针对实时视觉通信的图像序列自动提炼

黄沛杰[1,4],朱立华[2],刘学慧[1,4],吴恩华[1,3,4],王传铭[2]
1. 中国科学院软件研究所计算机科学国家重点实验室,北京 100190; 2. 汤姆逊北京研究院,北京 100192;
3. 澳门大学科技学院电脑资讯系,澳门;4. 中国科学院研究生院,北京 100190)

摘 要: 本文提出了一项新技术,它可以实现从任意的图像序列中自动提炼出简洁的表达方式,以便进行高效的视觉通信.我们认为,视觉通信的全过程可分为视频数据的传输和人眼对视觉信号的理解两个阶段.因此,本文以心理学中人对图像的认知规律的相关理论为指导,专注于研究如何同时提高图像的可压缩性和可理解性.我们借助一个缘提取算法来保留对人的视觉系统最为敏感的物体边界,再用一个非线性扩散算法来减弱无足轻重的细节信号.为了使最终生成的动画保持时间上的一致性,本文的技术方案是在整个时空域上设计的.而我们依然能够保持实时的处理速度,因为该方法可以方便地使用 GPU 作并行计算.为了演示新技术的实用性,我们还建立了一个以本文算法作为处理内核的完整的视觉通信系统并在该系统进行所有实验.统计数据表明,本文方法不仅可以明显地降低传输带宽,而且提高了图像序列的可理解性.

关键词: 非真实感绘制; 视觉通信; 图像提炼; 视觉感知
中图分类号: TP391 **文献标识码**: A **文章编号**: 0372-2112 (2009) 4A-042-09

Automatic Abstraction of Image Sequences for Real-Time Visual Communication

HUANG Pei-jie[1,4], ZHU Li-hua[2], LIU Xue-hui[1,4], WU En-hua[1,3,4], WANG Chuan-ming[2]
(1. *Institute of Software, Chinese Academy of Sciences, Beijing* 100190, *China*; 2. *Thomson Corporate Research Beijing, Beijing* 100192, *China*;
3. *University of Macau, Macau, China*; 4. *Graduate University, Chinese Academy of Sciences, Beijing* 100190, *China*)

Abstract: We present a novel technique for efficient visual communication, where a compact representation is abstracted from arbitrary image sequences automatically. We aim at improving the entire visual communication process, including transmission of video data and perception of visual signal by human eyes. Therefore, guided by some related theory in Psychology, our work increases both the compressibility and comprehensibility of images. This goal is achieved by two steps. One step is an edge extraction process for preserving object boundaries which are believed to be the most sensitive features to human vision, and the other step is a non-linear diffusion process for reducing extraneous details. Both are designed in a spatiotemporal manner to guarantee the temporal coherence of resulting animations, while real-time processing speed is maintained by facilitating parallel computation on a GPU. We additionally build a complete visual communication system with the proposed algorithm as its core to demonstrate the practicality of our technique. Experimental statistics collected on the system indicates that transmission bandwidth is obviously saved while perceptibility of image sequences is improved.

Key words: non-photorealistic rendering; visual communication; image abstraction; visual perception

1 引言

在日常生活中,人们经常使用视频来传递视觉信息.因此,研究人员开发了各种各样的视觉通信系统来帮助人们进行思想交流和信息共享.在这些系统中,通常需要对视频数据依次进行压缩、传输、解压等处理之后才能给用户观看.影响通信效率的关键因素有两个,一是视频数据的传输速度,二是人们理解视频信号所传递的内容的难易程度.对于第一个因素,虽然视频压缩领域的大量研究成果已经为快速视频传输打下良好的基础,但是大部分是针对通用的视频数据的,而并不考虑被压缩数据的应用场合.相信如果加以考虑视频的具体应用目的,将能够进一步提高压缩性能.对于第二个因素,尽管图像的可理解性正是非真实感绘制(NPR)领域的主要研究目的[1],但就目前而言,还鲜有工作能够深入讨论压缩后的视频图像的可理解性问题,也鲜有工

收稿日期:2008-08-06;修回日期:2008-12-23
基金项目:国家科技 973 计划项目(No. 2009CB320802);国家自然科学基金项目(No. 60833007);863 科技(No. 2008AA01Z301);澳门大学研究计划

作能够用详实的实验数据探索如何通过 NPR 技术来提高传输效率.

在很多实时应用场合,人们通常只关注图像中最有意义的部分,而不关心各种无足轻重的细节. 例如,在车辆监控系统所拍摄的视频中,只需要记录汽车型号和车牌,而可以忽略路面上的纹理. 在这样的前提下,我们只需要保留甚至扩大对人眼最为敏感的视觉信号,并同时减弱视频中的无关细节. 这个过程被称之为视频提炼. 本文证明了,提炼算法可以同时提高视觉通信中的传输和理解两个过程的效率.

本文提出了一个实时视频提炼的新方案来达到上述目标. 与 Winnemoeller 等人的工作[2]类似,本文算法通过建模视觉显著点来简化低对比度区和增强高对比度区. 然而,跟他们的方法不同的是,我们的技术可以在保持实时处理速度的情况下显著地去除时间上的抖动. 首先,我们使用时空域上的边缘提取技术来抽取物体的轮廓,使该轮廓在空域上和时域上都是一致的. 接着,我们依次使用空域扩散算法和时域扩散算法来分别达到简化图像和降低抖动误差两个目的. 为了进行快速的各向异性扩散,本文设计了一个可分离的反高斯双边滤波器,并在 GPU 上实现. 上述这几个步骤结合在一起,构成了一个新颖的时空一致的实时视频提炼解决方案.

利用本文技术,我们不仅节省了传输带宽,而且提高了信息传达的精度和效率. 之所以能够达到这个目标,主要取决于两个事实. 第一,实时视觉通信应用通常可接受简化的表达方式. 我们针对这一应用需求,设计新的处理方案来进一步降低压缩后的码率. 实际上,本文是第一个通过详实的实验数据(见第 4 节)来探索压缩性能与 NPR 算法之间的关系的工作. 第二,作为基于 NPR 的技术之一,本文方法旨在提高已经被压缩、传输和解压过的视频信号的可感知性. 众所周知,在有损压缩过程中,有些内容必然会丢失. 但是我们把视频转化为一种特殊的风格,因而得以在压缩过程中保护最具意义的图像特征.

除了关注传输与理解的效率以外,本文算法还能顺便将视频转变成卡通风格. 近年来,一些卡通电影,如 2001 年的《梦醒人生》和 2006 年的《盲区行者》(如图 1 所示),通过将实拍视频转化为卡通动画而给观众带来了全新且震撼的视觉体验. 就我们所知,制造这一类电影需要巨大的手工劳动量. 相比之下,本文系统可以完全自动地从实拍视频出发实时地产生卡通动画.

图1　电影《盲区行者》

为了演示本文技术的实用性,我们将新算法整合到一个名为 Toontalk 的完整的视觉通信系统中,如图 2 所示. 该系统以摄像机捕捉的图像序列作为输入,在服务器上运行本文算法对序列进行提炼,然后压缩为 MPEG2、MPEG4 等格式. 最后,服务器将压缩后的码流通过互联网、WIFI 或广播等各种通道传送出去. 在个人电脑、PDA 或数字电视等接收终端,码流被解压后播放. 我们在 Toontalk 上进行本文的所有实验. 实验结果表明,本文技术确实既大大节省了码率又为观众提供舒服而又易于理解的视频节目.

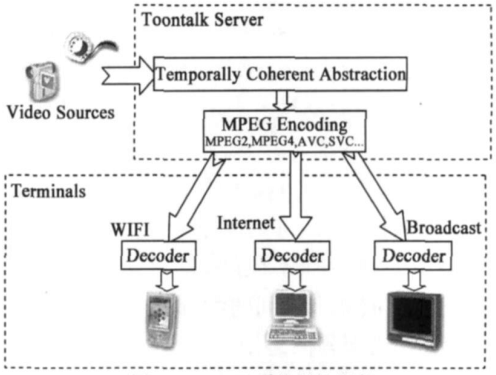


图2　Toontalk系统整体框架

需要指出的是,本文的技术是针对只关注主要信息结构的实时视觉通信的应用场合而提出的. 作者的视觉通信概念仅供学术探索之用. 除此,本研究领域中还有诸多优秀的工作可以进行高效的通信并保留细节,希望本文能作为一个有价值的补充.

2　相关工作

本文方法得以奏效的根源在于非真实感绘制技术和视频编码技术的共同发展. 所以,本文工作属于视频处理与图形学的交叉领域,而这一领域越来越受到关注. 同样,Pham 和 Vlient[3]的工作也属于该范畴,因为他们用 RMS 误差估计模型和 MPEG 质量分数来判断双边滤波后压缩率的提升. 但是他们只着眼于视频传输性能,而我们还考虑了人的感知过程. 另外,我们采用了新的双边滤波算子,使得压缩效果更为理想(见表 2). Collomosse 等[4]讨论了卡通表达法的紧凑性,只不过他们的方法在天性上无法支持实时视频编码的需求.

非真实感绘制的早期工作主要是针对三维场景的,如文献[5,6],但它们要求输入物体的几何属性. 而

本文是直接在图像序列上操作的.为了克服由于单纯的图像缺乏几何信息所带来的巨大困难,有人采用了一些特殊的硬件.Raskar 等[7]使用多闪光摄像机来从图片中恢复原始深度,而 Decarlo 和 Santella[8]则利用眼睛跟踪仪来指导图像简化过程.与他们不同的是,本文方法不需要任何特殊硬件,因此大大提高了应用的普遍性.

在普通单机上进行图像的非真实感绘制的技术可分为基于笔划的和基于像素处理两大类.前者通过模拟笔划效果来重绘图像,以生成各种艺术风格,如印象派画法[9],水彩画[10,11],中国水墨画[12]以及其他效果[13]等.还有人把这些方法从静态图片扩展到视频中[9,10,14,15],只是它们的处理速度往往很慢.这是因为,基于笔划的绘制系统要么需要用几遍来绘制一帧[13],要么需要对物理现象进行昂贵的模拟计算[11,12].因此,基于笔划的方法不适合实时视觉通信.而基于像素处理的算法的性能就要高得多,使得实时处理成为可能.Winnemoeller[2]引入了三种已有的图像处理算法来对视频进行实时提炼,分别是:双边滤波,亮度量化和边缘检测.尽管他们的方法与本文工作有一定的相似性,但两者至少在三个方面存在本质上的不同:时域噪声控制,边缘检测算子和双边滤波计算核.

把图像风格化算法应用到视频上所遇到的最大挑战,莫过于如何减少时域上的抖动.作为第一个将视频剪辑自动转化成油画序列的学者,Litwinowicz[9]根据光流场将笔划沿着像素的运动方向在帧间移动.Hertmann 等[15]也使用光流来评估连续帧之间的差异,以便只重绘变化足够大的区域.在这些油画绘制技术中,光流被用来引导笔划的修改.而本文使用光流场来引导连续帧上的时态滤波操作.

为了完全控制抖动,很多研究人员将视频当成图像数据的一个时空体,以便在时空体上进行整体分析,例如风格化视频立方体[16]、笔划曲面[4]和视频卡通[17]等.基于体的技术能够产生非常高质量的风格化视频,而且出现的抖动极少.但是由于它们的处理过程相当耗时,因此通常只能作为离线工具.而另一些方法在每一帧的内部维持时间连续性,这样就大大减少了处理时间,从而增加了交互性,如 snaketoonz[18]、跳帧技术[15]和量化技术[2].可惜的是,在遮挡关系比较复杂的情况下,单帧处理的方法都不是很鲁棒,从而导致闪动的出现.与以上基于体的和基于单帧的技术都不同的是,本文的解决方案提供了一个在交互性和视频质量之间的合理折衷.在本方案中,提炼和风格化过程是在一个连续帧的有限缓冲区上进行的,这样就减少了抖动效果又保证了实时的速度.这一策略实际上是借鉴了视频水彩化的相关工作[10],但我们的目标是视频提炼.

正如文献[1]所言,尽管大部分非真实感绘制技术旨在增强图像的可理解性,但是并没有清晰而且规范的方法来评估生成的 NPR 图像.早期的评估方法有问卷调查[19]或者版面设计反馈[20]等.近期不少方法将用户完成某些精心设计的任务的效率作为评价的标准.Gooch[21]在探索 NPR 虚拟环境中的空间感知问题时,让实验主体通过原始照片和 NPR 示意图[22]两种媒介来学习和识别不熟悉的脸孔,然后记录主体进行识别的速度和精确度.Winnemoeller[2]也借用了这种实验方法.但是,这些任务相关的评估是一种间接方法.很多外部因素会影响到参与者的反应,比如他们的精神状态等.与之不同的是,本文工作通过一个评分程序来协助专家直接测量提炼视频的可感知性.更进一步,我们是对被压缩过、传输过然后再解压过的视频数据做评估的.这些处理过程在真正的应用场合中都是很常见的.因此我们的评估结果更具实际意义.

3 算法描述

本文基于一个论断:物体的边界或轮廓构成了图像中最有意义的元素.该论断在人眼感知规律和视频压缩原理上都是成立的.对于人的感知而言,由于亮度和色彩对比度等图像特征在低层视觉中起着举足轻重的作用[23],而物体边界又是这些图像特征在空间上骤变的地方,因此边界是视觉上最显著的.又如 Tufte[24]所建议的,必须尽可能地淡化非边界区以突出对感知过程最敏感的特征.对于视频压缩而言,其最终目标是尽可能地节省压缩数据的码率,同时又要保持表征视频内容的必要信息.我们观察到物体的边界代表了图像内容的主要结构,所以它们必须在整个编码过程中得到保留.而那些与物体轮廓无关的像素可以被平滑地滤掉,以得到更大的相似性,这样也就增加了冗余度从而得到更高的码率.

因此,在这样的论断下,本文进行图像提炼的基本思想就是保留和强调物体的边界,同时去除无关紧要的纹理细节.首先,为了提取物体边界,我们采用边缘检测算法,因为边缘有很高的对比度,最有可能构成物体的轮廓或轮廓的一部分[8].接着,我们运行非线性扩散算法来滤掉高频特征从而去除细节.最后,我们把提取到的边缘和扩散后的结果进行叠加,得到最终提炼后的视频.为了避免边缘提取和像素扩散可能引进的抖动,我们在这两个过程中都应用了时空域滤波,以保证计算结果不仅在空间上而且在时间上都是一致的.时域滤波是在一个有限长度的连续帧缓冲区上执行的,执行过程借助了从原始视频中计算出来的光流场.这样,本文的计算框架由一个时空边缘提取过程和一个时空非线性扩散过程构成,如图 3 所示.在下面,我们

将在三个小节中分别介绍这两个步骤以及时域滤波的过程.

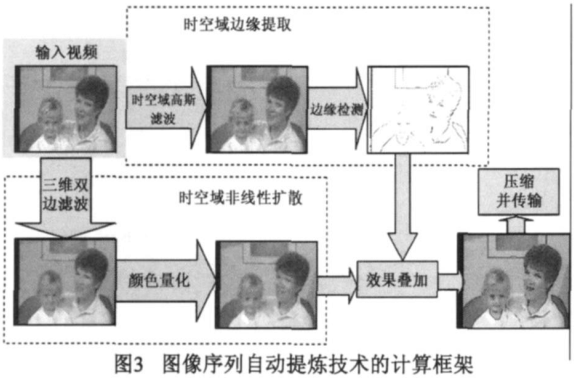


图3　图像序列自动提炼技术的计算框架

3.1　时空域边缘提取

该过程依次由时空滤波和边缘检测两个步骤来实现.

3.1.1　时空滤波

我们选择边缘来表达物体边界的动机源于一个基本事实:真实世界的物体在不同尺度的观察下具有不同层次的结构[25]. 精细的结构可能带有噪声而粗糙的结构只表示了高度抽象的模式. 因此,为了精确地把边缘点定位在物体边界上,我们需要选择一个适当的尺度以在边缘检测之前过滤视频帧. 我们不仅在空域上而且在时域上进行滤波,所以对空间维和时间维都采用相同的滤波尺度以使物体边界在时空上的精细度保持一致.

本文选择高斯滤波算子,因为该算子在缺乏先验知识的情况下,可以很好地寻找前端视觉中的模型[26]. 而且,由于高斯滤波算子的可分离性,我们可以很方便地用三遍一维高斯滤波来实现时空滤波,即依次在 x 方向,y 方向和时间轴上进行滤波. 通过实验,我们发现取方差为 1.4 这个尺度适合大多数情况.

3.1.2　边缘检测

研究人员已经设计出了很多边缘检测算法,各自带有不同的检测精度和计算复杂度. Canny[27] 发明的一个很著名的检测算子被广泛使用,如 Decarlo[8]. 但是,庞大的计算开销使 Canny 算子不适用于实时检测. Winnemoeller[2] 使用高斯相减法(Difference-of-Gaussians,DoG)来逼近 M-H 算法[28],即在一个宽高斯的滤波结果中减掉一个窄高斯的滤波结果. 他们利用高斯函数的可分离性来降低时间代价,只不过在某些情况下他们所得到的边缘会偏离真正的物体边界.

在本文的 GPU 实现中,我们发现基于高斯拉普拉斯法(Laplace of Gaussian,LoG)的 M-H 算法比基于高斯相减法的要快. 这是因为 LoG 只需要一次高斯滤波而 DoG 需要两次. 而且,我们证明了,LoG 也是可分离的,见附录 A. 给定一张输入图像 $I(\cdot)$,作用在像素 p 上的方差为 σ 的 LoG 算子 $L(\cdot)$ 由式(1)定义. 其中 ∇^2 对二阶非最大微分导数进行求和,而 Ω 代表高斯滤波核.

$$L(p,\sigma)=\frac{1}{2\pi\sigma^2}\sum_{q\in\Omega}\nabla^2\left(e^{-\frac{\|p-q\|^2}{2\sigma^2}}\right)I(q) \qquad (1)$$

当把 LoG 应用到离散图像上时,可得*

$$L(p,\sigma)=\frac{1}{2\pi\sigma^2}\sum_{q\in\Omega}\frac{\Delta x^2+\Delta y^2-2\sigma^2}{\sigma^4}e^{-\frac{\Delta x^2+\Delta y^2}{2\sigma^2}}I(q) \qquad (2)$$

其中:$\Delta x=x_p-x_q$,$\Delta y=y_p-y_q$

表 1　采用 DoG 和 LoG 算子的 M-H 边缘检测算法所处理帧率比较
(R 是滤波核的半径)

边缘检测算子	$R=4$	$R=5$	$R=10$
DoG	126fps	76fps	45fps
LoG	225fps	169fps	126fps

由附录 A 可知,LoG 算子可以依次在 X 方向和 Y 方向上分别执行. 我们在不同的核尺寸下比较了 DoG 和 LoG 的边缘提取性能,如表 1 所示. 实验环境的配置将在第 4 节讨论. 在这里,我们使用 Winnemoeller 的 Cg 代码来实现 DoG,而附录 A 给出了本文中 LoG 算子的 GPU 实现. 统计数据表明,基于 LoG 的 M-H 边缘检测算法比 Winnemoeller 的基于 DoG 的算法快一到两倍. 在 Toontalk 中,取 $\sigma=1$,$\Omega=3\times3$.

在检测得到的边缘图中,难免还存在一些噪声,尤其是当输入图像中带有复杂的高频纹理时,如美式足球序列上的草地. 为了除掉这类噪声,我们还在边缘图上采用了形态学的闭运算(即先膨胀后腐蚀)[29]. 如图 4 所示,相当多噪点被去掉了.


(a) Edge Image　(b) After Dilation　(c) After Erosion

图4　在边缘图上进行形态学闭运算. (a) 原始图像;(b) 在(a)上膨胀后的边缘图;(c) 在(b)上腐蚀的边缘图

3.2　时空域非线性扩散

为了减弱不重要的细节,本文在整个图像序列上应用了非线性扩散算法,因为非线性扩散可以平滑掉小的不连续区并锐化边缘[30]. Winnemoeller 等人也用非线性扩散来提炼视频,但是他们是在每一个单帧上做的,而并不考虑帧间的运动,所以结果中有明显的抖动误差. 相反,我们在时空域上扩散像素. 即对于一个像素,我们不仅用它所在帧上的周围像素,而且用相邻的

* $\nabla^2\left(e^{-\frac{x^2+y^2}{2\sigma^2}}\right)=\frac{\partial^2}{\partial^2 x}\left(e^{-\frac{x^2+y^2}{2\sigma^2}}\right)+\frac{\partial^2}{\partial^2 y}\left(e^{-\frac{x^2+y^2}{2\sigma^2}}\right)$
$=\frac{x^2+y^2-2\sigma^2}{\sigma^4}e^{-\frac{x^2+y^2}{2\sigma^2}}$

前后帧上的对应像素来对它进行滤波.

表2 使用 GS 和 IGS 处理后的图像序列的编码帧率(kbps)

	美式足球	工头	英式足球
GS	1238.68	630.63	645.11
IGS	1198.41	597.38	624.54

在起到扩散作用的各种滤波器中,本文选择迭代式的双边滤波器,因为双边滤波具有平滑图像并同时保留边界的特点. 而且,双边滤波可以用可分离核[3]来快速逼近. Tomasi[31]进一步指出,迭代地执行双边滤波可以得到卡通化的效果. 双边滤波的一个经典的计算核是高斯算子,如文献[2]中使用的. 但是,我们在经过高斯空间双边滤波器(Gaussian Spatial, GS)处理过的图像中发现,在大块的平滑区域里会有一些小斑点,会使观众很不舒服. 为了解决这个问题,我们修改成一个三维的反高斯双边滤波器(Inverted Gaussian Spatial, IGS),其细节将在 3.2.1 小节中介绍. 我们进一步发现,IGS 滤波器处理后的视频相比 GS 双边滤波处理的视频而言具有更高的可压缩性,如表 2 所示.

3.2.1 三维反高斯双边滤波

三维双边滤波器 $H(\cdot)$ 的一般形式可以定义如下:

$$H(p,\Omega)=\frac{\sum_{q\in\Omega} f\left(\|p-q\|\right) g\left(\|I(p)-I(q)\|\right) I(q)}{\sum_{q\in\Omega} f\left(\|p-q\|\right) g\left(\|I(p)-I(q)\|\right)} \tag{3}$$

在式(3)中,p 是输入图像 $I(\cdot)$ 上一个像素的位置,Ω 是 p 在时空域上相邻像素的集合,$f(\cdot)$ 称为空间滤波函数(SFF),计算基于像素位置之间的距离的权值,而 $g(\cdot)$ 称为亮度滤波函数(IFF),计算基于亮度差的权值. 经典的 $f(\cdot)$ 和 $g(\cdot)$ 都是高斯函数. Tomasi 等人[31]对双边滤波进行深入分析并指出,IFF 负责保留强边而 SFF 在像素亮度平整分布的区域里起主要作用.

当使用传统的高斯滤波器过滤大平滑区域上的一个小斑点时,斑点周围的像素的 IFF 权值($g(\cdot)$)很小,尽管其 SFF($f(\cdot)$)权值较大. 因此这些附近的像素所起到的整体作用($g(\cdot)f(\cdot)$)并不大,就跟外围那些 SFF 权值很小的像素一样. 于是,中间的斑点通常是没办法去除的. 为了对付这些斑点,我们把 SFF 改成反高斯的,以保证外围那些离斑点比较远的像素可以得到较大的权重. 这个含有高斯 IFF 和反高斯 SFF 的新双边滤波器由式(4)定义,其原理如图 5 所示.

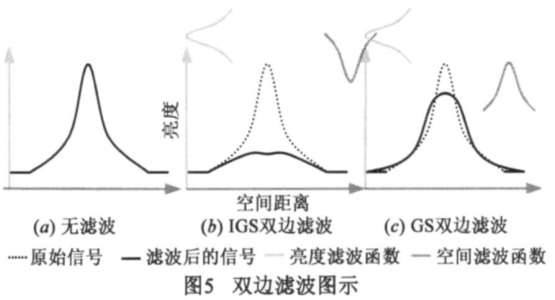


图5 双边滤波图示

$$H(p,\sigma_r,\sigma_d)=\frac{\sum_{q\in\Omega}\left(2-\mathrm{e}^{-\frac{\|p-q\|^2}{2\sigma_d^2}}\right)\mathrm{e}^{-\frac{\|I(p)-I(q)\|^2}{2\sigma_r^2}} I(q)}{\sum_{q\in\Omega}\left(2-\mathrm{e}^{-\frac{\|p-q\|^2}{2\sigma_d^2}}\right)\mathrm{e}^{-\frac{\|I(p)-I(q)\|^2}{2\sigma_r^2}}} \tag{4}$$

在式(4)中,$\left(2-\mathrm{e}^{-\frac{\|p-q\|^2}{2\sigma_d^2}}\right)$ 就是反高斯 SFF,σ_d 为它的滤波尺度. 通过提高 σ_d 的值可以得到更平滑的结果. 但如果 σ_d 太大时会出现一些荒谬的效果[31]. σ_r 是 IFF 即 $\mathrm{e}^{-\frac{\|I(p)-I(q)\|^2}{2\sigma_r^2}}$ 的滤波尺度. 在本文中,$\sigma_d=2.5$,$\sigma_r=4.5$. 式(4)所定义的滤波过程是在 CIE-Lab 颜色空间上的,沿 X 方向,Y 方向和时间轴依次执行. 沿着时间轴在相邻帧之间的滤波过程将在 3.3 小节中讨论.

在图 6 中,我们用美式足球序列比较了 IGS 双边滤波和 GS 双边滤波的结果. 可以看到,IGS(*b*)能够比 GS(*a*)滤去更多的琐碎细节. 为了使比较更为明显,我们将边缘提取过程中的高斯滤波替换为 IGS 或 GS 双边滤波. 同样,IGS(*d*)所产生的边缘比 GS(*c*)的更加干净.

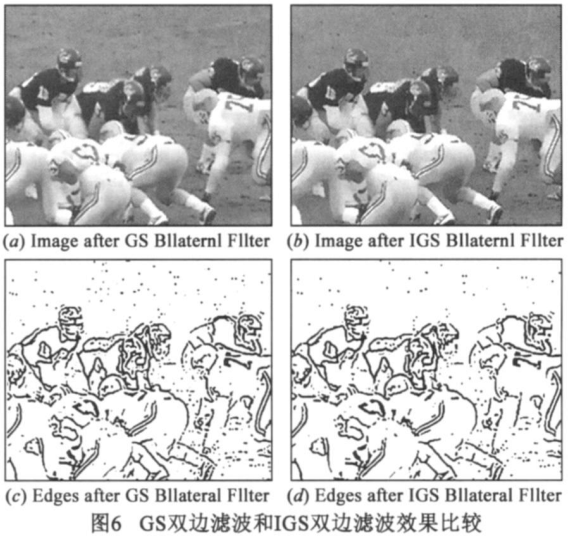


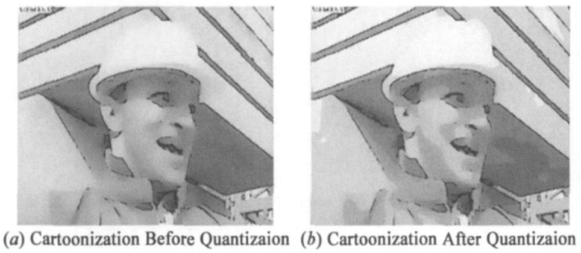
图6 GS双边滤波和IGS双边滤波效果比较

除了使用户的感觉更加舒服以外,本文的 IGS 双边滤波在视频压缩上也有它自己的优点. 如表 2 所示,将 IGS 双边滤波后的视频编码为 MPEG-4 所得到的码率很明显地比 GS 双边滤波过的要低得多. 究其原因,IGS 滤波器更趋向于创造大块的相似颜色的区域,使得编码冗余更大一些.

3.2.2 颜色量化

为了使观众的印象更加深刻,本文还使用了一个风格化算法来夸大被 IGS 滤波过的视频的特征. 这里我们借用了文献[2]的颜色量化算法. 量化方法最主要的特点是在梯度高的区域产生卡通化的强边而在梯度低的区域软化边缘,如图 7 所示. Winnemoeller 等人声称他

们的软伪量化法可以比标准量化法具有更好的时间连续性.可惜的是,在他们的结果中,抖动效应仍然相当明显.相比而言,本文的时空域计算框架可以很好的控制这些时域噪声.



(*a*) Cartoonization Before Quantizaion (*b*) Cartoonization After Quantizaion

图7　颜色量化前后的效果比较

3.3　时域滤波

我们在边缘提取和像素扩散两个过程中都要进行时域滤波.时域滤波包括三个主要步骤.第一步,用一个缓冲区来存储同一个场景中个连续帧,因此当场景变化时则需要重置缓冲区.第二步,使用最新的计算机视觉技术来计算相邻帧之间的像素对应关系.对应像素指的是在不同的帧中代表真实世界同一个空间点的像素(见图 8).因此,一个像素在时间轴上的相邻像素指的就是其在相邻帧上的对应像素.第三步,我们通过 GPU 上的一遍绘制来实现一维高斯滤波或者一维双边滤波,取决于时域滤波是用在边缘提取中还是用在像素扩散中.以上三步处理过程的细节将讨论如下.

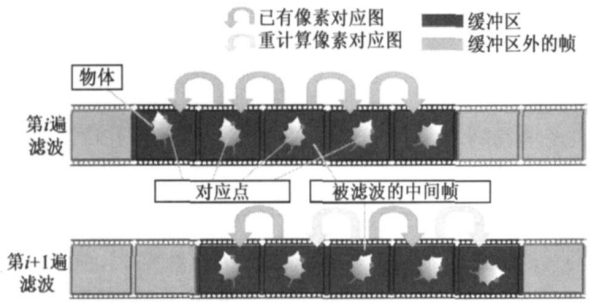


图8　时域滤波过程示意图.通过计算光流场来得到相邻帧之间的像素对应关系.从第i遍到第i+1遍时,缓冲区向前移动一帧,只需要重新计算两个新的对应图.

在第一步中,设置 $\tau = 2r + 1$,其中 r 是滤波核的半径.这样,就可以使整个缓冲区的视频帧刚刚好用来滤波中间的那一帧.在 GPU 实现中,由于硬件限制,r 不能过大,因为在一遍绘制中至少需要 $4r$ 张纹理(见第三步的讨论).本文取 $r = 2$.在第二步中,我们从两帧之间的光流场解出像素的对应关系,而光流场采用 Lucas-Kanade 方法[32]计算.不过,由于光流所估计的运动存在遮挡关系,某些像素可能失去对应像素.在这种情况下,我们标记该像素的对应关系无解,并在滤波过程中跳过它.在第三步中,为了在 GPU 上对中间帧进行时域滤波,我们把它前面和后面各 r 帧,还有 $2r$ 个相邻帧对之间的像素对应图都传给像素处理器.这样,可以并行地滤波中间帧上的所有像素,而每一个像素的输出值都由它在时间轴上前面和后面各 r 个邻居像素来决定.需要指出的是如果存在无解的对应关系,那么一个中间像素的对应像素数将少于 $2r$.

当前中间帧的滤波过程完成之后,缓冲区将在输入的图像序列中往前移动一帧.然后又重新执行上述三个处理步骤来滤波下一个中间帧(即上个中间帧的下一帧),一直到出现场景更替或整个序列处理完毕为止.虽然在算法中,计算像素的对应关系是最为耗时的,但是我们通过重用前面已计算好的对应图来减少这类计算的次数.由图 8 描述的过程可知,实际上当缓冲区前移一帧时,只需要重新计算两个新的像素对应图.通过这一系列时域滤波步骤之后,我们有效地控制了帧间的抖动,效果如图 9 以及演示视频所示.

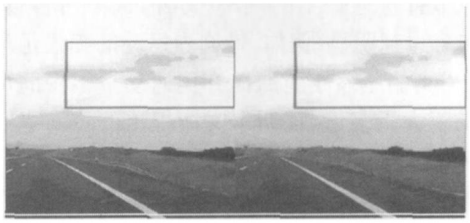


图9　执行时域滤波前(左)后(右)的提炼结果的比较

4　实验结果

与诸多已发表的非真实感绘制的工作不同的是,本文工作旨在提高整个视觉通信过程的效率,包括传输和感知两个部分.我们用不同的实验来分别测试新算法为这两个过程所带来的性能的提升.所有的测试都在 Toontalk 上进行.Toontalk 服务器带有主频为 2.8GHz 的 Intel Xeon 双核的 CPU 和显存为 2GB 的 ATI1950 显卡.表 3 给出了四个测试的视频剪辑,分辨率均为 710 ×576.在本节里,"提炼的"和"原始的"分别指已被和未被本文的提炼算法处理过的视频.图 13 以及所附的演示视频给出了提炼算法的最终处理结果.

表 3　实验中所有测试视频剪辑的信息

剪辑名称(简称)	特点描述	总帧数
工头(fm)	一般运动量,摄像机摇动	300
美式足球(fb)	高运动量,复杂内容	240
英式足球(sc)	一般运动量	300
母女(md)	低运动量	300

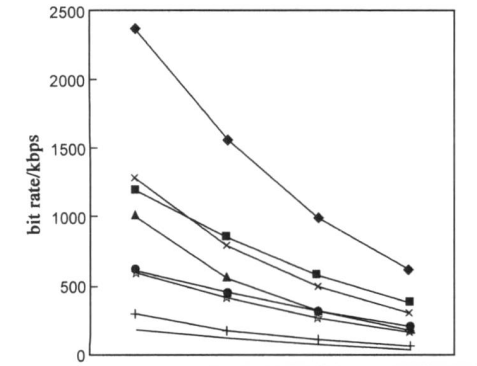
4.1　传输效率

本文借用视频编码研究中的通用工具来分析提炼后的视频在多大程度上可以节省压缩数据的传输带宽.我们引进量化参数(QP)和码率的概念来评估提炼前和提炼后的视频的压缩率.传输效率是通过 MPEG4 AVC 编码的压缩性能来测量的.编码过程的配置信息在表 4 中给出.

表 4 MPEG4 高级视频编码参数配置

参数名称	参数值
视频格式数	YUV 4:2:0,8 位
特征数据/等级	60(主要特征)/40
参考帧数	3
亚像素搜索时是否 Hadamard 变换	是
搜索范围	32 个像素
是否进行亚像素运动估计	是
是否进行反向搜索	否
是否进行 8×4,4×8,4×4 内部搜索	是
是否进行加权预测	否
画面组结构	IPPP

在图 10 中,我们在四个不同的 QP 值下比较了提炼视频和原始视频在编码后的码率. 显然,对于同一个 QP 值,提炼过的视频需要的码率要低得多. 而这种码率的节省在低 QP 时更加明显. 当 QP 为 24 时,提炼后的视频可以节省将近一半的码率. 另外,视频内容越复杂,这种节省也越明显. 例如,带剧烈运动的美式足球序列(fb)所得到的压缩性能的提升比运动量很少的母女序列(md)要大得多. 这证明本文所提出的技术更适用于处理复杂视频信号.



	QP=24	QP=28	Qp=32	Qp=36
fb_org	2373.48	1558.92	995.43	613.62
fb_ct	1198.41	856.52	584.6	388.05
fm_org	1005.95	561.8	326.64	193.99
fm_ct	597.38	419.47	275.78	163.4
sc_org	1286.62	795.17	494.61	312.93
sc_ct	624.54	452.74	312.25	207.61
md_org	299.6	176.94	111.13	69.21
md_ct	186.58	124.38	74.28	42.99

图10 原始视频(org)和用本文的提炼算法转化后的视频(ct)的压缩码率比较

有一些通信系统要求保持码率恒定,以使用稳定的数据流来传输视频. 这一类应用也能得益于本文的技术. 对于恒定的码率,降低 QP 可以得到更高的主观质量[33]. 另外,当人们观察视频时,往往希望质量是稳定的,而不是时好时坏. 所以我们最好把 QP 的变化限制在一个小范围之内. 我们在编码系统中修改 ρ 域模型[34]来进行恒定码率的实验,且实验中待编码的原始视频和提炼视频的帧数是一样的. 图 11 的实验结果表明,提炼视频所对应的 QP 的值以及 QP 的变化范围都比原始视频要小. 换句话说,在传输带宽恒定的情况下,本技术可以为观众提供更加舒服和更为稳定的视觉体验.

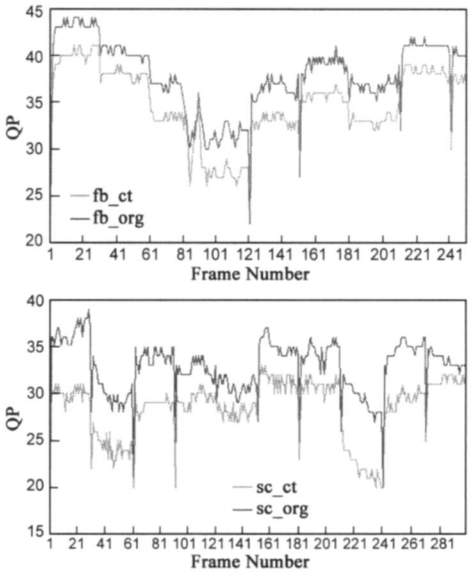


图11 "美式足球"和"英式足球"两个剪辑的原始视频(org)与提炼视频(ct)在恒定码率下压缩所对应的QP值比较

4.2 感知效率

与文献[2,22]相似,本文采纳了心理学研究中常用的用户学习法来评估图像的感知效率. 但是,与他们那种对普通用户进行间接实验的方法不一样的是,我们从同事中邀请了十五位具有视频处理经验的专家来直接评估视频. 专家在 Toontalk 的终端上观看测试视频并进行打分,这样就有效地避免了间接识别实验中由于普通用户的反应不稳定而带来的误差. 另外,在 Toontalk 系统的终端上进行感知实验意味着,所有实验视频都是被压缩过,被传输过以及被解压过的. 因此,相比以前那些在本地播放的未压缩视频上进行的实验,我们得到的数据更贴近于实际应用中的真实性能.

专家们根据视频的易理解性来对原始视频和提炼视频分别进行打分. 我们使用的是视频编码研究中广为引用的欧洲广播联盟发布的多媒体视频质量主观评估方法[35]. 对于表 3 中的一个剪辑,需要测试两组序列,一组是原始的而另一组是提炼过的. 每一组包含一个未压缩序列,以及四个分别以 QP 为 24、28、32 和 36 进行编码的序列. 这样,对于所有的四个剪辑,专家们总共需要对八组共 40 个测试序列进行打分. 我们将每一组中的无压缩序列当作参考数据.

专家们按组观察这些序列,根据捕捉到视频大概内容的难易度来给每个序列打分. 分数的范围是 0(最差)到 100(最好)之间. 参与者并不知道他们正在观看的视频所对应的 QP 值,但他们可以随时重新观察和重新评价同一组内的任意序列. 将每个专家所给的分数进行平均可以得到平均评价分数(Mean Opinion Score,MOS).

$$MOS_k = \sum_{i=1}^{N_{exp}} \frac{Mark_{i,k}}{N_{exp}} \quad (5)$$

在式(5)中,k 是图像序列的编号,N_{exp}是专家的人数,即 15. $Mark_{i,k}$是第 i 个专家对第 k 个图像序列的打分. 图 12 给出了四个测试剪辑关于不同 QP 值的 MOS 曲线. 可以看出,只要 QP 值足够大,也即视频被足够严重地压缩,提炼后的视频具有更高的可感知性. 换句话说,在压缩过程中,视频提炼技术能更好的保留甚至突出与人眼感知过程最相关的特征. 另一方面,如果视频只是被轻微地压缩(低 QP 值)甚至未被压缩过,视频提炼技术就没有太大的优势了. 然而,在真实的应用场合中极少用到未压缩或低压缩的视频,因为对它们进行传输和存储的负担太重. 所以,对于实际通信系统所经常使用到的视频来说,本文方法能够保持较高的可理解性. 这分别归因于我们所采用的边界保留策略和时域抖动去除方案.

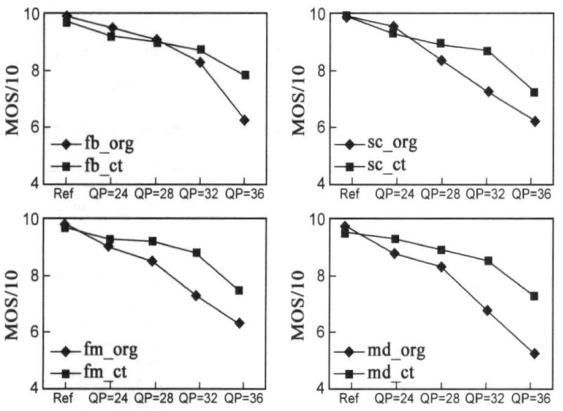


图12　专家对原始视频(*org*)和提炼视频(*ct*)进行打分所得的MOS比较

5　结论和未来工作

本文提出了一个新颖的视频数据的表达方式,以提高人与人之间的视觉通信效率. 这种表达方式突出图像序列中有意义的特征(定义为物体的边界)并去除无关紧要的细节. 本文通过两个主要的处理步骤来创建这种新的表达方式,一是旨在保留边界的边缘提取过程,二是旨在减弱不重要的信号的非线性扩散过程. 这两个步骤都是针对时空域的因而有效控制了抖动噪声. 我们把新算法整合到一个完整的、实用的视觉通信系统中. 在该系统上的实验结果表明,这种新的表达方式不仅节省了视频传输的带宽,还增加了被有损压缩过的视频的可理解性.

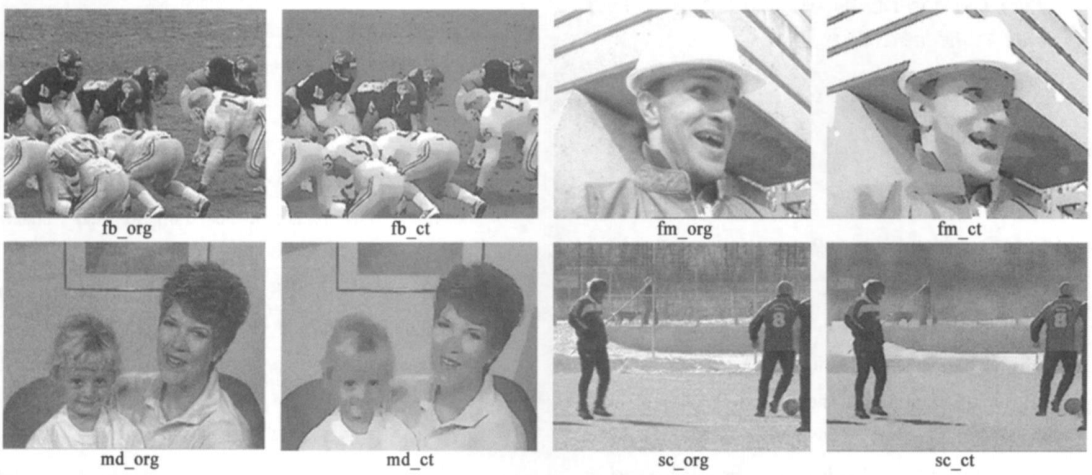


图13　四个测试剪辑的最终提炼结果. “org”和”ct”分别代表原始视频和提炼视频

仔细观察图 13 中的提炼结果,我们找到一些需要改进的地方. 首先,提炼后的边缘不如图 1 所示的人工创作的电影一样美观. 在下一步工作中,我们可以引入基于笔划的绘制技术来美化边界. 其次,被大片相似颜色所包围的区域上有一些重要的细节被减弱了. 为了保护这些区域,可以使用图像分割方法来优化计算. 另外,在视频编码实验中,我们发现边缘会带来额外的压缩负担. 如果强边太多,编码时间会急剧增加. 解决该问题的一个有用的尝试是将边缘图单独进行编码,这就需要借助多通道编码技术.

致谢　感谢参与感知效率评估实验的汤姆逊北京研究院的同事. 感谢张辉博士、朱广飞、杨继珩在本文工作上的帮助和讨论,以及何丹媛在 Demo 制作上的协助. 感谢 Winnemoeller, H. 提供的视频提炼的 Cg 源代码.

附录 A　基于 LoG 的可分离 M-H 边缘检测算法 GPU 实现

为了在 X 方向和 Y 方向上分离式地实现 LoG 算法,我们对每个方向都用两遍绘制来进行 M-H 边缘检测,分别以 $\frac{\Delta s^2 - \sigma^2}{\sigma^4}$ 和 $e^{-\frac{(\Delta s)^2}{2\sigma^2}}$ 作为乘法的系数. 其中 Δs 为 Δx 或 Δy. Cg 代码如下:

```
LoGPass1_PS(infloat2tex ,outfloat2color ,uniformsamplerRECTdatasource) {
    float2texel ;
    for(int i = - radius ; i <radius ; i ++) {
      texel. x = f1texRECT(data_Source ,tex + float2( i ,0)) ;
      texel. y = f1texRECT(data_Source ,tex + float2(0 , i)) ;
      float coef_one = ( i * i - sigma2) / (sigma2 *sigma2) ;
      coef_one * = exp( - i * i *0. 5/ sigma2) ;
      color + = texel *coef_one ;
    }
}
LoGPass2_PS(infloat2texel ,outfloat2color ,uniformsamplerRECTsampler) {
    float2texel ,sum = 0 ;
    for(int i = - radius ; i <radius ; i ++) {
      texel. x = f1texRECT(tex , uv + float2(0 , i)) . x ;
      texel. y = f1texRECT(tex , uv + float2( i ,0)) . y ;
      float coef_two = exp( - i * i *0. 5/ sigma2) ;
      sum + = texel *coef_two ;
    }
    color = (sum. x + sum. y) / (2 *pi *sigma2) ;
}
```

作者简介:

黄沛杰 男,1981 年生,中国科学院软件研究所博士研究生,香港中文大学研究助理. 研究方向为全局光照渲染技术、实时绘制、计算机非真实感绘制、及图像艺术化等.

E-mail :hpj @ios. ac. cn

朱立华 男,1977 年生,汤姆逊北京研究院研究工程师. 研究方向为视频编码技术与计算机非真实感绘制.

参考文献:

[1] Sayeed R, Howard T. State-of-the-art of non-photorealistic rendering (npr) for visualization [J]. Theory and Practice of Computer Graphics 2006.

[2] Winnemoeller H, Olsen S C, Gooch B. Real-time video abstraction [J]. ACM Trans Graph, 2006, 25(3) :1221 - 1226.

[3] Pham T Q, Van Vliet L J. Separable bilateral filtering for fast video preprocessing [A]. IEEE International Conference on Multimedia and Expo [C]. Los Alamitos, CA, USA: IEEE Computer Society, 2005. 454 - 457.

[4] Collomosse J P, Rowntree D, Hall P M. Stroke surfaces: temporally coherent artistic animations from video [J]. IEEE Trans. Vis. Comput. Graph, 2005, 11(5) :540 - 549.

[5] Anjyo K, Hiramitsu K. Stylized highlights for cartoon rendering and animation [J]. IEEE Computer Graphics and Applications, 2003, 23(4) :54 - 61.

[6] Saito T, Takahashi T. Comprehensible rendering of 3d shapes [J]. ACM SIGGRAPH Computer Graphics, 1990, 24(4) :197 - 206.

[7] Raskar R, Tan K H, Feris R, Yu J, Turk M. Nonphotorealistic camera: depth edge detection and stylized rendering using multi-flash imaging [J]. ACM Trans Graph, 2004, 23(3) :679 - 688.

[8] DeCarlo D, Santella A. Stylization and abstraction of photographs [J]. ACM Trans. Graph, 2002, 21(3) :769 - 776.

[9] Litwinowicz P. Processing images and video for an impressionist effect [A]. Owen G S. Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques [C]. New York: ACM, 1997. 407 - 414.

[10] Bousseau A, Neyret F, Thollot J, Salesin D. Video watercolorization using bidirectional texture advection [J]. ACM Transactions on Graphics, 2007, 26(3) :104.

[11] Curtis C J, Anderson S E, Seims J E, Fleischer K W, Salesin D. Computer-generated watercolor [A]. Owen G S. Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques [C]. New York: ACM Press/ Addison-Wesley Publishing Co, 1997. 421 - 430.

[12] Wang C M, Wang R J. Image-based color ink diffusion rendering [J]. IEEE Trans Vis Comput. Graph. 2007, 13(2) :235 - 246

[13] Hertzmann A. Painterly rendering with curved brush strokes of multiple sizes [A]. Cunningham Steve. Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques [C]. New York: ACM, 1998. 453 - 460.

[14] Hays J, Essa I A. Image and video based painterly animation [A]. Spencer S N. Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering [C]. New York: ACM, 2004. 113 - 120.

[15] Hertzmann A, Perlin K. Painterly rendering for video and interaction [A]. Fekete J D. Proceedings of the 1st International Symposium on Non-photorealistic Animation and Rendering [C]. New York: ACM, 2000. 7 - 12.

[16] Cohen M F, Colburn A, Finkelstein A, Klein A W, Sloan P J. Video cubism [R]. Redmond, Microsoft Research, 2001, MSR-TR-2001-45.

[17] Wang J, Xu Y, Shum H Y, Cohen M F. Video toning [J]. ACM Trans. Graph. 2004, 23(3) :574 - 583.

[18] Agarwala A. Snaketoonz: a semi-automatic approach to creating cel animation from video [A]. Finkelstein A. Proceedings of the 2nd International Symposium on Non-photorealistic Animation and Rendering [C]. New York: ACM, 2002. 139 - ff.

[19] Schumann J, Strothotte T, Laser S, Raab A. Assessing the effect of non-photorealistic rendered images in cad [A]. Nardi B. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground [C]. New York: ACM, 1996. 35 - 41.

(下转第 56 页)

ley & Sons ,1995.

[2] M Kyng. "Creating Contexts for Design" in Scenario Based Design[M]. New York :John Wiley & Sons ,1995.

[3] R O Briggs ,G-J De Vreede ,J F Nunamaker ,Jr. D Tobey. ThinkLets :Achieving predictable ,repeatable patterns of group interaction with group support systems (GSS) [A]. Proceedings of the 34th Annual Hawaii International Conference on System Sciences [C]. Hawaii :IEEE Computer Society ,2001. 436 - 444.

[4] Robert O Briggs ,Gert-Jan De Vreede ,J F Nunamaker. Collaboration engineering with thinkLets to pursue sustained success with group support systems [J]. Journal of Management Information Systems ,2003 ,19 (4) :31 - 64.

[5] J F Nunamaker ,Jr. R O Briggs ,D D Mittleman ,D R Vogel ,P A Balthazard. Lessons from a dozen years of group support systems research :A discussion of lab and field findings [J]. Journal of Management Information Systems ,1996 - 97 ,13 (3) :163 - 207.

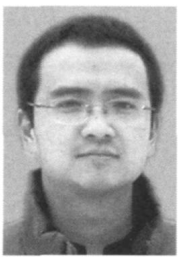
[6] D L Dean ,J D Lee ,R E Orwig ,D R Vogel. Technological support for group process modeling[J]. Journal of Management Information Systems ,1995 ,11 (3) :43 - 64.

[7] A M Hickey ,D L Dean ,J F Nunamaker Jr. Setting a foundation for collaborative scenario elicitation[A]. Proceedings of the 32th Annual Hawaii International Conference on System Sciences [C]. Hawaii :IEEE Computer Society ,1999 ,1 :1041 - 1051.

作者简介:

刘 锋 男,1977 年生于河南沁阳,2000 年在中山大学计算机科学系获得学士学位,现为北京大学计算机科学系硕士研究生,主要研究方向为:领域工程和需求工程.

Email :liufeng06 @sei. pku. edu. cn



张 伟 男,1978 年生于江苏徐州,2006 年在北京大学获得博士学位,现为北京大学信息科学技术学院讲师,主要研究方向为:领域工程、需求工程、软件复用与软件构件技术.

Email :zhangw @sei. pku. edu. cn

(上接第 50 页)

[20] Agrawala M, Stolte C. Rendering effective route maps : improving usability through generalization [A]. Pocock L. Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques [C]. New York : ACM, 2001. 241 - 249.

[21] Gooch B, Coombe G, Shirley P. Artistic vision : painterly rendering using computer vision techniques [A]. Finkelstein A. Proceedings of the 2nd International Symposium on Non-photorealistic Animation and Rendering [C]. New York : ACM, 2002. 83 - ff.

[22] Gooch B, Reinhard E, Gooch A. Human facial illustrations : creation and psychophysical evaluation [J]. ACM Trans Graph, 2004, 23 (1) :27 - 44.

[23] PALMER S. Vision Science : Photons to Phenomenology [M]. Cambridge, MA, USA : MIT Press, 1999.

[24] Tufte E. Envisioning Information [M]. Cheshire, CT, USA : Graphics Press, 1990.

[25] Orzan A, Bousseau A, Barla P, Thollot J. Structurepreserving manipulation of photographs [A]. Gooch B . Proceedings of the 5th International Symposium on Non-photorealistic Animation and Rendering [C]. New York : ACM, 2007. 103 - 110.

[26] Fischler M, Firschein O. Intelligence : The Eye, The Brain and the Computer [M]. Boston, MA, USA : Addison-Wesley, 1987. 331.

[27] Canny J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8 (6) :679 - 698.

[28] Marr D, Hildreth E. Theory of edge detection [A]. Proceedings of the Royal Society of London. Series B, Biological Sciences [C]. The Royal Society, 1980. B - 207, 187 - 217.

[29] Breen E J, Jones R, Talbot H. Mathematical morphology : A useful set of tools for image analysis [J]. Statistics and Computing, 2000, 10 (2) :105 - 120.

[30] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion [J]. IEEE Trans. Pattern Anal. Mach. Intell, 1990, 12 (7) :629 - 639.

[31] Tomasi C, Manduchi R. Bilateral filtering for gray and color images [A]. Proceedings of the Sixth International Conference on Computer Vision [C]. Washington, DC, USA : IEEE Computer Society, 1998. 839 - 846.

[32] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [A]. International Joint Conference on Artificial Intelligence [C]. Morgan Kaufmann, 1981. 674 - 679.

[33] Westerink P H, Rajagopalan R, Gonzales C A. Twopass mpeg-2 variable-bit-rate encoding [J]. IBM Journal of Research and Development, 1999, 43 (4) :471.

[34] He Z, Mitra S K. Optimum bit allocation and accurate rate control for video coding via-domain source modeling [J]. IEEE Trans Circuits Syst Video Techn, 2002, 12 (10) :840 - ff.

[35] Union E B. Samviq-a new ebu methodology for video quality evaluations in multimedia [R]. INIST-CNRS : European Broadcasting Union, BPN 056, 2003.