Frame rate conversion - a moving target

When do we need frame rate conversion? In the cinema when projecting 24Fps material at 48 or 72 Fps. During post-production when combining material shot at different frame rates or changing speed for effect. For high frame rate displays in order to reduce flicker. When delivering to different countries and different display devices.

The convention is that motion imaging has a native frame rate. This is the number of images created over a particular period of time. When a ramp-up or ramp-down is done during creation the native frame rate must be specified as it is no longer constant throughout the piece. Also time-lapse or high speed filming is done with a planned presentation frame rate different from the capture frame rate. How do we know what the "native" frame rate of a piece of video is supposed to be?

> Did you ever get a tape that was marked as 24Fps and was actually 23.976? Really not a problem as your VTR will tell you what's right. If however it was a file that needed new subtitles the timing would be off by the end, but a spot check at the beginning would not show this.

The simplest frame rate conversion presents the original frames at a different rate than that at which they were shot. Time lapse footage shot at 1 frame per minute and played back at 30 Fps or 25Fps for television is a typical example. The most common example is when moving from film at 24Fps to PAL at 25Fps, one simply plays back at 25Fps and shifts the audio pitch. The duration is reduced by 1/24 every second, but we cannot see this. "Gone with the Wind" is 238 minutes long, the same version on TV in Europe is 228.5 minutes, many people think that 10 minutes have been cut out!

The television frame rate in the Americas is 30Fps. Speeding up from film would definitely be visible, therefore 2:3 pulldown was invented, inserting additional frames so that the duration remains the same. In most theatres film is actually projected at 72Fps via a shutter showing the same frame 3 times before the next frame is moved into place. When the frame rate for US television was standardised (1941) the best that could be economically achieved was 30Fps interlaced at 2:1. This provided a satisfactory picture in the living room environment with minimal motion artefacts. By the 1930's the UK had already standardised on 25/50, so moving between these 2 different frame rates was also required. Today it is possible to synthesise frames based upon the information contained in adjacent frames. Originally this was done via frame blending, which was quite acceptable for broadcast television, but causes visible artefacts in still frame. Modern digital techniques allow much more accuracy in reconstructing the missing frames.

> The most complicated frame rate conversion I have been involved in was a recent shoot in 3D of a classical pianist. This was originally shot for European television at 1920x50i. It was our job to meet the needs for a 3D Blu-ray. Specification limitations require a 23.976 source due to the increased data rate required for simultaneous playback of 2 streams of 1920x1080 video. Pitch shifting the audio was not an option (the piano player would of shot us). Dropping video frames resulted in visible judder, unacceptable in such a prestige project. New frames had to be synthesised, and in sync for both the left and right eyes.

If duplication or deletion of frames is not an option (duration has to stay the same and the resulting judder is unacceptable), then new frames will need to by synthesised. How these frames (or fields) are generated is a trade-off between time, money and the resulting accuracy. No matter which method is used (blending or interpolation) the number of frames to be created is a primary factor. The smaller the frame rate change the more new frames required per second, ie going from 24Fps to 25Fps allows the reuse of only 1 frame per second whereas going from 24fps to 36fps allows for the reuse of 3 frames per second. This is only the case if frames have to stay in the same relationship. If the difference between adjacent frames is insignificant then more frames may be reused. This is a trade-off between spatial accuracy (reusing the original frame) and temporal accuracy (the objects in the new frame are where they are supposed to be in time).

The simplest method for frame syntheses is blending; this can actually be done in the analog domain by using a delay with multiple taps equalling the number of frames you wish to blend. Digital frame memories have made this the basis for the low cost standard converters available today. Add more memory and some processing to compare frames so that original frames may be used as often as possible and the results can be quite good. When the blending algorithm can use different percentages from adjacent frames this is called adaptive motion interpolation.

The next step is to cut and paste from adjacent frames only the portions which have changed. This is done by dividing each picture up into a number of blocks and then estimating where the changes would be in the new frame.
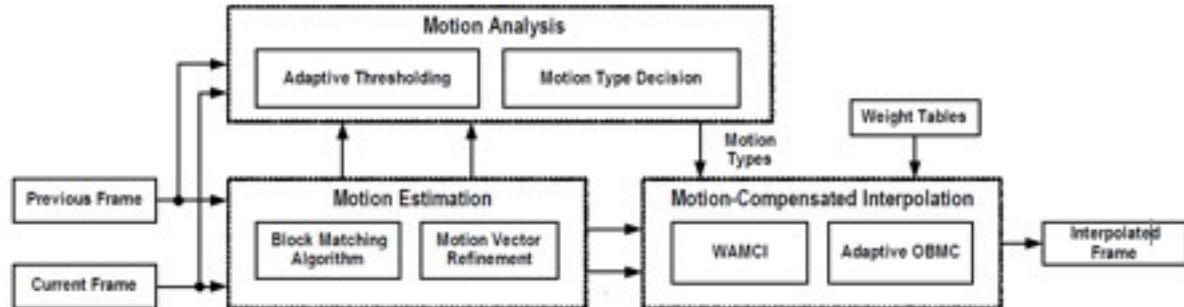
The most highly regarded method of frame rate conversion utilises phase correlation to estimate the motion and thus accurately define the missing pixels.

The difference is in the details, by translating the picture into the frequency domain using FFT it is possible eliminate irrelevant information from the motion vector calculation. Manufacturers of real time motion vector compensated frame rate converters for production applications can be counted on one hand, and each of these excels when converting different material.
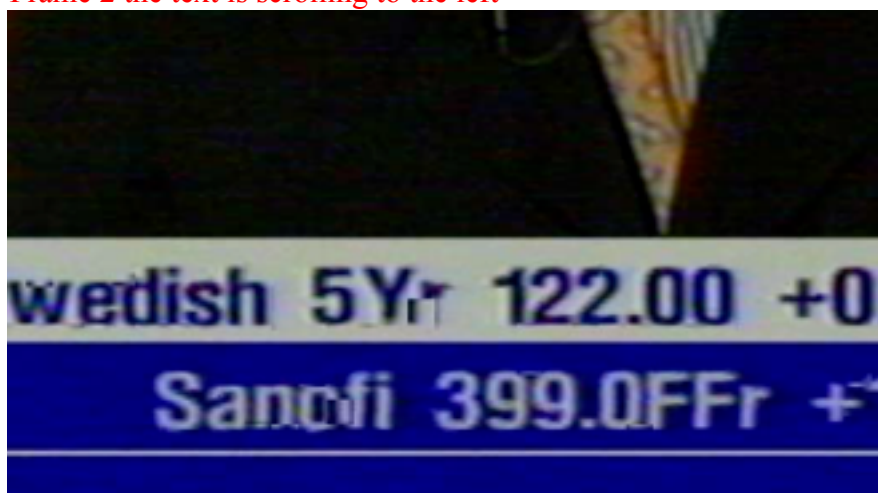
http://dasan.sejong.ac.kr/~dihan/dip/6a_WA_FRUC.pdf

Frame 1

Frame 2 the text is scrolling to the left

Frame between 1 and 2 Interpolated using standard methods, notice the blocking artefacts

Using Weighted-Adaptive Motion-Compensated Frame Rate Conversion

Algorithms to do this are a research topic at major universities world wide. Commercial and non commercial implementations of these algorithms are available. ASIC implementations of these algorithms are built into high frame rate TV sets under names such as Motionflow, Motion Picture Pro, Perfect Pixel, or HyperReal Engine. If render time is not a criteria then excellent results can be achieved using software implementations of these algorithms.

The underlying algorithms for frame rate conversion are the same as those used for video compression. It is hoped that the advancements in video compression and the very real possibility of applying these advancements to frame syntheses will result in more choices. More choices and falling prices will make for better programmes as the creative options provided by frame syntheses in post become available to a larger community of users.