Issue Faced in DTV Up-Conversion

By
Scott Ackerman
Director – Product Management
Teranex, Inc
http://www.teranex.com

Abstract

In the transition from standard definition (SD) to high definition (HD) television, broadcasters are faced with a need for HD programming for the new DTV channels. One solution to this is to use the existing SD program feed and up-convert it to an HD format.

As the transition progresses and more HD programming becomes available, the up-converter will still remain useful for broadcast of archival SD material in the HD domain.

There are several different techniques available for up-conversion. This paper will examine the different techniques, delineate a new technique and discuss its application in an up-conversion process. It will also review some of the issues faced in colorspace and aspect ratio conversion as appropriate in a DTV Up-Conversion.

Introduction

In the past, broadcast facilities where oriented around an analog composite signal (NTSC) path. Although the facility may have had several different types of tape machines, 1", 3/4", Betacam, each of these could be interconnected without the use of conversion equipment.

Today, with the transition to DTV, this is no longer the case. Now facilities are faced with having to interconnect or work with formats that have different numbers of pixels/line, lines/field, and in some cases a different number of fields or frames/second. In order to work with these new formats the broadcaster must rely on the use of a DTV format converter.

In the up-conversion process, in addition to changing the number of pixels and lines in input signals, it is also necessary to change the image colorimitry. The colorimetry of standard definition video is defined by ITU-R BT.601, while the DTV formats are defined by ITU-R BT.709. The conversion between the different colorspaces must be addressed in the format conversion process.

In addition to the above issues, there is also an issue of aspect ratio. The standard definition (SD) format is generally defined as a 4:3 aspect ratio while most of the high definition (HD) formats are in a 16:9 aspect ratio. The conversion of one aspect ratio to another is not only a technical one; it also presents a number of creative and artistic issues that must be addressed.

Up-Conversion Process

Moving images exist in three dimensions. In the horizontal direction they are made up of individual pixels. In the vertical direction they are made up of the lines contained in the field or frame. Pixels and Lines exist in what is referred to as the spatial domain. Finally there is the number of fields or frames per second, which is referred to as the temporal domain.

Up-conversion is a form of sample rate conversion in two or three of the above dimensions. It consists of expressing moving images sampled on one three-dimensional sampling lattice to a different lattice.

The process of interpolation is used to convert between these various spaces. Interpolation is defined as computing the value of a sample or samples, which lie off the sampling matrix of the source signal. In other words it is the process of computing the values of output samples that lie between the input samples.

The term up-conversion is generally associated with changing the number of pixels and lines in a format. Most format conversions do not involve temporal rate changes. Examples of format conversion include:

480i59.94 to 720p59.94
480i59.94 to 1080i59.94
576i50 to 576p50
576i50 to 1080i50

If we look at an up-conversion from 480i59.94 to 1080i59.94 we need to change two of the three parameters of the signal. The first is the number of pixels in each. A 480i signal has 720 pixels per line while the 1080i signal has 1920 pixels.

The second parameter that needs to be changed is the number of lines in each field (or frame). The 480i signal has 240 active video lines per field while the 1080i signal has 540 lines, as shown in Figure 1.1.

Teranex

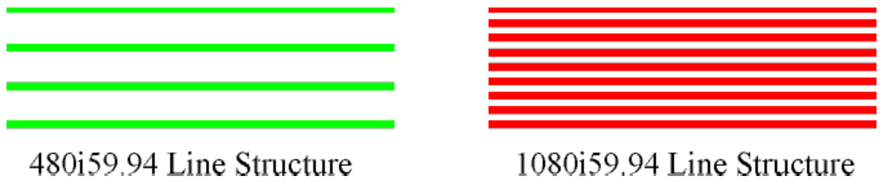

Figure 1.1 – 480i vs. 1080i Line Structure

The third parameter, the number of fields (or frames) per second does not change between the 480i signal and the 1080i signal.

One of the main issues faced in the up-conversion process is that of resolution. While it is possible to create pixels and lines based on input information, it is not possible to create resolution. It is therefore extremely important that the maximum resolution of the input signal be recovered and passed on to the output. One way to achieve this goal is to de-interlace the input signal to recover the full vertical detail of the input.

An Overview of De-Interlacing Techniques

Interlaced video signals consist of two video fields, one containing the odd lines and the other containing the even lines of the image. During the image capture process, the camera outputs the odd lines at one instant in time, and then 33 milliseconds later, outputs the even lines. This creates a temporal shift between the odd and even lines of the image, which must be addressed in frame-based processing systems. The de-interlace process attempts to overcome this problem by creating a clean frame from the two fields. There are two basic classes of de-interlacing algorithms: non-motion compensated and motion compensated. Both classes are described in the following sections.

Non-Motion Compensated De-Interlacing

There are two categories of non-motion compensated de-interlacing algorithms – linear and non-linear. Both categories contain spatial (or intra-field), temporal (or inter-field), and spatio-temporal algorithms. For brevity, only the most popular methods are examined.

Linear Techniques

The two most basic linear conversion techniques are called “Bob” and “Weave”. “Weave” is the simpler of the two methods. It is a linear filter that implements pure temporal interpolation. In other words, the two input fields are overlaid or “woven” together to generate a progressive frame; essentially a temporal all-pass. While this technique results in no degradation of static images, moving edges exhibit significant serrations, which is an unacceptable artifact in a broadcast or professional television environment.

“Bob”, or spatial field interpolation, is the most basic linear filter used in the television industry for de-interlacing. In this method, every other line (one field) of the input image is discarded, reducing the image size from 720x486 to 720x243. The half resolution image is then interpolated back to 720x486 by averaging adjacent lines to fill in the voids. The advantage of this process is that it exhibits no motion artifacts and has minimal compute requirements. The disadvantage is that the input vertical resolution is halved before the image is interpolated, thus reducing the detail in the progressive image.

A combination of the linear spatial and linear temporal methods is the linear vertical-temporal (VT) filter. The VT de-interlacer is the best performing of the linear filters. This method gradually reduces the vertical detail as the temporal frequencies increase. The contribution from the neighboring field is limited to the high vertical frequencies, such that motion artifacts can be minimized. Note that the vertical detail from the previous field is being combined with the temporally shifted current field, so some motion blur may occur.

Non-Linear Techniques

Linear interpolators work quite well in the absence of motion, but television consists of moving images, so more sophisticated methods are required. The field-weave method works well for scenes with no motion, and the field interpolation method is a reasonable choice if there is high motion. Non-linear techniques, such as motion adaptive de-interlacing, attempt to switch between methods optimized for low and high motion. In motion adaptive de-interlacing, the amount of inter-field motion is measured and used to decide whether to use the entire input frame (if no inter-field motion detected), or discard one field (if significant motion detected). Advanced motion adaptive techniques can vary the percentage of the previous field’s data that is retained as the inter-field motion increases. The challenge in implementing an effective motion adaptive algorithm is managing the trade-off between double images, due to the inclusion of too much uncorrelated previous field information, and lost vertical resolution, when not enough previous field information is used. Another challenge in varying the temporal aperture is avoiding “resolution pumping” as objects start or stop moving. The processing requirements for motion adaptive de-interlacing are higher than that of field interpolation, at the gain of somewhat higher output image quality.

Motion Compensated De-Interlacing

The most advanced method of de-interlacing is motion compensation. This technique, which at one time was thought to be too complex to implement in hardware, is currently used in advanced SDTV standards converters, and is beginning to appear in high-end

DTV format converters. Motion compensated de-interlacing measures the inter-field motion, and then aligns data between the two video fields, maximizing the vertical resolution of the image.

Motion compensation consists of shifting the pixels in the two temporally displaced fields to a common point in time. Determining the amount of shift for each pixel is referred to as motion estimation, in which motion vectors are identified and tracked from one field to another. This is typically implemented as a block-matching process. Typical block sizes range from 4x4 to 8x8. The implication of using blocks is that only these relatively large pixel blocks can be moved in time to the correct spatial position; with the further restriction of requiring every pixel in the block to move the same displacement.

The Teranex PixelCompä Algorithm

Teranex has developed an enhanced version of motion compensated de-interlacing called PixelComp™. The PixelComp™ motion estimator uses a full exhaustive search on 1x1 block sizes, generating a motion vector for every pixel in the image. This allows motion of any kind to be detected at pixel granularity. The pixel motion vectors are then used to precisely position the objects in the image to the correct spatial position, maintaining the full vertical resolution of the imagery even when significant motion is detected. PixelComp™ also includes advanced techniques for filtering motion vectors to insure their accuracy.

Colorspace Conversion

Up-Converters must adjust colorimitry as part of the conversion process, transforming the input colorspace to that of the output, making adjustments to both the luminance equation and the chromaticity coordinates as needed.

The history of colorspace goes back to the beginning of the century. A French organization, the Commission Internationale de L'Eclairage or CIE began to define all visible light. The CIE now regulates and defines color and practical color range for the print, photographic, and video industries.

In color theory the chromaticity diagram conveys color gamut, an abstract representation of "pure" color in the absence of brightness, as an 'x, y' plot. A sweep from a monochrome light source plotted on this diagram will result in a locus that encompasses all possible colors, as shown in Figure 2.1. Today's coding schemes, however, can only hope to represent a subset of these colors and as such the chromaticity coordinates are used to describe the bounds of this subset. The 'x, y' chromaticity coordinates then, define the possible colors of any coding scheme. For video, the coordinates of the RGB primaries define a triangular area on the chromaticity diagram that includes all the possible colors for an RGB signal.

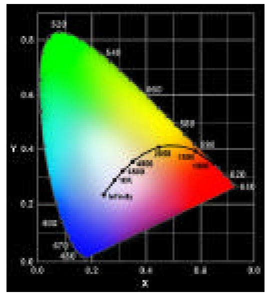

Figure 2.1 – CIE Coloespace

The CIE defines any color in terms of an XYZ tristimulus value. These are linear light values that embed the spectral properties of human color vision. For a video signal these XYZ values need to be expressed in terms of the reference RGB primaries and take into account the gamut restrictions imposed by the chromaticity coordinates of the specification. Accordingly a transformation matrix, Figure 2.2, is applied to the XYZ tristimulus, to obtain RGB, and this matrix is derived from the specified chromaticity coordinates of the RGB primaries and the white reference.

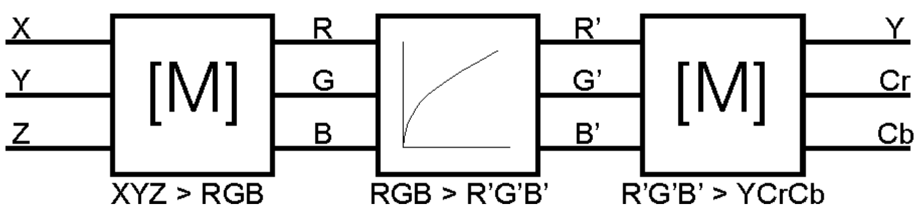

Figure 2.2 – Transformation matrix from CIE XYZ to YCrCb

In addition to the Luminance equation and chromaticity coordinates a signal's gamma must be taken in to consideration. Gamma correction is a non-linear transfer function that is applied to the reference RGB primaries and results in the non-linear primaries denoted by R'G'B'. Gamma correction is primarily used to correct for the non-linearity in the CRT display, but also has benefits for noise reduction and subjective image quality. Because gamma is a non-linear function it must be removed from the signal before any matrix changes can occur. Once the matrix change is complete the gamma may be re-applied to the signal.

Colorspace conversion is therefore the process of transforming YCrCb to XYZ and then from XYZ back to YCrCb. The steps used in this conversion process are listed below.

- Convert YCrCb to R´G´B´ using the matrix derived from the luma equation of the source color space.
- Convert R´G´B´ to RGB by applying the transfer function of the source color space on each component.
- Convert RGB to XYZ using the matrix derived from the primary chromaticities and white point of the source color space.
- Convert XYZ to RGB using the matrix derived from the primary chromaticities and white point of the destination color space.
- Convert RGB to R´G´B´ using the inverse transfer function of the destination color space.
- Convert R´G´B´ to YCrCb using the matrix derived from the luma equation of the destination.

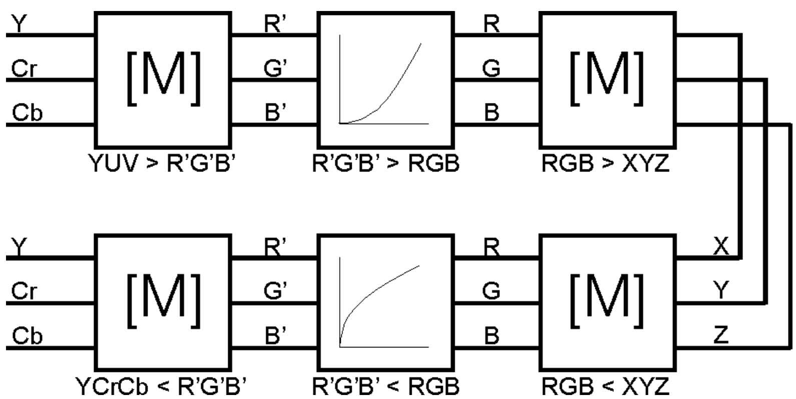

Figure 2.3 – Transformation matrix from YCrCb to CIE XYZ then back to YCrCb.

In practice the input is always a standard definition signal in an up-conversion, so the input transformation can be fixed. The Gamma correction is also the same for all standards and as such can be fixed as well. The two variables are the output XYZ > RGB and R'G'B' > YCrCb transformation matrices. The first of these is derived from chromaticity coordinates and the white reference and the second from the applied luminance equation.

Colorimity for video signals is governed by the video standards SMPTE 125M, SMPTE 240M, SMPTE 274M and SMPTE 296M; with reference to the two additional standards

SMPTE RP145 and ITU-R BT.709. These standards specify each of the three transforms required to convert from linear light to YUV.
SMPTE 125M – Rec. 601 (Standard Definition), uses SMPTE RP145 for chromaticity coordinates.

Chromaticity	**Red**	**x**	**0.6400**	**y**	**0.3300**	**Luma Eqtn**	**Y**	**0.2126R' + 0.7152G' + 0.0722B'**
	Green	**x**	**0.3000**	**y**	**0.6000**		**U**	**- 0.1146R' - 0.3854G' + 0.5000B'**
	Blue	**x**	**0.1500**	**y**	**0.0600**		**V**	**0.5000R' - 0.4542G' - 0.0458B'**
	White	**x**	**0.3127**	**y**	**0.3290**			

SMPTE 274M/296M – 1080i/720p (High Definition), uses ITU-R BT.709 for all colorimitry.

Chromaticity	**Red**	**x**	**0.6300**	**y**	**0.3400**	**Luma Eqtn**	**Y**	**0.2990R' + 0.5870G' + 0.1140B'**
	Green	**x**	**0.3100**	**y**	**0.5950**		**U**	**- 0.1690R' - 0.3310G' + 0.5000B'**
	Blue	**x**	**0.1550**	**y**	**0.0700**		**V**	**0.5000R' - 0.4190G' - 0.0810B'**
	White	**x**	**0.3127**	**y**	**0.3290**			

Aspect Ratio Conversion

Aspect ratio conversion is another factor that must be addressed in the up-conversion process. Most standard definition material is in a 4:3 aspect ratio while high definition material is in a 16:9 aspect ratio. It is important to remember that aspect ratio refers to the ratio of an image width to height, it is not related to image size.

The video industry generally expresses aspect ratios as whole numbers such as 4:3 or 16:9. Thus an image with an aspect ratio of 4:3 means that the image is 4 units wide by 3 units high. The value of the units is completely arbitrary. The film industry uses lowest common denominators to express aspect ratio. An image's height is assigned a value of 1 and its width is then a decimal unit of the height. The image expressed as 4:3 in the video industry would be expressed as 1.33:1 or 1.33 in the film community. See Figure 3.1.

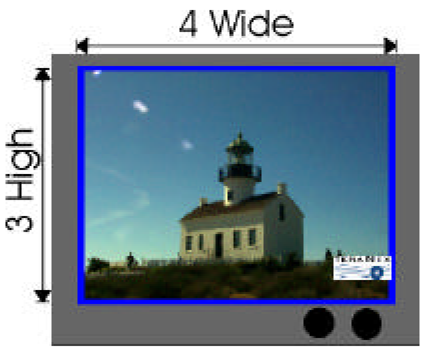

Figure 3.1 – Aspect ratio is the ratio of an image's width to height. It does not describe the image's size. In the video community this image would be 4:3, it is 4 unit wide by 3 units high. In the film community the same image would be 1.33:1 or just 1.33 as all aspect ratios in the film community use 1 as the height.

The technical issues of aspect ratio conversion are not very complex. The process does not require any information to be created, rather it is simply a process of cropping, stretching or squeezing the image. This manipulation does present a number of creative issues in how an image is resized and reshaped to change it aspect ratio. The issues discussed below will be limited to those faced in the up-conversion process of a 4:3 image.

If a 4:3 image is up-converted to a high definition format without any aspect ratio changes it is referred to as Common Top and Bottom. The top and bottom edges of the image match the top and bottom edge of the display device. This creates a pillarbox, as shown in Figure 3.2, or an image with black curtains on either side.

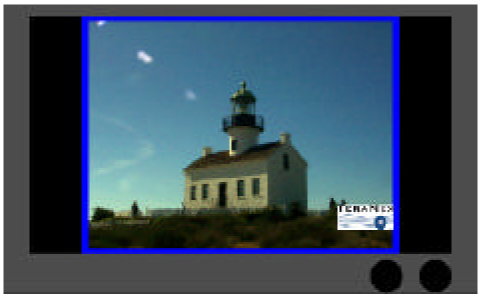

Figure 3.2 – Pillarbox or Common Top and Bottom

A derivation of this theme is to shift the actual image to the left or right thus creating one larger black bar, Figure 3.3 as opposed to two smaller ones on either side. This new larger area is then used as a background to key information in along side the video image. Information such as upcoming programs, weather and sports statistics, news, web pages, etc.

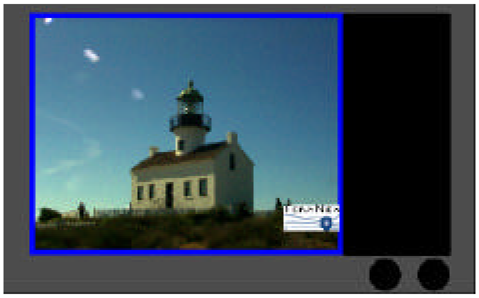

Figure 3.3 – Pillar Box with image shifted left

If the 4:3 image is stretched to fill the 16:9 output display it is referred to as 16:9 anamorphic, as in Figure 3.4. This mode is actually intended for material that was captured with an anamorphic lens. If it is used with standard material the result is a distortion of the geometry of the image (circles become ovals).

Figure 3.4 – 16:9 Anamorphic

As was seen above if the 4:3 image is stretched horizontally to fill the 16:9 display it results in a distortion of the geometry of the image. In order to correct this and continue to fill the output display the image must be stretched vertically as well. This will result in 'correct' geometry 16:9 but causes approximately 33% of the original input information to be lost. This mode is referred to as Common Left & Right or Common Sides, as seen in Figure 3.5.

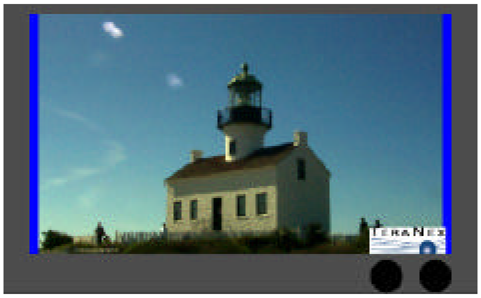

Figure 3.5 – 16:9 or Common Left and Right

The loss of the information in the vertical domain, in addition to its creative issues, does present a technical one as well. By cropping the lines off, less vertical information is made available to the interpolation process. This will effectively lower the overall resolution of the output image.

Another format, which has become known as the compromise format, is 14:9, shown in Figure 3.5. This format is generally shown with a correct geometry. It again requires both a horizontal and vertical stretch of the image. The horizontal stretch being 14 unit as opposed to 16 will result in small bars (pillarbox) on either side of the image. The vertical stretch used to maintain geometry is only 15% and thus results in less information being cropped. Like the 16:9 format above this will also have an effect on the overall vertical resolution.

Figure 3.5 – 14:9

Conclusions

The up-conversion process is important to broadcasters that are involved in the transition to DTV and HDTV. The up-converter allows broadcasters to put standard definition programming on their DTV channel now until HD originated programming is more readily available.

The use of a high quality de-interlacing process is very important in up-conversion. Since an up-converter must create more information than is available in the input format it is critical that the maximum amount of vertical detail is recovered from the input signal. The Teranex Xantus, with its PixelComp motion compensated de-interlacing process, can recover the full vertical detail of the input signal even in areas of motion.

Colorspace conversion is also important to maintain correct image color through the conversion process. It is important that this conversion be done in a gamma correct manner to eliminate and potential for error in the output signal.

Finally, the area of aspect ratio conversion is one that will impact both the engineering and the creative side of a broadcast facility. From an engineering standpoint one must ensure that the up-converter can address aspect ratio conversion. There may be a decision as to whether the facility should switch to 16:9 for all programming new and old. This decision would lead to changes in equipment such as cameras, and redesign of sets. From a creative side there is the issue of how to fill the new widescreen displays, what percentage of cropping or distortion will be acceptable.

References

Ackerman, Scott, "Film Sequence Detection and Removal in DTV Format and Standards Conversion", August 2000.

Deame, Jed, "Motion Compensated De-Interlacing: The Key to the Digital Video Transition", 1999.

Poynton, Charles, "Color Transforms for SDTV, HDTV, CGI, and Film", February 2000.