

基于内容划分和运动补偿的视频质量评价*

姚杰, 谭建明, 陈婧

(重庆通信学院 信息网格实验室, 重庆 400035)

摘要: 目前基于结构相似性的图像质量评价算法均是对图像进行整体质量分析,但人类视觉系统对图像中不同部分的敏感程度不同,而对图像进行整体质量分析无法有效反映出这些差异。鉴于此,提出了一种基于内容划分的图像质量评价算法,根据图像不同区域的梯度将图像分为四个部分,分别进行质量评价。之后,采用基于运动补偿的帧加权方式将上述方法扩展为视频质量评价。实验证明,所述算法与目前比较流行的几个算法相比具有较高的评价准确性。

关键词: 视频质量评价; 图像质量评价; 结构相似性; 运动补偿

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)10-3956-04

doi:10.3969/j.issn.1001-3695.2012.10.095

Video quality assessment based on content-partitioned approach and motion compensation

YAO Jie, TAN Jian-ming, CHEN Jing

(Information Grid Laboratory, Chongqing Communication Institute, Chongqing 400035, China)

Abstract: Current structural similarity based image quality assessment algorithm is generally the overall image quality analysis. However, for human visual system, different regions in image have different visual sensitivities, and the current method can not reflect these differences effectively. In this view, this paper proposed a content-partitioned image quality assessment algorithm, which partitioned an image into four regions according to their different gradient magnitudes and assessed their qualities respectively. Then, based on motion compensation, this paper proposed a frame-weighting approach, which extended the proposed algorithm to video quality assessment. The experiments show that the proposed algorithm is more accurate than several current popular algorithms.

Key words: video quality assessment; image quality assessment; structural similarity; motion compensation

0 引言

对于大多数视频应用来说,人都是最终消费者,所以评价视频质量最准确的办法是主观质量评价。但是,根据 ITU-T BT500-11^[1]所规定的方法进行主观质量评价,人力和财力的耗费巨大,而且无法嵌入到视频应用系统中。因此,如何获得高效的客观视频质量评价算法已经成为图像和视频处理领域亟待解决的问题。

由于计算简单,峰值信噪比 PSNR 一直都是评价视频质量的重要方法。但大量事实表明,基于像素误差的 PSNR 评价结果在很多情况下并不能准确反映人的真实感受。为了解决这个问题,许多能反映人真实感受的客观视频质量评价算法陆续被提出。Wang 等人^[2]根据人类视觉系统对自然图像的结构敏感性原理提出了基于结构相似性的 SSIM 视频质量评价算法。在此基础上,Wang 等人^[3]又通过对图像的反复低通滤波和降采样将原来的算法扩展为 MSSIM,性能上有了较大的提高。VSNR^[4]利用小波进行图像分析,将人类视觉系统可觉察到的自然图像失真通过两个阈值进行量化,并通过求均值的方

法扩展为视频质量评价。由美国国家电信管理中心开发的 VQM^[5]算法因其在 VQEG Phase 2 视频测试序列库上的良好效果,已被美国国家标准研究院制定为国家标准。V-VIF^[6]通过时间延伸的方法来度量失真视频的信息保真度,反映了视频在时间域上的失真程度。

目前,大多数视频质量评价算法的一般思路是对每帧图像分别使用图像质量评价算法得到其质量分值,再取所有帧的质量平均值作为整段视频的质量值。但在视频序列中,帧一般是以每秒 25~30 张的速度连续播放的,人在观看时能明显感觉到视频中物体的运动具有连续性和方向性。而在静态的图像中,物体运动的大小和方向是无法进行准确判断的。如果简单地将每一帧质量进行求和并取平均值作为整个视频的质量,无法反映物体的运动特性,所以是不合理的。因此,必须充分考虑视频中物体的运动特性,对其运动情况进行准确判断,才能更好地反映出人观看视频时的真实感受。基于这一点,本文首先对传统的基于结构相似性的图像质量评价算法进行改进,提出了基于内容划分的图像质量评价算法,并在此基础上提出了一种基于运动补偿的帧加权方法,将图像质量评价算法扩展为

收稿日期: 2012-03-19; **修回日期:** 2012-04-21 **基金项目:** 国家创新基金资助项目(11c26215115768);重庆市重点攻关基金项目(cstc2011ab2064)

作者简介: 姚杰(1986-),男,江苏盐城人,硕士研究生,主要研究方向为图像质量评价(yj511608130@163.com);谭建明(1966-),男,教授,主要研究方向为计算机网络;陈婧(1986-),硕士研究生,主要研究方向为计算机网络。

视频质量评价算法。

1 基于结构相似性的图像质量评价

Wang 等人利用图像中像素点之间的相关性提出了图像结构的概念,并基于此提出了新的图像质量评价指标——结构相似度 SSIM(structural SIMilarity)。设 x 和 y 为两个非负的信号,并且已经过校准(来自两幅图像中对应空间位置上的对应像素点作为算法运算的两个输入),那么 SSIM 算法的数学模型定义为

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (1)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (2)$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \quad (3)$$

其中: $l(x, y)$ 、 $c(x, y)$ 和 $s(x, y)$ 分别是亮度、对比度和结构比较函数; μ_x 和 μ_y 分别是 x 和 y 的样本平均值,表示图像的亮度; σ_x 和 σ_y 分别是 x 和 y 的样本标准差,表示图像的对比度; σ_{xy} 是 x 和 y 的样本相关系数,表示失真图像相对于原始图像的非线性改变程度(结构失真); c_1 、 c_2 和 c_3 是足够小的正常数,用来保证在以上三个比较函数中当分母趋近于 0 时的计算稳定性。需要注意的是,式(1)仅由亮度决定,式(2)仅由对比度决定。联合以上三个方程,两个图像的结构相似度 SSIM(x, y) 为

$$\text{SSIM}(x, y) = l(x, y)^\alpha \times c(x, y)^\beta \times s(x, y)^\gamma \quad (4)$$

其中: α 、 β 和 γ 是调整亮度、对比度和结构三个部分的权重系数,一般取 $c_2 = 2c_3$, $\alpha = \beta = \gamma = 1$,那么结构相似度的一种普通形式为

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (5)$$

SSIM 算法一般作用在一个 11×11 的滑动窗口内,滑动窗口逐像素地在整个图像空间内滑动。最后,对所有窗口的 SSIM 值取平均作为整幅图像的 SSIM 值。同时,从相应的 SSIM 映射图可以直观地看出图像的结构失真部分。令 MSSIM(x, y) 表示整幅图像的平均结构相似度,那么

$$\text{MSSIM}(x, y) = \frac{1}{M} \sum_{j=1}^M \text{SSIM}(x_j, y_j) \quad (6)$$

其中: x 和 y 分别代表原始图像和失真图像; x_j 和 y_j 分别代表第 j 个滑动窗口中的原始图像块和失真图像块; M 代表所有滑动窗口的个数。不难看出,结构相似度满足以下性质:

- 对称性: $\text{SSIM}(x, y) = \text{SSIM}(y, x)$;
- 有界性: $\text{SSIM}(x, y) \leq 1$;
- 有唯一最大值:当且仅当 $x = y$ 时, $\text{SSIM}(x, y) = 1$ 。

传统的 SSIM 算法从结构失真的角度考虑失真图像的质量,虽然在一定程度上反映了人类视觉系统的真实感受,但并没有进一步考虑到具体的图像内容。例如,边缘区域必然包含了描述图像中物体结构的重要信息^[7],而且人类视觉系统对结构信息的失真也是相当敏感的,如果将它和图像中的平坦光滑区域赋予同样的权值,显然是不合理的。因此,根据实际的内容将图像划分为不同的区域,并根据人对不同区域失真的敏感程度赋予不同的权值,能更真实地反映人的视觉感受。

2 基于内容划分的图像质量评价

基于人类视觉系统对图像不同区域敏感程度不同的思想,本文在 SSIM 算法的基础上采用内容划分的方法将图像分为四部分,并赋予不同的权值,得到了基于内容划分的 CPSSIM (content-partitioned SSIM) 算法。

2.1 算法整体流程

CPSSIM 算法的整体流程如图 1 所示,共分为五个步骤:

- 对原始图像和失真图像进行平滑滤波,减小随机噪声对梯度的影响;
- 计算失真图像的 SSIM 映射图;
- 在不考虑 SSIM 算法计算结果的情况下,根据梯度将参考图像和失真图像划分成四个区域;
- 根据人类视觉系统的敏感程度对不同内容的区域赋予不同的权值;
- 对所有区域的质量值进行加权求和。

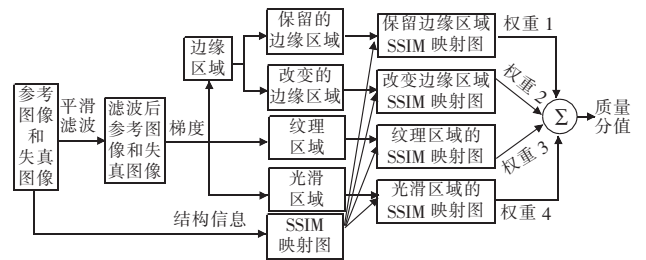


图 1 CPSSIM 算法流程

2.2 内容划分的方法

在 CPSSIM 算法中,将图像划分为四个部分:a)改变的边缘区域,即在参考(或失真)图像中存在,而在失真(或参考)图像中不存在的边缘区域;b)保留的边缘区域,即在参考和失真图像中相同的边缘区域;c)纹理区域,图像细节密集而人类视觉系统无法分辨的区域;d)光滑区域,即图像光滑,细节较少的区域。文献[8]给出了将图像划分为三个区域的方法,本文在此基础上进行了修改,将图像划分为四个部分,并根据实验结果对划分的梯度阈值进行了适当调整。具体划分步骤如下:

- 对原始图像和失真图像进行平滑滤波。图像的梯度,尤其是梯度的方向,对随机噪声是十分敏感的,因为随机噪声在图像中加入了大量的细节部分,很容易使有噪声的光滑或纹理区域被判断为边缘区域。因此,一般在计算梯度之前,都要先对图像进行平滑滤波,消除随机噪声对图像梯度计算的影响。在本算法中,使用二维高斯滤波器对图像进行平滑滤波。
- 在参考和失真图像的滑动窗口中通过 Sobel 算子计算图像的梯度。
- 确定阈值: $T_1 = a \times g_{\max}$, $T_2 = b \times g_{\max}$,其中 $g_{\max}$ 是参考图像中的最大梯度;一般取 $a = 0.1$, $b = 0.05$,并且图像中边缘区域较多时, a 的取值应当适当减小,光滑区域较多时, b 的取值应当适当增大。
- 令 $p_r(i, j)$ 表示参考图像在点 (i, j) 处的梯度, $p_d(i, j)$ 表示失真图像在点 (i, j) 处的梯度,根据如下规则确定每个像素点具体属于哪个区域:
 - 如果 $p_r(i, j) > T_1$ 且 $p_d(i, j) > T_1$,则此像素点属于保留的边缘区域;

(b) 如果 $p_r(i, j) > T_1$ 且 $p_d(i, j) \leq T_1$, 或 $p_d(i, j) > T_1$ 且 $p_r(i, j) \leq T_1$, 则此像素点属于改变的边缘区域;

(c) 如果 $p_r(i, j) < T_2$ 且 $p_d(i, j) > T_1$, 则此点属于光滑区域;

(d) 否则, 该点属于纹理区域。

2.3 各个区域的权值分配

边缘区域在人眼对图像的感知过程中起着重要的作用, 边缘的失真对导致图像质量的严重下降, 所以应给边缘区域赋予较高的权值。在本文中, 分别赋予两个边缘区域 0.3 的权值。根据人类视觉系统的失真遮蔽效应, 纹理区域的失真基本被掩盖了, 几乎无法觉察。在光滑区域中, 人类视觉系统除了对块效应和高频噪声等个别失真类型有细微觉察外, 其他类型的失真也不易被觉察, 因此本文给光滑和纹理区域赋予较小的权值。在本文中, 对纹理和光滑区域分别赋予 0.2 的权值。

3 基于运动补偿的帧加权

给定原始视频 R 和失真视频 D , 设它们的分辨率都为 $P \times Q$, 帧数都为 N 。若要评价失真视频 D 的质量, 首先采用上述基于内容划分结构相似性的图像质量评价算法评价 D 中每一帧图像的质量。对每一帧来说, 由于所使用的图像质量评价算法是在一个滑动窗口内进行计算的, 因此对每帧图像的每个滑动窗口中的图像块都能得到其相应的质量值。对于每帧图像, 只要综合考虑该帧中所有图像块的质量就能得到整帧图像的质量。

在考虑视频的运动特性对各个帧进行加权时, 采用如下的步骤:

a) 计算运动向量。采用一种作用在相邻帧之间的基于块的运动补偿算法, 第 i 帧的运动向量由第 $i-1$ 帧得到。由于失真视频的运动特性可能是失真之后产生的, 存在很大的不确定性, 所以只在原始视频上进行运动估计。在进行运动估计时, 采用文献[9]中提供的自适应方法进行快速块匹配运动估计, 获得运动向量及残差。块的大小为 $b \times b$ 。

b) 计算运动补偿块。对于原始视频中第 i 帧中的块 (m_R, n_R) , 其中 $m_R = \{1, 2, \dots, P/b\}$, $n_R = \{1, 2, \dots, Q/b\}$, 计算其对应的第 $i-1$ 帧中的运动补偿块 (m'_R, n'_R) 。所用的方法是通过将第 $i-1$ 帧中的对应块通过上述计算得到的运动向量进行运动估计后得到的预测块与第 i 帧中的对应块进行比较, 从而得到运动补偿块。同样地, 在失真视频中对应的块 (m_D, n_D) 本文采用相同的步骤计算, 从而得到运动补偿块 (m'_D, n'_D) 。需要注意的是, 在失真视频中计算运动补偿块时所使用的运动向量仍然是之前在原始视频中计算得到的运动向量, 而不是在失真视频中重新计算运动向量。

c) 对运动补偿块进行质量评价。对计算得到的运动补偿块仍采用上述基于内容划分结构相似性的图像质量评价算法进行质量评价, 从而对失真视频中的每个预测补偿块能得到表示其质量的分值。同时, 对每一帧都能得到一个质量映射图, 其大小为 $(P/b, Q/b)$ 。

d) 求整个视频的质量。YUV 格式的视频包含一个亮度分量 Y 和两个色度分量 C_r 和 C_b , 因此质量评价算法不仅要作用在 Y 分量上, 同时也要作用在 C_r 和 C_b 分量上。对于视频中的每个分量, 本文采用上述方法计算出其对应的运动向量。根据

文献[2]的描述, 给亮度分量 Y 赋予 0.8 的权值, 给色度分量 C_r 和 C_b 分别赋予 0.1 的权值。令 S_i^m 、 $S_i^{mC_r}$ 和 $S_i^{mC_b}$ 分别表示第 i 帧中第 m 个窗口的 Y 、 C_r 和 C_b 分量的空间质量 (图像质量), 令 T_i^m 、 $T_i^{mC_r}$ 和 $T_i^{mC_b}$ 分别表示第 i 帧中第 m 个窗口的 Y 、 C_r 和 C_b 分量的时间质量 (运动质量), 也即是对当前窗口赋予的相应权值。令 S_i^m 表示第 i 帧中第 m 个窗口的空间质量, 令 T_i^m 表示第 i 帧中第 m 个窗口的时间质量, 令 Q_i 表示第 i 帧在运动补偿帧加权后的质量, 令 Q_v 表示整段失真视频的质量, M 为每帧中所含的窗口数, N 为整段视频中所含帧数, 那么有

$$S_i^m = 0.8 \times S_i^{mY} + 0.1 \times S_i^{mC_r} + 0.1 \times S_i^{mC_b} \quad (7)$$

$$T_i^m = 0.8 \times T_i^{mY} + 0.1 \times T_i^{mC_r} + 0.1 \times T_i^{mC_b} \quad (8)$$

$$Q_i = \frac{1}{M} \sum_{m=1}^M S_i^m \times T_i^m \quad (9)$$

$$Q_v = \frac{1}{N} \sum_{i=1}^N Q_i \quad (10)$$

事实上, 人在观看视频时, 对背景中运动快的物体的失真并不是特别敏感。比如, 视频的主要画面是两个人在谈话, 一辆汽车在人身后的背景中快速驶过, 如果汽车有失真, 观察者也很难觉察到, 因为他们的注意力在两个谈话者身上, 而谈话者始终在画面的中心, 没有运动。如果在原始视频中通过运动向量预测出的块与失真视频中对应块的差异很大, 则说明该帧中的物体运动较为剧烈, 人眼对该帧的失真觉察程度就会很小, 因此该帧就被赋予了较小的权值, 这个权值可以用原始视频和失真视频中两个对应块通过运动向量进行运动估计后得到的两个预测块的 CPSSIM 值给出, 其取值范围为刚好为 $[0, 1]$ 。本节所述的基于运动补偿帧的加权方法也正是基于这种思想, 采用合适的运动向量计算方法, 实现了根据运动情况对帧进行加权。

4 实验与分析

实验采用美国德克萨斯大学 LIVE 实验室的视频测试序列库。序列库中包含了 10 个典型的自然场景, 每个场景用 MPEG-2 压缩、H.264 压缩、模拟 IP 传输和模拟无线传输四种视频失真方式进行失真处理, 生成了 15 个失真序列, 并进行了主观质量测试, 给出了每个失真序列的差异主观意见分值 DMOS (difference mean opinion score)。

本实验中, 采用目前被普遍认可的两个指标来度量算法的性能。一个是算法结果和 DMOS 之间的秩相关系数 SROCC (spearman rank order correlation coefficient), 用来度量算法结果和人的主观感受之间的单调性; 另一个是经过非线性回归处理后的算法结果和 DMOS 之间的线性相关系数 LCC (linear correlation coefficient), 用来度量算法的准确性。在进行非线性回归处理时, 本文采用 VQEG 的报告^[10]中提供的方程:

$$Q' = \beta_1 + \frac{\beta_2 - \beta_1}{1 + e^{-\left(\frac{Q - \beta_3}{|\beta_4|}\right)}} \quad (11)$$

其中: Q 代表算法结果, Q' 代表经过处理后的算法结果。为了获得以上方程中能使算法结果和预测值方差最小的 $(\beta_1, \beta_2, \beta_3, \beta_4)$, 本文采用 MATLAB 中的 nlinfit 函数进行非线性最小二乘法优化。为了便于数据收敛, 在进行优化之前用线性重调节的方法对算法的结果进行处理。最后, SROCC 和 LCC 在 Q' 和 DMOS 之间计算。图 2 给出了本文算法和引言中其他算法的

散点比较图,同时也画出了每个图中所有散点的最佳拟合曲线。表 1 给出了本文算法和引言中其他算法的 SROCC 和 LCC 值的比较。可以看出,除了 MSSIM 算法,与其他算法相比,本文算法在单调性和准确性上都有明显提高。MSSIM 图像质量评价算法在性能上要优于 CPSSIM 可能是其在视频质量评价方面的性能优于本文算法的原因,这需要进行进一步的实验分析和验证。

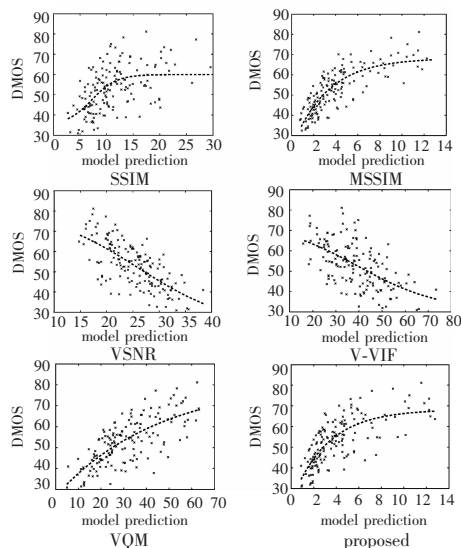


图 2 引言中各算法与本文算法的散点比较图
(图中曲线代表散点的最佳拟合曲线)

表 1 引言中各算法与本文算法的 SROCC 和 LCC 比较

算法	SROCC	LCC
SSIM	0.525 7	0.544 4
MSSIM	0.736 1	0.744 1
VSNR	0.675 5	0.689 6
V-VIF	0.571 0	0.575 6
VQM	0.702 6	0.723 6
Proposed	0.723 1	0.734 3

另外,关于不同区域权值的选取和块大小的选取有以下三点需要进一步说明:

a) 划分区域后不同区域权值的选取。在求最佳权值时,将改变的边缘区域和保留的边缘区域都作为边缘区域并赋予相同的权值,将光滑区域和纹理区域都作为纹理区域并赋予相同的权值,并在 LIVE 图像测试库上作出 CPSSIM 算法对库中所有图像的 SROCC 和边缘区域权重之间的散点图,如图 3 所示。可以看出,在边缘区域的权重取值达到 0.6 时,整个算法的 SROCC 值最高,算法性能最好。根据这一点,分别赋予两个边缘区域 0.3 的权值,同时对光滑区域和纹理区域赋予 0.2 的权值。但是,图像质量评价中的最佳权值是否是视频质量评价中的最佳权值,这个问题还需要更深入地研究。

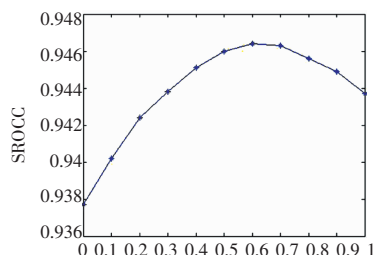


图 3 本文算法 SROCC 和边缘区域权值间的散点图

b) 视频 Y 、 C_r 和 C_b 分量权值的选取。本文没有用实验来选取合适的 Y 、 C_r 和 C_b 分量权值。根据文献[2]中的描述,

0.8、0.1 和 0.1 的权值设置取得了较好的效果。因此,在进行仿真实验时借鉴了上述权值,且最后的实验结果也说明这种借鉴使算法在性能上取得了较好的效果,具有一定的合理性。

c) 进行运动估计时的块大小的选取。在选取块大小时,本文选取了 $b=4$ 、 $b=8$ 和 $b=16$ 三个值,分别进行了视频质量评价仿真实验并分析其评价准确性。笔者发现,块的大小对算法的评价准确性几乎没有影响,用不同的块大小进行运动估计得到的算法性能几乎一样。所以,在本实验中折中地选择了 8×8 的块大小。

5 结束语

本文在图像结构相似性的基础上,考虑到图像中不同区域有着不同的感知重要性,通过对梯度进行分析将图像划分为不同的区域,得到了基于内容划分的图像质量评价算法,并通过运动补偿帧加权的方式将其扩展到了视频质量评价中。在今后的工作中,还有许多方面可以进一步提升算法的准确性。如果改变区域划分的方式,对图像中的不同区域进行更细致地划分,相信算法的效果会更加理想。另外,本文中的权值主要是根据实验结果不断调整得到的,并没有经过专业的优化处理。如果能在充分考虑人类视觉特性的基础上,选取先进的优化算法对不同区域的权值进行寻优,相信算法的性能会进一步提升。

参考文献:

- [1] ITU-R Rec. BT500-11, Methodology for the subjective assessment of the quality television picture[S]. [S. l.]: The International Telecommunication Union, 2002.
- [2] WANG Zhou, LU Li-gang, BOVIK A C. Video quality assessment based on structural distortion measurement[J]. *Signal Processing: Image Communication*, 2004, 19(2): 121-132.
- [3] WANG Zhou, SIMONCELLI E, BOVIK A C, et al. Multiscale structural similarity for image quality assessment[C]//Proc of IEEE Asilomar Conference on Signals, System and Computers. 2003: 139-140.
- [4] CHANDLER D M, HEMAMI S S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images[J]. *IEEE Trans on Image Process*, 2007, 16(9): 2284-2298.
- [5] PINSON M H, WOLF S. A new standardized method for objectively measuring video quality[J]. *IEEE Trans on Broadcast*, 2004, 50(3): 312-322.
- [6] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. *IEEE Trans on Image Processing*, 2006, 15(2): 430-444.
- [7] MEUR O L, NINASSI A. Overt visual attention for free-viewing and quality assessment tasks[J]. *Signal Processing: Image Communication*, 2010, 25(7): 547-558.
- [8] LI J L, CHEN G, CHI Z R. Image coding quality assessment using fuzzy integrals with a three-component image model[J]. *IEEE Trans on Fuzzy Systems*, 2004, 12(1): 99-106.
- [9] NIE Yao, MA Kai-kuang. Adaptive rood pattern search for fast block-matching motion estimation[J]. *IEEE Trans on Image Processing*, 2002, 11(12): 1442-1449.
- [10] Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment[EB/OL]. http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI.