ON THE DETECTION OF TEMPORAL FIELD ORDER IN INTERLACED VIDEO DATA

David M. Baylon

Motorola, Inc., Home & Networks Mobility, San Diego, CA

ABSTRACT

In traditional analog interlaced video such as NTSC [1], the timing of the two fields is determined within the analog video signal. However, in today's converging world of analog and digital video, it becomes increasingly possible that information about field order is lost or not known. The analog interlaced video may undergo digital sampling, editing, or processing which alters or removes the field order information. This paper presents a simple and efficient approach for detecting the correct field order from only the interlaced video data. The method is based on proposed measurements of "zipper" points and energy of the interlaced video. Simulations show that the approach provides very promising detection performance.

Index Terms—video signal processing, image motion analysis, filtering, displays

1. INTRODUCTION

For display, compression, or processing of interlaced material [2], it is important to maintain correct field timing. If the top and bottom (or even and odd) fields are displayed in reverse chronological order, visual artifacts can occur especially for high motion scenes. Video compression and processing with incorrect field order can result in a loss of compression efficiency and video quality.

In many video applications, the proper scan or display field order can be obtained from temporal side information transmitted with the video. However, when this video is digitally captured, edited, or stored in a file, the field order information may be lost or incorrect. This paper proposes a simple motion-based approach for detecting the field order using only the interlaced video data, where each successive pair of top and bottom fields of the interlaced data is interleaved into single frames. Although interlaced motion detection has been widely studied, such as in de-interlacing [3-4], its application to field order detection does not appear to have received attention.

Section 2 presents the approach and an implementation using a proposed "zipper" filter. The analysis of the detection algorithm in Section 3 provides insight into some variations. Simulation results are provided in Section 4, and Section 5 gives some concluding remarks.

2. PROPOSED METHOD FOR FIELD ORDER DETECTION

Interlaced video displayed in correct field order tends to exhibit a smooth motion flow. However, if displayed in the wrong order, the video will exhibit a jerky backward-and-forward motion. Although an analysis of both motion magnitude and direction would be useful for detection, this paper proposes a simpler method that uses only magnitude information.

Consider two consecutive frames, curr(0) and next(1), and their respective fields: curr_top(0t), curr_bot(0b), next_top(1t), next_bot(1b). Let "motion(i,j)" indicate the relative amount of motion magnitude between i and j. Assuming typical object motion, for top field first sequences it might be expected that due to the smaller temporal interval, there will be less object displacement between (1t,0b) than between (0t,1b). Therefore, one approach to detect the correct field order is to compare motion(1t,0b) with motion(0t,1b) and to use the following "rule" (where either field can be chosen for the case of equality):

top first: motion(1t,0b) < motion(0t,1b)
bottom first: motion(1t ,0b) > motion(0t,1b).

There are various ways to measure "motion", but many require significant computation. A simple approach to measure interfield motion can be made by observing that when the fields are pair-wise interleaved within a frame, moving objects exhibit the well-observed "zigzag" or "zipper" effects especially near object boundaries. For measuring interfield motion, consider applying the following simple N-point (N even) linear shift invariant vertical "zipper filter" $h_z[n_1,n_2]$ over an interlaced frame:

$$h_z[n_1,n_2] = \begin{cases} (-1)^{n_2}, & n_1 = 0 \quad and \quad -(N/2-1) \le n_2 \le (N/2) \\ 0, & otherwise \end{cases} \tag{1}$$

The value of N is chosen even so that the same number of lines are filtered from each field. Since the filtered output magnitude gives an indication of the amount of interfield motion, let a "zipper point" be defined as a filtered output point whose magnitude (or magnitude squared) is greater than some threshold T. If the filtered output of frame (0t,1b) has more zipper points than frame (1t,0b), it will be said to be more "strongly interlaced" than the latter, and the sequence will be detected to be top field first. An outline of this proposed approach can be summarized as follows:

1. Let the fields of the current frame 0 be (0t,0b) and those of the next frame 1 be (1t,1b). Assemble two new frames $x_0[n_1,n_2]$ = (1t,0b) and $x_1[n_1,n_2]$ = (0t,1b), corresponding to the current frame 0 and next frame 1, respectively.

2. Let the filtered output to $x_i[n_1,n_2]$ using the zipper filter $h_z[n_1,n_2]$ be $y_i[n_1,n_2]$, i = 0,1, and let $C_T(y_i[n_1,n_2])$ represent the number of "zipper" points in $y_i[n_1,n_2]$ which have a magnitude larger than a specified threshold T.

3. Then proposed method for field order detection uses the following decision rule (if the denominator is zero, then the numerator and denominator are compared):

$$R_N^T \equiv \left(\frac{C_T(y_1[n_1,n_2])}{C_T(y_0[n_1,n_2])} \right) \underset{\textit{bottom first}}{\overset{\textit{top first}}{\gtrless}} 1 \qquad (2)$$

where in R^T_N, N refers to the length N zipper filter used to generate $y_i[n_1,n_2]$, and T refers to the threshold. The next section explores this basic method in more detail and provides insight into some variations.

3. ANALYSIS AND EXTENSIONS OF THE DETECTION METHOD

It is possible to eliminate the threshold T in step 2 above by simply comparing the sum of absolute values (or squared values) of the filtered outputs $y_0[n_1,n_2]$ and $y_1[n_1,n_2]$. This "zipper energy" comparison leads to the following decision rule, where l = 1 or 2, corresponding to an L^1 or L^2 type norm, respectively:

$$R_N^l \equiv \left(\frac{\sum_{n_2}\sum_{n_1} |y_1[n_1,n_2]|^l}{\sum_{n_2}\sum_{n_1} |y_0[n_1,n_2]|^l} \right) \underset{\textit{bottom first}}{\overset{\textit{top first}}{\gtrless}} 1 \qquad (3)$$

It follows that if $X_0(\omega_1, \omega_2)$, $X_1(\omega_1, \omega_2)$, and $H_z(\omega_1, \omega_2)$ are the discrete-space Fourier transforms of $x_0[n_1,n_2]$, $x_1[n_1,n_2]$, and $h_z[n_1,n_2]$ respectively, then for l = 2 the decision rule in equation (3) can be expressed in the frequency domain as:

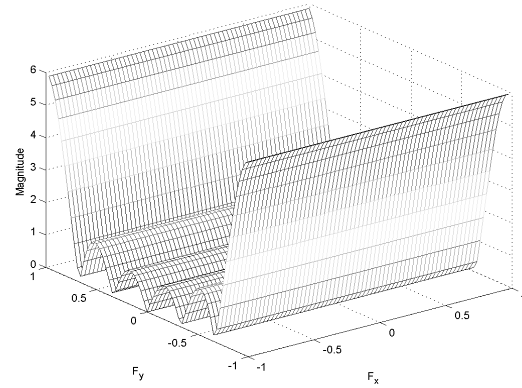

Figure 1. Magnitude response $|H_z(\omega_1, \omega_2)|$ for N = 6 (F_y corresponds to the normalized vertical ω_2 direction).

$$\int_0^{2\pi}\int_0^{2\pi} |H_z(\omega_1,\omega_2)|^2 \left(|X_1(\omega_1,\omega_2)|^2 - |X_0(\omega_1,\omega_2)|^2 \right) d\omega_1\, d\omega_2 \underset{\textit{bottom first}}{\overset{\textit{top first}}{\gtrless}} 0 \qquad (4)$$

Therefore, equation (4) uses a vertical frequency weighted energy difference between $|X_1(\omega_1, \omega_2)|^2$ and $|X_0(\omega_1, \omega_2)|^2$, where the frequency weighting is based upon $|H_z(\omega_1, \omega_2)|$ shown in Figure 1 for N = 6. Although as N gets large the spatial filtering becomes less localized, it can be shown that as $N \rightarrow \infty$, the condition in equation (3) can be expressed as:

$$R_{N\rightarrow\infty}^l \equiv \left(\frac{\sum_{n_1=0}^{N_c-1} y_1^l[n_1]}{\sum_{n_1=0}^{N_c-1} y_0^l[n_1]} \right) \underset{\textit{bottom first}}{\overset{\textit{top first}}{\gtrless}} 1 \qquad (5)$$

where:

$$y_i^l[n_1] \equiv \left| \sum_{n_2=0}^{N_r-1} (-1)^{n_2} x_i[n_1,n_2] \right|^l \qquad (6)$$

and $x_i[n_1,n_2]$ is N_c (columns) by N_r (rows, typically even), and is zero outside. Note that as N increases, the weighting is more towards $\omega_2 = \pi$. Let $X_i[k_1,k_2]$ and $H_z[k_1,k_2]$ represent the N_c x N discrete Fourier transform of $x_i[n_1,n_2]$ and $h_z[n_1,n_2\text{-}N/2+1]$, respectively. It can be shown that for large N, the condition in equation (3) with l = 2 can be written (ignoring aliasing) as:

$$\left(\frac{\sum_{k_1=0}^{N_c-1} |X_1[k_1,N/2]|^2}{\sum_{k_1=0}^{N_c-1} |X_0[k_1,N/2]|^2} \right) \underset{\textit{bottom first}}{\overset{\textit{top first}}{\gtrless}} 1 \qquad (7)$$

Equation (7) compares the total energy of the two composed frames $x_0[n_1,n_2]$ and $x_1[n_1,n_2]$ in the frequency samples only along the vertical high frequency $\omega_2 = \pi$ in Figure 1. Furthermore, it is interesting to note that the first frequency sample ($k_1 = 0$) corresponds to:

$$\left|X_i[0,N/2]\right|^2 = \left|\sum_{n_2=0}^{N-1}\sum_{n_1=0}^{N_c-1}(-1)^{n_2}x_i[n_1,n_2]\right|^2 \quad (8)$$
$$= \left|top\ field\ sum_i - bottom\ field\ sum_i\right|^2$$

where *top field sum*$_i$ and *bottom field sum*$_i$ correspond to the sum of pixels in the top and bottom fields in $x_i[n_1,n_2]$. Although it only represents one frequency sample, the L^2 based measure between the top and bottom fields in equation (8) can be viewed as a simple measure of interfield motion. This measure, or a corresponding L^1 measure between the two fields, can be used as a basis for field order detection as follows:

$$R^l_{fieldsum} \equiv \left(\frac{\left|X_1[0,N/2]\right|^l}{\left|X_0[0,N/2]\right|^l}\right) \underset{bottom\ first}{\overset{top\ first}{\gtrless}} 1 \quad (9)$$

Another interfield motion metric similar to equation (9) compares sums of magnitudes (or magnitudes squared) of the difference between the top and bottom fields as follows:

$$R^l_{field} \equiv \left(\frac{\sum_{n_2=0}^{N_r/2-1}\sum_{n_1=0}^{N_c-1}\left|x_1[n_1,2n_2]-x_1[n_1,2n_2+1]\right|^l}{\sum_{n_2=0}^{N_r/2-1}\sum_{n_1=0}^{N_c-1}\left|x_0[n_1,2n_2]-x_0[n_1,2n_2+1]\right|^l}\right) \underset{bottom\ first}{\overset{top\ first}{\gtrless}} 1 \quad (10)$$

For $l = 1$, the numerator and denominator each represent one term in the $N = 2$ case in equation (3), with the only other term being that generated by subtracting the bottom field from a shifted version of the top field.

4. SIMULATION RESULTS

This section presents the results of the proposed field order detection method based on the zipper point and energy ratios in equations (2), (3), (5), (9), and (10). In general, more than one frame of data is needed for detection, since a single frame contains only one time instance of each field. In the simulation experiments, a field order decision is made for each current frame based on the current and next frames, and a final decision is also made for the entire sequence based on all the frames. Although there are many possible ways to generate a final sequence decision based on many frame decisions (e.g. field order majority, average decision ratio R, etc.), the sequence decision in this study is based on

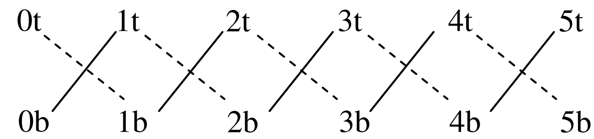

Figure 2. One method for zipper measurements over a sequence.

a single decision ratio R value generated from zipper measurements over the entire sequence. In particular, zipper measurements are computed for each successive pair of frames, where each frame in the sequence (except the last frame) is treated as the current frame in step 1. Then all the current frame zipper measurements (zipper points, L^1 or L^2 zipper energy) are added to obtain the denominator part of the ratio R, whereas all the next frame zipper measurements are added to obtain the numerator part of R. This is illustrated in Figure 2 where the zipper measurements for the solid lines are combined in the denominator, while those for the dashed lines are combined in the numerator. In this manner, the aggregate ratio R avoids being biased by a single zipper measurement.

The first experiment was based on a comparison of zipper points using the decision ratio R^T_6 in equation (2) with a fairly localized $N = 6$ zipper filter. A threshold $T = 3*75$ was chosen to give an average pairwise pixel edge step of $75/255 \approx 29\%$ for detection of a zipper point. The results in Table 1 show that the algorithm correctly detects the field order for each of the six interlaced sequences.

The R^T_6 results in Table 1 are based on equation (2) with $N = 6$. Table 2 plots the overall results R^T_N for $N = 2, 4, 6, 8, 10$ for the Bus and Football sequences. Table 3 plots the error rate for individual frame detection. The threshold T was scaled at $T = 75 * N/2$ in order to account for the different number of pairwise pixel comparisons. The method is effective at determining the correct field order for the entire sequence, although Bus had up to 19 frames incorrectly detected for $N = 10$ (12.75% error rate).

By comparing the L^1 or L^2 zipper activity of the filtered outputs directly as in equation (3), the need for threshold T is eliminated. Table 2 shows the results for Bus and Football using R^l_N from equation (3) with $N = 2, 4, 6, 8, 10$ and $l = 1, 2$. This metric continues to be effective at detecting the correct field order, with $l = 2$ yielding values farther from 1.0 than with $l = 1$, but with both generally closer to 1.0 than R^T_N.

Table 2 also shows the results for Bus and Football using the decision metrics $R^l_{N\to\infty}$, $R^l_{fieldsum}$, and R^l_{field} from equations (5), (9), and (10), respectively. These metrics remove the need for the parameters T and N, but still correctly detect the field order over the entire sequence. Surprisingly, $R^l_{N\to\infty}$ and $R^l_{fieldsum}$ yield correct sequence field order even though $R^l_{N\to\infty}$ has less local adaptivity and $R^l_{fieldsum}$ represents only one frequency sample. However,

Table 3 shows that these metrics are less reliable for detection on a frame basis. It is interesting that the simple DC field measurements in $R^{l}_{fieldsum}$ are able to capture the correct interfield motion over the entire sequence. Overall, the R^{l}_{field} metric with l = 1 had very good sequence and frame detection performance while requiring very little computation.

5. SUMMARY

This paper proposes a simple technique for detecting field order using only the interlaced video data. Simulations based on zipper points and energy yield very promising results, indicating the effectiveness of the approach. The proposed algorithms correctly detected the field order for all the sequences tested, and the simple R^{l}_{field} metric with l = 1 in equation (10) also performed well on a frame basis.

6. REFERENCES

[1] B. Grob,, *Basic television and video systems*, McGraw-Hill, New York, 1984.

[2] R. Beuker and I. Shah, "Analysis of interlaced video signals and its applications," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 501-512, Sept. 1994.

[3] Y. Chang, P. Wu, S. Lin, and L. Chen, "Four field local motion compensated de-interlacing," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 5, pp. 253-256, May 2004.

[4] C. Hentschel, "High quality noise insensitive motion detector using one field memory," *IEEE Trans. Consumer Electronics*, vol. 42, pp. 696-704, Aug. 1996.

Sequence	Format and Scan	Number of Frames	R^{T}_{6} ratio	Detected First Field	Correct Detection?
Bus	720x480i	150	1.28	Top	Yes
Football	720x480i	260	0.33	Bottom	Yes
Canoa	720x576i	220	5.55	Top	Yes
F1car	720x576i	220	2.26	Top	Yes
Rugby	720x576i	220	2.72	Top	Yes
Tempete	720x480i	260	0.33	Bottom	Yes

Table 1. Overall detection results using zipper points with N = 6.

R values (over sequence)	Bus (top field first)			Football (bottom field first)		
	R^{T}_{N}, $T=75*N/2$	R^{l}_{N}, $l=1$	R^{l}_{N}, $l=2$	R^{T}_{N}, $T=75*N/2$	R^{l}_{N}, $l=1$	R^{l}_{N}, $l=2$
$N=2$	1.29	1.18	1.28	0.44	0.68	0.51
$N=4$	1.34	1.21	1.31	0.37	0.65	0.47
$N=6$	1.28	1.23	1.31	0.33	0.62	0.44
$N=8$	1.23	1.23	1.30	0.29	0.61	0.43
$N=10$	1.21	1.23	1.29	0.27	0.60	0.41
$N=\infty$	N/A	1.15	1.55	N/A	0.55	0.32
N="fieldsum"	N/A	2.75	7.07	N/A	0.48	0.21
N="field"	N/A	1.22	1.33	N/A	0.69	0.53

Table 2. Overall sequence detection results using zipper points and energy for Bus and Football.

% Error (frame decisions)	Bus (top field first)			Football (bottom field first)		
	R^{T}_{N}, $T=75*N/2$	R^{l}_{N}, $l=1$	R^{l}_{N}, $l=2$	R^{T}_{N}, $T=75*N/2$	R^{l}_{N}, $l=1$	R^{l}_{N}, $l=2$
$N=2$	0.00	0.00	0.00	0.77	0.00	0.39
$N=4$	0.67	0.00	0.00	0.39	0.00	0.39
$N=6$	3.36	0.00	0.00	0.00	0.00	0.00
$N=8$	6.04	0.00	1.34	0.00	0.00	0.00
$N=10$	12.75	0.00	2.68	0.00	0.00	0.00
$N=\infty$	N/A	26.17	15.44	N/A	0.77	1.16
N="fieldsum"	N/A	12.75	12.75	N/A	36.29	36.29
N="field"	N/A	0.00	1.34	N/A	0.00	0.00

Table 3. % Error rates (frame decisions) for Bus and Football.