

ICS 33.160
M 63

YD

中华人民共和国通信行业标准

YD/T 2647-2013

IP 网络高清视频客观全参考质量评价方法

Objective perceptual HD video quality measurement
for IP video in the presence of a full reference

2013-10-17 发布

2014-01-01 实施

中华人民共和国工业和信息化部 发布

目 次

前 言	II
1 范围	1
2 规范性引用文件	1
3 术语、定义和缩略语	1
3.1 术语和定义	2
3.2 缩略语	3
4 应用场景	3
5 全参考方法描述	3
6 模型描述	4
6.1 概述	4
6.2 预处理过程	5
6.3 时间校准	6
6.4 空间信息校准	7
6.5 局部相似度和局部差异特征参数的计算	7
6.6 局部特征的分布分析	8
6.7 块效应参数	9
6.8 抖动特征参数的计算	9
6.9 MOS 值的映射	10
6.10 处理空间校正严重错的视频序列	12
6.11 源视频代码的执行	13
附录 A (资料性附录) 视频质量专家组的结论	14

前 言

本标准用翻译法修改采用ITU-T J.341:2011《Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference》。

本标准做了下列编辑性修改：

- 本标准将 ITU-T J.341:2011 中前言部分编辑到范围中，并按 GB/T1.1-2009 补充前言；
- 本标准范围简化 ITU-T J.341:2011 中范围部分；
- 本标准第 4 章修改 ITU-T J.341:2011 中范围部分；
- 本标准附录 A 修改 ITU-T J.341:2011 中第 7 章部分；
- 本标准第 6 章为 ITU-T J.341:2011 中规范性附录 A 的内容；
- 将 ITU-T J.341:2011 所有章节中的程序改为链接引用。

本标准按照 GB/T1.1-2009 给出的规则起草。

本标准由中国通信标准化协会提出并归口。

本标准起草单位：工业和信息化部电信研究院、华为技术有限公司、中兴通讯股份有限公司、中国移动通信集团公司。

本标准主要起草人：栗 蔚、罗 忠、聂秀英、杨黎波、黎家力。

IP 网络高清视频客观全参考质量评价方法

1 范围

本标准提出了一种非交互的IP网络高清视频感官质量全参考(FR)测量的方法,规定了有源信号可以参考的情况下,对IP网络高清视频的感官质量进行客观评价的方法。

本标准适用于但不限于以下场景:

- 源端的实时在线质量监测;
- 远程接收端,源端视频拷贝可用于质量测量的情况下,对此点的质量监测;
- 一个或多个利用视频解压缩技术的存储或传输系统的质量测量;
- 实验室视频系统的测试。

本标准不适用于视频会议的场景。

注1:此标准的视频质量测量模型不能替代主观测试。两个不同的实验室执行的主观测试相关度一般在 0.95 到 0.98 之间。如果用这个客观标准比较两个视频系统(比如两个编码器),建议用同一量化的方法(比如 ITU-T J.149)来确定模型的准确率。

注2:当出现视频停顿,测试条件允许视频停顿不超过 2s。此标准的模型不适用于测量重新缓冲的视频(这样会增加延迟或者停顿)。模型不用于其他帧率视频,只用于测试 TV 的帧率(一般 TV 的帧率为 29.97fps 和 25fps,逐行或隔行扫描)。

注3:需要注意的是对于新的编码和传输技术,会产生此评价方法中没有考虑到的新的损伤,客观模型可能会出现错误的结果。这时需要主观评价方法。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅所注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

ITU-T J.144: 2004 数字电视全参考客观质量评价(Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference)

ITU-T J.149:2004 视频质量度量准则统一准确率和方法(Method for specifying accuracy and cross-calibration of Video Quality Metrics)

ITU-T P.910:2008 多媒体应用主管视频质量评价方法(Subjective video quality assessment methods for multimedia applications)

3 术语、定义和缩略语

3.1 术语和定义

下列术语和定义适用于本文件。

3.1.1

主观评价(图像) Subjective Assessment (Picture)

一组评论人员在各自观看场景中对事先制作的图像的质量或损伤程度的意见。

[ITU-T J.144]

3.1.2

客观感官评价（图像） Objective Perceptual Measurement (Picture)

对事先制作的图像的客观（仪器）方法的评价，其测量性能指标是要接近主观评价的评分。

[ITU-T J.144]

3.1.3

提交评价模型者 Proponent

在ITU标准中，提交视频质量评价模型的一个组织或公司。

3.1.4

帧率 Frame Rate

每秒播放的帧数。

3.1.5

仿真传输损伤 Simulated Transmission Errors

在一个高可控环境中对数字视频比特流中增加错误。比如仿真丢包和误码。仿真传输损伤的参数已经有很好的定义。

3.1.6

传输损伤 Transmission Errors

视频传输中的各种损伤。损伤种类包括了实时网络状态的和仿真的传输损伤。

3.2 缩略语

下列缩略语适用于本文件。

ACR	Absolute Category Rating	绝对种类定级
ACR-HR	Absolute Category Rating with Hidden Reference	带隐藏参考信息的绝对种类定级
AVI	Audio Video Interleave	音频视频交织
DMOS	Difference Mean Opinion Score	平均主观评分之差
FR	Full Reference	全参考
FRTV	Full Reference Television	全参考电视评价
HRC	Hypothetical Reference Circuit	假设参考电路
ILG	VQEG's Independent Laboratory Group	视频质量专家组的独立实验室组织
MOS	Mean Opinion Score	平均主观评分
MOSp	Mean Opinion Score, predicted	预测的平均主观评分
NR	No (or zero) Reference	无参考
PSNR	Peak Signal-to-Noise Ratio	峰值信噪比
PVS	Processed Video Sequence	处理过的视频序列
RMSE	Root Mean Square Error	平均根方差
RR	Reduced Reference	部分参考
SFR	Source Frame Rate	源帧率

SRC	Source Reference Channel or Circuit	源参考信道或电路
VQEG	Video Quality Experts Group	视频质量专家组

4 应用场景

全参考测量方法使用时需要无损的原始视频信号在测量点可用，所以这种测量方法可以用于实验室或一个封闭环境（例如cable电视的发送和接收端）的单个或系列设备系统。预测方法的步骤包括矫正和客观视频质量预测。

测试的视频序列包含了H.264和MPEG-2两种编码损伤，以及传输误差（比如，误码，丢包等）。此标准的模型可以监控部署网络的质量以保证其战备完好性。视觉感官损伤包括空域和时域的损伤。此标准的模型也可以用于视频系统的实验室测试。当用于比较不同的视频系统时，建议用同一种量化的方法（比如ITU-T J.149）来确定模型的准确率。

标准适用于码率为1Mbit/s到30Mbit/s的通信业务。测试包括以下分辨率：

- 1080i 60 Hz (29.97 fps)
- 1080p (25 fps)
- 1080i 50 Hz (25 fps)
- 1080p (29.97 fps)

测试条件如下：

测试条件
视频分辨率：1920x1080逐行扫描和隔行扫描
视频帧率：29.97 fps和 25 fps
视频码率：1~30 Mbit/s
时域停顿（暂停或跳过）最大2秒
有丢包的传输损伤
原始视频SRC由1080p转换为720p，压缩，传输，解压，然后再转回1080p
编码技术
ITU-T H.264/AVC (MPEG-4 Part 10)
MPEG-2

注：720p的视频作为测试方案中测试序列HRC的一部分。因为目前720p一般可以放大作为显示的一部分，所以720pHRC能更好的匹配这种格式。

5 全参考方法描述

全参考的评价方法，是视频感官质量评价方法的一种，通过比较系统输入端的源无损视频信号和系统输出端的损伤信号评价系统的性能。评价系统性能的示例如图1所示。

输入信号和输出信号的比较，事先可能需要对图像时域或空域信息校正，然后再对图像水平或竖直方向的平移或切割做出补偿。有时也可能需要对亮度和色度信号的偏移或增益进行校正。然后用基于人类视觉的感官模型对图像的质量进行客观评分。

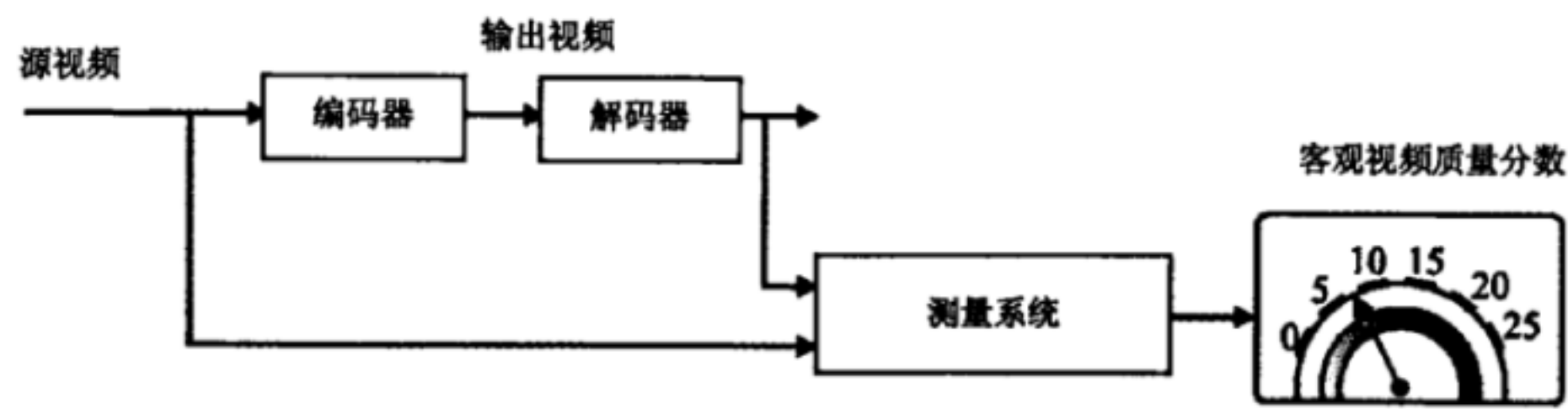


图1 实验室中全参考方法测试编码器性能的示例

校正和增益的调整被称为图像配准。这个过程是必需的，因为全参考方法，要对原始图像和损伤图像一个像素一个像素的比较。本标准中的视频质量评价模型包括了图像配准方法。

因为视频质量客观评价方法是要尽量逼近人类视觉感官的反应，而不是单纯计算编码的量化损伤，所以此类方法同时适用于数字和模拟的视频系统。此类方法可以应用于模拟和数字的混合系统，或者多个视频压缩系统的。

全参考方法测试传输信道应用的示例如图2所示。

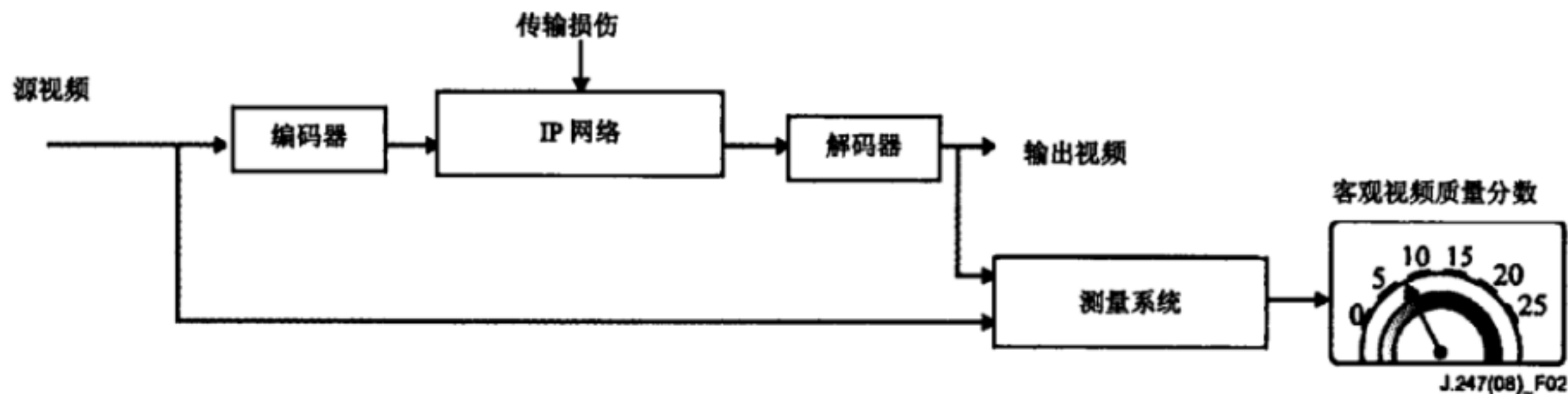


图2 全参考方法测试传输信道应用的示例

所以，解码器可以放在传输系统的不同的测试点，比如，解码器可以放在网络中的一个点，如图2所示，或者直接放在编码器的输出端，如图1所示。如果数字传输系统是透明的无损的，那么对于源端图像的客观评价和对之后系统任何一点图像的评价效果一致。

之前的研究显示，全参考方法能够获得与图像主观评价较高的相关度。此标准中的全参考方法与主观评价有较高的相关度，其中主观评价参照ACR-HR方法获得[b-ITU-T J.910]。本标准的全参考模型性能结果参见附录A。

6 模型描述

6.1 概述

此模型的方法可以估计主观评价视频质量的结果。此客观评价方法利用视觉心理和自我学习的模型来达到仿真主观感受的效果。因为是全参考方法，所以此方法的模型比较输入的高质量源视频和输出的损伤视频序列，损伤视频序列通过实验处理获得，其过程参如图3所示。

模型估计视频质量通过以下步骤：

- 第一步，视频序列预处理。包括降噪和像素的下采样。
- 第二步，源视频和处理后视频序列的时域校正。
- 第三步，源视频和处理后视频序列的空域校正。
- 第四步，局部空域特征参数计算：利用视觉效应模型计算局部相似度和局部差异。

- 第五步，局部相似度和差异度参数的分布分析。
- 第六步，计算图像全局的块效应参数。
- 第七步，图像全局时域损伤特征抖动参数的计算。此参数由计算局部和全局运动的剧烈程度，以及帧的显示时间得出。

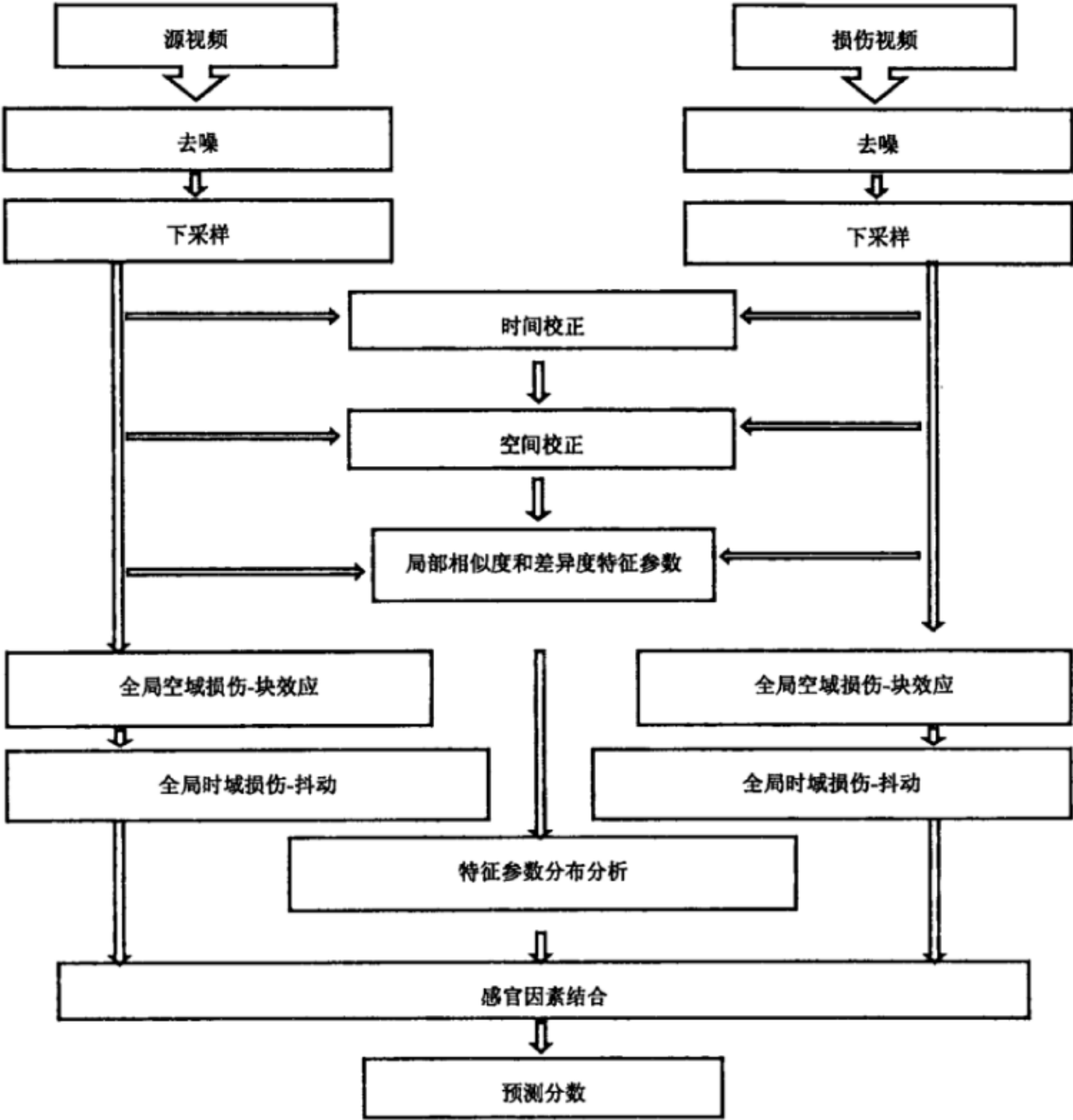


图3 模型评价过程的流程图

基于以上的特征参数，质量分数由非线性的函数得出。

为避免由于源视频和损伤视频序列的错误校正而获得错误的预测分数，以上步骤计算3个不同的视频序列竖直和水平方向的校正结果，最大的预测分数为最后的质量分数。

注：顶端输入的分别是源视频序列和损伤视频序列。模型通过一系列不同的处理过程输出质量分数，是流程图的最后一个图框。

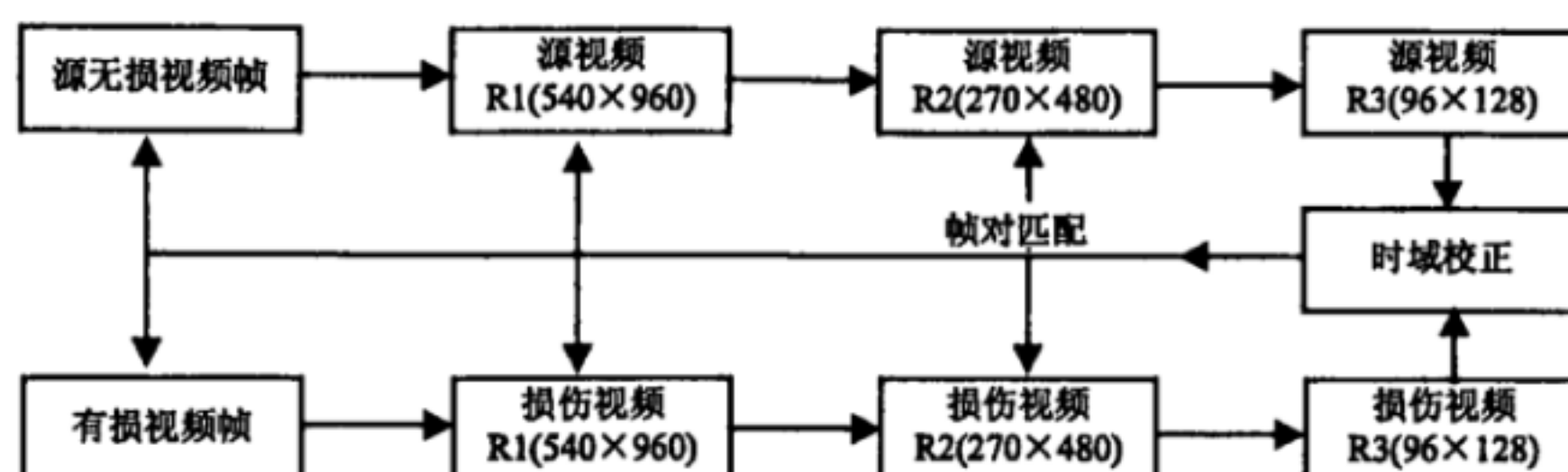
6.2 预处理过程

源视频和损伤视频的每一帧图像通过低通滤波器和下采样到以下R1、R2和R3等3个不同分辨率的图像。

	original frame		R1		R2		R3
height×width	1080×1920	→	540×960	→	270×480	→	96×128

R3图像的获得参见6.11, 程序方法: CFrameAnalysisFullRef::ContentTimeAlignment。图像R1和R2的获得参见方法: CFrameSeq::ReadFrame。

图像下采样的过程如图4所示。



注: 源视频和损伤视频的每一帧图像通过低通滤波器和下采样到以下3个不同分辨率的图像。R3用于帧的时间校准。匹配图像的结果列表可以匹配其他任何一个分辨率的帧。

图4 图像下采样的过程

6.3 时间校准

时间校准用于源视频和损伤视频的最低分辨率的子图像R3。

校准过程采用递归的方式, 如下:

- a) 在源视频中确定一个帧, 记为“anchor”(Ref_anchor);
- b) 在损伤视频中找到与“anchor”帧最匹配的一帧 (Deg_best_match)。

把损伤视频中的“Deg_best_match”帧放到源视频中, 与源视频中“anchor”帧附近的帧进行匹配。根据相似度, 在Ref_anchor周围试图找到一个与Deg_best_match更匹配的帧, 存为最匹配的帧对。相似度计算公式如下, 损伤视频帧为x, 源视频帧为y:

$$\text{sim} = \exp(-\text{mean_square_diff}(a \times x + b, y)) \quad (1)$$

参数a和b使均方差最小。参见6.11, 程序FrameSimilarity::similarity。

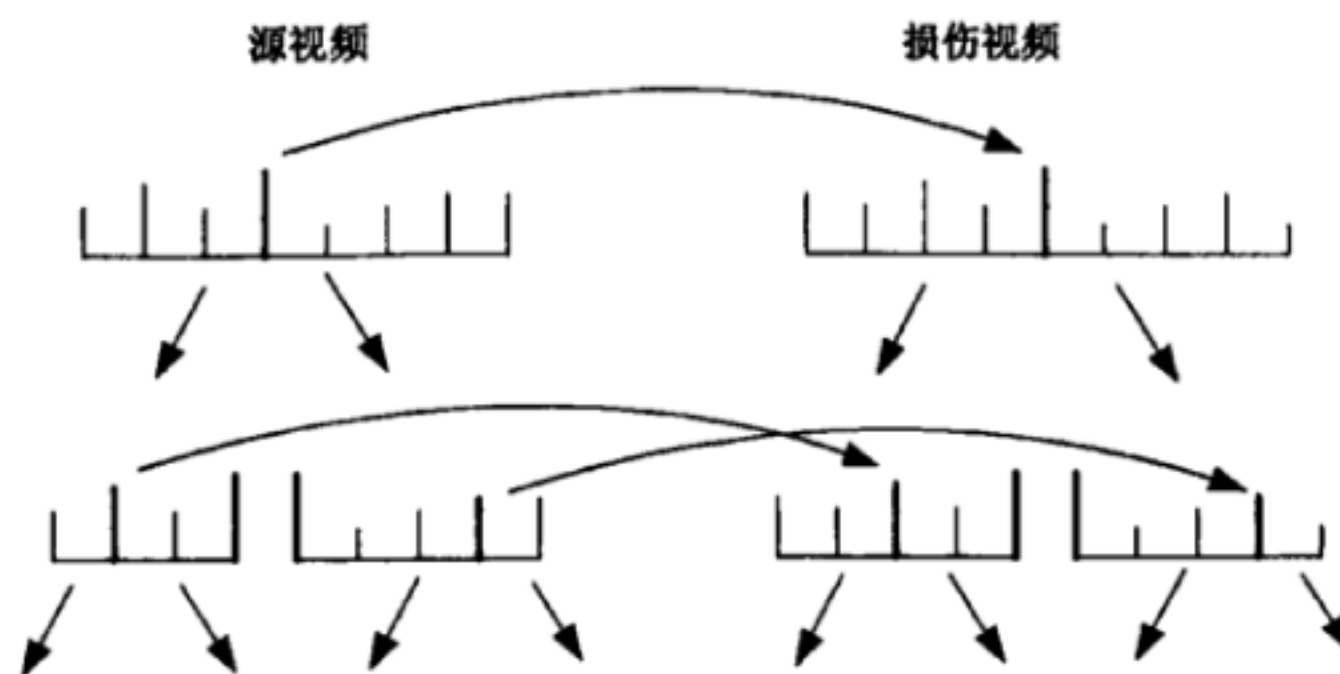
a) 如果以上找到的最匹配帧对符合要求, 即相似度大于一定的阈值, 那么分别将源视频和损伤视频中对应的最匹配帧的左右两侧分成两部分视频子序列。然后对于这两部分的子序列, 继续从第 a) 步的方法开始处理。

b) 如果最匹配帧不令人满意, 即相似度小于一定的阈值。那么从第 a) 步重新开始, 选择一个不同的“anchor”帧。最匹配帧对的相似度的阈值没有一个先验的值, 一般随着递归的次数, 阈值是逐渐降低的。根据大量的数据实验, 我们得出以下阈值: 第一次确定最匹配帧对的公式 (1) 阈值设为 0.98。10 个“anchor”帧都失败后, 阈值比 0.98 降低, 按依次类推, 10 次“anchor”帧再失败后, 阈值再次降低。这个过程一直到阈值最小值 0.1。细节参见 6.11, 程序 SQ_TimeAlignement::findAnchorAndDescend。

采用递归方法处理时间校准如图5所示。

时间校准的结果是一个匹配序列, 列出了损伤视频中的每一帧在源视频中最匹配的帧对。而对于损伤视频中某些帧, 按时间校准没有在源视频中找到最匹配帧, 在后续处理中, 这些帧与源视频中的两个帧进行比较, 分别是与此帧前后两帧最匹配的帧。需要注意的是, 之前时间校准中的相似度阈值已经设得非常低了, 所以只有损伤非常严重的视频序列才会存在找不到匹配的帧。

具体方法参见6.11, 程序: CFrameAnalysisFullRef::sqVTA_ContentFrameTimeAlignement_M。



注：如图，源视频中的“anchor”帧与损伤视频中的最匹配帧。然后这两帧左右两侧再分成两个视频子序列，源视频每个视频子序列再选择一个“anchor”帧，然后在对应的损伤视频子序列中寻找最匹配帧。

图5 递归方法处理时间校准

6.4 空间信息校准

空间信息校准的方法是对损伤视频的所有帧采用递归的方法，具体操作如下：

a) 如果此帧没有匹配帧，用前一帧的空间校准。如果此帧有匹配帧，那么根据时间校准的结果，在此帧和源视频中对应的匹配帧之间进行空间校准。

1) 对于第一帧，初始空间偏移量为 0（水平和竖直方向皆是）。对于后续的帧，使用事先以前的匹配帧的空间对齐。

2) 在第 b) 步规定的范围内，竖直方向和水平方向寻找可能的空间偏移量。根据 cost 函数，在损伤帧和源帧中找到一个使 cost 函数最小值的空间偏移量。Cost 函数如下：

$$\text{rmse}(Y(dv, dh), Y_{\text{ref}}) + \text{abs}(dv) + \text{abs}(dh)$$

Y 是 $R1$ 分辨率子图像的损伤帧， Y_{ref} 是 $R1$ 的源视频的匹配帧， $Y(dv, dh)$ 是 Y 的偏移量， dv 是水平偏移量， dh 是竖直偏移量。对于较小空间偏移量， $\text{abs}(dv)$ 和 $\text{abs}(dh)$ 项是有必要在 cost 函数里的。需要注意的是如果一帧的边界非常小，那么可以跳过 RMSE 计算，以避免更复杂的边界处理。

3) 这样，各个时间点的空间偏移量能够得到补偿。一帧中的错误校准可以在下一帧的校准中得到修正。

b) 第一步中空间偏移校准的区间为 ± 4 个像素。对于更大的空间偏移，参见 6.9。

c) 空间校准之后，根据时间校准和空间偏移量的校准，损伤视频中的每一帧都有一个对应的源视频帧。所以，损伤视频能准确的和源视频的帧进行比较。这是接下来的特征参数提取的基础。

所有此步骤的细节参见 6.11，程序 CFrameAnalysisFullRef::DetermineSpatialAlignment。第 b) 步中的 ± 4 个像素值可以根据更大的空间偏移需求而增大。

6.5 局部相似度和局部差异特征参数的计算

每一对已经校准过的视频匹配对，一组像素间质量特征参数通过以下步骤计算：

第一步，计算局部相似度和差异度。此特征用 $R2$ 子图像进行计算， $R2$ 的损伤视频帧和源视频帧中，以 13×13 像素块为单位计算。因为 $R2$ 所有的像素个数不能被 13 整除，所以有小部分边界像素忽略。

局部的 13×13 像素块被记做 s_{prc} 和 s_{ref} ，相似度 S 和差异度 D 计算公式为：

$$S = (\text{cor}(s_{\text{prc}}, s_{\text{ref}}) + 25) / (\text{var}(s_{\text{ref}}) + 25) \quad (2)$$

$$D = \text{sqrt}(\text{avg}((S \times (s_{\text{prc}} - \text{mean}(s_{\text{prc}})) - (s_{\text{ref}} - \text{mean}(s_{\text{ref}})))^2)) \quad (3)$$

其中 cor 是相关度, var 是两个像素块对应像素之间的差值, 函数 avg 计算像素块区域所有像素的平均值, $sqrt$ 是均方差。相似度 S 和差异度 D 是空间特征的主要参数。

这样, 每帧的相似度 S 和差异度 D 是一个矩阵, 矩阵的每一个数值对应相应的像素块。对于感官质量来说, 不仅相似度 S 和差异度 D 的均值很重要, 相似度 S 和差异度 D 的分布特征也很重要。

6.6 局部特征的分布分析

首先介绍几个公式。分位数 $quantile(X, c)$ 表示, 一个矩阵 X 所服从的概率分布的分位点为 c , 对一个矩阵 X 和一个常数 c , $0 \leq c \leq 1$, 它们的分位数记为 $q = quantile(X, c)$, 以 c 点为分界, X 的所有值小于或等于 q 。

函数 $trimmed_mean$ 定义如下: $trimmedMean(X, c)$ 表示 X 概率分布分位点 c 和 $(1-c)$ 之间的所有项的平均值。比如, $trimmedMean(X, 0.1)$ 是去掉 X 的10%最小的数和10%最大的数后, 剩下的 X 的所有项的平均值。

$X(X > c)$ 表示 X 中所有大于 c 的值。比如 $trimmedMean(X, c) = mean(X(X > quantile(X, c) \text{ and } X < quantile(X, 1-c)))$ 。

利用这些公式, 以下特征参数的值基于6.5节中的 S 和 D :

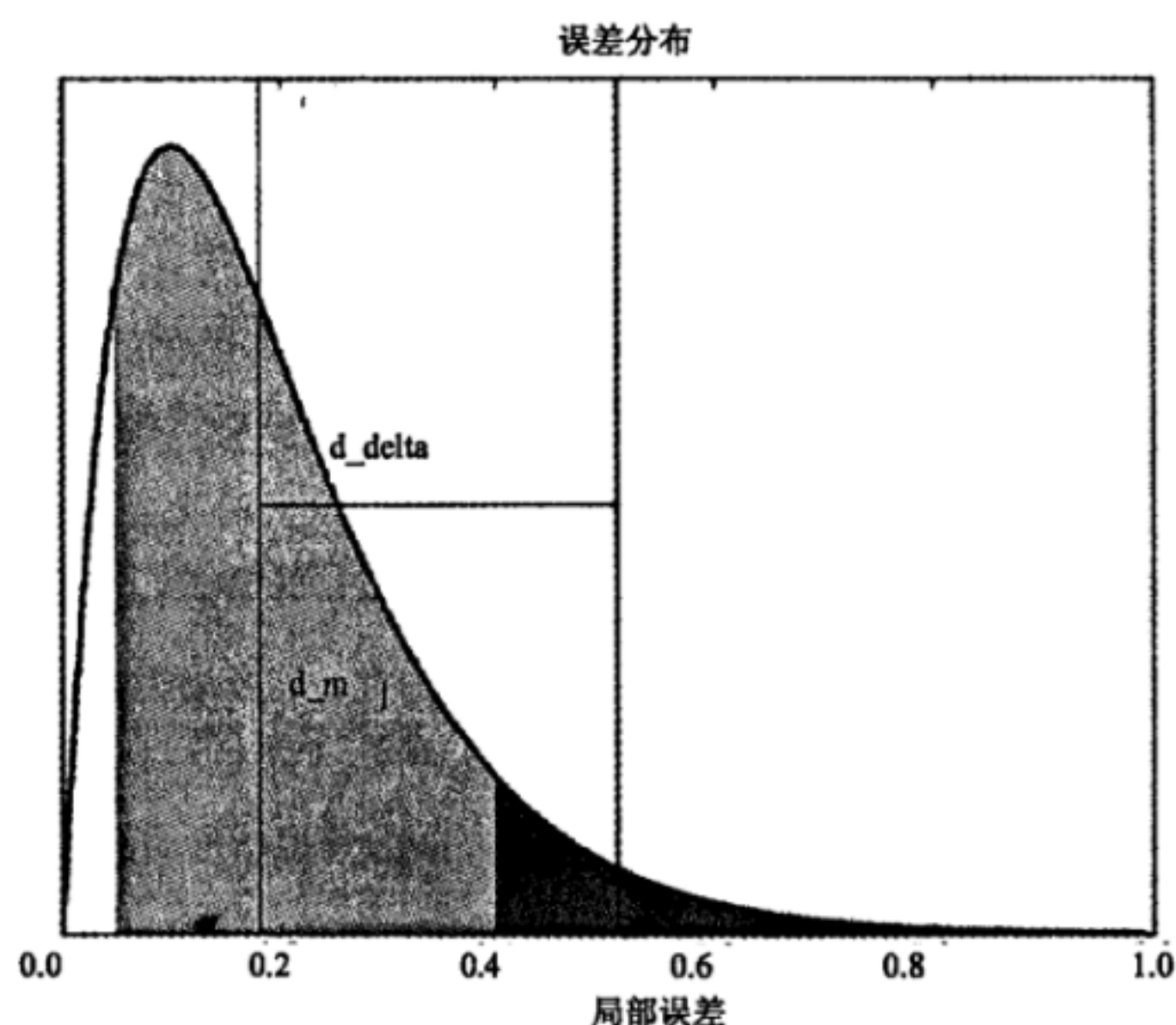
$$s_m = trimmedMean(S, c) \quad (4)$$

$$d_m = trimmedMean(D, c) \quad (5)$$

$$s_delta = s_m - mean(S(S < quantile(S, c))) \quad (6)$$

$$d_delta = mean(D(D > quantile(D, 1-c))) - d_m \quad (7)$$

其中 $c=0.2$, d_m 和 d_delta 如图6所示。



注1: $trimmedMean\ d_m$ 是图中浅灰色区域的平均值(黑色竖直线)。

注2: d_delta 的值是深色灰色区域平均值和浅色灰色区域平均值的差(水平方向的黑色线)。

图6 局部差异度 D 的概率分布

相似度 S 和差异度 D 的具体计算方法参见6.11, 程序`CFrameAnalysisFullRef::ComputeSimilarity`。

6.7 块效应参数

用分辨率为R1子图像计算块效应参数。这个特征参数是度量由于编码器或传输错误导致的块边界的可视损伤。主要步骤如下：

对图像用 sobel 等边缘算子，获得竖直方向和水平方向的边缘图像。视频序列的每一帧，结果是两个矩阵，一个是水平方向边缘点组成的图像，一个是竖直方向边缘点组成的图像，在下面的程序中分别记做 verGrad_n 和 horGrad_n。

分别计算水平方向和竖直方向边缘点对数的和，结果为两个向量，一个是水平方向边缘点的和，另一个是竖直方向边缘点的和，在下面记做 sumW 和 sumH。

sumW 和 sumH，在块大小为 n ，不同偏移量 m 的条件下，计算其平均值，下面称为函数 vq_AvgSubsample。

比如，源视频帧中为 4×4 大小的块，R1中对应的块就是 2×2 ，如果块效应比较严重，所以计算 vq_AvgSubsample($x, 2, 0$) 和 vq_AvgSubsample($x, 2, 1$) 结果会有很大不同。为避免过分基于内容，实验用大量的视频序列，计算损伤视频和源视频的差值。

更多的计算步骤在以下程序中显示。horGrad和verGrad是一帧竖直方向和水平方向的导数。

verGrad_n(i, j) = $Y_n(i+1, j) - Y_n(i, j)$, 和

horGrad_n(i, j) = $Y_n(i, j+1) - Y_n(i, j)$,

其中 $Y_n(i, j)$ 是第 n 帧坐标为(i, j)的像素值。

vq_AvgSubsample($x, \text{step}, \text{offset}$) 函数计算向量 x 在块大小为 step 条件下，所有偏移量 offset 的像素平均值。

详细内容参见6.11，程序vquad_hd::vq_BlockinessPhaseDiff 和CQualityModelFullRef::Blockiness。

6.8 抖动特征参数的计算

抖动参数是根据相关显示时间和运动剧烈度的非线性映射函数。运动剧烈程度主要是根据相邻两帧之间特定区域的差，显示时间就是一帧在屏幕上显示的时长，以毫秒为单位。注意应该确定每帧的显示时长再计算运动剧烈程度。因为损伤视频有时一帧是前一帧的重复。

抖动特征还考虑到了损伤视频播放时丢失的一些信息。如果播放比较流畅，那么抖动特征参数就低，如果出现停顿或者在帧率变低，那么抖动就大。另一方面，对于一个固定的时域损伤，抖动度量需要较大运动剧烈度的视频序列的较大值。

以下程序显示细节过程。其中运动剧烈程度的输入向量motionInt，帧重复的概率向量repFrame，帧的显示时间向量displayTime。输出为抖动向量，jerkiness是损伤视频序列每帧的值。motionInt是使用R2子图像进行计算，相邻两帧之差的均方根。repFrame是帧重复的概率，取决于运动的剧烈程度，每一帧都有可能是前一帧的重复，这样为了完美的重复前一帧，实际上每一帧是前一帧的重复的概率为1。如果运动剧烈程度较大，那么每一帧是前一帧重复的概率是0。如果运动剧烈程度非常小，但是非零值，那么帧重复的概率是0到1之间的中间值，具体如下：

$$\text{repFrame}(i) = \begin{cases} 0 & \text{if } m(i) < p/2 \\ m(i) - p/2/p & \text{if } p/2 \geq m(i) < 3/2 \times p \\ 1 & \text{if } 3/2 \times p \geq m(i) \end{cases} \quad (4)$$

其中 $m(i)$ 是第 i 帧的运动剧烈程度，根据经验， $p=0.01$ 。

详细过程参见 6.11，程序 `vquad_hd::vq_ProbOfRepeatedFrame` 和 `vquad_hd::vq_CalcJerkiness`。

6.9 MOS 值的映射

以上特征参数：相似度 s_m 和 s_{Δ} ，差异度 d_m 和 d_{Δ} ，块效应 `blockiness` 和抖动 `jerkiness`，结合帧显示时长向量，是分数评估的基础。这些特征参数都是向量，向量的长度就是损伤视频序列帧的长度。

将这些特征参数映射到感官度量，采用一个 S 型的函数：

$$S: \mathbb{R} \rightarrow \mathbb{R}, y = S(x).$$

函数 S 有三个系数 (p_x, p_y, q) 。其中 (p_x, p_y) 是 $\mathbb{R} \times \mathbb{R}$ 的坐标， q 是函数曲线在拐点的斜率，具体如下：

$$S(x) = \begin{cases} a \times x^b & \text{if } x \leq p_x \\ d / (1 + \exp(-c \times (x - p_x))) + 1 - d & \text{else} \end{cases} \quad (5)$$

其中

$$a = p_y / p_x^{(q \times p_x / p_y)}$$

$$b = q \times p_x / p_y$$

$$c = 4 \times q / d$$

$$d = 2 \times (1 - p_y)$$

如图 7 所示，不同系数下， S 函数示例。 S 函数从原点出发，非线性增长至拐点，然后趋近于 1。

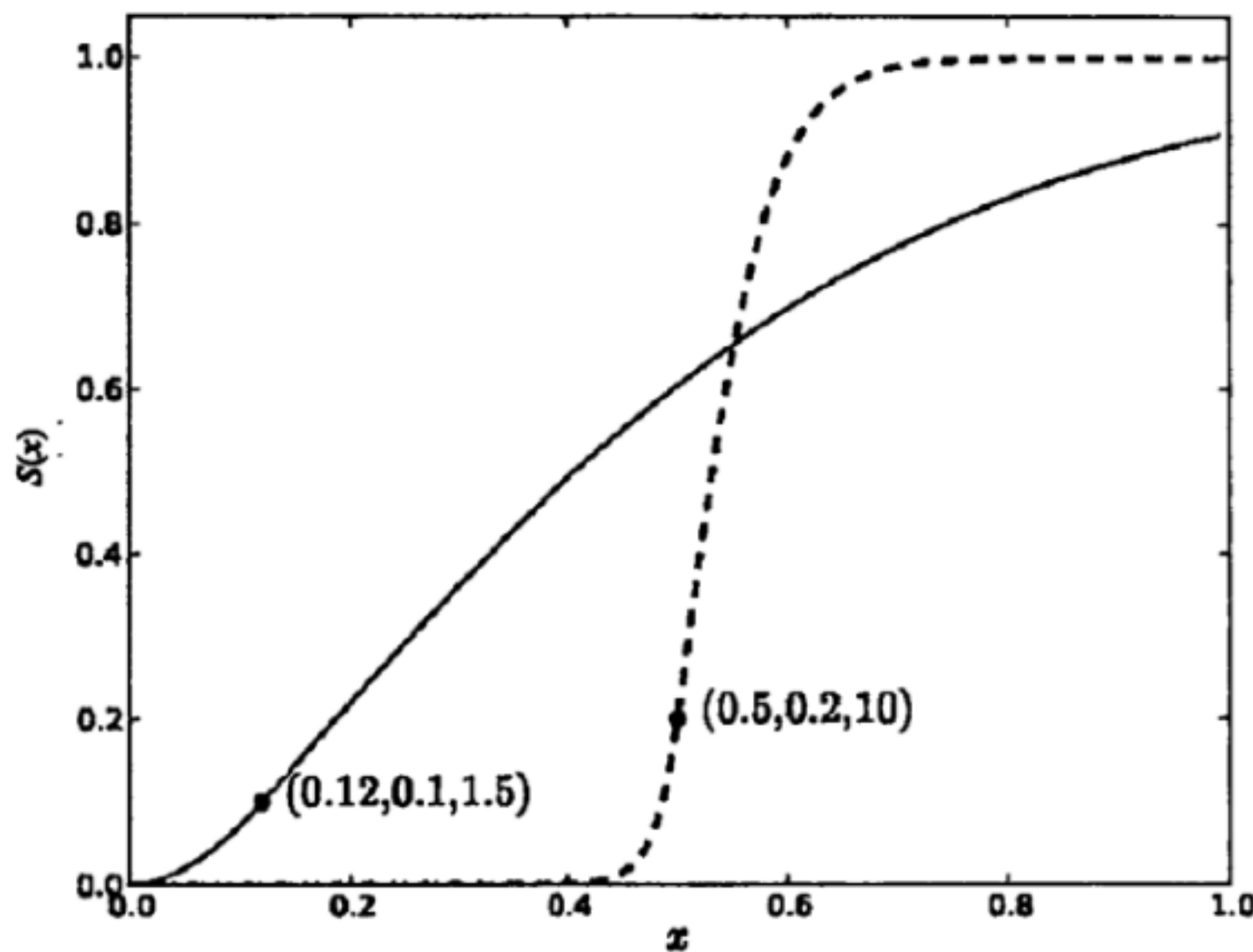


图7 不同系数下， S 函数示例。两个不同坐标和拐点斜率的 S 函数

在常量系数下， S 函数，将一组特征参数值映射到感官度量。用 S 函数，最好把系数压缩到较小值，并且拉伸系数到较大区间，即归一化，这样可以避免短暂损伤特征参数对整体的影响，即避免极值的影响。如图 8 所示。

在图 8 中，图 (a) 显示 S 函数在两组不同的系数组合下的示例；图 (b) 示例一个信号向量，将里面的 x 值都转换后，再用图 (a) 的 S 函数进行映射，结果分别为图 (c) 和图 (d)；图 (c) 所示是在转换后的系数下， S 函数将一个特征 x 映射到感官度量；图 (d) 是用图 (d) 中虚线表示的 S 函数，此函数更有利于发现瞬间损伤。

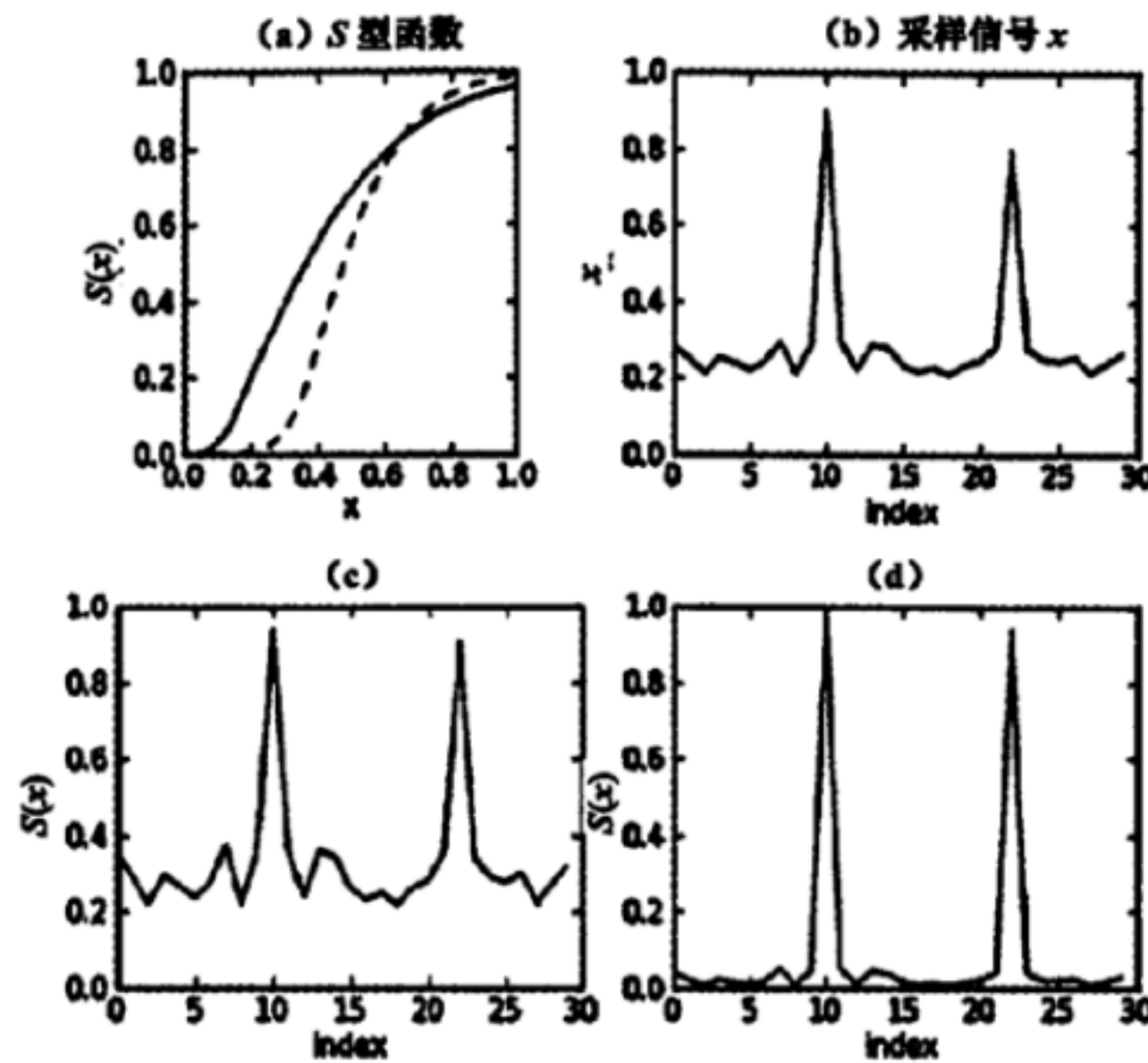


图8 基于 S 函数的数据

继续评分模型的介绍，首先，定义基于相似度的损伤，如下，对于损伤较大的给予较大的权重：

$$d_s = 1 - s_m + 1.5 s_{\Delta}$$

接下来，用两个S函数，一个是用一组固定的系数cod_par，将d_s映射到感官度量，反映编码器对视频的损伤。

第二个S函数，用转换的系数，将d_s映射到感官度量，反映传输错误对视频的损伤。

详细程序参见 6.11 程序，主函数为 SplitCodTrans，输入为向量 d_s 和帧显示时间向量 disp_time，输出为 d_cod 和 d_trans：

```
SplitCodTrans(d_s, disp_time, d_cod, d_trans)
```

全部过程细节，参见6.11，程序CQualityModelFullRef::SplitCodTrans。

其中d_diff_cod和 d_diff_trans基于d_m和d_delta，：

$$d_{\text{diff}} = d_m + 1.5 d_{\Delta}$$

函数SplitCodTrans(d_diff, disp_time, d_diff_cod, d_diff_trans),用3个系数和两个S函数：

$$\text{cod_par} = (4.0f, 0.05f, 0.2f)$$

$$\text{trans_par} = (0.5 \times (q + 4.0f), 0.1f, 0.4f)$$

其中 根据之前的函数，q是中间位数的平均值。输出为d_diff_cod和d_diff_trans。

下一步，用S函数计算和瞬时较大抖动有关的一个特征参数：

$$d_{t_trans} = S(\max(0, \text{jerkiness} - q))$$

其中系数为 (max(0.048, q), 0.2, 40.0)，q表示抖动向量的 分位数0.55和0.65的平均数。系数被映射到较大区间的数据。

全部细节过程参见6.11，程序CQualityModelFullRef::SplitTempTrans。

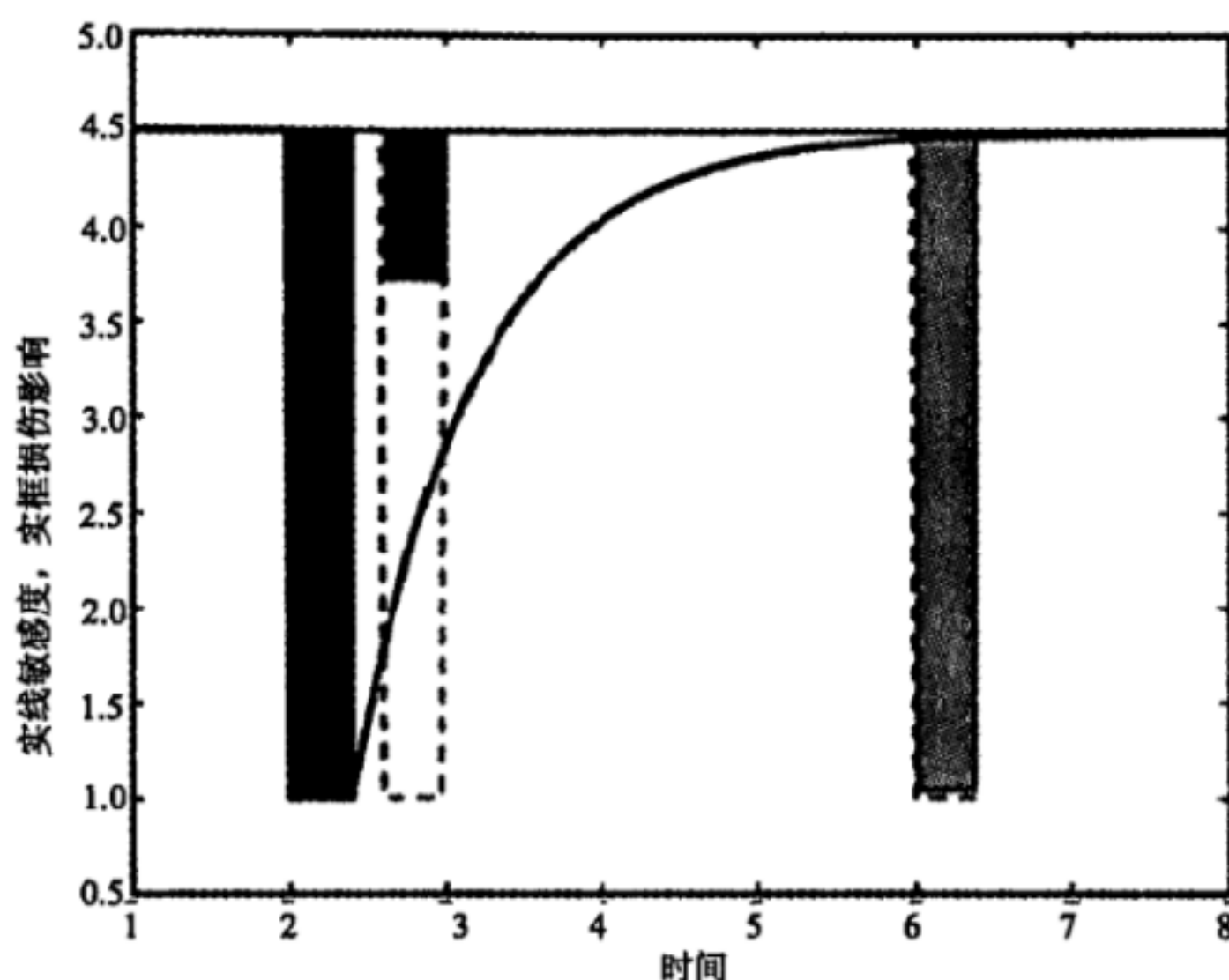
下一步，基本质量 q_cod 定义如下：

$$q_{\text{cod}} = (1 - d_{\text{cod}}) \times (1 - d_{\text{diff_cod}}) \times (1 - \text{blockiness}) \quad (6)$$

瞬时质量定义如下：

$$q_{trans} = (1 - d_{trans}) \times (1 - d_{diff_trans}) \times (1 - d_{t_trans})$$

额外的损伤影响如果是短暂出现的,在此简化。如果考虑其影响,则 q_{trans} 转换为 q_{fq} 。其主要思想如图9所示。



注:黑色长条块是一个视频的序列的第一个损伤。黑色的曲线代表了人对接下来的损伤的敏感程度。如果在第一次损伤后出现的损伤延续时间较短,那么如第一个灰色长条块所示,人对其的敏感程度可以忽略。如果接下来的损伤与第一次损伤之间的间隔时间较长,如浅灰色长条块所示,那么其影响没有改变。

图9 损伤影响的简化

计算的细节在接下来的程序中有详细描述。 $1 - q_{trans}$ 是输入向量 v 。向量 $disp_time$ 帧显示时长。延迟时间参数 $dT=1000\text{ ms}$,然后:

$$q_{fq} = 1 - w \quad (7)$$

其中 w 是函数的输出向量:

$disp_time(i)$ 表示第 i 帧的显示时间, $jerkiness(i)$ 是第 i 帧的抖动参数。 q_{cod} 由公式equation (6)获得, q_{fq} 由公式(7)获得。然后设:

$$Q_t = 1 - 1/T \sum_i [jerkiness(i)]$$

$$Q_{fq} = 1/T \sum_i [q_{fq}(i) \times disp_time(i)]$$

$$Q_{cod} = 1/T \sum_i [q_{cod}(i) \times disp_time(i)]$$

其中 $T = \sum_i [disp_time(i)]$, \sum_i 表示所有项的和 $i=0, \dots, \text{number_of_frames}$ 。

最后的视频分数在区间[1,5]内,由下式获得 mos 值:

$$s = 4 \times Q_t \times Q_{cod} \times Q_{fq} + 1 \quad (8)$$

详细过程参见6.11,程序CQualityModelFullRef::PredictScoreCodTrans。

6.10 处理空间校正严重错的视频序列

为避免由于源视频和损伤视频之间较大区域的空间错误校正导致的错误评分,以上步骤计算3个不同的视频序列的竖直方向和水平方向的空间校正步长(像素块),其中哪种像素块计算的分数最大的,作为最终的质量分数。

根据块效应一节中的介绍, 当step的大小为 4×4 像素块时, 那么对应的原始图像搜索的块大小范围是 8×8 大小。这个包括了最大的块偏移量 ± 5 像素。那么这个增大块区域的情况在评分模型的高阶函数中包括, 其中的校正范围可以很容易的调整较大的偏移量或者如果为了节省计算资源就忽略掉(比如 ± 4 像素)。

参见6.11, 程序vquad_hd::vq_vquad08。

6.11 源视频代码的执行

本标准中的视频序列样本是VQEG提交给研究组的可用视频。

C++程序包括了评分模型实现部分和其他需要的部分。所有本标准中涉及的实际实现部分都是指的这个源码程序, 具体见<http://www.itu.int/rec/T-REC-J.341>。

附 录 A
(资料性附录)
视频质量专家组的结论

以下数据由VQEG HDTV组发布。报告包括对每个实验的Pearson相关度和RMSE的计算，每个实验还统计了置信区间以及统计显著性测试，部分损伤图像（比如只用ITU-T H.264编码的图像）、散点图以及拟合函数的分析。

此标准中FR模型性能结果见表A.1。PSNR是与FR模型相比较的评价方法，根据ITU-T J.340计算获得。“全集RMSE”是根据所有实验视频集合（6个实验集合）获得的“性能最优实验组”是指被统计模型，在0~6组实验中，表现最优或与最优模型统计指标相当的对应的实验组号。“比PSNR性能更好”是指被统计模型，在0~6组实验中，统计指标比PSNR更好的对应的实验组号。“比全集PSNR更好”列出了被统计模型是否在全集实验中表现比PSNR好。“全集相关度”是通过全集实验获得的Pearson系数。

表 A.1 FR 模型性能

度量	PSNR	FR 模型
全集 RMSE	0.71	0.56
性能结果最好的测试集数	1	5
比 PSNR 性能好的测试集数	—	4
是否比全集的 PSNR 结果好	—	Yes
全集相关度	0.78	0.87

中 华 人 民 共 和 国
通 信 行 业 标 准
IP 网络高清视频客观全参考质量评价方法
YD/T 2647-2013

*

人民邮电出版社出版发行
北京市丰台区成寿寺路 11 号邮电出版大厦
邮政编码：100164
宝隆元（北京）印刷技术有限公司印刷
版权所有 不得翻印

*

开本：880×1230 1/16 2014 年 2 月第 1 版
印张：1.25 2014 年 2 月北京第 1 次印刷
字数：32 千字

15115 • 456

定价：15 元

本书如有印装质量问题，请与本社联系 电话：(010)81055492