

Geneva, 2-5 December 1994

Question: 22/12

SOURCE: RAPPORTEUR FOR QUESTION 22/12

TITLE: TWO CRITERIA FOR VIDEO TEST SCENE SELECTION

ABSTRACT:

This contribution describes some important considerations relevant to video test scene selection and offers two measurements that can be used during such selection. Measurements of spatial information content (SI) and temporal information content (TI) are presented and their use as criteria for test scene selection is demonstrated.

We recommend that this contribution be included in Recommendation P.VQ in Section 4.3.

Introduction

This contribution presents two criteria which can be used in video test scene selection. Measurements are defined which can be used to characterize the relative coding difficulty among a group of test scenes. The spatial content and temporal content of test scenes are not the only criteria which can be useful in test scene selection. They are important perceptually, and the relative extent of each content type may be assessed among a group of scenes, as suggested here. While some issues remain regarding the use of these criteria in comparing scenes of different video formats, the basic measurements presented herein are a useful start.

We recommend that the following text be included in Recommendation P.VQ in Section 4.3.

4.3.1 Test Scene Selection.

The selection of test scenes is an important issue. In particular, the spatial and temporal information content of the scenes are critical parameters. These parameters play a crucial role in determining the amount of video compression that is possible, and consequently, the level of impairment that is suffered when the scene is transmitted over a fixed-rate digital transmission service channel. Fair and relevant video test scenes must be chosen such that their spatial and temporal information content is consistent with the video services that the digital transmission service channel was intended to provide. The set of test scenes should span the full range of spatial and temporal information content of interest to users of the devices under test.

4.3.2 Spatial and Temporal Information Measures.

This section presents methods for quantifying the spatial and temporal information content of test scenes. These methods for evaluating the spatial and temporal information content of test scenes are applicable to video quality testing—both now and in the future. The location of the video scene within the spatial-temporal matrix is important because the quality of a transmitted video scene (especially after passing through a low bit rate codec) is often highly dependent on this location. The spatial and temporal information measures presented here can be used to assure appropriate coverage of the spatial-temporal matrix.

4.3.2.1 Spatial Information Measurement

The Spatial Information content, SI, is based on the Sobel filter. Each video frame (luminance plane) at time n (F_n) is first filtered with the Sobel filter ($Sobel(\bullet)$). The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is then computed. This operation is repeated for each frame in the video sequence and results in a time series of spatial information values. The maximum value in the time series (max_{time}) is chosen to represent the spatial information content of the scene. This process can be represented in equation form as,

$$SI = \max_{time} \{ std_{space} [Sobel(F_n)] \} \quad (1)$$

4.3.2.2 Temporal Information Measurement

The Temporal Information content, TI, is based upon the motion difference feature,

$M_n(i,j)$, which is the difference between the pixel values (of the luminance plane) at the same location in space but at successive times or frames. $M_n(i,j)$ as a function of time (n) is defined as,

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j) \quad (2)$$

where $F_n(i,j)$ is the pixel at the i^{th} row and j^{th} column of the n^{th} frame in time.

The measure of temporal information content, TI, is computed as the maximum over time (\max_{time}) of the standard deviation over space ($\text{std}_{\text{space}}$) of $M_n(i,j)$ over all i and j .

$$TI = \max_{\text{time}} \{ \text{std}_{\text{space}} [M_n(i,j)] \} \quad (3)$$

More motion in adjacent frames will result in higher values of TI. Note: For scenes that contain scene cuts, two values may be given: one where the scene cut is included in the temporal information measure, and one where it is excluded from the measurement.

4.3.3 Example

Figure 1 shows the relative amounts of spatial and temporal information for some representative test scenes and how they can be placed on a spatial-temporal information matrix. Along the $TI=0$ axis (along the bottom of the plot) are found the still scenes and those with very limited motion (such as l, f, and a). Near the top of the plot are found scenes with a lot of motion (such as p, q and i). Along the $SI=0$ axis (at the left edge of the plot) are found scenes with minimal spatial detail (such as l, k, x, u, and f). Near the right edge of the plot are found scenes with the most spatial detail (such as h and s). The values of SI and TI were obtained using the above equations and video which has been spatially sampled according to CCIR 601 specifications. Table 1 lists the example test scenes by scene content category.

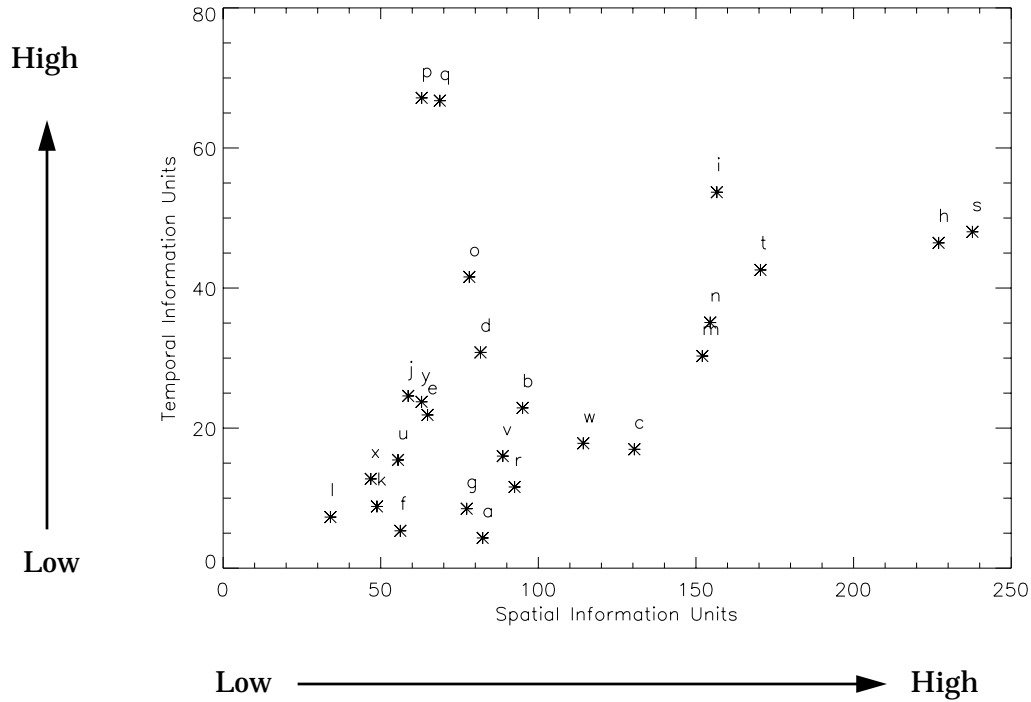


Figure 1 Spatial-Temporal Plot for Example Test Scene Set

Table 1: Scene Content Categories

Content Category and Description	Scene Name and Letter
One person, mainly head and shoulders, limited detail and motion.	vtc1nw(f), susie(j), disguy(k), disgal(l)
One person with graphics and/or more detail.	vtc2mp(a), vtc2zm(b), boblec(e), smity1(m), smity2(n), vowels(w)
More than one person.	3inrow(d), 5row1(g), intros(o), 3twos(p), 2wbord(q), split6(r)
Graphics with pointing.	washdc(c), cirkit(s), rodmap(t), filter(u), ysmite(v), inspec(x)
High object and/or camera motion (Examples of Broadcast TV).	flogar(h), ftball(i), fredas(y)