Committee T1 Performance
Standards Contribution

Document Number:	T1A1.5 / 96-110
TIBBS File:	6a151100.doc
DATE:	May 31, 1996
STANDARDS PROJECT:	Analog Interface Performance Specifications for Digital Video Teleconferencing / Video Telephony Service (T1Q1-12)
SUBJECT:	Measuring digital video transmission channel gain, level offset, active video shift, and video delay
SOURCE:	NTIA / ITS
CONTACT:	Stephen Wolf Voice: (303) 497-3771 Fax: (303) 497-5323 e-mail: steve@its.bldrdoc.gov Margaret Pinson Voice: (303) 497-3579 Fax: (303) 497-5323 e-mail: margaret@its.bldrdoc.gov
KEY WORDS:	video, gain, level offset, shift, delay
DISTRIBUTION:	Working Group T1A1.5 (announced via t1a15@t1.org)
ABSTRACT:	This contribution presents a computerized search method for determining the gain, level offset, active video shift, and video delay of a digital video transmission channel. The method uses digitized input and output NTSC video fields that have been time tagged with SMPTE time code. The primary applications of the method for T1A1 are (1) to produce estimates of channel gain, level offset, and active video shift when data from ANSI T1.801.03-1996 calibration test patterns are not available, and (2) to provide a dynamic method to measure these quantities in conjunction with video delay.

1. Introduction

ANSI T1.801.03-1996 specifies robust methods for measuring gain, level offset, and active video shift (i.e., spatial registration of input and output video frames). These methods require the use of still video and in the case of the gain and level offset calculations, that still video is a test pattern defined in the standard. However, a researcher analyzing video test scenes from an experiment where the ANSI T1.801.03 calibration frames have not been included must devise an alternative method for performing these calibration measurements. This contribution presents an adaptation of the methods in ANSI T1.801.03 for calculating gain, level offset, and active video shift using natural motion video. The method has the added advantage of being able to track dynamic changes in gain, level offset, and active video shift. This will be useful for channels which change their calibration characteristics on a scene by scene basis (e.g., an MPEG channel that is re-tuned for each scene to optimize quality).

The calibration method described in this contribution is fully compatible with and supplements the video delay measurement techniques presented in the Draft ANSI Standard on Visual Channel Delay and Frame Rate Measurement (T1A1.5/96-101). The calibration procedure could be performed once before beginning visual channel delay measurements, and perhaps reapplied during the measurement process if a temporally varying gain and/or active video shift is suspected. Calibration is an important issue whenever input and output video frames are being directly compared. Neglecting calibration can produce large video delay measurement errors. For example, a horizontal shift of the output video picture (which is not calibrated out) can look like a video delay when the scene contains horizontal motion. In addition, non-unity channel gains and non-zero level offsets distort mean square error (MSE) calculations that are used for input and output video field alignment.

The calibration algorithm presented here has been implemented on the ITS Video Quality Measurement System Laboratory workstations using the C++ programming language. This system performs automated video frame digitization according to ITU-R Recommendation BT.601 and stores these sampled frames on a random access read-write optical jukebox. Frames are time tagged according to SMPTE time code for easy reference and to allow measurement of absolute video delay.

2. Calibration Algorithm for NTSC Video

2.1 Overview

The basic calibration algorithm is applied to a single field from the output video stream. For NTSC video, field comparisons are used rather than frame comparisons because of the advantages described in contribution T1A1.5/95-152. For each selected output field, the following quantities are computed:

1. The closest matching field from the input video stream.
2. The estimated gain and level offset between the output field and the closest matching input field.
3. The estimated active video shift (horizontal and vertical spatial shift) between the output field and the closest matching input field.

The interdependence of the above listed quantities produces a "chicken or egg" measurement problem. Calculation of the closest matching input field requires that one know the gain, level offset, and active video shift. However, one cannot determine these quantities until the closest matching input field is found. If there are wide uncertainties in the above three quantities, a full exhaustive search would require a tremendous number of computations. The approach taken here is to reach the solution using an iterative search algorithm. For robustness, the basic calibration algorithm can be independently applied to several output fields and the results averaged.

2.2 Description of Basic Calibration Algorithm for One NTSC Output Field

The basic calibration algorithm for one selected output field will be described in this section. The next section will discuss how multiple applications of this basic calibration algorithm can be used to track dynamic changes in the calibration quantities or to obtain robust estimates of static calibration quantities.

2.2.1 Inputs to the Algorithm

The following is a list of quantities that must be pre-specified in order for the search algorithm to work. The initial search limits should be generous enough to include the correct calibration point. *A priori* knowledge of the transmission channel behavior may be used to help define the initial search limits (e.g., minimum and maximum video delay may be used to specify the range of input fields to search).

1. o_m, the current output field on which to perform the calibration, sampled according to ITU-R Recommendation BT.601 (horizontal extent: 0 to 719 pixels, vertical extent: 0 to 242 active video lines). The image pixel at vertical and horizontal coordinates (v=i, h=j) will be denoted by $o_m(i, j)$, where (0,0) is the top-left pixel in the image. If this calibration algorithm is being used in conjunction with the techniques in the Visual Channel Delay Standard (T1A1.5/96-101), an output field which is active should be used.

2. $\{i_L, \ldots, i_n, \ldots, i_U\}$, the range of contiguous input fields (lower, …, current, …, upper) to examine for a match with output field o_m, sampled according to ITU-R Recommendation BT.601 (horizontal extent: 0 to 719 pixels, vertical extent: 0 to 242 active video lines).

3. *ROI* = {*top*, *left*, *bottom*, *right*}, the input field sub-region (region of interest) over which to perform the comparison, *left* and *right* are in ITU-R Recommendation BT.601 pixels (numbered 0 to 719), *top* and *bottom* are in lines (numbered 0 to 242). Note: *ROI* may be a manually determined input to the calibration algorithm or an appropriate *ROI* could be automatically calculated using the considerations given in step 1 of section 2.2.3 below.

4. $\{h_L, \ldots, h_s, \ldots, h_U\}$, the range of possible horizontal shifts (lower, …, current, …, upper) of the output field in ITU-R Recommendation BT.601 pixels, where a positive shift indicates that the output is shifted to the right with respect to the input.

5. $\{v_L, \ldots, v_s, \ldots, v_U\}$, the range of possible vertical shifts (lower, …, current, …, upper) of the output field in lines, where a positive shift indicates that the output is shifted downward with respect to the input.

6. g, an initial guess for the transmission channel gain as defined in ANSI T1.801.03 (nominally set to 1.0).

2.2.2 Comparison Function

Given the above definitions, a standard comparison function for comparing output field o_m to input field i_n is defined as:

$$\mathbf{var}(o_m, i_n, h_s, v_s, g) = \left\{ \frac{1}{P} \sum_{i=top}^{bottom-1} \sum_{j=left}^{right-1} \left[\frac{1}{g} o_m(i+v_s, j+h_s) - i_n(i,j) \right]^2 \right\} - \left[\mathbf{mean}(o_m, i_n, h_s, v_s, g) \right]^2$$

where

$$\mathbf{mean}(o_m, i_n, h_s, v_s, g) = \left\{ \frac{1}{P} \sum_{i=top}^{bottom-1} \sum_{j=left}^{right-1} \left[\frac{1}{g} o_m(i+v_s, j+h_s) - i_n(i,j) \right] \right\},$$

$$P = (bottom - top)(right - left),$$

and h_s, v_s, and g are some hypothesized horizontal shift, vertical shift, and gain of the output field. The point (i_n, h_s, v_s, and g) where the comparison function is minimized is defined as the global calibration point for output field o_m. Using the variance instead of the mean square error for the comparison function has several advantages. One advantage is the reduction of time alignment errors resulting from changes in scene brightness levels. The variance comparison function is more likely to use true scene motion for time alignment of the input and output images rather than changes in scene lighting conditions or transmission channel level offset. The variance comparison function also eliminates the transmission channel level offset from the search, and allows this calibration quantity to be directly computed after the other calibration quantities are determined.

2.2.3 Algorithm Description

Figure 1 presents a flow diagram of the search algorithm that is used to find the desired global calibration point for output field o_m. The algorithm uses the following steps which are applied as shown in the figure.

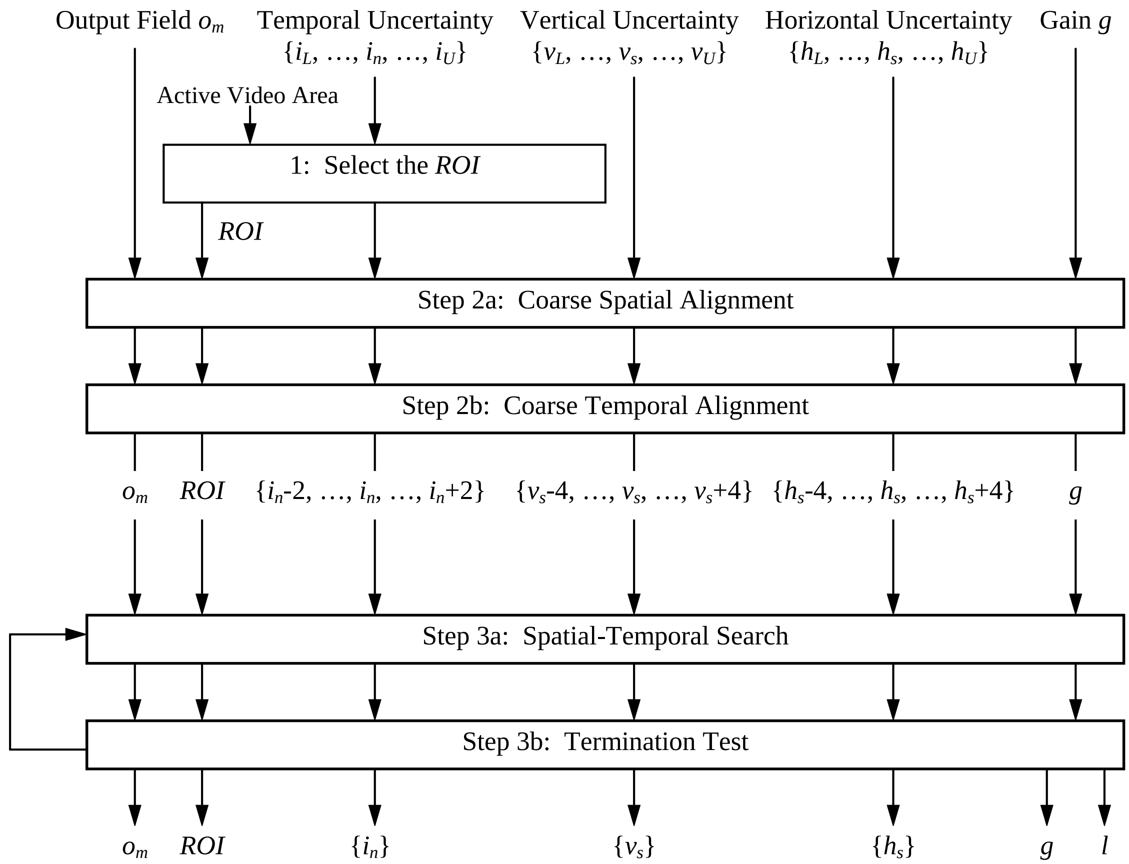

Figure 1 Calibration Algorithm Flow Diagram

STEP 1 - Select the Region of Interest (*ROI*)

The first step is to select a region of interest (*ROI*) upon which to base the comparison function calculations. This is an important step to assure that the comparison function is minimized at the true global calibration point. The *ROI* can be manually or automatically selected depending upon the following important considerations:

1. The *ROI* should be chosen such that it is contained within the active video area.[1]
2. The *ROI* should include both horizontal and vertical edges to assure proper spatial registration of the input and output fields. The spatial information (SI) features in section 6.1.1.1 of ANSI T1.801.03 can be applied to the input sequence to determine if horizontal and vertical edges are present.
3. The *ROI* should include both still and motion areas to assure proper temporal registration of the input and output fields. The temporal information (TI) features in section 6.1.1.2 of ANSI T1.801.03 can be applied to the input sequence to determine if motion and still areas are present.

[1] The active video area is defined in section 5.3 of ANSI T1.801.03-1996 as that rectangular portion of the input active video that is not blanked by the transmission service channel. Technically, the active video area cannot be calculated before the active video shift is known. However, one can choose a conservative *ROI* well within the estimated active video area.

4. The size of the *ROI* should be carefully considered. Input to output field comparisons will be faster if a smaller *ROI* is selected. Too small of a *ROI* might miss important alignment information while too large of a *ROI* might create difficulties in temporal registration for scenes that contain small amounts of motion.

5. The *ROI* should contain only the valid scene area or that portion of the input scene that contains picture. For example, the *ROI* should be reduced for scenes that are in the letterbox format.

6. The *ROI* must be no larger than the intersection of the active video area (point 1 above) and the valid scene area (point 5 above), and must account for the horizontal and vertical shift uncertainties (i.e., $\{h_L$ to $h_U\}$, $\{v_L$ to $v_U\}$).

STEP 2 - Coarse Spatial and Temporal Alignment

Since images are often oversampled from Nyquist both spatially and temporally, a coarse spatial and temporal alignment search (i.e., a search that does not include every pixel and field) can be used to effectively reduce the initial spatial and temporal uncertainties (i.e., $\{h_L, \ldots, h_s, \ldots, h_U\}$, $\{v_L, \ldots, v_s, \ldots, v_U\}$, and $\{i_L, \ldots, i_n, \ldots, i_U\}$). The course search parameters are selected to be fine enough so that the search algorithm will not miss the global calibration point (i.e., the point at which the comparison function is a minimum). Coarse registration to within (and subsequent fine registration over) ±4 pixels, ±4 lines, and ±2 fields is sufficient to insure that the desired global calibration point is achieved.[2]

For efficiency, the coarse spatial and temporal search is itself performed as a two step process as follows:

a) Coarse Spatial Alignment

Coarse spatial alignment of output field o_m is performed using the current best guess for the matching input field. The comparison function is computed for: output field o_m, input field i_n (current best guess) [3], horizontal shifts $\{h_L, \ldots, h_s-4, h_s, h_s+4, \ldots, h_U\}$, vertical shifts $\{v_L, \ldots, v_s-4, v_s, v_s+4, \ldots, v_U\}$, and g equal to the current guess for the transmission channel gain. The horizontal and vertical shifts (h_s and v_s) are updated to that point which minimizes the comparison function. An updated estimate for the transmission gain g is then computed using the calibration equations in section 5.1.2 of ANSI T1.801.03 and the updated spatial alignment.

[2] The spatial search limits of ±4 pixels and lines are based on scenes with a moderate amount of motion. To assure that the fine registration algorithms converge to the proper input field, these spatial search limits should be chosen to include the maximum amount of motion between two sequential fields (i.e., field 1 and the next field 2). A temporal uncertainty of ±2 fields allows for the possibility of being off by one field of the same type as the current field (for example, consider the case where o_m is an NTSC “field 1”, the current i_n is an NTSC “field 1”, but the correct input time alignment is an NTSC “field 1” at time location i_n-2).

[3] Caution should be observed near a scene cut to assure that input field i_n is the same scene as the output field o_m. One could examine the input sequence for scene cuts using the techniques presented in T1A1.5/93-152 and T1A1.5/94-110. These techniques locate large changes, or spikes, in the temporal information (TI) sequences which are indicative of scene cuts.

b) Coarse Temporal Alignment

Coarse temporal alignment of output field o_m is performed using the spatial alignment and gain found in step 2a. The comparison function is computed for: output field o_m, input fields $\{i_L, \ldots, i_n\text{-}2, i_n, i_n+2, \ldots, i_U\}$, the updated horizontal shift h_s from step 2a, the updated vertical shift v_s from step 2a, and the updated gain g from step 2a. The best matching input field i_n is updated to that field which minimizes the comparison function. An updated estimate for the transmission gain g is then computed using the calibration equations in section 5.1.2 of ANSI T1.801.03 and the updated input field.

<u>STEP 3 - Fine Spatial and Temporal Alignment</u>

Fine spatial and temporal alignment of output field o_m is performed using the coarse calibration estimates and reduced uncertainties (±4 pixels, ±4 lines, ±2 fields) from step 2. The fine search algorithm uses the comparison function of section 2.2.2 to examine all possible spatial and temporal shifts within the reduced uncertainties. The fine search algorithm is applied repeatedly until convergence is reached (i.e., i_n, h_s, and v_s remain the same from one iteration to the next).

a) Spatial-Temporal Search

The comparison function is computed for: output field o_m, input fields $\{i_n\text{-}2, i_n\text{-}1, i_n, i_n+1, i_n+2\}$, horizontal shifts $\{h_s\text{-}4, \ldots, h_s\text{-}1, h_s, h_s+1, \ldots, h_s+4\}$, vertical shifts $\{v_s\text{-}4, \ldots, v_s\text{-}1, v_s, v_s+1, \ldots, v_s+4\}$, and transmission channel gain g. The horizontal and vertical shifts (h_s and v_s) are updated to that point which minimizes the comparison function over the above range of inputs. An updated estimate for the transmission gain g is then computed using the calibration equations in section 5.1.2 of ANSI T1.801.03 and the updated spatial-temporal alignment.

b) Termination Test

The value of i_n, h_s, and v_s at the end of step 3a is compared to what their values were at the beginning of step 3a. If there is any difference, then step 3a is repeated with the new calibration values. Otherwise, stop because the search algorithm has finished. The level offset l is then calculated using the current values of i_n, h_s, v_s, g, and the equations in section 5.1.2 of ANSI T1.801.03-1996.

2.3 Multiple Application of the Basic Calibration Algorithm

The basic calibration algorithm shown in Figure 1 and described in section 2.2 can be applied to more than one output field. The two primary reasons for doing this are to:

1. Compute more robust estimates of the calibration quantities for static (i.e., not time varying) transmission systems.
2. Continuously update the calibration quantities for transmission systems that change their behavior over time (e.g., the calibration changes from one scene to the next).

When the calibration quantities are static, the calibration algorithm in section 2.2 can be applied to multiple output fields o_m (m=1, 2, 3, …, M) and the results can be filtered to produce robust estimates for the gain g, level offset l, horizontal shift h_s, and vertical shift v_s. In this case, one should select a set of output fields that are active (see T1A1.5/96-101 for a definition of an active field) so that each selected output field provides new information. A median filter is recommended for gain *g* and level offset *l* since the median is generally more robust than the mean and not as sensitive to outliers. A mean filter can be used for the horizontal shift (h_s) and the vertical shift (v_s) if one desires to estimate sub-pixel or sub-line shifts in the output image. If nearest pixel or nearest line registration is desired, a median filter should be used.

A digital video system may vary its contrast and color saturation levels over time. This might result from system drift or from scene dependent behavior of the digital coding system. Time varying changes in the calibration quantities can be tracked by repeated application of the algorithm given in section 2.2. Once again, active output fields should be used. Scene dependent changes have been observed for MPEG systems and thus the set of output fields used to form the composite estimate should not cross scene cut boundaries.

3. Test Results

Table 1 gives a summary of recent results produced by the calibration algorithms in this contribution for a set of MPEG systems. This analysis has revealed that it is quite common for digital video systems to have substantial non-unity gains, level offsets, and horizontal and vertical shifts of the output video. In particular, note that active video shifts up to 8 horizontal pixels and 9 vertical <u>field</u> lines (i.e., 18 vertical <u>frame</u> lines) were measured.

Table 1 Measured Calibration Quantities for MPEG Systems

MPEG System	Gain, *g*	Level Offset, *l*	H shift, h_s (pixels)	V shift, v_s (field lines)
1	.95	-0.2	0	-8
2	.96	-0.9	-7	-8
3	.95	-1.4	3	-9
4	1.17	8.3	-7	1
5	1.17	7.7	-8	1
6	.90	-3.8	4	-8
7	.98	2.6	-7	1
8	.99	2.0	-7	1
9	.99	2.2	-7	1

Without proper calibration of the sampled output video, it has been observed that the frame matching algorithms for the visual channel delay standard (T1A1.5/96-101) can produce significant video delay measurement errors and poor output to input frame matching. Figure 2 is a plot of the square root of the comparison function for 1 second of input fields for an output field before calibration (light line) and after calibration (heavy line). This particular case occurred for MPEG system 2 in Table 1 for the scene “Being There”. The minimum in the uncalibrated curve occurs at input field 9 while the minimum in the calibrated curve occurs at input field 43. The scene "Being There" contains two men looking at each other and talking. One man is moving his head in a horizontal direction (to the right). Since the uncalibrated output field has been spatially shifted to the left, the man's head in the uncalibrated output field aligns to an input field that is too early in time by about half of a second (34 fields). The shallow minimum at input field 9 and the high residual error is due to incomplete cancellation of the background. When the output field is properly calibrated, the comparison function produces a well defined minimum at the correct matching input field 43.

The field to field ripple in the comparison function plots is because the output field matches one type of input field (e.g., NTSC field 1) better than the other type of input field (e.g., NTSC field 2). This ripple is to be expected since the two types of input fields (NTSC field 1 and NTSC field 2) are from slightly different spatial locations and the output field should match one type of field better than the other type.

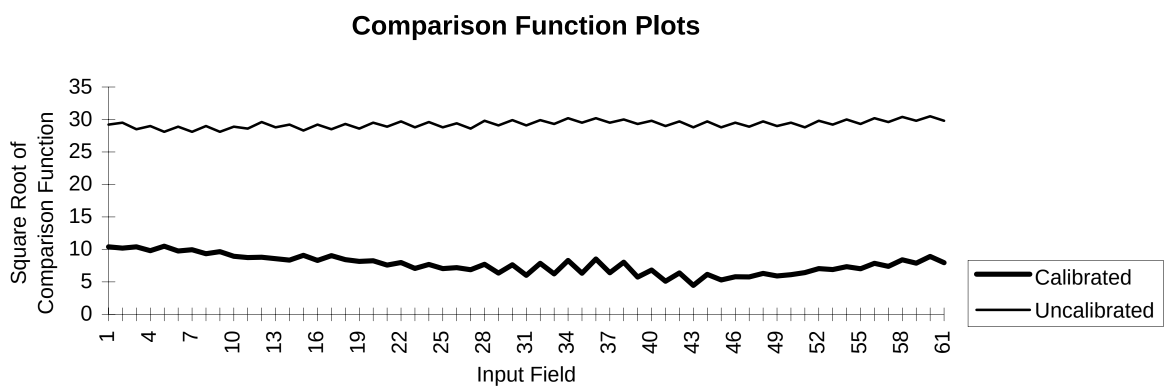

Figure 2 Comparison function results before and after calibration

4. Summary

A calibration algorithm has been presented that can simultaneously measure the gain *g*, level offset *l*, horizontal shift h_s, vertical shift v_s, and video delay of a digital video transmission system. The calibration algorithm in based on the methods presented in ANSI T1.801.03-1996, except that the input and output video can be from real scenes rather than deterministic test patterns. The algorithm is therefore anticipated to be useful if ANSI T1.801.03 test patterns are not available or if scene dependent variability of the channel calibration is present.

This contribution has demonstrated that using uncalibrated output video can result in measurement errors for visual channel delay. Since the computational resources required to implement the calibration procedures in ANSI T1.801.03 appear to be minor in comparison to the mean square error (MSE) frame matching methods being proposed for the visual channel delay standard, it would seem reasonable to specify these methods as a normative requirement in section 6.1.2.3 of the visual channel delay standard (T1A1.5/96-101).