

# 一种基于视觉感知的非线性幅型比变换方法

胡彦婷 高新波 路文 王体胜

(西安电子科技大学工程学院, 西安 710071)

**摘要** 随着高清电视的出现,幅型比变换的应用范围越来越广泛。本文提出了一种基于人的视觉感知特性的非线性幅型比变换方法。该方法利用运动估计技术获得视频帧的运动矢量,采用基于修正划分模糊度的模糊聚类算法对运动矢量进行聚类并估算每一帧的运动重心(MCOG)。同时针对幅型比变换后由视频序列前后帧不连续引起的视觉上的抖动现象,采用基于运动重心和中心区域的非均匀区域划分法,并针对不同的区域进行非线性比率扩展。实验结果表明,该方法很好地反映了人的视觉对视频的感知情况,变换后的视觉效果明显优于传统的幅型比变换方法。

**关键词** 幅型比变换 模糊聚类算法 感兴趣区域 视觉感知特性

中图法分类号: TN919; TP391 文献标识码: A 文章编号: 1006-8961(2009)06-1082-06

## A Non-uniform Aspect Ratio Conversion Algorithm Based on Visual Perception

HU Yan-ting, GAO Xin-bo, LU Wen, WANG Ti-sheng

(School of Electronic Engineering, Xi'an University, Xi'an 710071)

**Abstract** With the emergence of HDTV, the application of aspect ratio conversion (ARC) is getting wider in related fields. In this paper, a new aspect ratio conversion algorithm is proposed based on human visual perception. To start with, the fuzzy cluster algorithm based on the modified partition fuzzy degree algorithm is applied to cluster motion vectors adaptively, and then the image motion center of gravity (MCOG) is estimated based on the clustering result. Moreover, to solve the dithering artifact, a splitting region approach based on MCOG and center region of image is developed. Finally, different regions of image are scaled with different ratios. Experimental results illustrate that the proposed method has good consistency with the human visual perception and show its superiority.

**Keywords** aspect ratio conversion, fuzzy cluster algorithm, region of interest, visual perception

## 1 引言

随着高清电视(HDTV)和多媒体系统的发展,人们不仅可以欣赏到丰富多彩的电视节目,还可以得到各种信息资讯。但是,HDTV与传统电视系统的视频信号格式不兼容,视频帧的宽高比不一致,当用HDTV接收传统电视节目或用传统电视接收

HDTV节目时均需要进行显示的格式转换<sup>[1]</sup>。由于幅型比为4:3的节目源与宽高比为16:9的HDTV还会共存很长一段时间,因此,对不同视频格式之间转换的研究具有重要的现实意义<sup>[2]</sup>。

目前幅型比变换的方式主要有两大类:一类是无图像失真,但损失图像部分内容或浪费屏幕部分资源的方式,如贴黑边,裁剪和移位<sup>[3-4]</sup>,贴黑边和裁剪组合;另一类是产生图像失真的方式,如线性幅型比变换<sup>[4-5]</sup>和非线性幅型比变换<sup>[2,4-5]</sup>。从保持原

基金项目:教育部长江学者和创新团队支持计划(IRT0645);新世纪优秀人才支持计划(NCET-04-0948);国家自然科学基金项目(60771068)

收稿日期:2007-06-22;改回日期:2007-11-13

第一作者简介:胡彦婷(1980~),女,西安电子科技大学信号与信息处理专业硕士研究生。主要研究方向为图像、视频处理与通信。

E-mail: hyt229@163.com

视频的完整性、图像画面的利用率、图像是否变形以及画面的和谐度等几个方面综合考虑,以上方法均有一定的缺陷,因此需要一种能够避免上述缺陷的方法。图像的视觉质量主要取决于视觉感兴趣区域(ROI)的质量,而人的视觉对视频各帧的运动区域和中心区域最为敏感。

鉴于此,根据人的视觉对图像序列不同运动区域的感知程度以及中心区域的位置,提出了一种基于视觉感知特性的非线性幅型比变换方法。该方法首先结合视频帧的运动信息,利用自适应模糊聚类技术估算视频帧的运动重心(MCOG),再针对视频各帧的运动重心及中心的位置进行区域划分,并选择合适的变换比率进行幅型比变换。该方法大大减小了视频帧的视觉感兴趣区域的失真,反映了人的视觉感知特性,从而提高了视频的整体视觉效果。本文方法框图如图 1 所示。

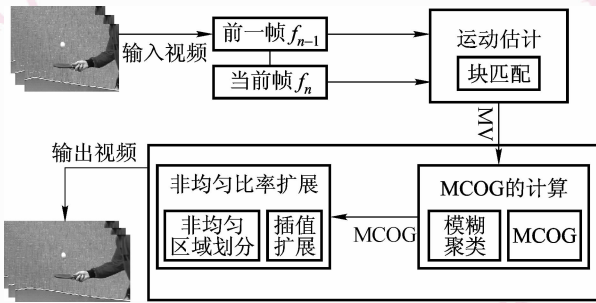


图 1 本文方法框图

Fig. 1 The block diagram of the proposed method

## 2 基于视觉感知特性的非线性幅型比变换方法

首先利用全搜索块匹配的运动估计方法<sup>[6-7]</sup>对视频各帧进行运动估计;再利用自适应模糊聚类技术对运动矢量进行聚类,进而根据聚类结果估算各帧的运动重心;最后,利用视频帧的运动重心和中心位置信息,对各帧进行非均匀区域划分和比率变换,从而得到相应幅型比的视频序列。以下对运动重心估计以及区域划分和比率变换分别进行介绍。

### 2.1 运动重心估计

在估算视频帧的运动重心之前,需要对所得到的运动矢量进行聚类分析,再根据聚类结果,利用运动重心计算公式获得视频各帧的运动重心。

#### 2.1.1 运动矢量的聚类

本文采用自适应模糊 C 均值(FCM)聚类算法

对运动矢量进行聚类。该算法首先利用修正划分模糊度参数优选算法,自适应地确定视频帧运动矢量的最佳聚类数;然后根据最佳聚类数,采用 FCM 聚类算法完成对运动矢量的聚类。

基于修正划分模糊度参数优选的算法<sup>[8]</sup>是一种自适应地确定最佳聚类数的算法。对于给定的类别数  $c$  和模糊划分矩阵  $U$ ,待聚类分析的数据集  $X$  的修正划分模糊度定义为

$$M_{pf}(U, c) = \frac{P_f(U, c)}{\tilde{H}(U, c)} \quad (1)$$

式中,  $\tilde{H}(U, c) = Smooth(H(U; c))$  是平滑后的模糊划分熵,  $P_f(U, c)$  为划分模糊度。如果记  $\Omega_c$  为所有“最优”分类矩阵的有限集,如果存在  $(U^*, c^*)$  满足

$$M_{pf}(U^*, c^*) = \min_c \{ \min_{\Omega_c} M_{pf}(U, c) \} \quad (2)$$

则  $(U^*, c^*)$  是最有效的聚类结果,  $c^*$  是最佳的分类数。

以各帧像素的运动矢量  $MV = (mv_x, mv_y)^T$  作为待聚类分析的数据集,并预先设定运动矢量的最大聚类数  $c_{max}$ ,则利用基于修正划分模糊度参数优选算法获得运动矢量最佳聚类数的步骤如下:

(1) 初始化运动矢量的聚类数  $c_0 = 2$ ,同时令  $n = 0$ 。

(2) 利用 FCM 聚类算法<sup>[9]</sup>,计算当聚类数为  $c_n$  时,对应的运动矢量划分矩阵  $U_n$ 。

(3) 计算运动矢量的模糊划分熵  $H(U_n; c_n)$  和划分模糊度  $P_f(U_n; c_n)$ ,并利用式(1)获得聚类数为  $c_n$  时的修正划分模糊度  $M_{pf}(U_n; c_n)$ ,并将其保存待用。

(4) 如果  $c_n < c_{max}$ ,则令  $c_{n+1} = c_n + 1, n = n + 1$ ,转到步骤(2)。如果  $c_n \geq c_{max}$ ,再依据式(2),从已经得到的  $M_{pf}(U_n; c_n) (0 \leq c_n \leq c_{max})$  中寻找最小值  $M_{pf}(U^*; c^*)$ ,其对应的聚类数就是运动矢量的最佳分类数  $c^*$ ,则终止程序。

图 2 是视频序列“football”和“tennis”的修正划分模糊度随运动矢量的类别数增加的变化曲线,修正划分模糊度的最小值为最优值,对应的类别数就是最佳类别数  $c^*$ 。由图 2 可以看出,视频序列“football”和“tennis”的运动矢量的最佳聚类数分别为 3 和 2。

获得视频帧运动矢量的最佳分类数  $c^*$  后,再利用 FCM 算法将运动矢量划分为  $c^*$  类,并获得每一类的中心运动矢量。从而,自适应 FCM 聚类算法可根据视频的运动特性自适应地对运动矢量进行分类。

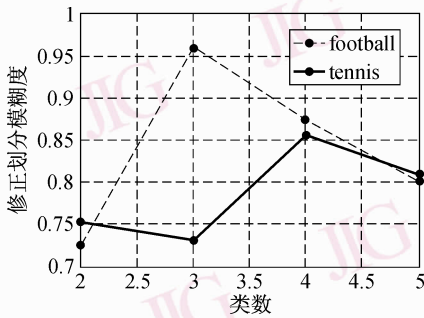


图 2 修正划分模糊度随类别数变化的曲线图

Fig. 2 Modified Partition Fuzzy Degree as function of the class number for “football” sequence and “tennis” sequence

### 2.1.2 运动重心估计

在对每一帧的运动矢量聚类完成之后,把所有像素的运动矢量赋值为所属类别的中心运动矢量值  $mv_{x,y}^{(c_{fcm})}$ ,那么视频每一帧的运动重心<sup>[2]</sup>可以表示为

$$MCOG = \frac{\sum_x mv_x^{(c_{fcm})} \cdot x}{\sum_x mv_x^{(c_{fcm})}} \quad (3)$$

式中,  $mv_x^{(c_{fcm})} = \sum_y mv_{x,y}^{(c_{fcm})}$  表示对水平坐标  $x$  所在的垂直( $y$ 轴)方向的运动矢量值求和。

本文之所以采用自适应模糊 C 均值聚类算法对运动矢量进行聚类,是因为此算法不仅能自适应地确定类别数,而且抗噪声能力优于 K 均值聚类算法<sup>[4]</sup>。同时,此聚类算法能较客观地反映实际运动矢量的分类,进而使运动重心估计更准确稳定。

## 2.2 幅型比变换的区域划分和插值扩展

针对人的视觉系统对视频帧的运动区域和中心区域较敏感的特性,本文采用了基于运动区域和中心区域的非均匀区域划分法,并且对各区域采用不同的变换比率进行扩展。这里,区域面积的分配以及各区域的扩展比率大小是预先确定的,其原则是综合利用视觉感知特性,对不同的视频序列进行实验和对比分析,选择能提高大部分视频序列变换后视觉效果的面积分配方案和扩展比率。

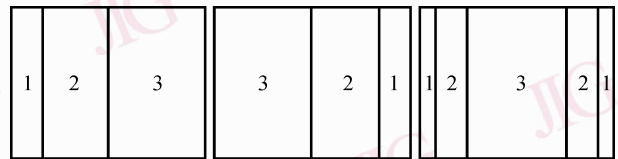
### 2.2.1 非均匀区域划分法

非均匀区域划分法是一种综合考虑视频帧的运动重心和中心区域的分区方法。分区方式及各区域面积的大小是根据视频帧的运动重心和中心的位置及运动场的分类数,并通过实验确定的。对重心和中心所在区域分配相对较大的面积,而对其余区域分配较小的面积。下面对运动场的分类数分别为 3

和 2 时,所对应的区域划分情况作介绍。

当运动场的分类数为 3 时,首先判断视频帧的运动重心和中心的位置,然后根据其位置进行分区。当运动重心居于视频帧的右部时,视频帧扩展前的区域划分如图 3(a) 所示。区域 3 则为运动相对剧烈的区域,根据确定的面积分配方案,在实验中将其分配为整个视频帧的 60% 的面积。对靠近区域 3 的区域 2 则分配整个视频帧的 30% 的面积,对于远离运动剧烈的区域并靠近边缘的区域 1,则分配最小的面积。当运动重心居于视频帧的左部时,也进行同样的分区方法,如图 3(b) 所示。越靠近视频帧的运动重心和中心的区域分配越大的面积,相反,越远离运动区域而靠近图像边界的区域则分配越小的面积。当运动重心居于视频帧的中部时,其区域划分如图 3(c) 所示。实验中区域 3 的面积仍为整个视频帧面积的 60%,位于左右两部分的区域 2 的面积均占整个视频帧面积的 15%,边缘区域 1 为视频帧的 5% 的面积。

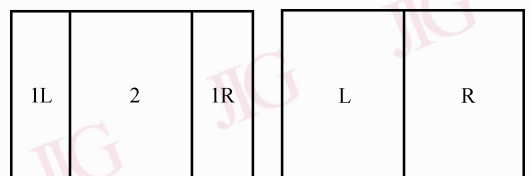
当运动场的分类数为 2 时,区域划分如图 4 所示。在运动重心位于图像中心区域时,直接按图 4(a) 的分割方式进行区域划分。运动相对剧烈的区域 2 同样分配了整个图像帧的 60% 的面积,边缘的 1L 和 1R 区域则分配较小的面积;当运动重心位于靠近边缘区域 1L 或 1R 时,则按图 4(b) 所示的方式进行划分,这时运动重心所在区域分配稍大的面积,本文分配了视频帧 60% 的面积。



(a) 重心居右 (b) 重心居左 (c) 重心居中

图 3 基于运动重心和中心区域的视频帧分区示意图

Fig. 3 Splitting regions based on motion and center regions



(a) (b)

图 4 基于运动重心和中心区域的视频帧分区示意图

Fig. 4 Splitting regions based on motion and center regions

### 2.2.2 非线性比率扩展法

为了减小变换后视频序列感兴趣区域的失真,并能满足相应幅型比变换的比例要求,在区域划分后,对运动剧烈区域以较小的比率进行扩展,对其他区域进行较大比率的扩展。在对每个区域进行插值扩展时,根据待插值点的 4 个最近邻像素的色度以及待插值点和这 4 个相邻像素的距离关系,利用双线性插值算法获得待插值点的色度,进而使各区域得到相应比率的插值扩展。

在进行从 4 : 3 到 16 : 9 的幅型比变换时,对于运动矢量的分类数为 3 的视频帧,在运动剧烈的区域 3,在实验中采用宽高比为 1.08 的扩展比例进行插值扩展,使视频的视觉感兴趣区域失真最小。相应地,对于区域 2,则要增大扩展比率,采用 1.25 的宽高比例进行插值扩展,同时,以 1.42 的宽高比例插值扩展区域 1。当运动矢量的分类数为 2 时,同样,对于运动重心所在的区域以接近于 1 的宽高比例进行扩展,同时为了使扩展后视频帧的幅型比为 16 : 9,其余区域则以较大的比例进行扩展。

这种非线性比率扩展的方法,不仅能使变换后的视频序列满足所需幅型比的要求,而且有效地减小了幅型比变换对视频帧运动区域和中心区域的影响,具有较好的视觉效果。

值得注意的是,当运动场的分类数为 2 时,在插值前,视频的一帧被分为两部分。如果只考虑运动重心所在的区域,便只根据运动重心位置,沿着中线将图像分割为左右两部分,并对两部分进行不同比率的扩展,如图 5(a)所示。对于运动区域集中在某一部分的视频序列,它具有较好的效果。但对于一般的视频序列来说,运动重心不可能始终在同一区域,这将导致幅型比变换后的视频产生视觉效果上的抖动现象,原因是前后帧的变换方式不一致,不同比率的变换会使得定义的分界线在变换后的视频帧之间进行移位,如图 5(b)和图 5(c)所示。

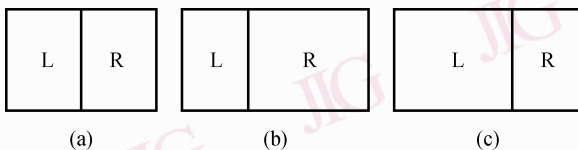


图 5 只基于运动的视频帧分区扩展示意图

Fig. 5 Extending regions based on motion region

“football2”序列进行从 4 : 3 到 16 : 9 的幅型比变换,并且通过自适应 FCM 聚类分析,获得运动场的分类数为 2。序列的其中两帧如图 6(a)、图 6(b)所示。图 6(c)和图 6(d)是只基于运动重心,直接将视频帧分成左右两部分并进行非均匀扩展方法的实验结果,从图中可看出,由于前后帧的运动重心从右边区域变化到左边区域,导致左右两部分的扩展比率在相邻帧之间发生改变,从而使变换后视频的前后帧发生视觉的不连续现象。如果同时考虑运动重心和中心区域,按照图 4(a)的方式划分区域,则可解决此问题,使变换后的视频保持了连续性,如图 6(e)和(f)所示。

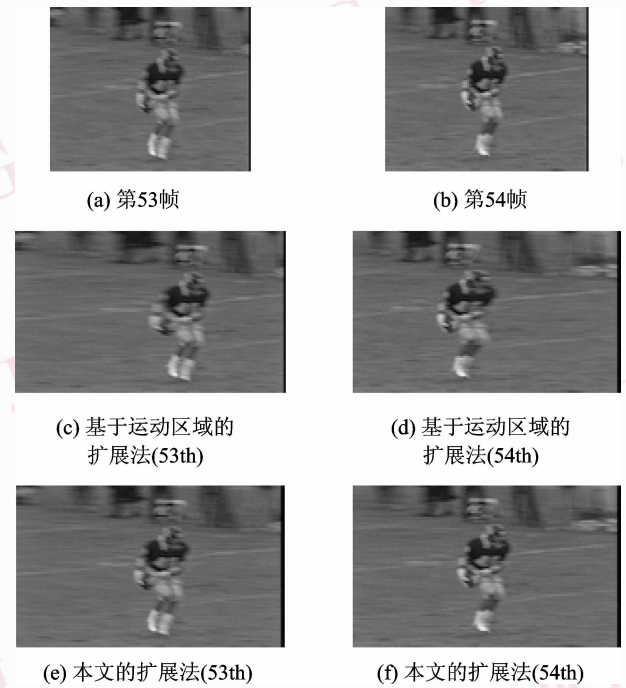


图 6 视频连续性方面的不同方法实验对比结果图

Fig. 6 Comparison of different methods on continuity of video

## 3 实验结果与分析

为了验证本文方法的有效性和正确性,采用了主观和客观两种评测方案。实验数据采用运动重心居中的“tennis”序列和运动情况较复杂的“football”序列,帧速率均为 30 帧/秒。实验中,将本文方法分别与横向线性拉伸的幅型比变换方法和基于中心区域的非线性幅型比变换方法<sup>[5]</sup>进行了比较。每个测试序列的实验代表图如图 7 所示。

图 6 是对上述问题的一个实验结果。将

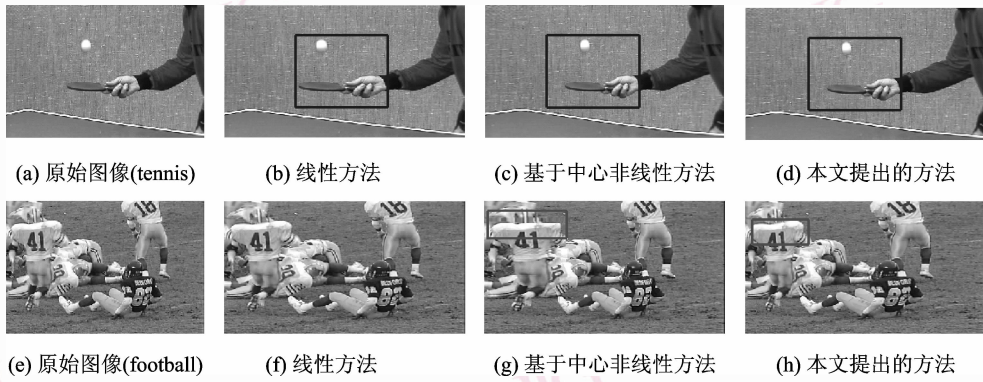


图 7 测试序列的变换结果图

Fig. 7 Results of two test sequences with three methods

### 3.1 主观评测实验

由图 7 可以看出,在图像的细节保真方面,与其他各种变换方法相比,本文的方法明显减小了感兴趣区域的细节失真,如“tennis”序列中运动的乒乓球、球拍,以及运动员的手部,线性方法有明显的横向拉伸的失真效果,而本文方法减小了上述区域的失真;在视频序列的整体性和连续性方面,通过观察变换后的视频发现,本文方法很好地解决了只基于运动区域的非均匀扩展变换法所造成的视频序列前后不连续的问题;而且对于包含复杂运动的视频序列,本文方法也优于基于中心区域的非线性扩展变换法和线性变换法,很好地保持了画面的和谐性。比如在“football”序列中,对比图 7(g)和图 7(h)可看出,图 7(g)中的 41 号运动员的左面肩部有明显的失真,与右肩整体看来不协调,而在图 7(h)中这个问题得到了很好地解决。

### 3.2 客观评测实验

#### 3.2.1 客观测试方法设计

在对视频序列客观测试过程中,由于变换前后视频帧的大小不一致,无法用传统的 MSE,PSNR 来测试比较视频帧的质量。考虑到幅型比变换是依据人的视觉感知特性进行,本文采用了一种基于视觉兴趣的图像质量评价方法<sup>[10]</sup>,该方法近似地认为人眼对感兴趣区的兴趣程度与其面积成反比。

为简单起见,假设测量图像中只有一个感兴趣区域  $A_1$ ,其面积为  $S_1$ ,不感兴趣区域  $A_2$  的面积为  $S_2$ , $S$  为原图像的总面积( $S = S_1 + S_2$ )。在测试中,忽略变换后图像中  $S$  以外的最不敏感部分区域,同时假定人眼对感兴趣区域内各像素点的兴趣程度相同,并且设感兴趣区域和不感兴趣区域的兴趣加权

值分别为  $\lambda_1$  和  $\lambda_2$ ,加权值越大,表示人的视觉对该区域感兴趣的程度越高。测量图像的均方误差为

$$IMSE = \frac{1}{S} \left[ \lambda_1 \sum_{(m,n) \in A_1} (f_{m,n} - f'_{m,n})^2 + \lambda_2 \sum_{(m,n) \in A_2} (f_{m,n} - f'_{m,n})^2 \right] \quad (4)$$

式中,  $\{f_{m,n}\}$  和  $\{f'_{m,n}\}$  分别表示原图像和变换后的图像,  $\lambda_1, \lambda_2, S_1, S_2$  满足式(5):

$$\lambda_1 S_1 + \lambda_2 S_2 = \lambda_1 S_1 + \lambda_2 (S - S_1) \quad (5)$$

则 ISNR (improved signal-to-noise ratio) 定义为<sup>[10]</sup>

$$ISNR = 10 \lg \frac{256 \times 257}{IMSE} \quad (6)$$

本文测试时的区域划分如下图 8 所示。其中,  $S_1 = 0.7S, \lambda_1 = 1.42852, S_2 = 0.3S, \lambda_2 = 0.00012$ 。

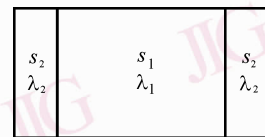


图 8 ISNR 的测试参数示意图

Fig. 8 Schematic diagram of parameters in ISNR

#### 3.2.2 客观测试结果分析

测试结果如表 1 所示。通过对比发现,本文方法更符合人的视觉感知特性,两组 ISNR 的值均高于线性变换方法和基于中心区域的非线性变换方法<sup>[5]</sup>;对于“tennis”序列,本文的方法分别高于线性方法 6.128 dB 和基于中心非线性方法 0.059 dB。由于“tennis”序列的运动剧烈部分靠近画面的中心

区域,所以与基于中心的非线性方法相比,本方法变换结果的 ISNR 提高得相对较少。而对于包含复杂运动的序列而言,本文方法具有更好的变换效果,比如“football”序列,其运动剧烈部分并非总靠近画面的中心区域,因此用本文方法得到的视频帧的 ISNR 明显高于其他的方法。

表 1 3 种方法在测试视频序列时的性能 (ISNR) 比较

Tab. 1 Comparison of ISNR for two sequences in three methods

序列	ISNR		本文的方法 ISNR
	线性方法 ISNR	基于中心非线性方法 ISNR	
football	15.826 1	15.513 9	17.597 7
tennis	19.584 6	25.654 1	25.712 9

## 4 结 论

针对从幅型比为 4:3 的 SDTV 电视信号格式到 16:9 的 HDTV 信号格式的转换问题,提出了基于人的视觉特性的非线性幅型比变换方法,根据人的视觉系统对视频帧的运动区域和中心区域具有较强感知度的特性,首先利用运动信息和运动场的聚类结果来估算视频帧的运动重心,再依据各帧的运动重心和中心的位置,对视频帧进行分区,并采用非线性扩展比率对各区域进行插补扩展。实验结果表明,该方法解决了视频的整体失真一样大、画面不和谐的问题,避免了变换后视频前后帧之间的跳跃现象,能较好地反映人的视觉对视频的感知情况,与传统方法相比,变换效果有明显提高。为了进一步扩展本文方法,可以通过估算多个运动重心的方法来减小多个运动区域的失真。也可以采用水平处理与垂直处理相结合的方法,即在垂直方向裁剪掉很小的一部分次要信息,同时减小水平方向的扩展比率,进一步减小变换后视频帧的失真。

## 参考文献 (References)

- Jeong J, Jeon B. A multiplierless letter box converter for displaying 16:9 images on 4:3 screen [J]. IEEE Transactions on Circuits Systems for Video Technology, 1995, 5(4): 363-366.
- Zhang Guang-lie, Zheng Nan-ning, Liang Feng, et al. Image extension algorithm with non-uniform ratio based on motion adaptation [J]. TV Engineering, 2003, (3)16-18. [张光烈, 郑南宁, 梁峰等. 基于运动自适应的非均匀比率图像扩展算法[J]. 电视技术, 2003, (3): 16-18.]
- Auty S J, Dabner S C. To and from the widescreen: aspect ratio conversion [A]. In: Proceedings of International Broadcasting Convention[C], Amsterdam, Netherlands, 1995, 413: 554-558.
- 萧小笑. 4:3和16:9两种幅型比[EB/OL]. <http://www.tencent.com/Article/ShowArticle.asp?ArticleID=166,2004-12-28>.
- Hsia S C, Liu B D, Yang J F, et al. A parallel video converter for displaying 4:3 images on 16:9 HDTV receivers [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(6): 695-699.
- Dufaux F, Moscheni F. Motion estimation techniques for digital TV: A review and a new contribution [J]. Proceedings of the IEEE, 1995, 83(6): 858-876.
- Wang Y, Ostermann J, Zhang Y Q. Video Processing and Communication [M]. Beijing: Publication House of Electronics Industry, 2003: 119-124. [(美)王尧等著. 侯正信等译. 视频处理与通信[M]. 北京:电子工业出版社, 2003, 119-124.]
- Li Jie, Gao Xin-bo, Jiao Li-cheng. A new cluster validity function based on the modified partition fuzzy degree [A]. In: Proceedings of the 4th International Conference on Rough Sets and Current Trends in Computing[C], Uppsala, Sweden, 2004: 586-591.
- Gao Xin-bo. Fuzzy Cluster Analysis and Its Applications [M]. Xi'an: Publishing House of Xidian University, 2004: 50-56. [高新波. 模糊聚类分析及其应用[M]. 西安:西安电子科技大学出版社, 2004: 50-56.]
- Wang Kong-qiao, Shen Lan-sun, Xing Xin. A quality assessment method of image based on visual interests [J]. Journal of Image and Graphics, 2000, 5(4):300-303. [汪孔桥, 沈兰荪, 邢昕. 一种基于视觉兴趣性的图像质量评价方法[J]. 中国图象图形学报, 2000, 5(4): 300-303.]