国 际 电 信 联 盟

ITU-R

国际电联无线电通信部门

ITU-R BT.1866 建议书

(03/2010)

在具有全参考型信号情况下用于采用普通清晰度电视广播应用的客观感知视频质量测量技术

BT 系列

广播业务

(电视)

ITU 国际电信联盟

前言

无线电通信部门的职责是确保卫星业务等所有无线电通信业务合理、平等、有效、经济地使用无线电频谱，不受频率范围限制地开展研究并在此基础上通过建议书。

无线电通信部门的规则和政策职能由世界或区域无线电通信大会以及无线电通信全会在研究组的支持下履行。

知识产权政策（IPR）

ITU-R的IPR政策述于ITU-R第1号决议的附件1中所参引的《ITU-T/ITU-R/ISO/IEC的通用专利政策》。专利持有人用于提交专利声明和许可声明的表格可从http://www.itu.int/ITU-R/go/patents/en获得，在此处也可获取《ITU-T/ITU-R/ISO/IEC的通用专利政策实施指南》和ITU-R专利信息数据库。

ITU-R 系列建议书

（也可在线查询 http://www.itu.int/publ/R-REC/en）

系列	标题
BO	卫星传送
BR	用于制作、存档和播出的录制；电视电影
BS	广播业务（声音）
BT	广播业务（电视）
F	固定业务
M	移动、无线电定位、业余和相关卫星业务
P	无线电波传播
RA	射电天文
RS	遥感系统
S	卫星固定业务
SA	空间应用和气象
SF	卫星固定业务和固定业务系统间的频率共用和协调
SM	频谱管理
SNG	卫星新闻采集
TF	时间信号和频率标准发射
V	词汇和相关问题

说明： 该ITU-R建议书的英文版本根据ITU-R第1号决议详述的程序予以批准。

电子出版
2011年，日内瓦

© ITU 2011

版权所有。未经国际电联书面许可，不得以任何手段复制本出版物的任何部分。

ITU-R BT.1866 建议书

在具有全参考型信号情况下用于采用普通清晰度电视*广播应用的客观感知视频质量测量技术

（2010年）

范围

本建议书规定了在具有全参考型信号情况下用于采用普通清晰度电视（LDTV）广播应用的客观感知质量测量技术的评估方法。

国际电联无线电通信全会，

考虑到

a) 自动测量广播视频质量的能力长期被视为业界的一种宝贵财富；

b) ITU-R BT.1683建议书描述了在具有全参考型信号情况下用于采用普通清晰度电视广播应用的客观感知视频质量的客观方法；

c) ITU-R BT.1833建议书描述了通过手持接收机移动接收多媒体和数据广播应用的多媒体系统；

d) 普通清晰度电视（LDTV）正越来越广泛地用于供移动接收的多媒体和数据广播应用；

e) ITU-T J.247建议书[1]规定了在具有全参考型信号情况下适用于LDTV应用的客观感知视频质量的测量技术；

f) LDTV视频质量的客观测量可能会补充主观评价方法，

认识到

a) LDTV的使用主要用于在手持和移动接收机那些可用的小屏幕上进行观看，

做出建议

1 附件1中给出的指导原则、范围和限制应用于做出决议2中所述的客观视频质量测量模型的应用；

2 ITU-T J.247建议书中给定的客观感知视频质量测量模型应用于在具有全参考型信号情况下采用LDTV的广播应用。

说明1 – 附件2至附件5中给出了测量模型的摘要，供提供信息之用。详情请参见ITU-T J.247建议书。

---

* 普通清晰度电视（LDTV）指像素数量低于ITU-R BT.601建议书中所规定像素数量的视频清晰度。正在审议一项与LDTV有关的ITU-R建议书。

[1] ITU-T J.247建议书可查阅 <http://www.itu.int/rec/T-REC-J.247-200808-P/en>。

附件 1

1 引言

本建议书规定了评估在具有全参考型信号情况下采用LDTV的广播应用的感知视频质量方法。

本建议书中的评估方法适用于：

– 编解码器评估，指标和验收测试；

– 在视频源进行可能的实时、不中断服务的视频监测；

– 当可以获得节目源的副本时，远程质量监控；

– 采用单向或串连视频压缩和解压缩方法来监控存储或传输系统的质量测量；

– 视频系统的实验室测试。

当在测量点已经可以获得未受损伤的参考视频信号时，可以采用全参考测量方法，在单个设备或在实验室或封闭环境的一系列过程中进行的测量也可采用该方法。评估方法基于在VGA、CIF和QCIF清晰度下处理普通清晰度视频。

验证测试材料包括多重编码退化和各种传输误差条件（如比特误差和丢失的封包）。

在考虑视频信号编码失真的情况下，编码器可采用各种压缩方法（如MPEG-2、H.264等）。可采用本建议书所述的各种模型来监测所部署网络的质量，以确保整个操作就绪程度。退化的视频效率可能包括空间和时间退化（如帧重复、跳帧和帧频降低）。本建议书中的模型也可用于视频系统的实验室测试。

本建议书可以认为适合于在移动接收机上以4 Mbit/s或更低速率传递的业务。在各种分辨率的验证测试中采用了以下的条件并发现是合适的：

– QCIF（四分之一通用中介格式（176 × 144像素））：16至320 kbit/s。

– CIF（通用中间格式（352 × 288像素））：64 kbit/s至2 Mbit/s。

– VGA（视频图形阵列（640 × 480像素））：128 kbit/s至4 Mbit/s。

表 1

模型评估中采用的要素

测试要素
有封包丢失的传输误差
视频清晰度QCIF、CIF和 VGA
视频比特率 QCIF：16至320 kbit/s CIF：64 kbit/s至 2 Mbit/s VGA：128 kbit/s至4 Mbit/s
最大2秒的时间误差（跳跃停顿）
从5到30 fps的视频帧频

表 1 （完）

编码方案
H.264/AVC（MPEG-4第10部分）、MPEG-4第2部分以及三种其它专利编码方案。 （见以下说明1。）

说明 1 – 模型的验证测试包括采用15种不同视频编码进行编码的视频序列。表1所列的五种编解码器是最常见的用于编码测试序列的编解码器，且对任何推荐的模型都可以视为适合于评估这些编解码器。除这五种编解码器以外，采用以下编解码器生成了一小部分测试序列：H.261、H.263、H.263+[2]、JPEG-2000、MPEG-1、MPEG-2、H.264 SVC以及其它专利系统。请注意，这些编解码器的一部分只用于CIF和QCIF分辨率，因为它们被认为主要用于这些分辨率的领域。在将一种模型适用于采用这些编解码器的一种进行编码的序列之前，用户应仔细研究其预测性能，以判断模型是否达到了可接受的预测性能。

2 应用

本建议书中所述的评估模型应用包括但不限于：

1 编解码器评估，指标和验收测试；

2 在视频源进行可能的实时、不中断服务的视频监测；

3 当可以获得节目源的副本时，远程质量监控；

4 采用单向或串连视频压缩和解压缩方法来监控存储或传输系统的质量测量；

5 视频系统的实验室测试。

3 模型使用

本建议书包括了表2所示的客观计算模型。模型性能的简述见表3。更详细的信息述于附录1。

表 2

客观计算模型

模型标识	提议者	国家	附件号
A	NTT	日本	2
B	OPTICOM	德国	3
C	Psytechnics	英国	4
D	延世大学	韩国	5

所有四种模型都明显地胜过峰值信噪比（PSNR）。

[2] H.263+是H.263（1998年）的一种特别配置。

在部分清晰度下，模型B和C似乎在效果上要略微好于模型A和D。模型B和C通常在统计上得出相当的结果。对于QCIF，模型A经常在统计上相当于模型B和C。对于VGA，模型D在统计上典型地相当于模型B和C。

下表提供了模型性能的概述。尽管可以采用四种模型来充分满足业界的不同需求，但对于VGA，在绝大多数情况下强烈建议采用模型B、C和D来获得略微更好的性能。出于同样的原因，对于CIF，强烈建议采用模型B或C；对于QCIF，采用模型A、B或C。

模型B显示了最好的总体最小相关性。模型B、A、D和C的最小相关性系数分别为0.68、0.60、0.59和0.57。

模型C获得了最高的在最优表现组中出现的反复次数。对于模型C，在最优组出现的反复次数总数为37，模型A为25，模型D为24。

表 3

模型性能概述

VGA	**模型A**	**模型B**	**模型C**	**模型D**	**PSNR**
平均相关性	0.786	0.825	0.822	0.805	0.713
最小相关性	0.598	0.685	0.565	0.612	0.499
等级1的出现频率	8	10	11	10	3
等级分析	二等	最好	最好	最好	–

CIF	**模型A**	**模型B**	**模型C**	**模型D**	**PSNR**
平均相关性	0.777	0.808	0.836	0.785	0.656
最小相关性	0.675	0.695	0.769	0.712	0.440
等级1的出现频率	8	13	14	10	0
等级分析	二等	最好	最好	二等	–

QCIF	**模型A**	**模型B**	**模型C**	**模型D**	**PSNR**
平均相关性	0.819	0.841	0.830	0.756	0.662
最小相关性	0.711	0.724	0.664	0.587	0.540
等级1的出现频率	9	11	12	4	1
等级分析	最好	最好	最好	二等	–

4 局限性

本建议书所述的评估方法不能完全替代主观测试。两个认真设计并实施的主观性测试（在两个不同的实验室中）的相关值通常在0.95至0.98之间。

本建议书中的模型采用显示帧凝固高达2秒的测试视频进行了验证。

本建议书中的模型并未经过具有平稳增加时延（如在帧凝固之后并不丢帧的视频）的测量视频的验证。

请注意，在新的编码和传输方法产生伪影（并未包含在本评估中）的情况下，客观模型可能会产生错误结果。此时需要主观性评估。

附件1的
附录1

视频质量专家组的结论

称为视频质量专家组（VQEG）的非正式小组进行了感知视频质量测量的研究，该小组向ITU-T第9和第12以及无线电通信第6研究组报告。最近完成的VQEG多媒体第一阶段评估了QCIF、CIF和VGA格式的建议全参考感知视频质量测量算法的性能。

根据目前的事实，在当前阶段ITU-R可以建议四种方法。他们是：

模型 A（附件 2）－VQEG，提议者日本NTT

模型 B（附件 3）－VQEG，提议者德国OPTICOM

模型 C（附件 4）－VQEG，提议者英国Psytechnics

模型 D（附件 5）－VQEG，提议者韩国延世大学。

这些模型的技术描述可分别见附件2到附件5。请注意，附件的排序不分先后，并不意味着质量预测性能的高低。

表4提供了VQEG多媒体第一阶段测试中一个模型性能的详细情况。

表 4

a) VGA清晰度：VQEG多媒体第一阶段测试中模型性能 – 14项主观测试的平均值

度量	模型A	模型B	模型C	模型D	PSNR(1)
附件	2	3	4	5	
皮尔逊相关	0.786	0.825	0.822	0.805	0.713
均方根误差	0.621	0.571	0.566	0.593	0.714
离散率	0.523	0.502	0.524	0.542	0.615

表 4（完）

b) CIF分辨率：VQEG多媒体第一阶段测试中模型性能 – 14项主观测试的平均值

度量	模型A	模型B	模型C	模型D	PSNR[1]
附件	2	3	4	5	
皮尔逊相关	0.777	0.808	0.836	0.785	0.656
均方根误差	0.604	0.562	0.526	0.594	0.720
离散率	0.538	0.513	0.507	0.522	0.632

c) QCIF分辨率VQEG多媒体第一阶段测试中模型性能 – 14项主观测试的平均值

度量	模型A	模型B	模型C	模型D	PSNR[1]
附件	2	3	4	5	
皮尔逊相关	0.819	0.841	0.830	0.756	0.662
均方根误差	0.551	0.516	0.517	0.617	0.721
离散率	0.497	0.461	0.458	0.523	0.596

(1) 此处报告的PSNR值摘自于VQEG多媒体第一阶段的最终报告（见：http://www.its.bldrdoc.gov/vqeg/projects/multimedia/）。这些值由NTIA/ITS计算。

根据每一个度量标准，每种FR VGA模型在表现最好模型组出现的次数如下：

统计数据	模型A	模型B	模型C	模型D	PSNR
相关	8	10	11	10	3
RMSE[1]	4	8	10	6	0
离散率	9	11	12	8	4

(1) RMSE：均方根误差。

根据每一个度量标准，每种FR CIF模型在表现最好模型组出现的次数如下：

统计数据	模型A	模型B	模型C	模型D	PSNR
相关	8	13	14	10	0
RMSE[1]	6	10	13	9	0
离散率	10	13	12	11	1

(1) RMSE：均方根误差。

根据每一个度量标准，每种FR QCIF模型在表现最好模型组出现的次数如下：

统计数据	模型A	模型B	模型C	模型D	PSNR
相关	9	11	12	4	1
RMSE(1)	7	10	11	2	1
离散率	10	11	12	8	4

(1) RMSE：均方根误差。

说明 1 – 作为一条一般性指导原则，这些总数值中微小的差别并不代表性能上整体的不同。

次要分析

次要分析将与各种视频系统（或条件）有关的所有视频序列进行平均并由此反映出模型跟踪平均假设基准电路（HRC）性能的满意程度。以下表格显示了次要分析中各种模型和分辨率的平均相关性。

VGA相关性

	模型A	模型B	模型C	模型D	PSNR
平均	0.891	0.914	0.903	0.864	0.809

CIF相关性

	模型A	模型B	模型C	模型D	PSNR
平均	0.915	0.919	0.913	0.892	0.817

QCIF相关性

	模型A	模型B	模型C	模型D	PSNR
平均	0.942	0.937	0.920	0.893	0.882

附件 2

模型 A

模型A分为三个软件模块：视频匹配模块，时间/空间特征量推导模块和主观视频质量评估模块（图1）。

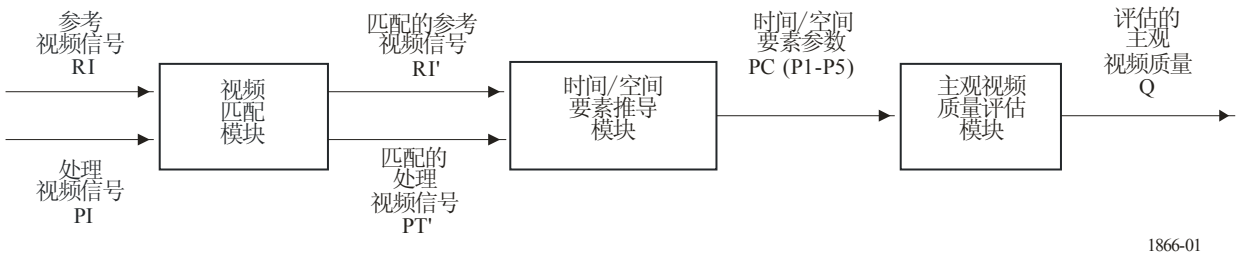
视频匹配模块分为宏观匹配程序和微观匹配程序。宏观匹配程序在空间-时间方向上匹配参考视频信号（RI）和处理视频信号（PI）之间的像素并在考虑解码器视频捕获和后期处理影响的情况下过滤视频序列。微观匹配程序待宏观匹配程序结束后，在考虑视频帧跳跃和帧凝固影响的情况下匹配参考和处理视频序列的帧。

时间/空间特征量推导模块通过采用匹配的参考视频信号（RI’）和匹配的处理视频信号（PI’）来计算空间退化参数和时间退化参数（PC）。空间退化参数基于反映总噪声、寄生边缘、局部运动失真或局部空间失真等四个参数。由加权的凝固长度总和计算的时间退化参数反映了帧凝固和帧频的变化。

主观视频质量评估模块通过前述的参数计算客观视频质量（Q）。

图 1

模型 A的结构图



模型A的完整描述请参见ITU-T J.247建议书（08/2008）的附件A。

附件 3

模型 B

模型B的基本概念见图2。

在预处理步骤，从参考和测试信号中提取空间感兴趣区（ROI）。所有的后续计算都只在这个ROI上进行，后者由裁切的信号S_P和P_P表示。在预处理之后，在空间和亮度区域进行输入序列的粗匹配（注册）。这些模块获得的“亮度和匹配信息”用于随后的“详细序列匹配”过程。该过程通过两个视频序列的帧匹配，空间漂移补偿以及在柱状图评估基础上色彩和亮度差异的补偿等进行逐帧的时间匹配。“详细的序列匹配”的结果为“匹配信息”（matching info），后者用于判定时间退化以及裁切并匹配的序列S_A和P_A的视觉影响。

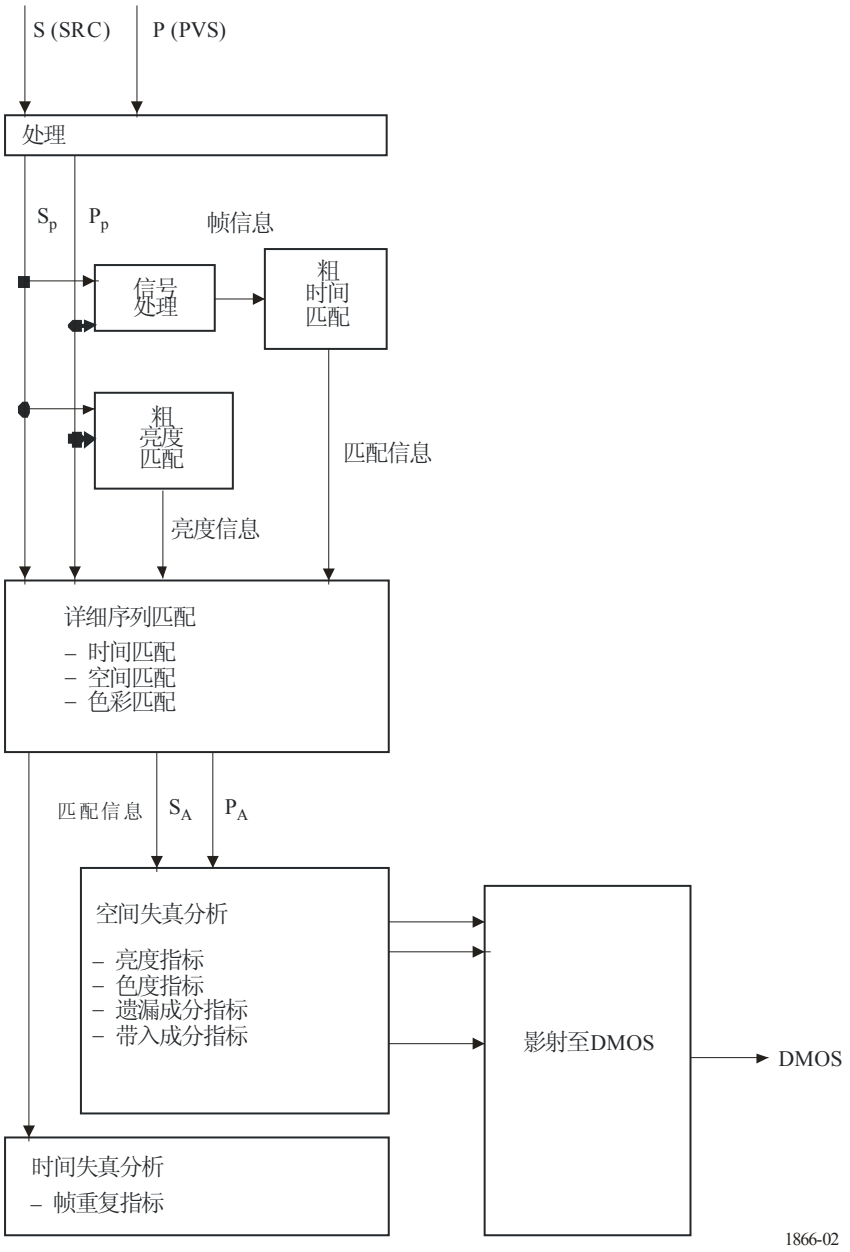
空间失真进一步通过“空间失真分析”单元进一步进行分析，该单元计算空间与序列之间的视觉差异，得出四种失真指数。

“匹配信息”进一步通过视觉“时间失真分析”进行处理，后者产生一个代表帧重复和其它时间失真的指标。

在模型C的最后一步中，上述获得的五种指标通过逻辑函数进行加权并合并，以形成最终的PEVQ（视频质量的视觉评估）分数，该分数与从主观测试中获得的MOS（平均意见得分）高度相关。

图 2

模型 B的概述



模型B的完整描述请参见ITU-T J.247建议书（08/2008）的附件B。

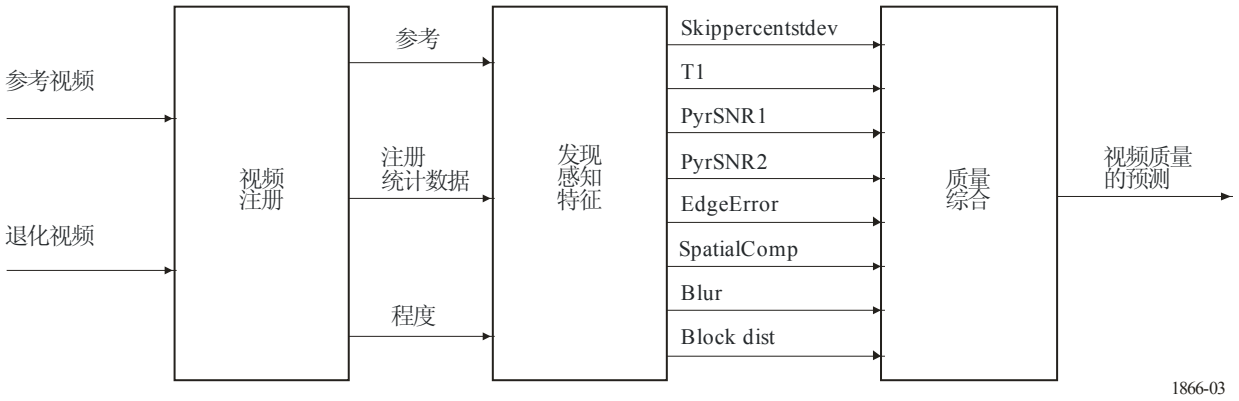
附件 4

模型 C

模型C的概述见图3。模型包括三个主要阶段：(1) 视频注册，(2) 发现感知特征以及(3) 将这些特征综合到一个整体的质量预测分数。

图 3

模型 C的概述



第一步包括视频注册，该步骤将退化视频中每一个帧临时与参考视频中最佳的匹配帧进行配对。需要视频注册来提供准确的质量测量。从初始的参考和退化视频获得匹配的参考视频序列和匹配的退化视频序列以及注册信息。匹配的参考视频包含所有已经与退化视频的帧进行匹配的参考帧。匹配的退化视频包括已经确定了参考帧的所有退化帧。视频注册算法可以与参考和退化视频之间时变的时间和空间偏移相匹配。

分析的第二阶段包括直接从退化视频或参考与退化视频时间比较所提取一套有意义的感知特征。这些感知特征并不是基于关于视频中内容类型的任何假定或视频如何编码并传送到最终用户。感知特征作为一种感觉层发挥作用，过滤掉人类观众感觉不到的图像部分。这是因为视频编解码器降低了编码的数据量，通过采用图像处理技术来去除观众最不可能注意到已经丢失的视频信号部分。通过将人类视觉系统的模型包括在内，算法的该步骤可以判定这些过程的有效性并确定可见误差。当编码方案不能成功地实现该目标时，最终用户将会看到退化并因此从感知层形成输出的一部分。

在最后的阶段，合并感知参数来生成单一的视频质量总体预测。通过将模型运用于大量主观实验（训练集）并在一组未知实验（验证集）上验证其性能来获得积分函数的最优形式。

如VQEG多媒体测试方案所述，模型的软件实施所支持的、并提交给VQEG多媒体第一阶段评估的输入参考和退化视频的格式主带有UYVY (YUV 4:2:2)色空间格式的非压缩AVI。但是，质量评估模型与该格式是独立的，因此也同样适用于其它格式（如未压缩的AVI RGB24），条件是首先采用了适当的输入滤波器（如色空间转换滤波器或文件阅读器）。

参考和退化视频必须都是逐行格式。退化视频的绝对帧频必须与参考视频的帧频相同（如25 fps），其中绝对帧频为每秒的帧数量（逐行）。但是，退化视频有效帧频可与参考视频的帧频不同，其中有效帧频为每秒单帧的（平均）数量。退化视频的有效帧频也可以是时变的。提供了以下示例，以供说明。参考视频（A）具有25 fps的绝对帧频且采用12.5 fps (B)的有效（目标）帧频进行了编码。编码的视频进行播放（即解码）并按照25 fps (C)进行捕获。因此，C中每一个相隔的帧都与前一个帧相同。质量评估模型的输入参考和退化视频分别为A和C。

模型C的完整描述请参见ITU-T J.247建议书 (08/2008)的附件C。

附件 5

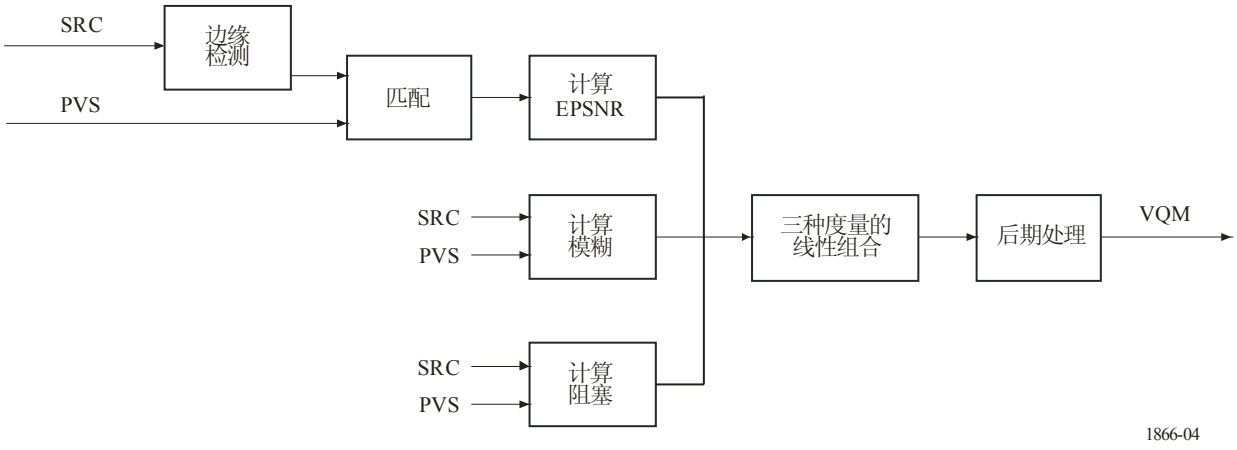
模型 D

研究发现，人类的视觉系统对边缘的退化敏感。进一步研究发现，视频压缩算法趋向于在边缘地区产生更多的伪像。根据这一观察，模型提供了一种可以衡量边缘周围退化的客观视频质量衡量方法。在该模型中，首先将边缘检测算法适用于原始序列，以确定边缘区域。然后，通过计算均方误差测量这些边缘地区的退化。根据该均方误差计算边缘PSNR (EPSNR)。此外，模型计算与EPSNR合并的两个其它特征，以生成最终的视频质量度量（VQM）。

图4显示了全参考（FR）边缘退化基础上的模型D的结构图，该结构图采用了两种输入：源视频序列（SRC）和处理视频序列（PVS）。

图 4

模型 D的结构图



模型D的完整描述请参见ITU-T J.247建议书 (08/2008)的附件D。