

2002年上海大学博士学位论文 35

基于VO分割的视频对象 平面形状编码研究

作者：石旭利
专业：通信与信息系统
导师：张兆扬



上海大学出版社

· 上海 ·

N533
Y302
2002

2002 年上海大学博士学位论文

基于 VO 分割的视频对象 平面形状编码研究

作 者：石旭利

专 业：通信与信息系统

导 师：张兆扬

上海大学出版社

• 上海 •

533
1302
2002

Shanghai University Doctoral Dissertation (2002)

Study of Shape Coding Algorithm of VOP Based on VO Segmentation

Candidate: Shi Xu-li

Major: Communication and Information System

Supervisor: Prof. Zhang Zhao-yang

Shanghai University Press

• Shanghai •

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学博士学位论文质量要求。

答辩委员会名单：

- | | | |
|---------------|--------------------------|--------|
| 主任：余松煜 | 教授，上海交大电子信息工程学院 | 200030 |
| 委员：张一钧 | 教授，上海广电集团 | 200030 |
| 戚飞虎 | 教授，上海交大计算机系 | 200030 |
| 王朔中 | 教授，上海大学电子信息工程系 | 200072 |
| 杨根庆 | 研究员，中科院上海微系统与
信息技术研究所 | 200031 |
| 导师：张兆扬 | 教授，上海大学 | 200072 |

评阅人名单:

戚飞虎	教授, 上海交大计算机系	200030
莫玉龙	教授, 上海大学通信工程系	200072
孙 军	教授, 上海交大电子信息工程学院	200030

评议人名单:

朱维乐	教授, 电子科技大学	610054
余松煜	教授, 上海交大电子信息工程学院	200030
杨根庆	研究员, 中科院上海微系统与 信息技术研究所	200031
薛向阳	教授, 复旦大学计算机系	200080

答辩委员会对论文的评语

论文《基于 VO 分割的视频对象平面形状编码研究》是国家自然科学基金 (60172020) 和上海市自然科学基金 (992D14035) 的资助项目中的重要内容之一。课题围绕基于对象的视音频压缩标准 MPEG-4 中的核心内容——形状编码算法和分层编码算法展开深入的研究, 是对 MPEG-4 标准的改进和完善。对于国民经济和社会发展具有现实意义。其实用化能对多媒体在互联网中的应用起到一定促进作用。

论文从多尺度形状描述的观点出发, 对基于对象的编码算法进行了分析和总结, 提出了一种新的形状编码算法。重要贡献如下:

用跟踪算法对 CSS 算法进行了改进, 使之能用于 MPEG-4 的形状编码和便于特征点的提取。

在上述基础上进一步提出了自适应算术编码算法。其中帧内用改进的 CSS 法, 帧间用 CSSI (曲率尺度空间图) 的曲线运动估计算法和基于弧长索引的编码。此种方法比 MPEG-4 中的 CAE 法在率失真性能上有显著提高。

在上述所提出的新的形状编码算法的基础上, 提出了对形状曲线的可分层描述和分层编码算法。这种方法特别适合于基于对象的分层压缩编码, 且便于实现基于互联网的流视频。

论文选题紧扣国民经济和社会发展的实际, 在坚实理论的基础上进行了深入的研究, 提出的算法新颖, 非常具有创造性, 行文流畅, 条理清晰。在答辩过程中, 回答问题正确。

答辩委员会表决结果

经答辩委员会表决,全票同意通过石旭利同学的博士学位论文答辩,建议授予工学博士学位。

答辩委员会主席: **余松煜**

2002年6月26日

摘 要

传统的编码标准, 如 H.261、H.263、MPEG-1、MPEG-2, 虽然可以提供较高的压缩比而且正被广泛应用, 但却不能满足多媒体用户对视音频内容快速、灵活和交互访问的要求.

MPEG-4 标准旨在提供一种在多媒体环境中有效存储、传输和操作视频数据的标准化的核心技术. 这个目标的实现依赖于对视频数据进行基于内容的描述. MPEG-4 从这种基于对象的描述方式出发, 将与该对象有关的纹理信息、运动信息和形状信息分别编码形成该对象的一个比特流. 解码端对几个对象所对应的各自的 比特流分别进行解码, 并可按照码流中所包含的合成信息合成原有的场景, 也可以对其中的某几个对象进行存取和操作. 即这种基于内容的编解码方式是各种多媒体应用中实现交互性的关键. MPEG-4 标准的输入序列的每一帧被分割成任意形状的视频对象平面 (VOP). 编码/解码是针对 VOP 序列组成的视频对象 (VO) 的. 这使得交互存取成为可能. 为实现基于对象的编码功能, MPEG-4 包含了众多的编码算法: 形状编码算法、运动估计和补偿算法、纹理编码算法、静态纹理编码算法、错误恢复、精灵编码和可分层编码等等. 其中, 形状编码算法是其最核心的内容之一. MPEG-4 成员通过大量的核心实验, 对所提交的基于上下文的算术编码算法, MMR 法、基线法、多边形逼近法以及色键法进行了综合评价, 最中将基于上下文的算术编码算法选入校验模型中.

这种编码算法向解码器传递视频序列中每一个视频对象的形状及其在图象中的空间位置. 这显然会使编码的比特数增大. 因此, 对于运动补偿和残差图象编码以及形状编码本身提出了非常高的要求. 形状编码算法的简洁性、准确性、高效性和易于硬件实现性作为评价其优劣的综合标准.

本文从多尺度形状描述的观点出发, 对基于对象的编码算法进行了分析和总结, 并将视野从现存的形状编码算法扩展到已经存在的众多形状描述的领域中, 从中发现了 CSS 多尺度形状描述的许多优点. 对此种形状描述法进行了更为深入的探讨和改进, 以此为基础, 提出了一种新的形状编码算法. 重要贡献如下:

用用于形状检索的 CSS 算法进行曲线平滑时必然造成曲线总长度的缩减, 这使得曲线上点的位置偏离其在原始曲线上的位置, 从而因位置信息无效而不能用于形状编码, 本文中我们用跟踪算法对 CSS 进行改进来获得位置信息使之能用于形状编码.

由于 CSS 将局部曲率最大点作为特征点, 必然使形状曲线中由圆弧组成的部分所提取的特征点数为零, 我们用棋盘距离对 CSS 进行改进并提取逼近点.

我们采用自适应算术编码算法对[4]中基于八分圆的固定长度的顶点编码方法进行改进, 并用改进的八分圆算法对所提取出来的特征点进行熵编码从而进一步提高压缩效率.

在帧间模式中, 提出了一种新的基于曲率尺度空间图 (CSSI) 的曲线运动估计算法; 采用了一种新的基于弧长索引的编码方法对匹配部分进行编码而用帧内模式中的 CSS 方法对不匹配部分进行编码, 从而提高了编码的效率.

由于基于对象的分层编码算法在流视频网上传输中的重要性. 本文在所提出的新的形状编码算法的基础上, 对形状曲线的分层描述进行了深入的研究, 用数种规则得出除了基层 (第 0 层) 以外, 所有高层象素可能存在的位置, 从而实现了基于 CSS 的分层编码. 这种方法特别适合于头肩序列中形状序列的分层压缩编码.

由于上述算法中 CSS 算法所造成的原始形状曲线的缩减问题增加了计算的复杂度, 故本文在第六章研究了基于周期小波变换的形状曲线的分层表示问题, 以作为今后进一步工作的基础. 由于小波变换在多尺度图象表示和图象压缩编码上的成功运用, 本文今后的工作将侧重于研究一种基于小波变换的任意形状运动对象的分层编码算法, 以满足于流视频网上传输的要求.

本课题是上海市自然科学基金 (992D14035) 和国家自然科学基金 (60172020) 的资助项目.

关键词 MPEG-4, 形状编码, 曲率尺度空间, 曲率尺度空间图, 分层编码

Abstract

Traditional video standards such as H.261, H.263, MPEG-1, and MPEG-2 can achieve high compression ratios and are suitable for a wide range of application. However, multimedia users demand fast, flexible and interactive access to audio-visual content.

MPEG-4 video aims at providing standardized core technologies allowing efficient storage, transmission and manipulation of video data in multimedia environments. To this end, the approach taken relies on a content based visual data representation. Each frame of the input sequence is segmented into arbitrarily shaped image regions (VOP's). the sequence of VOP is called video object (VO), which is the encoded/decoded content. Following the object-based approach, MPEG-4 visual transmits texture, motion and shape information of on VO within one bit stream. The bit streams of several VOs and accompanying composition information can be multiplexed such that the decoder receives all the information to decode the VOs and arrange them into a video scene which is either the original scene or a new one including only some VOs. It is believed that such a encoding/decoding approach is a key to enable interactivity with objects for a variety of multimedia applications.

To realize content-based encoding, MPEG-4 is a platform with a definition of encoding/decoding algorithm: Shape coding, Texture

encoding, Motion estimation and compensation, Error resilience, Sprite coding, Generalized scalable encoding and so on. Shape coding algorithm plays an important role on the MPEG-4 platform.

So far, MPEG-4 video group has put CAE into VM by carrying out plenty experiments on the CAE shape coding, MMR shape coding, Baseline shape coding, Vertex-based shape coding, Chroma-key shape coding and so on.

In this thesis, we analyse and synthesize the content-based coding algorithm in the view of shape description. We find many good features of CSS-based shape representation from many methods of shape description. So we deeply study this method, modify it and propose a new shape coding algorithm by using the modified CSS method.

The achievements of the dissertation include:

The smoothing operator makes the curve shrink when CSS is used in shape retrieval. It leads to the movement of each point of the curve. The movement results in the location information of point invalid. We modify the CSS by using tracing algorithm so as to obtain the location information of the point which is used in shape coding.

The local maximal curvature point is coded by shape coding. This smoothed shape representation will not very efficient when the shape consists of arcs. So we further modify CSS theory to extract feature point that lies in arc.

The feature points are encoded by adaptive arithmetic encoding algorithm and the modified octan-based vertex coding algorithm.

A new CSSI-based motion estimation algorithm is proposed in inter mode.

The arc-length-index-based algorithm is adopted to encode the matched contour segments and the mismatched contour segments are coded by using CSS-based algorithm which is used in intra mode.

The object-based scalability is very important to transmit video bit stream in internet. We research the multiscale representation of shape contour and obtain the probable locations of points in higher layers by using some rules. Then, we realize the CSS-based scalable shape coding algorithm which encode the base layer points using the new shape coding algorithm. The scalable shape coding algorithm is very efficient to scalable encode the head and shoulder sequences.

Shrinkage of contour smoothed by CSS algorithm strictly decrease the speed of encoding algorithm. In chapter 6, we study the periodized wavelets based shape description, which is the further study in the future. For the successful application of waveler transform in image representation and compression, from now on, we will focus our attention on the scalable coding of arbitrarily shaped motion object to meet the demand of transmitting video stream in internet.

This thesis is supported by the National Natural Science Foundation of Shanghai (No. 992D14035) and the National Science Foundation (No. 60172020).

Key words MPEG-4, shape coding, curvature scale space, curvature scale space image, scalable algorithm.

目 录

第一章 绪 论	1
1.1 课题研究的意义	1
1.2 论文的结构	6
第二章 视频对象编码算法	9
2.1 简 介	9
2.2 基本的数据压缩理论——熵编码	9
2.3 基于帧的图象和视频压缩理论	12
2.4 基于对象的压缩编码理论	20
2.5 图象编码标准	32
第三章 MPEG-4 的实现	35
3.1 MPEG-4 视频编码框架	35
3.2 静态纹理编码	57
3.3 实验结果	61
第四章 基于运动对象的形状编码新算法	63
4.1 引 言	63
4.2 任意形状曲线的表示理论	64
4.3 多尺度空间的曲线表示法	66
4.4 CSSAS 形状编码算法	78
4.5 实验结果和分析	103
第五章 可分层形状编码算法	113

5.1 基于流视频网上传输的基本原理	114
5.2 基于 SI 的形状分层编码算法	115
5.3 基于 CSS 的形状分层编码算法	119
5.4 实验结果	126
第六章 基于周期小波变换的形状描述	128
6.1 小波变换和多分辨分析	128
6.2 离散周期小波变换 DPWT	132
6.3 基于周期小波的形状描述	135
6.4 实验结果	135
6.5 结 论	137
第七章 结论和进一步的工作	138
参考文献	140
致 谢	149

第一章 绪 论

1.1 课题研究的意义

多媒体通过 Internet 网络使人们的生活内容更加丰富多彩。多媒体促进了远距离通信,消费电子和计算机工业的迅猛发展,但同时远距离通信、消费电子和计算机工业又对多媒体技术提出了越来越高的要求。各种与多媒体技术有关的应用领域需要一个统一的标准对其提供支持。在这些应用领域中,包括静止图象、立体图象、视频等数据类型;包括自然图象、计算机生成图象、文本/图形、医学图象等信源类型;包括点到点,点到多点等通信类型;包括对象操作、在线编辑、渐进传输、误差恢复等功能类型。如此之多的通信内容和通信方式通过 Internet 网达到完美的融合和交互是现有的编码标准所不能满足的。

传统的视频标准中, H.261^[1]是针对视频电话和视频会议应用的视频压缩标准,其比特率在 2 Mb/s 以上。它的传输率是 64 Kbit/s 的倍数,因此适用于 ISDN 线路的传输。H.263^[2]标准致力于低比特率通信,主要应用在视频会议领域。MPEG-1^[3]是运动图象专家组(MPEG)即 ISO/IEC JTC1/SC29 WG11,发布的第一个国际标准,最初应用于在 CD-ROM 上存储信息,采用混合的 DCT/DPCM 压缩方案。MPEG-2^[4]在更高的比特率上产生更高质量的视频,适用于基于电视的应用,如数字卫

星广播或 CATV.

因此,自 1993 年以来, MPEG 致力于发展给用户具有新的交互性的视觉内容的 MPEG-4 标准^[5],其应用范围从数字电视、比特流视频到移动多媒体和电子游戏. 编码比特率可以分为 64 kb/s 以下, 64~384 kb/s 和 384 kb/s~4 Mb/s 三个档次. 其最为突出的改变是给出了基于内容而不是基于像素的观察、存取和操作对象的技术, 在大的比特率范围内具有很强的误差坚韧性. 实现上述功能和比特率的编码算法包括形状编码, 运动估计和补偿, 纹理编码、误差恢复、精灵编码等.

其中, 形状编码算法是进行基于运动对象的编码算法中必须的. 由于每一个视频对象的形状及其空间位置的表述和编码的存在, 总体编码的数据量就会增大. 因此, 形状编码算法必须首先是高效的, 而后还需是易实现的.

在 MPEG-4 核心实验中^[6], 有五种形状编码算法:

- (1) 基于上下文的算术编码算法^[7,8];
- (2) MMR 方法^[9];
- (3) 顶点法^[10];
- (4) 基线法^[11,12];
- (5) 色键法^[13].

其中, 基于上下文的算术编码算法和 MMR 算法是基于位图的方法而顶点法、基线法和色键法是属于基于轮廓的方法.

在现存的形状编码算法中, 主要分为两大类, 如上所述的基于位图和基于轮廓的两类. 基于轮廓的形状编码算法有 Freeman 链码法^[14,15]、基线法、顶点法和多项式拟合法等四种. Freeman 方法是最早的基于轮廓的形状编码算法, 它首先定义空间每一个方向的代码, 通过对区域边界点沿一定顺序 (顺时