

Video-conversion techniques ensure a sharper image

BRUCE INTIHAR, GENESIS MICROCHIP INC

In this era of digital video, viewers pay much more attention to image quality. The old interlaced-video standards no longer meet the quality levels that many viewers demand. Deinterlacing offers a way to improve the look of interlaced video. Although converting one video format to another can be relatively simple, keeping the on-screen images looking good is another matter. With the right deinterlacing techniques, the resulting image is pleasing to the eye and devoid of annoying artifacts. Also, for the growing number of viewers demanding wide-screen displays, flexible aspect-ratio conversion is a must. The optimum approach gives users the choice of how to display the original image on the wide screen. Together, these video-processing techniques will enable any video source—whether it's a playoff football game, newscast, classic film, or recent blockbuster—to look the best.

Despite the resolution of digital-TV-transmission standards and the market acceptance of state-of-the-art video gear, a staggering amount of video material is still recorded, broadcast, and retrieved in the ancient interlaced formats. Every day, millions of people enjoy watching their favorite TV programs, and every one of these programs is broadcast in an interlaced-video format. Interlaced video has been around for many years, and although the circuitry in TV receivers has greatly improved, interlaced video has not changed in decades. In contrast, an increasing number of displays for PCs and other new video-graphics technologies—including those technologies employing the filmlike 16-to-9 aspect ratio—use progressive-scan techniques (see box “Interlaced video vs progressive scan”).

Low resolution and artifacts degrade image

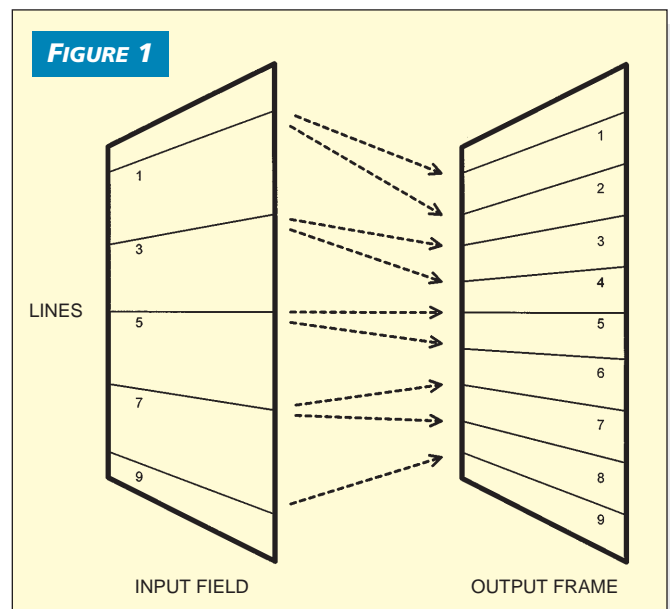
Interlaced video suffers from several problems: The basic vertical resolution is inadequate for displaying filmlike quality, and interlaced video generates visual artifacts, including edge flicker, shimmering, and diagonal jaggedness. Today's video broadcasts have low vertical resolution. NTSC systems,

Interlaced video is still everywhere, but it fails to meet the increasing demands of viewers for high-quality. Carefully chosen conversion techniques—for both deinterlacing and modifying the aspect ratio—result in an image of flexible size with no annoying artifacts.

for example, have just 480 active horizontal lines/frame. For static images, this vertical resolution is comparable with that of a VGA display. The resolution of PAL/SECAM systems, at 576 active lines/frame, is not much higher. Until a few years ago, these numbers were acceptable, but the success

of computer displays has demonstrated the viewing public's desire for higher resolutions and sharper images. Most desktop PCs shipped now have display resolutions of at least extended-graphics-adapter (XGA, 768 lines/frame) levels, and laptops feature at least SVGA resolution (600 lines/frame).

Low vertical resolution leads to visible scan lines in the image. When you combine this low resolution with inter-



The line-replication deinterlacing technique simply repeats each horizontal line in a field to create a complete frame.

VIDEO-CONVERSION TECHNIQUES

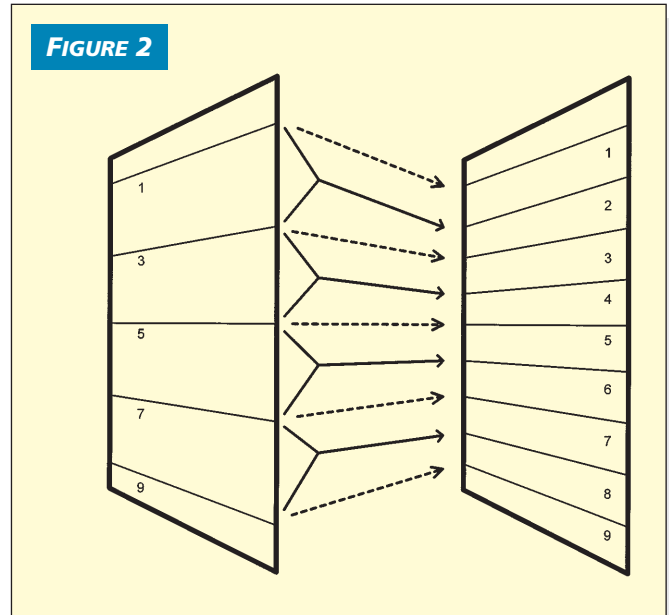
lacing, unpleasant artifacts appear on screen. These artifacts include flicker, shimmering, and diagonal-edge effects. Flicker is a video artifact in which image material appears to flash on and off at a high rate, similar to a star twinkling. Both wide-area and edge flicker are common.

Wide-area flicker appears over large areas of the screen, typically when the image content is bright and static and has little detail. Wide-area flicker results not from the interlaced video but from the low frame rate. You can also see wide-area flicker with a progressive-scan image at 60 Hz. If video-image data updates (refreshes) at only 60 or 50 Hz, your eyes notice the flicker. (This observation is common among people accustomed to NTSC when viewing a PAL or SECAM display.) Wide-area flicker is usually unnoticeable in areas of low intensity and fine detail because the eye is more interested in image content.

Edge flicker is a phenomenon related to fine detail and horizontal edges. A horizontal edge is an edge running horizontally across the image, such as the top of a door frame. Because of the nature of interlaced video, a fine horizontal edge may be present in only one of the video fields and not in the other. This fine detail is then displayed at the frame rate alone, either 30 or 25 Hz. A viewer finds this type of flicker extremely noticeable and annoying. If the image includes a little motion, the edge twinkles prominently as it moves vertically. Images such as shuttered windows are an interlaced-video nightmare, because the entire shutter scintillates during a slow pan.

Shimmering resembles wide-area flicker

Shimmering is an interlacing artifact that appears in images with little detail and no motion, similar to wide-area flicker. This phenomenon is more noticeable when the view-

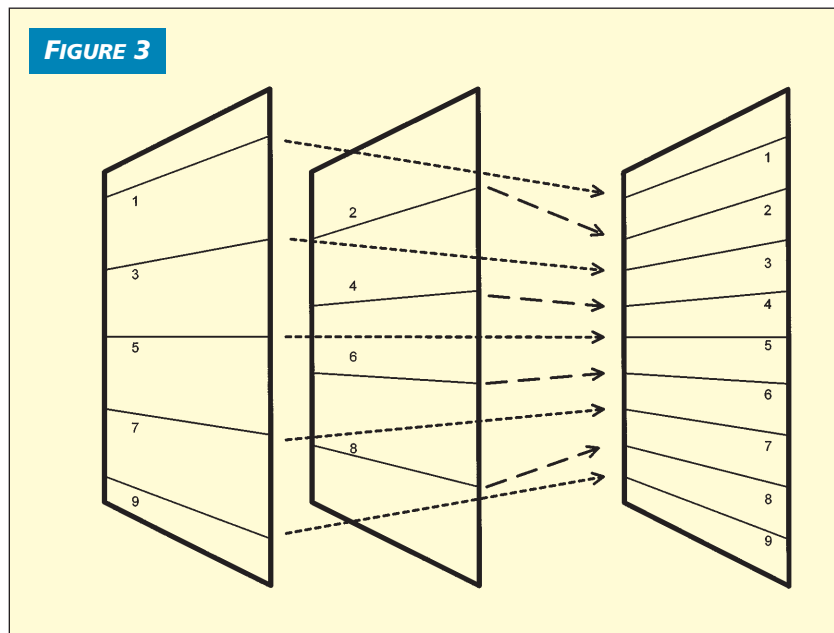


Vertical filtering creates missing lines by filtering a number of nearby lines.

er is close to the display. Because each scan line updates just once per frame, the eye can sometimes latch on to a group of lines on the screen. In the next field, new lines appear between the old scan lines. It can appear that the group of lines has moved up or down to the new position. In the following field, the group then appears as it did originally. The overall effect is that the group of lines rapidly moving up and down causes a sort of shimmering. This artifact is less noticeable when the viewer is farther from the display, because the eye tends to average out the interlaced lines.

Diagonal-edge effects, sometimes called "jaggies," are another common problem. These artifacts appear on both stationary and moving diagonal edges as a jagged saw-toothed pattern. A combination of interlacing and low vertical resolution causes jaggies. In static images or those with little motion, poor vertical resolution results in a stair-stepping effect as the edge displays across several lines. Interlaced edge flicker also affects these edges, causing a visible twinkling.

With motion, the problem becomes much worse. In this interlaced image, only half the lines of the edge update every field, resulting in a loss of detail at the edge. Furthermore, as the edge moves, the next field shows a displacement. The jagged edge from one field shifts slightly with a jagged edge from the next field, and so on. For this reason, broadcast video can make objects such as hockey sticks look like zippers.



Field merging takes lines from the previous field and inserts them into the current field to construct a frame.

VIDEO-CONVERSION TECHNIQUES

A number of reasons exist for deinterlacing video. First, the process can remove or greatly reduce interlacing artifacts. The resulting video is far more pleasing to watch without visible scan lines, flicker, and edge effects. Also, many applications, such as PC displays, projection systems, and videoconferencing, exclusively employ progressive-scan technology because you can more easily display, store, transmit, and manipulate progressive-scan video in these systems.

Broadly speaking, two fundamental approaches exist for converting interlaced video to progressive scan: static and dynamic (also called adaptive). Static techniques use the same overall conversion method regardless of the image source or content. Dynamic systems adapt and optimize the deinterlacing method based on the image content.

Static deinterlacing techniques use several methods, including line replication, vertical filtering, field merging,

and vertical-temporal filtering. Line replication simply repeats each horizontal line in a field to create a complete frame (**Figure 1**). Vertical filtering is a more sophisticated approach in which filtering a number of nearby lines, rather than simply copying the nearest neighbor, creates missing lines (**Figure 2**). Field merging takes lines from the previous field and inserts them into the current field to construct the frame (**Figure 3**). A better option is to use a hybrid approach of both field merging and vertical filtering. This vertical-temporal-filtering technique filters information both vertically across several lines and temporally from one or more adjacent fields (**Figure 4**).

Dynamic deinterlacing techniques, such as adaptive motion compensation and adaptive field pairing, use motion analysis to choose the optimum method. Adaptive-motion-compensation systems usually apply the same basic approaches as static deinterlacing, but these systems com-

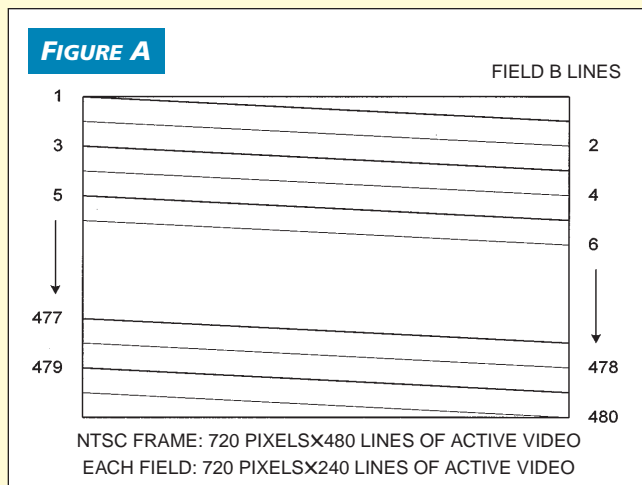
INTERLACED VIDEO VS PROGRESSIVE SCAN

In the early days of TV, transmitting a 525-line, 60-Hz video program in a 6-MHz broadcast channel was impossible; there was just too much information to cram into a 6-MHz allotment. To reduce transmission bandwidth requirements, engineers and broadcasters devised a method in which half of each frame (called a field) transmits every $\frac{1}{60}$ sec, followed by the other field $\frac{1}{60}$ sec later. Then, the transmission system weaves the scan lines of the second field between the lines of the first in an "interlaced" pattern (**Figure A**). The first field contains all the odd-numbered horizon-

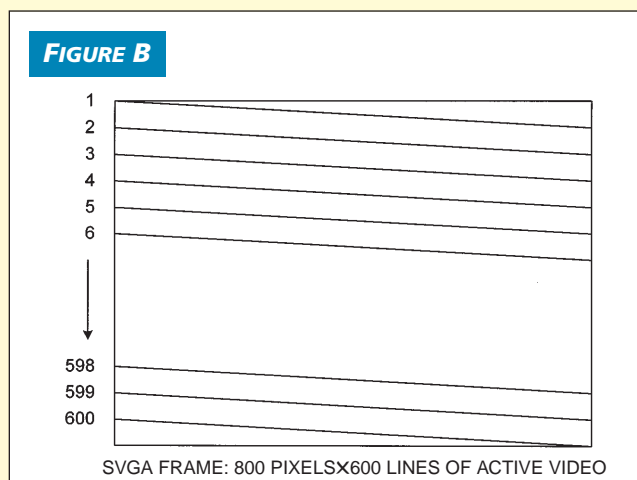
tal lines; the second field contains the even-numbered lines. Every image pixel updates every $\frac{1}{30}$ sec, but the eye tends to filter the change, giving a relatively pleasing image without exceeding bandwidth limitations.

All broadcast video—except for a few experimental high-definition stations—is interlaced. The major color-TV standards—NTSC, PAL, and SECAM—are all based on an interlaced transmission format. NTSC uses a 525-line frame (or $262\frac{1}{2}$ lines/field, of which 240 are active) at 60 Hz. PAL and SECAM use a 625-line frame (or $312\frac{1}{2}$ lines/field, of which 288 are active) at 50 Hz.

Progressive scan is a contrasting technique, in which all the lines of each frame transmit together (**Figure B**). Most PC monitors, which are often called noninterlaced monitors, use the progressive-scan method. A number of similar progressive-scan image formats exist, many of which were defined by VESA (Video Electronics Standards Association). A common progressive-scan display format is SVGA, with 800 pixels by 600 lines and a 72-Hz frame rate. Other formats range from 640 pixels by 480 lines (VGA) to 1600 pixels by 1200 lines (UXGA), each with a variety of frame rates from 60 to 85 Hz.



Interlaced video comprises odd and even field lines, which each transmit at $\frac{1}{60}$ sec. When the system weaves, or interleaves, these field lines together, the overall image's update rate is $\frac{1}{30}$ sec.



In progressive-scan video, all of the lines of each frame transmit together.

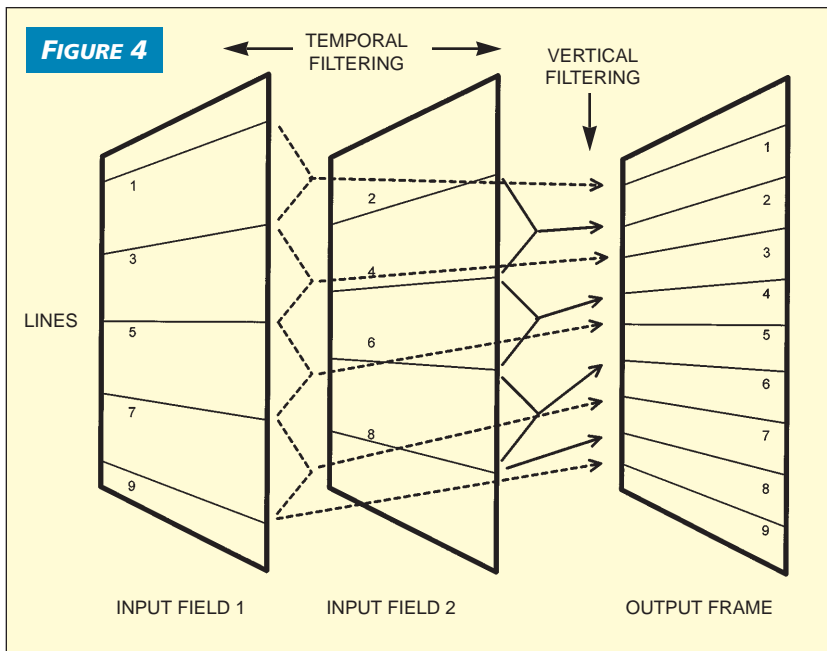
VIDEO-CONVERSION TECHNIQUES

bine the approaches on a pixel-by-pixel basis, depending on the amount of motion in the image sequence and within each image. Such a system always has a motion analyzer that looks at both the current and one or more previous fields to determine the motion content. Most analyzers have a motion threshold below which the system deems an object static.

Adaptive-field-pairing systems use a slightly different approach. In image sequences with little or no motion, a field-pairing system switches to a field merge. A field merge achieves the maximum vertical resolution by using information from both fields. In image sequences with plenty of motion, the system switches to vertical filtering. Vertical filtering avoids the “smearing” of objects with high motion from one field to another. This system can further enhance adaptive deinterlacing switching methods within a field as well as between fields. For example, in a static landscape with a bird flying, the still images should be merged, and the motion areas should be vertically filtered.

Note that these techniques apply to live video streams, usually captured with a video camera. Each field is spatially and temporally displaced from the previous field. In the case

of a video sequence converted from film, a special approach is necessary. You first convert the film to NTSC video by taking each film frame and creating either two or three fields in



The vertical-temporal-filtering technique filters information both vertically across several lines and temporally from one or more adjacent fields.

FILM ON VIDEO

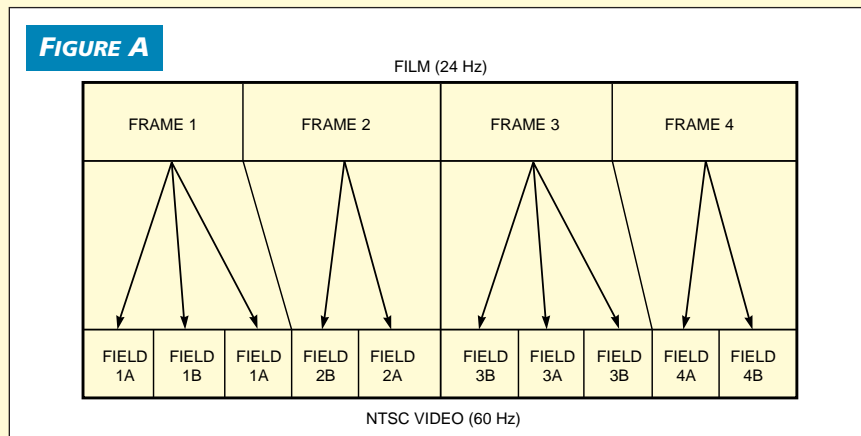
Cameras capture movies on film at a 24-Hz frame rate. This frame rate differs from that of NTSC video, which has a 30-Hz frame rate and 60-Hz field rate. To convert the 24-Hz film to 60-Hz interlaced video first requires scanning and separating the individual film frames into A and B fields. Unlike live video, these two fields have no temporal displacement; you can display fields A and B in either order because they both contain material shot at the same time.

Next, display the film fields one after the other, alternating between A and B with an extra field inserted every four fields (**Figure A**). This process is called “3-to-2 pull-down,” so named for the number of video fields “pulled down” from each film frame: three fields from frame 1, two fields from frame 2, and so on. The resulting field rate is now 60 Hz, which is easy to transfer to videotape.

You can convert film to PAL/SECAM video (25-Hz frame rate, 50-Hz field

rate) in one of two ways. The first method is to simply split the film frames into two fields and show the result as is. The result is a slight but imperceptible increase in the apparent speed of the

film scenes (essentially, viewing 24-Hz material at 25 Hz). The other method involves adding an extra field every 24 fields to align the 24-Hz film to a 50-Hz field rate.



Converting from film’s 24-Hz rate to 60-Hz interlaced video involves 3-to-2 pull-down, a process that inserts an extra field every four fields. A and B are identical copies of each film frame.

VIDEO-CONVERSION TECHNIQUES

a process called “3-to-2 pull-down” (see box “Film on video”). In this case, the fields created from a particular frame have no temporal displacement.

The correct approach to deinterlacing this type of video is to exclusively use field merging and to merge only fields coming from the same film frame. This process, “inverse 3-to-2 pull-down,” may require a merge with the previous field or the next field, depending on the incoming pattern. This process differs slightly from static field merging, in which the system always merges a given field with the previous field. Inverse 3-to-2 pull-down gives the best results with this type of video. You can use static and dynamic techniques with video converted from film, but the resulting image quality is not as high as that obtained from inverse 3-to-2 pull-down.

Carefully weigh deinterlacing options

All deinterlacing techniques have advantages and disadvantages. Static deinterlacing approaches are inexpensive, are relatively simple to design, and need no elaborate motion computation and analysis logic. The system cost is low because only a limited amount of video data storage is necessary. Line replication requires just one line store, vertical filtering needs a few line stores, and field merging requires only one field store. However, these options generally give mediocre output quality. Line replication and vertical filtering do not remove edge flicker because they use no image data from adjacent fields, and flicker is one of the most annoying interlacing artifacts.

These static deinterlacing approaches also suffer from poor vertical resolution, again because they use no data from adjacent fields. Field merging provides acceptable results with static images, greatly reduces flicker, and offers higher vertical resolution. Still, in motion sequences, the edges of objects exhibit terrible jaggies as the data shifts in the merged fields. This problem can be as objectionable as the flicker problem with line replication and vertical filtering.

Vertical-temporal deinterlacing is the best static option and gives the best overall performance. A well-designed vertical-temporal filter can yield excellent results over a range of image sequences. This approach reduces flicker and

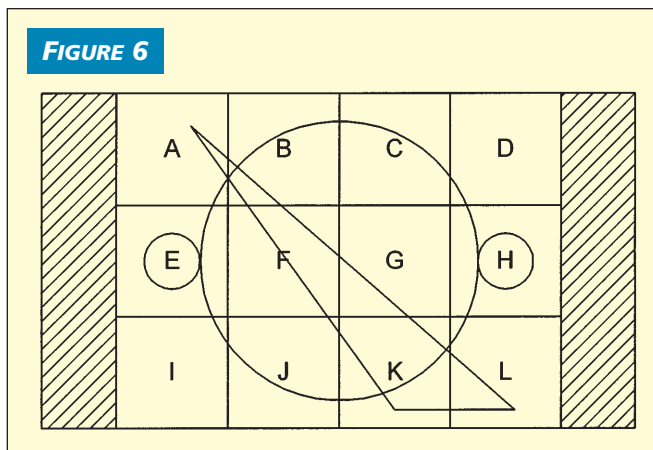


FIGURE 6 A direct copy of a 4-to-3 image onto a 16-to-9 display leaves unused areas on the right and left sides of the image.

increases resolution because of input from adjacent fields. Edge effects still exist but are much less evident than when using pure field merging. Vertical-temporal filtering is the method Genesis Microchip (Markham, ON, Canada) uses in its patent-pending series of single-chip video-line doublers.

Dynamic or adaptive deinterlacing approaches can also work. As stated, these systems try to select the best option or combination of options for the image sequence to be processed. However, these approaches suffer from high cost and complexity. The motion-analysis component is the foundation of the entire system. To adequately perform this function, you usually must store several fields of video for field comparisons and computations, which entails a high cost for external memory. The logic performing this analysis can be complicated. Also, an adaptive-motion-compensation system can sometimes incorrectly determine the speed and direction of an object’s motion. Thus, the system may not optimally filter the object, and artifacts appear at the object’s edges.

A big problem with adaptive-field pairing is that switching from one type of deinterlacing to another—field merging to vertical filtering, for example—is often visible. In cases in which the image sequence hovers around the motion threshold, the system can rapidly switch back and forth, causing objectionable artifacts. Even worse, the motion-analysis subsystem can make the wrong choice for certain image sequences and switch the deinterlacing process the wrong way. The result is an image full of artifacts.

Change the aspect ratio

Another challenge for digital-video engineers is aspect ratio, which is the ratio of the width to the height of an image or display. All broadcast video and most computer displays offer a 4-

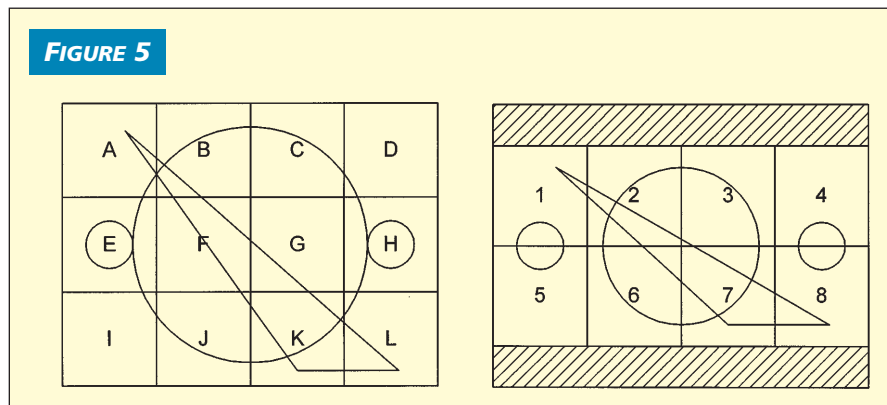
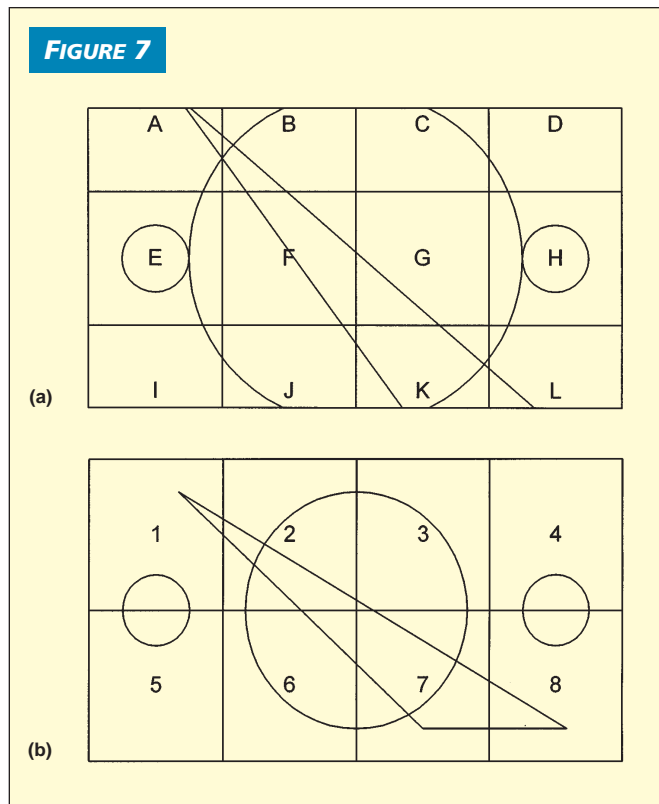


FIGURE 5 Aspect-ratio conversion techniques have different effects on an original 4-to-3 source image (a) and a letterbox 4-to-3 source image (b), the result of conversion to 4-to-3 from 16-to-9 film.

VIDEO-CONVERSION TECHNIQUES

FIGURE 7



A crop-and-zoom of the original 4-to-3 image in Figure 5a onto a 16-to-9 display removes vertical content from the original image (a). A crop-and-zoom of the letterbox source in Figure 5b shows no loss of information (b).

to-3 aspect ratio, in which the display is four units wide by three units high. In the push for higher video quality, both high-definition-TV and digital-video-disk sources are targeting an aspect ratio of 16-to-9, which is close to that found in movie theaters.

The video industry has often found conversion from 16-to-9 film to 4-to-3 to be a nuisance. You must convert films shot in wide angle (16-to-9) to 4-to-3 for viewing on a standard TV, hence the familiar phrase, "This material has been formatted to fit your screen." The two most common conversion techniques are "letterbox" video and "pan-and-scan." Letterbox video, which displays the film in full horizontal detail with black spaces above and below, suffers from a lack of vertical detail. Pan-and-scan systems capture and display only a certain central portion of the film, losing some of the original film content at the left and right edges of the image. Currently, pan-and-scan is more common because it fills the TV screen, but letterbox is becoming more common as new 16-to-9 displays become available.

With the rapid growth of 16-to-9 displays, the problem of converting 4-to-3 to 16-to-9 is getting some attention. Figure 5a shows an original 4-to-3 image, and Figure 5b shows a letterbox 4-to-3 image to illustrate the various conversion techniques. You can handle the conversion in a linear or nonlinear manner. Three common linear techniques exist

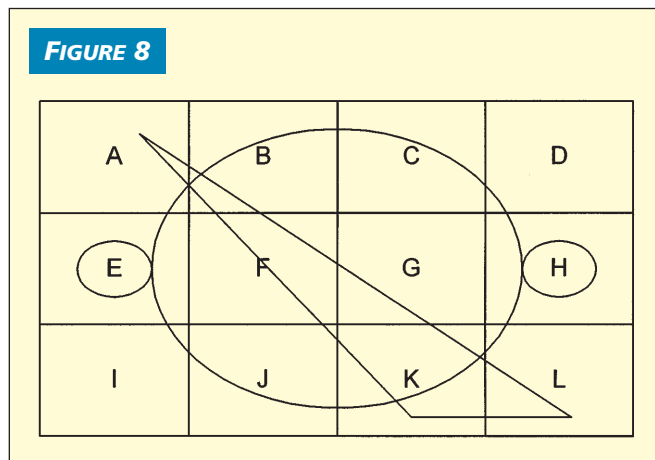
for viewing a 4-to-3 image on a 16-to-9 display. The first is to simply use a direct copy of the image with unused areas on the left and right parts of the display (Figure 6). The second method is "crop-and-zoom," in which the system removes the vertical content until the image is 16-to-9. The system can then display the result of crop-and-zoom of the original 4-to-3 source from Figure 5a (Figure 7a) and of the letterbox source from Figure 5b (Figure 7b). The third approach is to linearly stretch (zoom) the source material horizontally until the width gives a 16-to-9 ratio (Figure 8).

The direct-copy technique produces acceptable results because it does not distort or crop any part of the original image, but it does not give a filmlike viewing experience. The crop-and-zoom approach provides a full 16-to-9 experience and does not distort the image, but it can be undesirable when useful image content, such as the top of a newscaster's head, is lost. Still, crop-and-zoom is the best method when displaying letterbox video (Figure 7b). The horizontal-zoom approach retains all of the original image in a 16-to-9 format, yet it distorts image content, making people and objects look fat. None of these techniques is the best in all cases, and often the user has the choice of how to do the conversion. Many users do some of both, for example, a little crop-and-zoom followed by a little horizontal stretch.

Nonlinear techniques involve "panoramic zoom" (Figure 9). This technique is similar to linear zoom except that you don't stretch the image by the same amount in all places. Instead, you don't stretch the image at all in the center and increasingly stretch the image toward the edges. This stretching can occur only horizontally or both horizontally and vertically. The rationale for this type of stretching is that people most often watch the center of the screen and so the least distortion in this area is desirable. Material to the side is less important, so more distortion is tolerable. Panoramic zoom is basically a subtle image warp and can produce some unpleasant artifacts, such as curved walls and changes in the apparent speed of an object moving across the screen.

Ultimately, there is no textbook-correct method for solving the aspect-ratio problem. The only correct approach is a

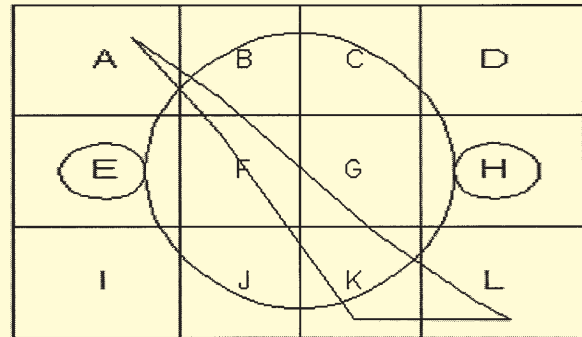
FIGURE 8



A horizontal zoom of a 4-to-3 image onto a 16-to-9 display linearly stretches the images horizontally.

VIDEO-CONVERSION TECHNIQUES

FIGURE 9



Panoramic zoom of a 4-to-3 image onto a 16-to-9 display is a nonlinear technique and is basically a subtle image warp.

subjective one chosen by the viewer, based on whatever personal criteria the viewer uses. Any system that supports aspect-ratio conversion should offer the widest range of choices for the user—ideally all of the linear and possibly some nonlinear approaches.

In all conversion techniques, scaling—usually image zoom—is necessary to change the source resolution to the target display and is critical to maintaining good image quality (Reference 1). Simple scaling techniques, such as pixel and line replication, are inexpensive but yield a “blocky” final image, which is generally inadequate for a high-quality system. Linear interpolation looks better but gives a softer image. A more effective approach is to use multitap filters in both the horizontal and vertical dimensions, giving crisp, clear results. □

Reference

1. Ngo, Calvin, “Image resizing and enhanced digital video compression,” *EDN*, Jan 4, 1996, pg 145.

Author’s biography

Bruce Intihar is a principal design engineer at Genesis Microchip Inc (Markham, ON, Canada, www.genesis-video.com), where he has worked for the past 5½ years. He is currently project leader for the next generation of Genesis Microchip’s deinterlacing-IC products. He has a BSc from the University of Waterloo (Waterloo, ON, Canada) and enjoys baseball, golf, and cycling in his spare time.

VOTE

Please use the Information Retrieval Service card to rate this article (circle one):

High Interest
598

Medium Interest
599

Low Interest
600