-time.

Keywords: intelligent surveillance; event detection; occurrence of event; visual attention model; visual attention rhythm; key frame extraction

随着社会对公共安全要求的不断提升,监控系统在银行、商场、高速公路、公交车、地铁站等各种公共场所的应用越来越普遍.监控系统的智能化将有助于遏制一些恶性事件或是危险情况的进一步发展.对突发事件做出及时的处理,对整个社会的公共安全有着非常重要的意义.但是,在现阶段监控系统的智能化还未达到一定程度的情况下,大多数监控视频中的事件检测,主要是靠保安人员不间断地观看来实现.特别是在大型监控平台的环境中,通常配备专职监控人员对监控视频进行实时监控,监控人员需要同时监视一定数量的监控屏幕,这使得监控人员很难长时间保持较高的注意力,并且达到多个屏幕无遗漏的人工监控.在这样的情况下,如何能够使监控系统对所有监控场景中发生的受关注事件不遗漏地、实时地发出警报,以提示监控人员进行特别

收稿日期:2012-03-02. 网络出版日期:2012-07-12.
基金项目:国家自然科学基金资助项目(61001180).
通信作者:孙建德. E-mail:jd_sun@sdu.edu.cn.

关注,成为了智能监控中非常具有实际应用价值的研究问题.

近年来,随着智能视频监控越来越多地受到重视,监控视频中的事件检测技术研究已经取得很大进展. Jiang 等采取了对目标的检测和追踪的方法,将单一物体的运动轨迹定义为视频的一个事件,通过轨迹线聚类来区分正常事件和非正常事件. 但是这种定义忽略了视频的一些时间和空间信息[1]. Liu 等提出了一种基于运动方向统计的异常事件检测方法,他们通过对视频中的连续帧进行分析,采用颜色、纹理、运动等低级特征对视频中的事件进行描述,进而从监控视频中辨别出异常事件. 但是,仅仅靠低级特征表征视频中的事件是远远不够的,需要采用基于人类视觉系统的高级语义特征才能对视频内容进行最准确的描述[2]. 因此,研究者们开始从视觉关注的角度来弥补视频的高级语义特征和低级语义特征之间的鸿沟. Jiang 等提出由视觉关注值表征人眼对视频内容的关注程度的方法,他们采用 K 均值聚类,选出视觉关注值最高的帧作为关键帧,以关键帧来描述视频中事件的内容[3]. Lai 等则通过运动、颜色、纹理等特征,形成静止和运动相结合的显著图,从而形成视觉关注曲线;然后利用时间限制的聚类方法来提取关键帧从而表征视频中的事件[4]. 这些方法在视频中的事件检测和事件描述方面都取得了较好的结果,但是这些方法大多需要对整段视频进行分析,因此很难达到事件的实时检测.

针对上述问题,文章从人的视觉关注特性出发,将视觉关注的转移作为事件检测的依据,针对现有的大多数算法在突发事件检测实时性方面的局限性,提出了一种关键帧触发的、能够对视频内容的变化准确描述的事件检测方法. 该方法将视觉上动态和静态的视觉关注模型进行融合,通过视觉关注模型来提取视频帧中的受关注区域,根据连续帧中的最受关注区域的变化来确定人眼视觉关注点的转移,形成视频关注节奏,根据关注节奏的变化强度来选取关键帧,通过关键帧表明事件发生的时刻,从而触发对受关注事件的提示. 实验结果表明,根据文章提出的算法提取的关键帧能够准确地表示监控视频中特征事件的开始时刻,并且关键帧的提取是实时的.

1 事件检测算法

关键帧是指从原始视频中提取的能够代表视频内容的图像集合,它常被用于视频检索及视频摘要领域[5],而文中提取的关键帧则是用来表征监控视频中事件的发生. 图 1 是提出的监控视频事件检测算法框图.

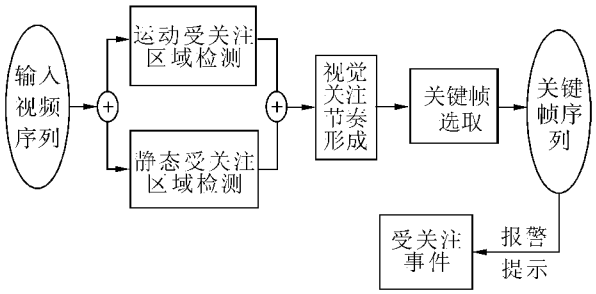


图 1 监控视频事件检测算法

Fig. 1 The framework of the proposed event detection algorithm of surveillance video

2 算法分析

2.1 视觉关注模型

视觉关注是视觉信息处理过程中一个非常重要的方面. 当视网膜拥有整个场景时,注意力只集中在一个或为数很少的几个区域,这几个区域就是图像的受关注区域,它们代表了人眼的视觉关注点,而受关注区域的提取是建立在视觉关注模型基础之上的. 视觉关注模型是根据人的视觉注意机制而建立的模型,它利用视觉注意机制得到图像中最容易引起注意的显著区域,并通过将这些显著区域的显著性用灰度值表示来构成显著图.

研究表明,人类的视觉感知系统对于静止图像中差异明显的区域更为关注,对于图像序列中运动的部分也更加关注. 因此,将视频的动态关注模型和静态关注模型结合形成最终的视觉关注模型,通过视觉关注模型来提取视频帧中的受关注区域.

2.1.1 *动态关注模型*

由于人眼往往对运动的目标更加关注,同时运动目标能够体现视频在时域上的信息,所以动态关注模型的提取过程实际上就是对视频序列中运动目标的检测过程. 在运动目标的检测过程中,使用文献[3]中所提到的基于块的 LK 光流算法计算视频各帧的光流,同时为了弥补运动目标检测中出现的阴影问题,在 LK_{motion} 光流算法的基础上,应用混合高斯背景建模技术,得到每帧的运动前景 G_{fg}.

将 LK 光流算法得到的运动区域跟混合高斯背景建模技术获得的运动前景进行膨胀和腐蚀运算,然后进行归一化形成最终的动态关注模型 T_{sm}:

$$T = \text{dilatation}(LK_{motion}, G_{fg}), \quad (1)$$

$$T_{sm} = \text{erosion}(T). \quad (2)$$

式中:T_{sm}是二值图像,运动区域的像素值为 1,静止区域的像素值为 0.

2.1.2　静态关注模型

静态关注模型的加入主要是从视频的空间信息考虑的. 监控视频中各帧的静态关注模型 S_{sm} 的提取方法为:首先对视频帧进行多尺度变换,在不同尺度上提取颜色、亮度、纹理等低级特征的局部对照特征,然后根据局部对照特征形成特征图,并将形成的特征图进行全局归一化,最后将归一化后的特征图进行线性结合,得到最终的显著图[6].

2.1.3　运动优先融合

动态关注模型和静态关注模型需要结合在一起,得到最终视频帧符合人眼视觉关注的显著图. 由于相比于对静态的关注,人眼往往对于运动的物体更加关注,所以两者在融合的时候采用的权重不能相同,在这里用文献[3]中提到的权重计算方法. 定义动态关注模型和静态关注模型的权重如式(3)、(4):

$$w_T = T'_{sm} \times \exp(1 - T'_{sm}), \tag{3}$$

$$w_S = 1 - w_T. \tag{4}$$

式中:w_T 和 w_S 分别是动态关注模型和静态关注模型的权重. 式(3)中的 T'_{sm} 如式(5)所示:

$$T'_{sm} = \max(T_{sm}) - \mathrm{mean}(T_{sm}). \tag{5}$$

式中:$\mathrm{mean}(T_{sm})$ 反映了视频中运动所占的比例,视频中运动区域面积越大,它的值就越大. 因为 T_{sm} 是二值图像,所以当视频序列中不存在运动时,$\mathrm{mean}(T_{sm})$ 和 $\max(T_{sm})$ 都为 0,此时 w_T 等于零,w_S 为 1,即视频帧中不存在运动,此时视频帧的视觉关注区域取决于静态关注;但两者都为 1 时,说明运动区域遍布整个视频帧,运动无法为视觉关注提供参考,此时视频帧的视觉关注由静态关注决定. 当 $\mathrm{mean}(T_{sm}) \in (0, 0.768]$ 时,$w_T > w_S$,视频序列中的动态关注优于静态关注,这时动态关注在视频帧的视觉关注中占主导地位;当 $\mathrm{mean}(T_{sm}) \in (0.768, 1]$ 时,$w_T < w_S$,静态关注优于动态关注,这时视频帧的视觉关注由静态关注主导.

这样就可以得到符合人眼视觉关注的视频帧的最终视觉显著图 Saliency:

$$\mathrm{Saliency} = w_T \times T_{sm} + w_S \times S_{sm}.$$

显著图 Saliency 的像素值归一化到 0 ~ 1 内. 其中 1 表示受关注度最高,0 表示受关注度最低.

2.2　视觉关注转移

根据人眼的转移机制,人眼关注点的转移意味着有特征事件的出现. 因此,首先要找到视觉显著图中局部最受关注的区域,然后再根据连续帧中的最受关注区域的变化来确定人眼视觉关注点的转移,从而选定关键帧的候选帧. 具体步骤如下:

1)将视频帧的最终视觉显著图分成无重叠的、大小为 8 × 8 的块.

2)计算每块的均值,并找到均值最大的块. 在这里用图像块平均灰度值代表这个图像块的受关注程度,均值越大,这个块的受关注程度越高.

3)以均值最大的块为中心,通过区域扩展形成一个最优的矩形区域作为视觉关注区域. 这个矩形区域要满足矩形面积最小且局部平均像素值最大的要求. 这样就可以获得第 1 个视觉受关注区域.

在选择最优矩形区域 R_k 时,采用文献[8]中的方法:

$$\text{如果} \frac{Sa(x_i)}{N_{R_k}} \geqslant T_k, \text{扩展 } R_k, \text{直到} \frac{Sa(x_i)}{N_{R_k}} < T_k.$$

$$T_k = \partial_k \times \mathrm{avg}(\mathrm{maxblock}_k\_Sa(x)).$$

式中:$k \in \{1, 2\}$ 代表被选定的受关注区域的数目,$x_i \in R_k$ 表示矩形 R_k 中的像素,N_{R_k} 表示矩形 R_k 中总像素数,T_k 是选择最优矩形区域时设定的阈值,$\mathrm{avg}(\mathrm{maxblock}_k\_Sa(x))$ 是第 k 个最大均值块的均值,$\partial_k \in (0, 1)$ 是一个经验数据. 在实验中,选用 $\partial_k = 0.9$,即将最大均值块均值的 90% 设定为选择最优矩形区域时的阈值. 每一个受关注区域可以表示为[7]:

$$R_1 = \mathrm{Re}\ ct(c_1, r_1, W_1, H_1),$$

$$R_2 = \mathrm{Re}\ ct(c_2, r_2, W_2, H_2).$$

式中:$\mathrm{Re}\ ct(\cdot)$ 代表设定矩形区域的函数,(c_k, r_k) 是矩形区域左上顶点的坐标,W_k 和 H_k 分别是矩形区域的宽度和高度.

当第 1 个视觉受关注区域确定以后,将这部分区域的像素值置为零,然后返回到 2)进行,直到第 2 个受关注区域被选定为止.

4)分别计算视频帧所选 2 个区域的均值,用均值的变化来表征视觉关注的转移.

当 $av_1(i) > av_2(i)\ \&\ av_2(i+1) > av_1(i+1)$ 或者 $av_2(i) > av_1(i)\ \&\ av_1(i+1) > av_2(i+1)$ 时,表示在第 i 帧时刻发生了视觉转移,此时将第 $i+1$ 帧选出作为关键帧的候选帧,这里 $av_1(i)$ 和 $av_2(i)$ 表示第 i 帧选出的 2 个受关注区域的均值.

5)获得视觉关注节奏曲线. 定义视觉转移量来表示视觉转移程度的大小. 视觉转移量是指视觉转移之前,视觉关注保持在某一关注区域的时间,它可以用视频中没有视觉转移发生的这段时间内视频的帧数来表示:

$$\delta_t(i) = \int_0^T \mathrm{d}t,\ T \in N.$$

式中:T 是没有视觉转移发生的一段时间内出现的视频帧数,N 是正整数或零,$\delta_t(i)$ 为第 i 帧出现时刻的视觉转移量.

视频当前帧选出的 2 个关注区域的均值与前一帧选出的 2 个关注区域的均值相比没有发生变化时,视觉转移量为零;若发生变化,视觉转移量累加,

以此获得视觉关注节奏曲线.

6)获得视觉关注节奏后,根据视觉转移量选定关键帧.

2.3 受关注事件报警或提示

视频中往往会出现一些与场景中大部分人的共性行为不一致的行为,或是场景中突然新增加一些人、事物以及发生一系列新的动作的情况,比如人群在步行前进却有一个人跑步前进或突然有一个人闯入一个环境中等. 这些与人们的共性行为不一致的行为的发生往往会引起观察者的注意,因此,它们往往被认为是视频中的受关注事件.

根据人眼的观看习惯,视觉转移的形式主要有2种:1)视觉注视点频繁地发生转移;2)视觉转移隔一段时间才发生一次. 第1种形式通常发生在事件进行过程中. 在这种情况下,由于视觉关注点转移的时间间隔短,所以视觉转移量往往不大;而第2种形式往往出现在新事件发生的时刻. 这种情况通常是在视觉关注点在某一区域保持了一段时间后才发生的,因此,视觉转移量往往较大. 所以,只要设定合适的视觉转移量门限值,就可以从关键帧的候选帧中选定表征事件发生的关键帧. 在所选关键帧对应的时刻进行报警或提示,以标示受关注事件的发生,就可以达到事件检测的目的.

3 实验结果及其分析

实验采用的计算机配置为 Pentium Dual-Core CPU E 5300@2.60 GHz,2GB 内存,Windows XP 系统中,仿真程序通过 Matlab 编程实现. 在进行实验时选用了标准的"Hall Monitor"视频,这个视频包括300帧. 对每一帧进行动态关注模型、静态关注模型的提取并融合,根据局部灰度像素均值最大的方法选择最受关注区域,进而得到图2所示的监控视频的视觉关注曲线,根据对"Hall Monitor"视频的运动量进行统计计算,设定转移量的门限值为15.

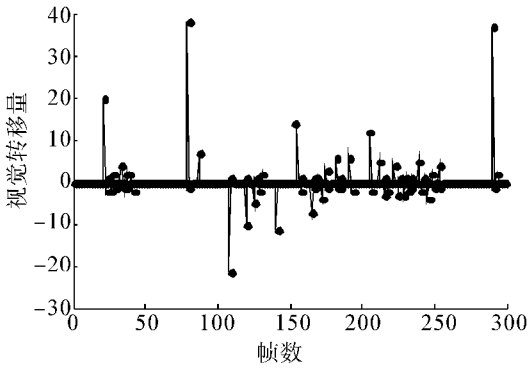


图2 视觉关注曲线

Fig.2 The curve of visual attention

从图2可以看出,视觉转移量大于15的帧中,第1帧和第2帧以及第2帧和第3帧之间基本上没有视觉转移量的起伏变化,符合2.3所述的第2种形式,说明此时刻事件的突发性. 根据图2选出的关键帧如图3所示,图3包括第1个人进入监控(图3(a))、第2个人进入监控(图3(b))、第1个人离开(图3(c))、第2个人从监控中离开(图3(d))的关键帧.

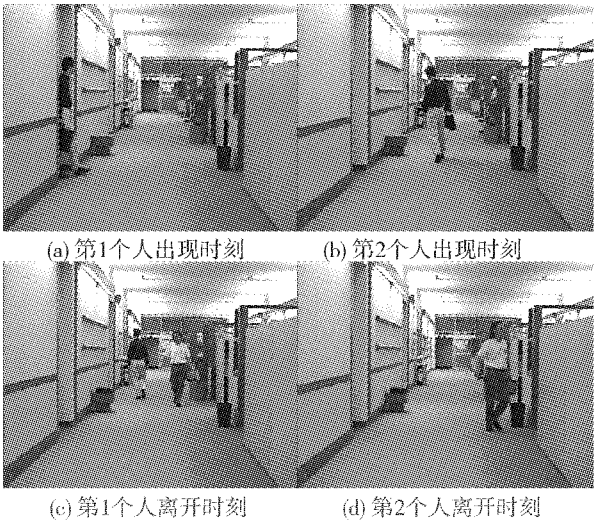


图3 门限值为15时视频"hall"中选出的关键帧

Fig.3 The extracted key frames when the threshold is 15

由图3可以看出根据视觉关注节奏曲线选出的关键帧能够代表视频中主要受关注事件出现的时刻,在这样的时刻进行报警,就可以达到事件智能检测的目的. 实验中视频相邻2帧之间的变化所需的时间大约是0.004 11 s,能够证明该方法能够实时地检测事件. 当设定门限值为10时,所选的关键帧如图4所示,此时2个人进入监控后的一系列连续的动作(图4(c)~(f))被检测出,这时选出的帧处于2个事件之间,表明了事件发展过程中的主要动作,即第1个人放下物体(图4(c)、(d)),第2个人拎起另一个物体(图4(e)、(f)).

(e) 第2个人开始拿起物体　(f) 第2个人拿起物体

(g) 第1个人离开时刻　(h) 第2个人离开时刻

图 4　门限值为 10 时视频"hall"中选出的关键帧

Fig.4　The extracted key frames when the threshold is 10

此外,又选用了实验室监控视频对算法进行验证. 这个视频各帧的 mean(T_{sm}) 比"Hall Monitor"视频的大,说明这个视频中运动区域较多,运动更为复杂,这里通过对视频运动量的统计计算,设定视觉转移量的门限值为 10. 实验结果如图 5 和图 6.

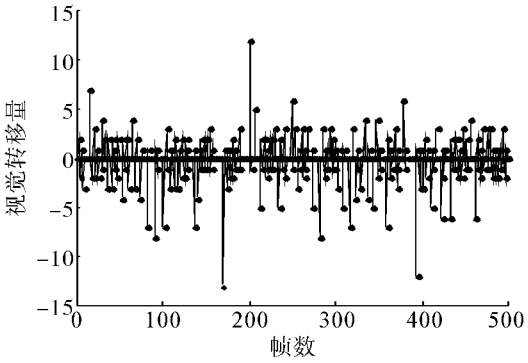
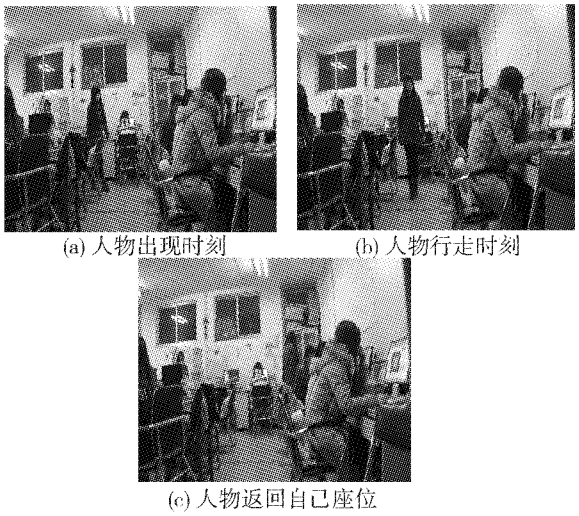


图 5　实验室监控视频的视觉关注曲线

Fig. 5　The curve of visual attention



(a) 人物出现时刻　(b) 人物行走时刻

(c) 人物返回自己座位

图 6　实验室监控视频中选出的关键帧

Fig. 6　The extracted key frames

图 5 是视频所对应的视觉关注节奏曲线,可以看出,在选定的帧中,各帧之间有比较频繁的视觉转移量的起伏变化,说明视觉转移一直在发生,事件一直在进行,符合 2.3 中提到的第 1 种形式. 在这种情况下,到达某个时刻时视觉转移量突然增加,表明这个时刻可能会有新事件的出现. 这个视频相邻 2 帧之间的转移所花时间大约是0.004 14s,也能证明该方法的实时性. 图 6 是根据图 5 选择的关键帧. 可以看出人物出现(图 6(a))、行走(图 6(b))、返回自己座位(图 6(c))等一系列关键动作都被检测出来.

4　结束语

提出了一种基于视觉关注转移的关键帧提取的事件检测算法,首先采用 LK 光流算法和混合高斯建模技术检测视频帧中的运动区域,进而获得视频帧的动态关注模型;同时,对视频中各帧提取静态关注模型;将这 2 类关注模型融合得到视频帧的视觉关注模型,通过视觉关注模型来提取视频帧中的受关注区域;然后根据连续帧中的最受关注区域的变化来确定人眼视觉关注点的转移,获得视觉关注节奏;最后根据视觉关注节奏中视觉转移量的大小选定关键帧,在关键帧出现的时刻进行报警或提示,以此来表征受关注事件的发生. 在下一步工作中,将研究视觉转移量阈值的自适应设定方法. 文中提出的方法能够较好地表征受关注事件的发生,但是需要进一步改进以应对更加复杂的视频监控场景.

参考文献:

[1] JIANG F, WU Y. A dynamic hierarchical clustering method for trajectory-based unusual video event detection[J]. IEEE Transactions on Image Processing, 2009, 18(4): 907-913.

[2] LIU C, WANG G J, NING W X, et al. Anomaly detection in surveillance video using motion direction statistics[C]// IEEE International Conference on Image Processing. Hong Kong, China, 2010: 717-720.

[3] JIANG P, QIN X L. Keyframe-based video summary using visual attention clues[J]. IEEE Transactions on Multimedia, 2010, 17(2): 64-73.

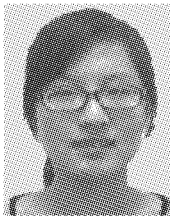
[4] LAI J L, YI Y. Key frame extraction based on visual attention model[J]. Journal of Visual Communication and Image Representation, 2012, 23(1): 114-125.

[5] AMIRI A, FATHY M, NASERI A. Key-frame extraction and video summarization using QR-decomposition[C]// IEEE International Conference on Multimedia Technology and Applications. Wuhan, China, 2010: 134-139.

[6] ZHANG J, SUN J D, YAN H, et al. Visual attention model with cross-layer saliency optimization[C]//IEEE International Conference on Intelligent Information Hiding and Mul-

timedia Signal Processing. Dalian, China, 2011: 240-243.
[7] 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.
[8] RUBINSTEIN M, SHAMIR A, AVIDAN S. Multi-operator media retargeting [J]. ACM Transactions on Graphics, 2009, 23(3): 1-8.

作者简介:



张丽坤,女,1986年生,硕士研究生,主要研究方向为多媒体信号处理.

孙建德,男,1978年生,副教授.主要研究方向为基于内容的多媒体分析、基于视觉关注模型的图像/视频分析、图像/视频复制检测、多媒体信息内容安全和数字水印等.承担及完成科研项目10余项,授权发明专利8项,发表学术论文40余篇.

李静,女,1979年生,主要研究方向为多媒体通信.

第9届国际机器学习和数据挖掘会议
9th International Conference on Machine Learning and Data Mining (MLDM) 2013

The aim of the conference is to bring together researchers from all over the world who deal with machine learning and data mining in order to discuss the recent status of the research and to direct further developments. Basic research papers as well as application papers are welcome.

All kinds of applications are welcome but special preference will be given to multimedia related applications, biomedical applications, and webmining. MLDM'2013 is the 9th event in a series of MLDM events that have been originally started out as a workshop.

Web site: http://www.mldm.de/.