

# Digital Video Format Conversion

RAYMOND WESTWATER, PH.D.

MITCHELL TALISMAN

Futureware, Inc.  
Princeton, NJ USA

## ABSTRACT

The proliferation of digital standards has made video format conversion, once the exclusive province of exotic post-production houses, an essential component of virtually every editing and broadcasting studio. If content is king within the broadcast industry, surely quality is queen; the market for inexpensive yet highest-quality format conversion has become ubiquitous. A survey of existing technology will assist the engineer in making an informed choice for this critical component in the distribution chain.

## BACKGROUND

The coming of digital television, particularly high-definition television, has led to a proliferation of video format standards which must be handled by each broadcasting or editing studio. The Advanced Television Systems Committee (ATSC), charged with Federal responsibility for codification of digital video standards, has ratified forty-eight distinct formats, and Europe will presumably add more. While there may be good reasons for each individual standard, the burden of handling multiple data formats, particularly within the broadcasting studio, is an unfortunate side-effect of widespread adoption of this technology.

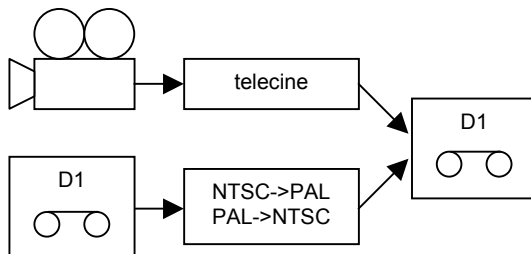


Fig 1. Video format conversion was once an off-line process performed exclusively in editing studios.

When once the primary format conversions were telecine and NTSC to PAL conversion in the post-production studio (figure 1), today's broadcasters

internally must support SD and HD content, newscasters must be able to accept HD content in any format from the field, and legacy libraries must be playable in either SD or HD format (figure 2). Many technologies have evolved for performing these types of conversions, each with its own set of advantages and disadvantages. It is incumbent upon the broadcast engineer to understand these technologies in order to assist management in the selection of high-quality conversion equipment at an acceptable price.

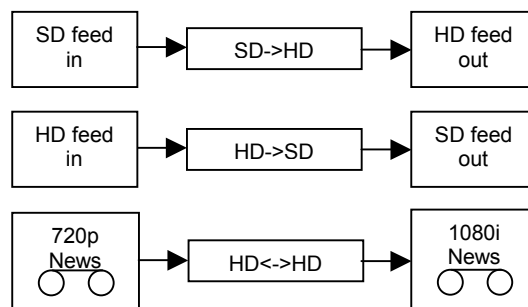


Fig 2. Video format conversion is now an integral part of the broadcast / news distribution chain.

## INTERLACING

Digital television bridges the gap between analog television based upon 1940's technology and modern technology displays. One important distinction can be seen in phosphor technology - analog television phosphors embrace interlaced format while many modern displays prefer progressive format. As illustrated in figure 3, analog phosphors illuminate brightly on being struck by a scanning electron beam and their intensity slowly diminishes; this property lead to the development of the interlaced video format which minimizes intensity changes from top to bottom of the screen by alternating the scan of even and odd lines (a field is composed of all even or all odd lines; a frame is composed of an odd field followed by an even field, figure 4.)

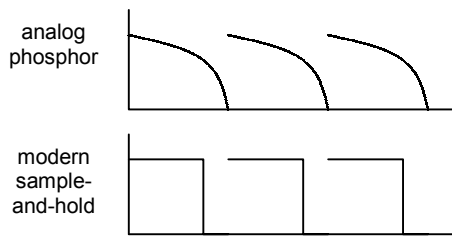


Fig 3. Analog phosphor intensity diminishes over a single frame time; modern screen technology (DLP) may hold the full intensity until the next frame time.

As further illustrated in figure 3, modern display technology may support longer persistence than phosphors. The difficulty this presents for interlaced format video is the simultaneous display of both fields, leading to a blurred picture (adjacent scan lines are separated in time by one field time.) Thus conversion from interlaced format (a frame is composed of two fields) to progressive format (a frame is composed of all scan lines, figure 4) must often be performed to show analog video on a digital screen.

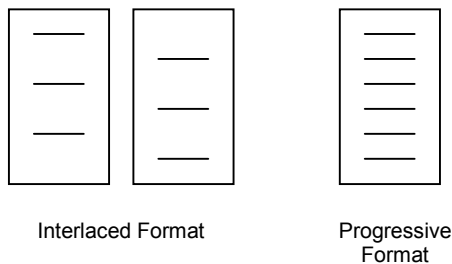


Fig 4. Interlaced format paints the odd scan lines in the first field, followed by the even scan lines in a second field. Progressive format sequentially paints all scan lines.

## FRAME RATES

The rate at which frames are animated is an important part of the video format. Several important frame rates have come to be adopted in the broadcast industry, particularly 24, 25, 29.97, 30 and 60.

The frame rate of 29.97 was adopted at the introduction of color by the United States as a  $1\%$  reduction from the frame rate of 30 used for black and white transmissions (for transmission quality purposes.) All content for US SD broadcast (NTSC) must eventually be rendered to play at this rate.

The European community chose the frame rate of 25 for its video transmission standard (PAL). Content mastered in the US must be converted to this frame rate for European broadcast and vice-versa.

Modern film is generally recorded at 24 frames per second. This rate must be adapted for transmission in NTSC or PAL format; conventional special-purpose conversions are embedded in telecine technology.

Finally, HD signals in the United States are transmitted at 30 frames per second or 60 frames per second.

## SCREEN FORMATS

Compelling new screen geometries have entered the home with the introduction of HD transmission in the United States. Screens come in two aspect ratios, legacy 4:3 and wide-screen 16:9.

Legacy aspect ratio screen support two display geometries, legacy 480X704 and square-pixel 480X640.

Wide-screen aspect also supports two display geometries, 720X1280 and 1080X1920.

## CONVERSION TECHNIQUES

One may consider that each digital video format may be specified by its screen resolution in scan lines by pixels, its frame rate, and its interlaced property. Common formats are enumerated in table 1.

Name	Lines	Pixels	Rate	Interlaced
NTSC	480	704	29.97	i
PAL	576	704	25	i
480p	480	704	60	p
720p	720	1280	60	p
1080i	1080	1920	30	i
film 2K	1080	1920	24	p

Table 1. Common video formats.

Format conversion is the process of rendering a video stream from one format to another while preserving as much of the subjective visual experience as possible. Two basic techniques have been explored for this purpose: interpolation (line-doubling, filtering, curve-fitting, and statistical prediction) and feature extraction (motion adaptation and motion estimation.) These

techniques are applicable in different contexts and have different tradeoffs in their use.

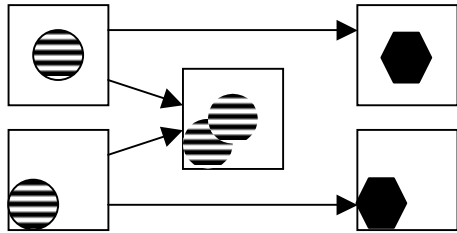


Fig 5. Motion between fields prevents simple combination of fields into progressive frames. Line doubling may be used to create a progressive frame from each individual field.

### Line doubling

Line doubling is used to upconvert from one scan rate to another. A simple example is upconversion from 480i to 480p. Interlacing odd and even fields into a single frame results in objectionable artifacts ("feathering" or "double imaging", figure 5.) Line doubling is a simple and effective means of upconverting - the lines of the each field are simply displayed twice to form a progressive frame. While this technique does increase sawtoothing, it does present a clear and compelling picture.

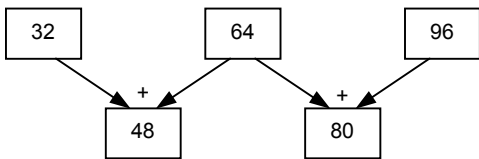


Fig 6. Filtering generates a weighted average of a pixel sequence to produce intermediate pixel values.

### Filtering

Filtering is another computationally inexpensive means of format conversion. A filter effectively generates a moving weighed average of samples; such averaging may be used to upsample the number of pixels on a scan line or change the number of scan lines (figure 6.)

Filtering notoriously reduces the clarity of samples, and may be combined with special-purpose filters to create a sharper feeling. These enhancement filters tend to fail on certain video sequences, and cannot be relied upon for faithful reproduction of video quality.

### Curve-Fitting

Curve fitting fits a pixel sequence to a predefined curve (figure 7). Cubic interpolation and fractal interpolation are probably the most popular techniques.

Curve-fitting techniques give very smooth, clear and compelling renderings. However, the results are often cartoonish, as the curve chosen may not correspond to the expected behavior between sampled points.

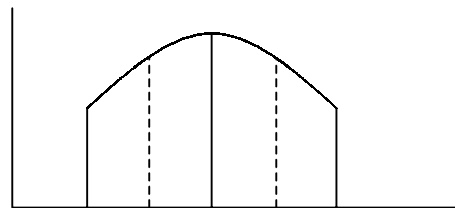


Fig 7. Interpolation under curve-fitting. A cubic is fitted through three pixel values (solid vertical lines) and intermediate values are calculated by evaluating the resulting curve.

### Motion Adaptation

A simple feature extraction technique is motion adaptation. In this technique, pixels whose values are stationary over time are identified (figure 8.) These pixels are used in lieu of interpolated pixels in certain contexts (particularly deinterlacing.)

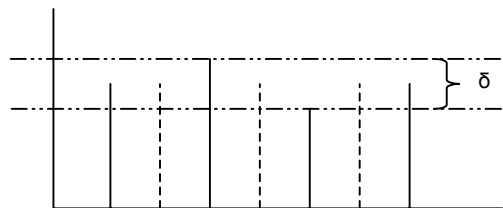


Fig 8. Motion adaptation. Pixel sequences the vary within a small range ( $\delta$ ) are replaced with the mean value in the interpolation.

The definition of "stationary over time" is somewhat problematical, as quantization noise and background variations may influence pixel values to a small degree. Thus some deviation within the pixel values must be built into the definition. Also,

behavior of nearby pixels may be considered in making the decision to declare a pixel stationary.

### Motion Estimation

A far more computationally complex means of feature extraction is motion estimation. This technique attempt to locate moving objects within a sequence of frames (figure 9). In this technique, a foreground object is mapped on top of a background (the background may also be moving!)

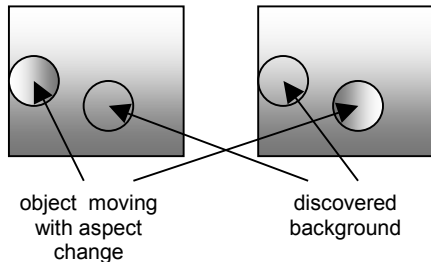


Figure 9. Motion estimation attempts to identify moving objects and the background each reveals.

Motion estimation is by no means a panacea for format conversion. It is not only expensive, but cannot account for changes in the aspect of the moving object (rotations, perspective change, illumination change, etc.) Motion estimation must also be used in conjunction with an interpolation method in order to scale the resulting image.

### Statistical Prediction

Mathematically ideal interpolation is based upon a statistical analysis of the video data. Data are transformed into a "basis" of statistically independent functions, which are then interpolated to produce intermediate pixel values (figure 10).

Statistical prediction yields the best possible guess for interpolated pixel values, giving visually compelling and accurate renderings.

### DEINTERLACING

The first step in converting from an interlaced format is deinterlacing. This process attempts to determine the best pixels to fill in the missing lines from each field, producing two full frames from the two fields of a single interlaced frame.

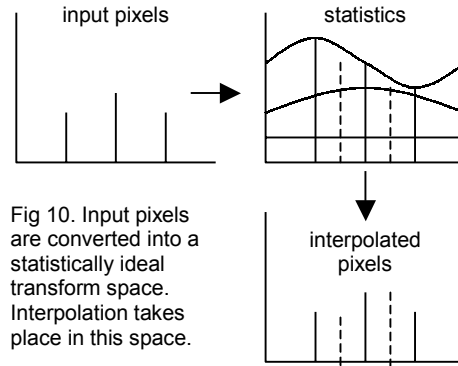


Fig 10. Input pixels are converted into a statistically ideal transform space. Interpolation takes place in this space.

The missing lines are generated by some combination of interpolation and feature extraction. In simple implementations, interpolation alone may be used to generate the missing data. This interpolation may be made across scan lines or fields, or both (figure 11.)

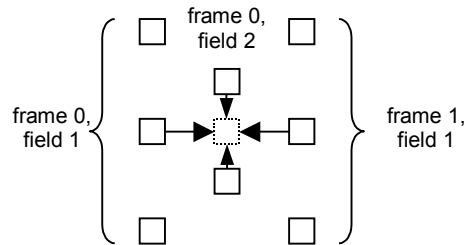


Fig 11. Deinterlacing via interpolation.

As most means of interpolation introduce visual distortions (loss of detail, color aberrations, etc.), deinterlace often combines motion adaptation with scan-line interpolation to improve clarity. Line-doubling does not require this fix. Statistical prediction, when used in the field direction, also gives good results without motion adaptation.

### FRAME RATE CONVERSION

Frame rate conversion will normally be applied to progressive frames (statistical prediction can perform frame rate conversion directly upon interlaced fields.)

A technique analogous to line-doubling may be applied to create intermediate frame (i.e., frame-doubling). This technique is known as "pulldown" (figure 12.)

Pulldown is performed by choosing, at each frame time of the rendered stream, the input frame closest in time. Examples of common pulldown applications are telecine and NTSC to PAL conversion.

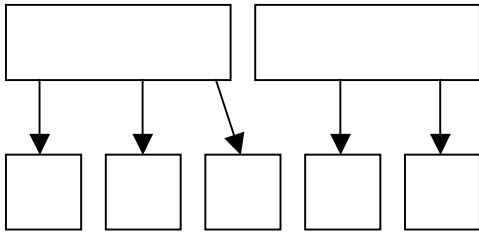


Fig 12. An example of pulldown converting 24 fps film to 60 fps NTSC. This process is known as "3:2 pulldown" as each frame of film is alternately assigned to between 3 and 2 fields of the NTSC video stream.

Interpolation may be used to create intermediate frames. Once again, with the exception of pulldown and statistical prediction, this technique is best supplemented with motion adaptation.

The filtering interpolation algorithm introduces multiple images when motion exceeds a certain threshold. (figure 13.)

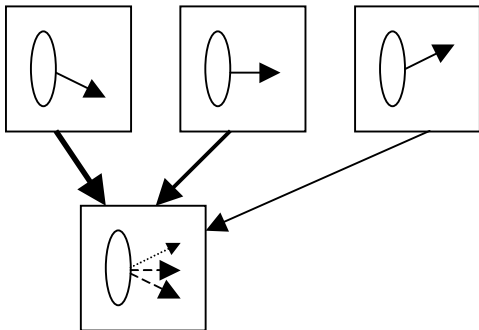


Fig. 13. Filtering causes ghosting of multiple images to occur in interpolated frames when motion is excessive.

Motion estimation is the technique used to remove ghosting in filter interpolation. The object moving and its background are determined by measuring interframe correlation. The object is then placed at its expected position in each interpolated frame.

Statistical prediction support "continuous pulldown" by predicting frame content at any discrete interval. Novel applications are supported by this approach including fractional frame

synchronization, conversion from 29.97i to 60p, and time base correction.

### SCREEN GEOMETRY CONVERSION

Once frame rate conversion has taken place, the screen may be sized to the display geometry. One of the interpolation techniques is used to perform this action.

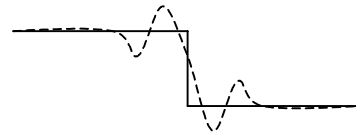


Fig 14. Interpolation across an object boundary causes ringing near the boundary.

Edge conditions can cause visible artifacts (figure 14.) Continuous curve fits tend to fail in the vicinity of abrupt discontinuities; detection and special handling of these discontinuities is an important refinement to all interpolation techniques.

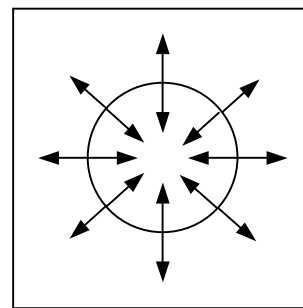


Fig 15. Directed interpolation proceeds away from the boundary of the object, eliminating ringing.

These artifacts can be mitigated with directed interpolation at an object's boundaries (figure 15.) This technique is used as an adjunct to high-end motion estimation algorithms. The boundary of the detected object is used as the origin for directed-interpolation techniques.

Statistical prediction incorporates edge analysis as an integral part of its processing. The visibility of ring artifacts is assessed, and directed interpolation only at boundaries where induced artifacts would otherwise be visible. This guarantees best possible results irrespective of visual content.

## SUMMARY

The popular format conversion technologies are summarized in table 2.

Problem	Technique	Comments
Deinterlace	Line Doubling	Clear results but increased aliasing
	Filtering + Motion Adaptation	Good results, inexpensive
	Statistical Prediction	Best results, expensive
	Motion Estimation	Best results, most expensive
Frame Rate Conversion	Pulldown	Jerky motion
	Filtering + Motion Adaptation	Ghosting
	Statistical Prediction	Best results, expensive
	Motion Estimation	Best results, most expensive
Screen Resolution	Interpolation	Artificial look and feel
	Object-Directed Interpolation (Motion Estimation)	Better, most expensive
	Statistical Prediction	Best results, expensive

Table 2. Summary of rendering techniques.

A comparison of price versus performance is shown in figure 16.

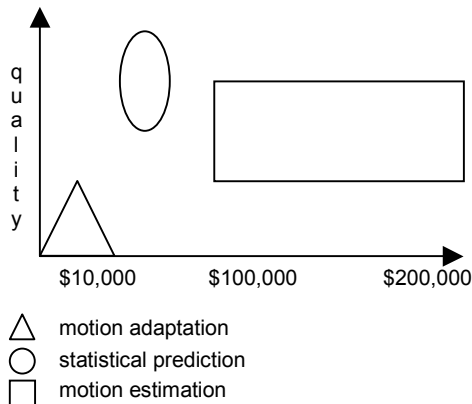


Fig 16. Price-performance comparison of video format conversion techniques.

## CONCLUSION

Format conversion is an important part of the modern broadcast studio. Provision of good high-definition rendering is perhaps most affected by this single piece of studio equipment, and the general look and feel of a multi-million dollar

facility may be most significantly affected by this single unit.

Furthermore, time-sensitive (newsworthy) content from the field may be received in any format, with more and more of the content coming from overseas. High-quality format conversion removes the post-production studio from the new distribution chain, shortening time-to-air and reducing critical dependencies.

Finally, the transition to HD support introduces its own conversion problems. Storage of media in its original format requires real-time in-studio conversion for its practical use while storage of converted content will overwhelm already taxed storage facilities.

Once the need for high-quality in-house format conversion is acknowledged, the problem of technology choice arises. The budget-minded will choose motion-adaptive solutions and accept quality loss in its HD offerings while the quality-conscious have, until recently, been restricted to six figure systems. Recent advances in the field of statistical prediction have made possible a moderately priced system that may better fit studio requirements.