

# 结构相似度在 AVS 帧间模式选择中的应用

陈云善<sup>1,2\*</sup>, 高慧斌<sup>1</sup>, 苏宛新<sup>1</sup>, 王 旻<sup>1</sup>

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033; 2. 中国科学院 研究生院, 北京 100039)

**摘要:**传统的音视频编码标准(AVS)率失真优化(RDO)算法在帧间模式选择中使用偏差平方和(SSD)作为失真度量,不能很好符合人的主观视觉,而最近提出的结构相似度(SSIM)图像质量评价方法更符合人眼视觉系统(HVS)的特性。为了有效地提高编码效率,本文将SSIM引入拉格朗日代价函数,对失真度量的表示进行修正,并在大量实验的基础上建立了拉格朗日参数的经验公式。实验结果表明,与AVS模式选择算法相比,本文算法平均可节约13.22%的比特率,尤其对于静止块较多的序列,在量化参数QP=10时,节省码率达到30%以上。与此同时,重建的图像质量仅下降0.14%,可忽略不计。

**关键词:**结构相似度(SSIM); 音视频编码标准(AVS); 率失真优化(RDO); 帧间模式选择

**中图分类号:**TP919.81 **文献标识码:**A **文章编号:**1005-0086(2011)03-0435-05

## Application of structural similarity in inter mode selection for AVS

CHEN Yun-shan<sup>1,2\*</sup>, GAO Hui-bin<sup>1</sup>, SU Wan-xin<sup>1</sup>, WANG Min<sup>1</sup>

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; 2. Graduate University, Chinese Academy of Sciences, Beijing 100039, China)

**Abstract:** The sum of squared differences (SSD) is used as the distortion metric for inter mode decision in traditional audio video coding standard (AVS) rate distortion optimization (RDO), but it is not consistent with human vision system (HVS) quite well after a lot of investigations. Compared with the other methods, the structural similarity (SSIM) proposed recently for assessing image quality accords with HVS much more well. Therefore, in order to improve coding efficiency, the SSIM is introduced into Lagrangian cost function to correct the expression of distortion metric. Based on experiments, the empirical formula of Lagrangian parameter is established in this paper. Experimental results indicate that the bit rate using the proposed algorithm is averagely reduced by 13.22% than that using the mode selection algorithm of AVS. Especially, when the sequences have massive stationary blocks, the bit rate is saved more than 30% with  $QP=10$ . At the same time, the quality of the reconstructed image is only decreased by 0.14% which could be ignored.

**Key words:** structural similarity (SSIM); audio video coding standard (AVS); (RDO); inter mode decision

## 1 引言

我国自主制定的音视频编码技术标准(AVS),包括系统、视频、音频和数字版权管理4个主要技术标准和一致性测试等支撑标准<sup>[1]</sup>。其中AVS-P2是AVS系列标准的第2部分,主要针对高清晰度数字电视广播和高密度存储媒体应用。AVS-P2的率失真优化(RDO)算法在模式选择中采用偏差平方和(SSD)作为失真度量,最后采用峰值信噪比(PSNR)来衡量视频图像的质量。虽然SSD和PSNR计算简单且物理意义清晰,但其仅是对像素点之间误差的纯数学统计,没有考虑到像

素点间的相关性和人眼视觉系统(HVS)的感知特性,在很多情况下不符合人的主观感受。

而近几年提出的结构相似度(SSIM)图像质量评价方法<sup>[2,3]</sup>,从图像组成角度将结构信息定义为独立于亮度、对比度反映场景中物体结构的属性,并将失真建模为亮度、对比度和结构3个不同因素的组合。由于SSIM的算法相对简单、准确性好,优于传统的PSNR等方法,将SSIM引入H.264率失真函数中,在保证压缩质量不下降的前提下,极大的降低了码率<sup>[6,7]</sup>。为此,本文将SSIM引入AVS-P2帧间模式选择中,用SSIM取代SSD作为重建图像相对原始图像的失真度量,称

\* E-mail: yiyunsn@163.com

收稿日期:2010-06-21 修订日期:2010-09-15

基金项目:国家“863”计划资助项目(2008AA7034320B)

为基于 SSIM 的率失真优化算法(BSSRDO)。本文在大量实验的基础上,通过曲线拟合建立了拉格朗日参数  $\lambda_{MODE}$  与量化参数  $QP$  的经验公式,实验表明,在保证重建图像质量基本不变的前提下,BSSRDO 可有效的降低码率。

## 2 BSSRDO

### 2.1 AVS 的模式选择过程

AVS 采用多种块模式的帧间预测,在可选的宏块类型中遍历每种模式,计算出每种模式的图像失真和所需的比特数,利用

$$J(s, c, MODE | QP) = D(s, c, MODE | QP) + \lambda_{MODE} R(s, c, MODE | QP) \quad (1)$$

比较各模式率失真代价,选取率失真代价最小的模式作为最佳编码模式。

其中:  $QP$  为量化参数;  $MODE$  是宏块候选编码模式;  $D$  为原始信号  $s$  与其重建信号  $c$  的差值平方和,作为失真度量;  $R$  为量化参数  $QP$  和模式  $MODE$  下编码此宏块所需的比特数(包括宏块头信息、运动矢量及残差变换量化后的系数等);  $\lambda_{MODE}$  为拉格朗日参数为

$$\lambda_{MODE} = 0.85 \times 2^{(QP-11)/4} \quad (2)$$

对应  $P$  帧,宏块编码模式的候选集合为

$$MODE \in \{P-Skip, P-16 \times 16, P-16 \times 8, P-8 \times 16, P-8 \times 8, I-8 \times 8\} \quad (3)$$

对于每种预测模式,需要通过运动估计为当前编码块选择最佳匹配块,最佳匹配块由

$$M_{COST}(s, c) = SA(T)D(s, c) + \lambda_{MOTION} b(\Delta_{MV}) \quad (4)$$

所示的率失真代价函数选择。

其中:  $SA(T)D$  表示  $SAD$  或  $SATD$ ,  $SAD(s, c)$  是当前编码块  $s$  与候选匹配块  $c$  的像素值差的绝对值之和,用于整数像素运动估计;  $SATD(s, c)$  表示当前编码块与候选匹配块  $c$  的像素值差经哈达玛变换后所得系数绝对值之和,用于分数像素运动估计;  $\lambda_{MOTION}$  为运动估计的拉格朗日参数,  $\Delta_{MV}$  是预测运动矢量和实际运动矢量的差值,  $b(\Delta_{MV})$  表示用来编码  $\Delta_{MV}$  所用的比特数。对于每种预测模式,具有最小  $M_{COST}$  的候选块为当前块的最佳匹配块。

### 2.2 SSIM

SSIM 是近年出现的新的图像质量评价标准,相比其他图像质量评价方法,SSIM 更符合主观评价标准和 HVS<sup>[2-5]</sup>。设  $x, y$  分别为参考图像和待评估图像,则它们间的 SSIM 定义为

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (5)$$

其中:  $\alpha, \beta, \gamma > 0$ , 分别用来调整亮度、对比度和结构信息的权重;  $l(x, y)$  是亮度比较,用参考图像和待评估图像的均值  $\mu_x$  和  $\mu_y$  来计算;  $c(x, y)$  是对比度比较,用参考图像和待评估图像的方差  $\sigma_x^2$  和  $\sigma_y^2$  来估计;  $s(x, y)$  是结构比较,用参考图像和待评估图像的协方差  $\sigma_{xy}$  来度量。这三个分量分别定义为

$$l(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (6)$$

$$c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (7)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (8)$$

其中,参数  $C_1, C_2$  和  $C_3$  是为避免分母为零时出现的不稳定情

况而加上的小常数,定义为

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2, C_3 = C_2 / 2 \quad (9)$$

式中:  $K_1, K_2 \ll 1$ ;  $L$  为像素亮度值的范围。参考相关文献[3]本文取

$$\alpha = \beta = \gamma = 1, K_1 = 0.01, K_2 = 0.03, L = 255 \quad (10)$$

SSIM 值一般为 0~1 间的浮点数,它反映的是两个图像块的相似程度,SSIM 值越大说明这两个图像块越相似。

### 2.3 基于 SSIM 的模式选择过程

由于 SSIM 取值越大失真越少,而 SSD 刚好相反。因此,将 SSIM 作为失真度量引入帧间模式选择,拉格朗日代价函数修正为<sup>[6-7]</sup>

$$J(s, c, MODE | QP) = \omega(1 - SSIM(s, c)) + \lambda_{MODE} R(s, c, MODE | QP) \quad (11)$$

其中,  $\omega$  为调整式子平衡的调节因子。在实验中发现,  $\omega$  的取值与  $QP$  相关。为了计算方便,本文将模式选择代价函数修正为

$$J(s, c, MODE | QP) = \lambda_{MODE}(1 - SSIM(s, c)) + R(s, c, MODE | QP) \quad (12)$$

将 SSIM 引入帧间模式选择后,在帧间宏块模式选择过程中,其余部分保持不变,运动搜索仍使用式(4)选择各模式最佳匹配块,仅在模式选择阶段改用式(12)进行模式选择。模式选择具体方法如下:

针对一个宏块,首先,利用式(5)~(10),按  $16 \times 16$  亮度块和  $8 \times 8$  色度块直接计算每个宏块各模式的 SSIM 值。由于使用 RDO 技术时要综合考虑亮度和色度分量,因此根据相关理论,式(12)中的 SSIM 值为

$$SSIM = 0.7SSIM_y + 0.15SSIM_u + 0.15SSIM_v \quad (13)$$

其中,  $SSIM_y, SSIM_u$  和  $SSIM_v$  分别为  $16 \times 16$  亮度分量和  $8 \times 8$  色度分量的 SSIM 值。

然后,利用式(12)计算各模式率失真代价,具有最小率失真代价的为最佳模式。式(12)中,拉格朗日参数  $\lambda_{MODE}$  是关于  $QP$  的函数,通过对大量标准测试序列的实验<sup>[8,9]</sup>,利用最小二乘法的拟合曲线如图 1 所示,得出的  $\lambda_{MODE}$  经验公式为

$$\lambda_{MODE} = 10^{0.0028QP^2 - 0.039QP + 5.34} \quad (14)$$

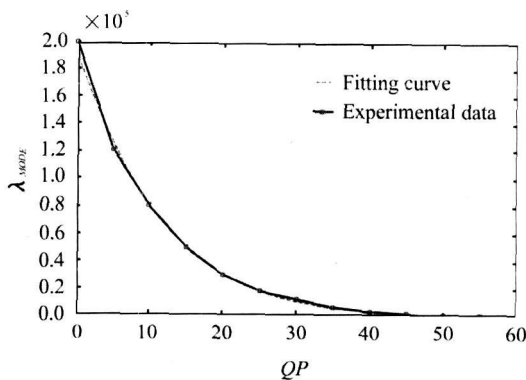


图 1  $\lambda_{MODE}$  与  $QP$  关系曲线

Fig. 1  $\lambda_{MODE}$  vs.  $QP$

## 3 实验结果与分析

为了检验 BSSRDO 算法和经验式(14)的有效性,将本文算法在 AVS 编解码器 rm52j 上实现。实验选取具有不同的空间细节、运动程度、镜头移动以及背景的 12 个标准测试序列的前

50 帧进行测试, 各序列均为 CIF 格式, 帧率为 30 f/s。图 2 所示为各序列的第 1 帧。QP 值分别选取为 10、20、30, 编码帧结构为 IPPP, 使用全搜索, 最大搜索范围为 16×16, 开率失真优化。为了比较 BSSRDO 和 AVS 标准算法的帧间编码性能, 二者在帧间编码时均不使用帧内模式。

采用亮度分量和色度分量的加权 ( $MSSIM$ ) 作为视频质量的客观评价标准, 计算公式为

$$MSSIM = 0.7SSIM_V + 0.15SSIM_U + 0.15SSIM_V \quad (15)$$

每一帧的  $MSSIM$  使用 16×16 滑动窗计算, 窗口从左上角开始, 向下向右重叠滑动(每次滑动一个像素), 直到该帧的右下角, 每一个窗口的  $SSIM$  用式(15)计算, 最后取平均值。

仿真环境为 AMD Athlon(tm) 64 X2 Dual Core Processor 4800+ 2.51 GHz, 2 GB, VS2005。

虽然  $SSIM$  的计算量略高于  $SSD$ , 但 BSSRDO 算法仅在

帧间模式选择阶段改用  $SSIM$  作为失真度量, 整体编码时间变化不明显, 在此仅列出压缩质量、压缩比的性能对比。QP 取值为 10、20 和 30 时的  $MSSIM$  值和平均每帧比特率的性能对比见表 1。其中, 码率表示所有 P 帧比特数的平均, 用来衡量压缩比;  $\Delta S$ 、 $\Delta B$  分别为 BSSRDO 算法的  $MSSIM$ 、比特率与 AVS 标准算法相比的变化百分比。

从表 1 可见, BSSRDO 算法与 AVS 标准算法相比, 在获得相近重建图像质量的前提下, 码率均有不同程度降低。对应于 QP= 10、20、30 的 12 个序列码率平均降低为 14.65、12.85 和 12.17%, 而  $MSSIM$  平均仅下降 0.14%。在 QP= 30 时, Tempete、Harbor 以及 Flower 序列图像质量下降较大。为了更直观的比较编码质量, 在图(3)~(5)中给出了各序列第 50 帧原始图像和相应的重建图像。可见, 在 QP 相对较大的情况下, BSSRDO 算法的重建图像仍与原图像差异很小。

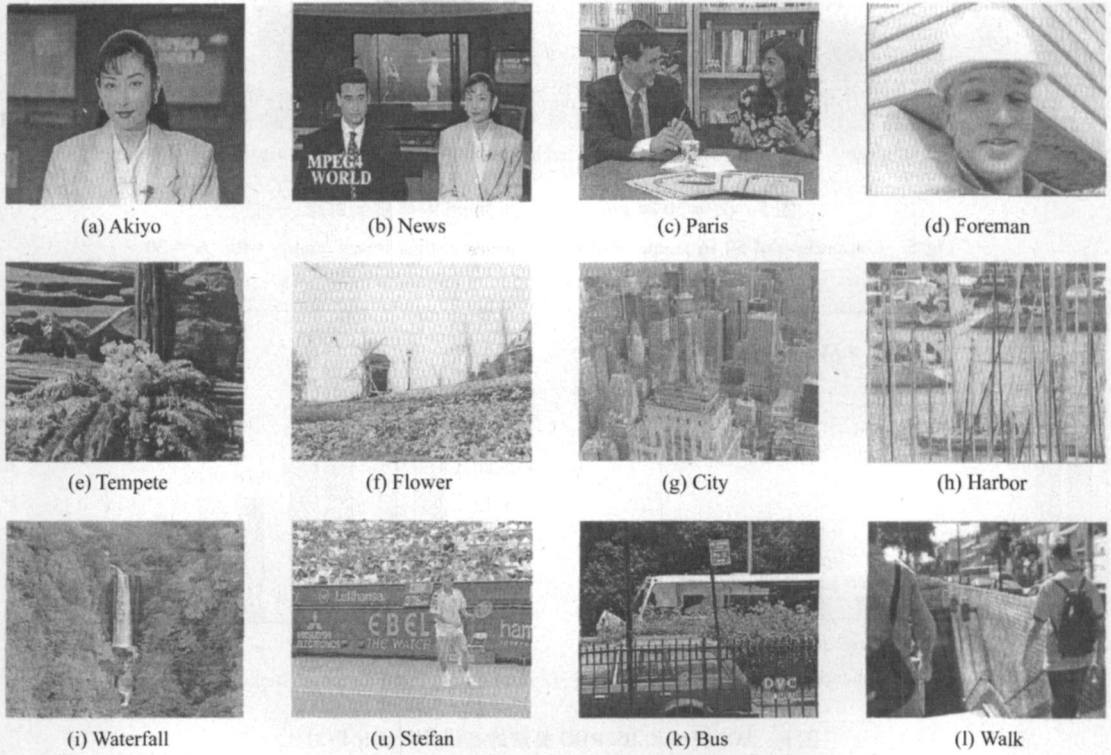


图2 各测试序列的第1帧

Fig. 2 The 1st frame of each sequences



图3 QP= 30 时 Tempete 序列第 50 帧编码质量对比

Fig 3 Comparison of 50 th frame of Tempete sequence coding image quality with QP= 30

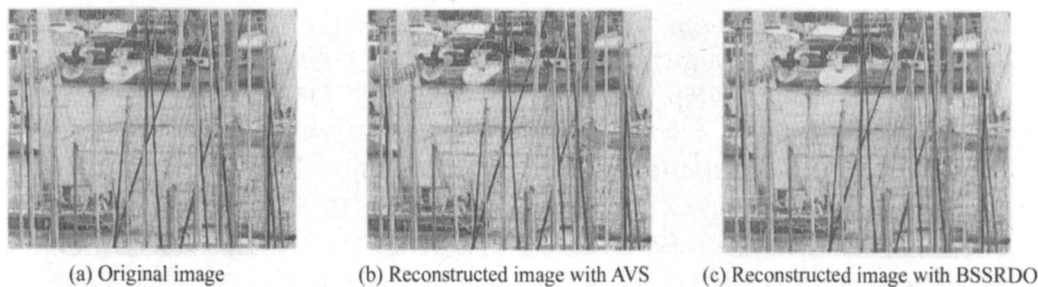


图4  $QP=30$  时 Harbor 序列第 50 帧编码质量对比

Fig 4 Comparison of 50 th frame of Harbor sequence coding image quality with  $QP=30$



图5  $QP=30$  时 Flower 序列第 50 帧编码质量对比

Fig 5 Comparison of 50 th frame of Flower sequence coding image quality with  $QP=30$

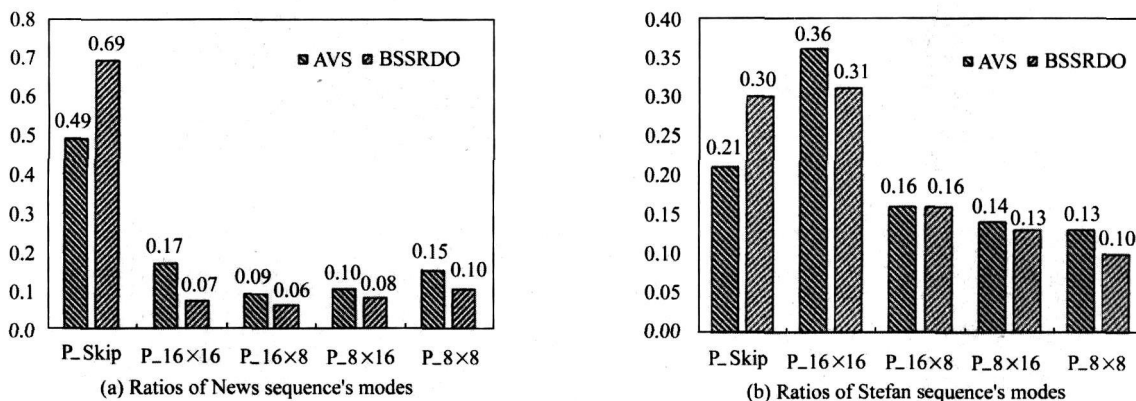


图6 AVS 算法和 BSSRDO 算法的各模式所占比例对比

Fig 6 Comparison of mode ratios between AVS and BSSRDO

此外,由表 1 还可以看出,BSSRDO 算法对于运动较小或者具有较多静态背景的序列(Akiyo 和 News),压缩效率改善更加突出,在  $QP=10$  时,News 序列压缩比提高了 36.93%。图 6 为 News 和 Stefan 序列使用 AVS 标准算法和 BSSRDO 算法编码时各模式所占比例对比图。由图 6 可以看出,BSSRDO 算法与 AVS 标准算法相比,无论是具有较多静止背景的 News 序列还是运动剧烈的 Stefan 序列,都有更多的宏块被选择为 P\_Skip 模式,而选择为其他 4 种模式的宏块则均有不同程度的减少。可见,BSSRDO 算法压缩比的提高,主要是由于采用 SSIM 作为失真度量来选择最佳模式,可以更好的保留图像的结构信息,使更多的宏块被选为 P\_Skip 模式。这也是 BSSRDO 算法对于具有较多静态背景序列效果更好的原因。

### 4 结论

将 SSIM 作为失真度量引入 AVS 模式选择过程中,实验结果表明,与 AVS 标准算法相比,本文提出的 BSSRDO 算法以较少的时间和质量下降为代价,极大的节约了码率,压缩比平均提高 13.22%,而压缩质量平均下降仅 0.14%,对主观视觉的影响可以忽略不计。该算法对于固定背景较多、运动量小的序列具有更好的效果,非常适用于视频会议、远程视频监控等场合。实验所使用的参数都是通过大量实验得出的经验值,仍有很大的优化余地。现有的 RDO 算法以及码率控制模型都是以方差为失真度量的,因此深入研究基于 SSIM 的 RDO 算法,提出基于 SSIM 的率失真模型,有望进一步提高编码效率。

表1 编码性能对比

Tab.1 Comparison of encoding performance

QP Sequence	AVS		BSSRD0		Comparison	
	MSSIM	Bitrate / kbit/p	MSSIM	Bitrate / kbit/p	$\Delta S/\%$	$\Delta b/\%$
10	Akiyo	0.9952	70.76	0.9948	48.52	-0.04 - 31.43
	News	0.9955	113.49	0.9938	71.58	-0.17 - 36.93
	Paris	0.9966	220.26	0.9949	153.07	-0.17 - 30.50
	Foreman	0.9959	256.84	0.9934	222.85	-0.25 - 13.23
	Flower	0.9967	297.91	0.9963	275.70	-0.04 - 7.46
	Tempete	0.9973	378.18	0.9963	333.97	-0.10 - 11.69
	City	0.9968	281.32	0.9963	260.01	-0.05 - 7.58
	Harbor	0.9971	369.92	0.9964	342.73	-0.07 - 7.35
	Waterfall	0.9976	277.68	0.9959	221.36	-0.17 - 20.28
	Stefan	0.9969	352.26	0.9964	337.00	-0.05 - 4.33
20	Bus	0.9968	346.98	0.9962	330.65	-0.06 - 4.71
	Walk	0.9966	321.31	0.9966	320.13	0.00 - 0.37
	Average					-0.10 - 14.65
	Akiyo	0.9900	16.25	0.9897	12.35	-0.03 - 24.00
	News	0.9888	25.85	0.9884	22.01	-0.04 - 14.85
	Paris	0.9918	72.31	0.9899	61.28	-0.19 - 15.25
	Foreman	0.9884	77.18	0.9858	69.03	-0.26 - 10.56
	Flower	0.9937	162.95	0.9918	139.41	-0.19 - 14.45
	Tempete	0.9923	174.89	0.9898	146.40	-0.25 - 16.29
	City	0.9912	98.11	0.9903	89.00	-0.09 - 9.29
30	Harbor	0.9934	175.63	0.9913	153.43	-0.21 - 12.64
	Waterfall	0.9898	83.44	0.9879	69.46	-0.19 - 16.75
	Stefan	0.9924	158.79	0.9910	141.40	-0.14 - 10.95
	Bus	0.9915	155.28	0.9901	143.17	-0.14 - 7.80
	Walk	0.9908	142.81	0.9907	140.85	-0.01 - 1.37
	Average					-0.15 - 12.85
	Akiyo	0.9806	5.44	0.9805	4.54	-0.01 - 16.54
	News	0.9775	9.57	0.9775	8.53	0.00 - 10.87
	Paris	0.9773	29.17	0.9753	24.88	-0.20 - 14.71
	Foreman	0.9701	22.54	0.9694	21.52	-0.07 - 4.53
Flower	0.9816	74.11	0.9771	59.74	-0.46 - 19.39	
Tempete	0.9754	64.68	0.9710	50.81	-0.45 - 21.44	
City	0.9739	30.95	0.9738	29.54	-0.01 - 4.56	
Harbor	0.9811	71.51	0.9766	60.19	-0.46 - 15.83	
Waterfall	0.9629	19.93	0.9628	18.27	-0.01 - 8.33	
Stefan	0.9809	59.60	0.9780	49.85	-0.30 - 16.36	
Bus	0.9753	61.17	0.9733	55.13	-0.21 - 9.87	
Walk	0.9760	55.47	0.9756	53.49	-0.04 - 3.57	
Average					-0.18 - 12.17	

参考文献:

[1] GB/T 20090. 2-2006. Information technology advanced coding of audio and video-part 2: video[S].

[2] WNAG Z, LU L, BOVICK A C. Video quality assesment using structural distortion measurement[A]. Proceedings of the IEEE International Conference on Image Processing[C]. Rochester: 2002: 65-68.

[3] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans on Image Processing, 2004, 13(4): 600-612.

[4] YU Tong, MA She-xiang, LIU Tie-gen. Image quality assessment based on multiscale edge structure similarity[J]. Journal of Optoelectronics • Laser, 2009, 20(6): 843-846.  
于彤, 马社祥, 刘铁根. 基于多尺度边缘结构相似性的图像质量评价[J]. 光电子 • 激光, 2009, 20(6): 843-846.

[5] WANG Yu-qing, LIU Wei-ya, WANG Yong. Image quality assessment based on local variance and structure similarity[J]. Journal of Optoelectronics • Laser, 2008, 19(11): 1546-1553.  
王宇庆, 刘维亚, 王勇. 一种基于局部方差和结构相似度的图像质量评价方法[J]. 光电子 • 激光, 2008, 19(11): 1546-1553.

[6] YANG Chun-ling, WANG Hua-xing. Fast motion estimation algorithm based on structural similarity for H.264 video coding[J]. Journal of South China University of Technology: Natural Science Edition, 2008, 36(8): 28-32.  
杨春玲, 王华兴. 基于结构相似度的 H.264 快速运动估计算法[J]. 华南理工大学学报: 自然科学版, 2008, 36(8): 28-32.

[7] ZHAO Lian-feng, GUO Bai-long, HAN He-min. Fast motion estimation method based on structural similarity[J]. Application Research of Computers, 2007, 24(11): 306-308.  
赵连凤, 郭宝龙, 韩合民. 基于结构相似度的快速运动估计算法[J]. 计算机应用研究, 2007, 24(11): 306-308.

[8] Wilegard T, Girod B. Lagrangian multiplier selection in hybrid video coder control[A]. Proceedings of the IEEE International Conference on Image Processing[C]. Thessaloniki: 2001: 542-545.

[9] Sullivan G J, Wiegand T. Rate-distortion optimization for video compression[J]. IEEE Signal Processing Magazine, 1999, 15(6): 74-90.

作者简介:

陈云善 (1984-), 男, 四川成都人, 博士研究生, 主要研究方向为达芬奇技术, 视频编码算法等.