Chroma Upsampling From Motion

John V. Sandusky
EE368 Digital Image Processing, Spring 2010
Electrical Engineering Department, Stanford University
Stanford, United States of America
jvsndsky@stanford.edu; jvsandusky@comcast.net

***Abstract*—Decompression of 4:1:1 YCbCr video by linear or nearest-neighbor interpolation renders color detail at low resolution and can alias color patterns, resulting in distracting artifacts. Advanced decompression engines minimize aliasing, but do not draw on sampled chroma typically available from neighboring video fields or frames. In this article, I demonstrate high-fidelity chroma decompression by motion registration on the luma channel and subsequent interframe median-filtering of the registered chroma samples.**

Keywords-4:1:1 YCbCr; chroma ; motion; registration; luma

I. Introduction

The great variety of color video compression engines currently in commercial use typically share chroma decimation as a core technique because human observers are more sensitive to luma detail than chroma [1]. The simplest chroma decimation scheme is consistent with with the raster-scanning format implemented on cathode-ray-tube displays, performing chroma downsampling independently along each video line. For example, 4:1:1 YCbCr compression samples the two chroma channels Cb and Cr at ¼ of the rate at which it samples the luma Y channel, and sampling is independent along each row, as shown in Figure 1. Display of such compressed color video on a pixilated monitor requires that the missing chroma values be filled in, which is typically accomplished in consumer electronics by simple linear interpolation. Such interpolation yields poor color detail and can result in aliasing.

Here I demonstrate substantial improvement of the color detail and fidelity of decompressed 4:1:1 YCbCr video, compared to interpolation, by using sampled chroma values from previous frames aligned by luma registration. While the 4:1:1 format may be subsiding in general use, its format provides a useful schema for exploring the extent to which chroma upsampling from motion might be applied to modern chroma compression schemes such as performed by the MPEG standards.

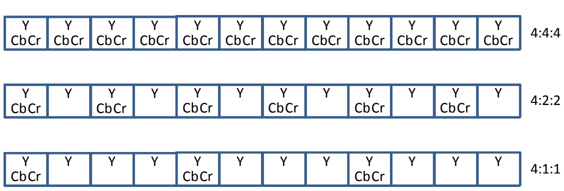

Figure 1. Sampling of luma (Y) and chroma (Cb and Cr) content.

II. Prior and Related Work

MPEG-2 encoding, for example, often makes use of interframe data to achieve even more compression than sampling chroma independently on each line can provide. Such interframe encoding schemes have proven performance and have become the standard for high-definition consumer video. Video in the 4:1:1 YCbCr format should first be restored to its most detailed and faithful color content to receive the most benefit from conversion to MPEG. If quality rather than speed is paramount, the additional time required to perform chroma upsampling from motion may be warranted.

III. Method of Chroma Upsampling from Motion

Figure 2 illustrates the concept of chroma upsampling from motion as the copying of chroma samples from neighboring frames by first aligning the frames using luma feature tracking. Chroma upsampling from motion is fundamentally different than upsampling by interpolating along the sampled row positions in that the chroma samples of neighboring frames are used. In this manner much of the original chroma content can be restored to the frame with high fidelity, provided that registration can be performed with the necessary accuracy.

Generalized motion registration is an enormous task requiring the segregation of moving objects from background scenery, interframe tracking of such objects, and perspective transform. Here I restrict motion registration to simple whole-frame translation, to demonstrate the concept of chroma upsampling from motion in its simplest implementation.

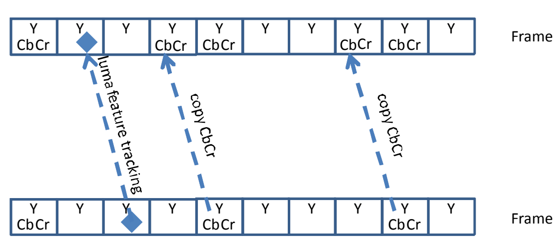

Figure 2. Chroma upsampling from motion copies chroma samples.

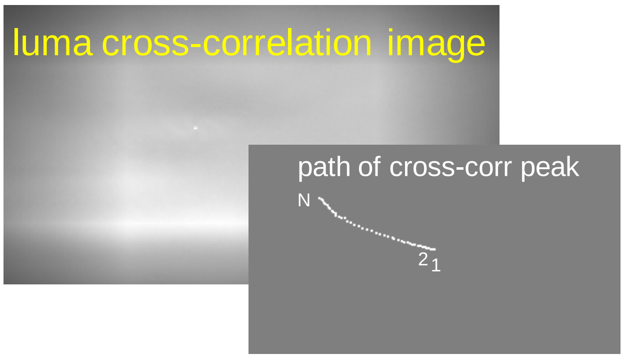

Figure 3. Interframe registration by luma cross-correlation.

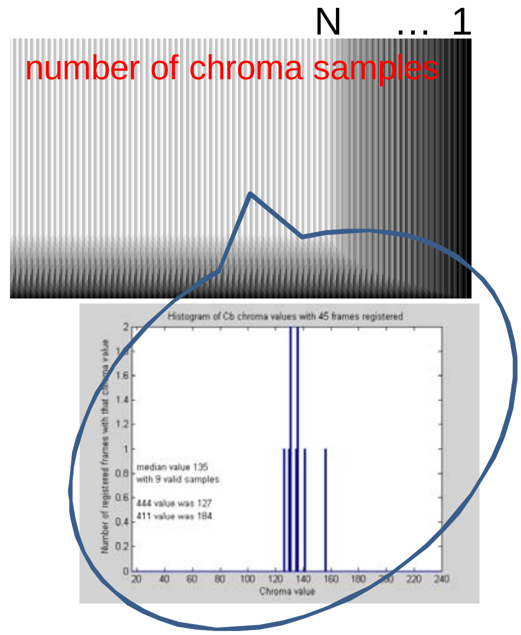

Figure 4. Overlapping chroma samples form a histogram for each pixel.

While the 4:1:1 YCbCr encoding scheme is most specifically applied to interlaced video, the results developed here are based on the application of 4:1:1 to non-interlaced video. This approach simplifies the demonstration without being overly restrictive, as the set of even or the set of odd fields can be thought of as independent images to which chroma upsampling from motion could be independently applied.

A. Luma Registration

Cross-correlation is a well-established technique for registering images to one another. Figure 3 shows the cross-correlation map of the luma channel from the frame to which chroma is to be restored (the 'base' frame) and the central half of the subsequent video frame, after first filtering each image by the Laplacian-of-Gaussian method to mitigate the brightness bias of correlation. Using only the central half of the subsequent video frame for registration was found empirically to be a good compromise between speed and reliability. Each frame is 180 rows by 320 columns, which is a small-scale version of the 16:9 aspect ratio in common use for high-quality displays. I define registration as the location of the peak of the cross-correlation pattern and do not attempt to interpolate registration to subpixel accuracy, even though the interframe motion is subpixel.

The lower image in Figure 3 shows the motion of the peak of the cross-correlation map from one frame to the next, revealing the motion of the scene. About 50 frames have been registered, as enumerated in the map by the numbering (1) of the base frame with respect to itself, (2) the second frame registered to the base frame, and (N) the final frame registered to the base frame. At each successive registration I check the magnitude of the interframe motion and, if it is more than five pixels, the registration result is discarded and registration ceases. This precaution guards against introducing vastly erroneous registrations to the results.

B. Chroma Resampling

As luma registration proceeds, an increasing number of chroma images overlap the chroma samples of the original frame. The brightness of the upper image in Figure 4 is proportional to the number of overlapping chroma samples from the registered sequence of motion video. Due to the direction of the simulated camera motion, more frames overlap in the upper-left of the image than in the lower-right. Indeed, at the extreme right edge no chroma samples from subsequent frames overlay the base frame. In these locations, chroma upsampling from motion fails to improve the color quality and the chroma values must still be filled in by interpolation in the usual fashion.

In the regions of overlap, however, a large number of overlapping frames generates the set of chroma values which are plotted in the lower image of Figure 4 as a histogram. Figure 4 shows that, although 45 frames were registered by their luma pattern to the base frame, only 9 of them have Cb samples that overlay the particular pixel indicated by the callout. Since the 4:1:1 YCbCr sampling method provides a Cb value for only one out of every 4 pixels, 11 to 12 of the 45 registered frames may have been expected to provide overlaying Cb values. Only 9 were provided, due to the particular simulated camera motion.

Figure 4 further shows that the original chroma value of 127 for this pixel had been eliminated by the 4:1:1 sampling method and that its linearly interpolated value would have been 184. Rather than linear interpolation, the method of chroma upsampling from motion generates the chroma value for this pixel as the median of the histogram of registered chroma samples, which is 135. This chroma value is much closer to the original value than linear interpolation would have produced, and is indicative of the overall color quality improvement that this method has the potential to deliver.

The use of the median value of several valid samples implies that the method is robust against minor disturbances in the registration quality. Subpixel registration accuracy is not necessary if enough samples are present in the histogram since, on average, the median value should be the value that subpixel registration would have obtained at greater computational expense and complexity.

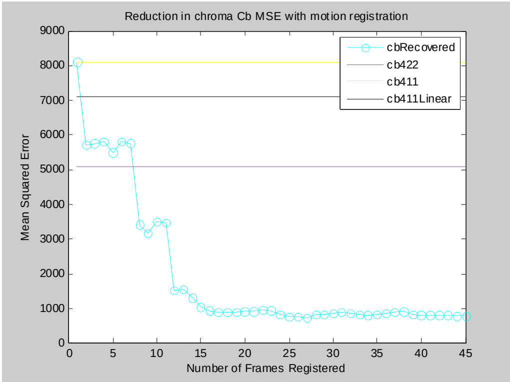

Figure 5. Cb MSE decreasing with the number of registered frames.

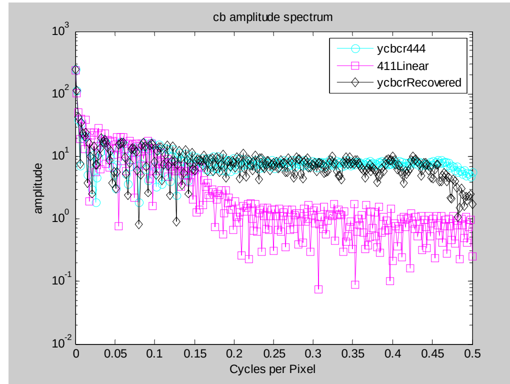

Figure 7. Cb amplitude spectrum of the decompressed color zoneplate.

C. *Recovering the Video*

The luma values of each frame are never disturbed in this process. Combining these luma values with the median value of the registered chroma samples restores 4:4:4 color quality to the base frame without interpolation, except in regions where no chroma samples registered from subsequent frames. Repeating the process by considering each frame of the 4:1:1 YCbCr movie to be a base frame converts the movie to 4:4:4 YCbCr while largely avoiding artifacts from linear interpolation.

IV. Experimental Results

I have tested the chroma upsampling method by manufacturing motion test patterns in Matlab and by generating motion video from RGB stills through cropping and subsequent sub-pixel interpolation along a user-interactive path. While the motion video is sub-pixel, registration remains at the whole-pixel level. RGB representations of the intermediate products are retained along the way for the sake of visualization. The registration process and chroma resampling all occur in the YCbCr color space.

A. *The Mean-Squared-Error Metric*

Mean-Squared-Error (MSE) is the mean of the sum of the squared differences between frames recovered from 4:1:1 YCbCr sampled video and original 4:4:4 YCbCr video frames, a common metric for gauging the quality of image processing. In these test cases the original 4:4:4 YCbCr values are all known, so the accuracy of the recovery of Cb values can be related as MSE. Figure 5 shows the MSE of the Cb values for the Zoneplate test pattern (Figure 6) as the number of registered frames increases. The yellow reference bar at the top is the MSE that would be obtained by simple nearest-neighbor interpolation from 4:1:1 to 4:4:4, and the black line is the MSE that would be obtained by linear interpolation. The cyan circles plot the MSE for the method of chroma from motion, connected with a cyan line as a guide for the eye. As soon as two frames have been registered, the MSE for chroma from motion is already better than linear interpolation. After eight frames have been registered, the chroma from motion MSE is less than the MSE that upsampling 4:2:2 by nearest-neighbor interpolation would have produced. After 15 frames have been registered, the MSE for chroma from motion ceases to improve significantly.

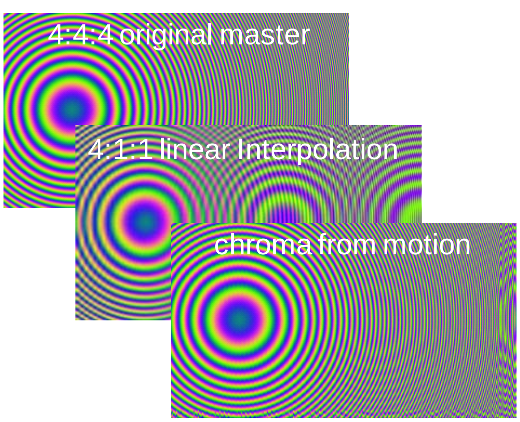

Figure 6. Color zoneplate test pattern.

This result implies that a practical system to perform chroma upsampling from motion might need to buffer only the previous four to eight frames. While the method is computationally intensive, the number of buffered frames is not far from the four fields buffered in the recent motion-adaptive de-interlacing offered by the Altera corporation [2].

B. *The Color Zoneplate Test Pattern*

Figure 6 shows a color zoneplate test pattern comprised of sinusoidally varying luma, Cb, and Cr values with local spatial frequency that increases linearly from the center of the pattern, such that the Nyquist rate is reached along the center row at the right-hand-edge of the frame as shown. The luma, Cb, and Cr patterns have the same spatial frequency and a 90 degree phase shift with respect to one another. The luma channel is here plotted as red, the Cb channel as green, and the Cr channel as blue. This test pattern is convenient for quantitative study of the accuracy of fine color detail, by plotting Fourier transform along the central row.

The top image in the figure is the full original zoneplate image. The second image in the figure is the image produced by linear interpolation from the 4:1:1 YCbCr video, which is the typical method employed by consumer-grade video decompression. Note the strongly aliased pattern in the center of the image and again at the right-hand-edge. Also, a block-

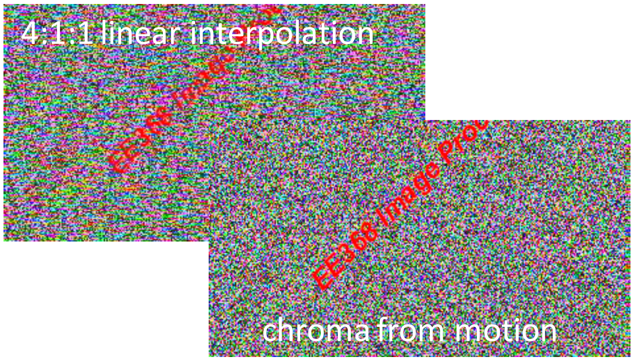

Figure 8. Colored text demonstration image.

pattern has developed along the rows due to the interpolation. The image at the lower end of the figure is produced by chroma upsampling from synthetic motion spanning of about 1/3 of the width and 1/5 of the height, with random subpixel steps. About 50 frames were registered and used to increase the sampling of the chroma channels. Note that the aliased pattern present at the center of the linearly interpolated image is greatly suppressed. The aliasing at the right-hand edge of the image remains, however, since the motion video was such that no additional chroma samples overlaid that edge.

C. Chroma Spatial Frequency Response

Figure 7 shows the amplitude of the spatial frequencies of the Cb values for the original 4:4:4 zoneplate image, the 4:1:1 downsampled zoneplate reconstructed using linear interpolation, and the 4:1:1 reconstruction using chroma upsampling from motion. Chroma upsampling from motion far outperforms linear interpolation, restoring frequency content with much higher amplitude beyond 0.15 cycles per pixel. The technique provides marginal improvement at Nyquist, where the limitation of whole-pixel registration is expected to play a dominant role.

D. Qualitative Comparison

Figure 8 shows red text ovelaid on a field of random luma and chroma values that have been converted to RGB. The upper image is a single frame of a synthetic motion video converted from 4:1:1 YCbCr to 4:4:4 YCbCr using linear interpolation, and then converted back to the RGB color space. The text is largely illegible.

The lower image shows the same frame converted from 4:1:1 YCbCr to 4:4:4 YCbCr by the chroma upsampling from motion method. In this image the text legibility is greatly improved. Whereas the color zoneplate provided a quantitative comparison of the two conversion methods, the text image provides a visually more familiar, yet qualitative, comparative restoration metric.

V. Conclusoins

Chroma upsampling from motion has the potential to substantially improve the color detail and fidelity of motion video converted from 4:1:1 YCbCr to 4:4:4 YCbCr, compared to conversion using traditional linear interpolation. The technique provides much greater frequency response and much lower MSE in the chroma channels, and the qualatative visual improvement is strong when viewed as isolated frames. However, much work would remain to implement this method for generic video of moving objects and scenery, owing to the inherent requirement to track motion between frames. The method may have its best application in achieving the highest quality restoration of 4:1:1 video clips before subsequent conversion to modern MPEG-2 or better standards.

References

[1] S. Qin, S. Ge, H. Yin, J. Xia, and I. Heynderickx, “Just noticeable difference in black level, white level and chroma for natural images measured in two different countries”, Displays, vol. 31, pp. 25-34, January 2010.

[2] Altera Corporation’s “Video Image Processing (VIP) Solution” marketing material, 2008.

Appendix

This paper, the corresponding Matlab code and movies produced therefrom, and the accompanying poster are the sole work of the author.