VIDEO DEINTERLACING WITH CONTROL GRID INTERPOLATION

[1]*Ragav Venkatesan,* [2]*Christine M. Zwart and* [1,2]*David H. Frakes*

[1]School of Electrical, Computer and Energy Engineering,
[2]School of Biological and Health Systems Engineering,
Arizona State University, Tempe, AZ.

ABSTRACT

Video deinterlacing is a key technique in digital video processing, particularly with the widespread usage of LCD and plasma TVs. This paper proposes a novel spatio-temporal video deinterlacing technique that adaptively chooses between results from segment adaptive gradient angle interpolation (SAGA), vertical temporal filter (VTF) and temporal line averaging (LA). The proposed method performs better than several popular benchmarking methods in terms of both visual quality and PSNR and requires minimal computational overhead. The algorithm performs better than existing approaches on fine moving edges and semi-static regions of videos, which are recognized as particularly challenging deinterlacing cases.

Index Terms- Deinterlacing, Interpolation, Saliency.

1. INTRODUCTION

The human visual system is more sensitive to temporal resolution than spatial resolution. Without increasing the overall transmission bandwidth, video interlacing can effectively double the video frame rate. A drawback of interlaced videos is that they are only suited for analog display technologies such as CRT. Modern displays require progressive or deinterlaced videos, necessitating a conversion.

The most straightforward way of deinterlacing a video is the weave algorithm (line-doubling or field insertion) [1]. This inter-field deinterlacing algorithm merges two fields, preserving the horizontal resolution while reducing the vertical resolution by half. The weave algorithm leads to a line crawling effect in regions that describe motion.

Line averaging (LA) and edge-based line averaging (ELA) [2, 3] are intra-plane deinterlacing algorithms. While LA is simply the average of two lines of data, ELA calculates directional distances over pixel neighborhoods and defines the interpolated pixel as the average of the pixel neighbors with the minimum intensity change. Modified versions of ELA are an area of active research [4].

Vertical temporal filtering is a spatio-temporal approach to deinterlacing. The filter includes the known vertical and temporal neighbors of the missing pixel being deinterlaced. The VTF method can be described as:

$$\widehat{F_n}(i,j) = \begin{cases} F_n(i,j)\,, & (j\ mod 2 = n\ mod 2) \\ \sum_m \sum_k F_{n+m}(i,j+k)\ h_m(k)\,, & (otherwise) \end{cases}, \quad (1)$$

where $F_n(i,j)$ represents the pixel value at position (i,j) in the n[th] field, $\widehat{F_n}(i,j)$ denotes the reconstructed pixel value at the same position, and $h_m(k)$ denotes the VT filter weights for field index m and vertical index k. The filter weights $h_m(k)$ as proposed by Weston [5] are:

$$h_m(k) = \begin{cases} \frac{1}{2}, \frac{1}{2} & , (k = -1,1\ \&\ m = 0) \\ -\frac{1}{16}, \frac{1}{8}, -\frac{1}{16} & , (k = -2,0,2\ \&\ m = -1,1) \end{cases}. \quad (2)$$

Like ELA, VTF is a popular benchmarking algorithm and VTF variations are still actively researched. One example is content-adaptive VTF (CAVTF) [6].

Oh, et. al proposed another inter-field deinterlacing technique: spatio-temporal edge based median filtering (STELA) [7]. STELA divides the video frame into low-pass and high-pass components. Six directional, inter-field distances are calculated on a low-pass version of the image. Referencing Figure 1, 'X' is the point to be estimated in frame '*n*'. The associated distances are,

$$C1 = |a - f|, C2 = |b - e|, C3 = |c - d|$$
$$C4 = |g - l|, C5 = |h - k|, C6 = |i - j|\,. \quad (3)$$

The de-interlaced point in the low-pass filtered image is estimated as $y = Med\{A, b, e, h, k\}$, where A is the average value of the two points that yield the minimum directional change among $C1$ to $C6$. A line-doubled version of the high pass image is added to *y*. STELA attempts to preserve both the horizontal and vertical frequencies of the image.

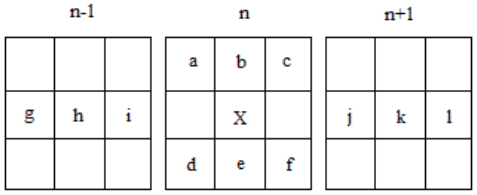

Figure 1. Example video and neighborhood description.

The remainder of this article is organized as follows: Section 2 describes the proposed algorithm. Section 3 presents results of the new algorithm. Section 4 provides concluding remarks.

2. PROPOSED ALGORITHM

The proposed algorithm performs an intra-frame deinterlacing similar to ELA and incorporates inter-frame information selectively based on a novel measure of spatial saliency and temporal difference. The algorithm buffers three interlaced frames at a time and deinterlaces the middle frame. A difference map d_n is derived from the intensity differences between frame $n-1$ and frame $n+1$ and a spatial saliency map S_n is calculated. The choice between deinterlacing estimates from SAGA [11], VTF and a pure temporal average is then made according to:

$$\widehat{F_n}(i,j) = \begin{cases} F_n(i,j), & (j\ mod 2 = n\ mod 2) \\ \dfrac{F_{n-1}(i,j) + F_{n+1}(i,j)}{2}, & d_n(i,j) < t \\ \sum_m \sum_k F_{n+m}(i, j+k)\, h_m(k), & S_n(i,j) < b;\ d_n(i,j) \geq t \\ SAGA(i,j), & (otherwise) \end{cases} \quad (4)$$

where $h_m(k)$ is as in Equation (2) and $SAGA(i,j)$ is the interpolated value from the spatial interpolator SAGA. The parameters used in Equation (4) are discussed in the following subsections.

2.1 Difference map.

The static regions of a video, where there are no temporal differences in pixel values, are best deinterlaced by temporal averaging. Accordingly, a difference map d_n is calculated for the n^{th} frame by finding the difference between frame number $n-1$ and frame number $n+1$. Any region that is temporally static ($d_n < t$, where t is a 1 bit difference) is interpolated by averaging temporally.

2.2 Spatial Saliency map.

The procedure for identifying saliency is well presented in [9]. A quaternion version of the same is implemented here. A color image can be represented using quaternions of the form:

$$q^n = Ch_1^n + Ch_2^n \mu_1 + Ch_3^n \mu_2 + Ch_4^n \mu_3, \quad (5)$$

where μ_i, $i = 1,2,3$ satisfies, ${\mu_i}^2 = -1, \mu_1 \perp \mu_2, \mu_2 \perp \mu_3, \mu_1 \perp \mu_3$. The three color channels of the images are allocated to Ch_2, Ch_3 and Ch_4. Ch_1 is set to zero. The Quaternion Fourier Transforms (QFT) of the frame n are defined in [10] as:

$$Q_i^n[u,v] = \frac{1}{\sqrt{WH}} \sum_{y=0}^{W-1} \sum_{x=0}^{H-1} e^{{\mu_1}^{2\pi\left(\frac{yv}{W}+\frac{xu}{H}\right)}} q_1^n(x,y) \quad (6a)$$

and

$$q_i^n[x,y] = \frac{1}{\sqrt{WH}} \sum_{v=0}^{W-1} \sum_{u=0}^{H-1} e^{{\mu_1}^{2\pi\left(\frac{yv}{W}+\frac{xu}{H}\right)}} Q_1^n[u,v], \quad (6b)$$

where (x,y) are the pixel locations of individual pixels and $[u,v]$ describe the frequencies. W and H are the width and height of the frame. The phase spectrum of $Q^n[u,v]$ is given by $Q_p = Q/||Q||$. The spatial saliency map, S_n is obtained by Gaussian smoothing ($\sigma = 8$) the L_2 norm of the inverse QFT of Q_p, as defined by Equation (6b). The spatial saliency map is then used as part of the decision making process in Equation (4) with a threshold level of b where b is 4% of the bit depth.

$$S_n(i,j) = g * q_p \ . \quad (7)$$

2.3 SAGA.

ELA and STELA both use edge-based information to improve deinterlacing results. Interpolation along rather than across edges avoids jagged or blurring artifacts known to be visually distracting. Using simple pixel subtraction to identify the direction of minimal change is an extremely rapid and reasonably effective approach to determining the edge direction. Notable short-comings include a restriction to a quantized subset of possible edge orientations, a sensitivity to noise, and problems detecting weaker edges. While a number of more effective edge-directed interpolators have been established for static image interpolation, the use of those methods in deinterlacing has been limited. Many of the complex methods for edge-directed or adaptive interpolation are computationally prohibitive in the context of video. Furthermore, many image interpolation methods include kernels with expectations for a uniformly (horizontally and vertically) under-sampled lattice.

We previously introduced one-dimensional control grid interpolation (1DCGI), an interpolator based on reducing the brightness constraint of optical flow to a one-dimensional edge constraining equation:

$$I(x,y) = I(x+\alpha, y+1), \quad (8)$$

where points on a line of constant intensity (isophote) can be parameterized in terms of a displacement variable α [8]. The control grid formulation of optical flow uses nodes to parse an image into sub-grids. Displacement parameters are determined explicitly at nodes and interpolated within the sub-grids. Using this formulation and the Taylor series expansion of Equation (8), we define the matching error for 1DCGI to be a function of the displacements at the nodes. If basis functions θ_1 and θ_2 are used to define the displacements between points x and $x+k$ (where k is the

node spacing), the error associated with selection of the displacements α_x and α_{x+k} can be formulated as:

$$E(\alpha) = \sum_{i=0}^{k}\left[\frac{\partial I(x+i,y)}{\partial x}[\alpha_x \quad \alpha_{x+k}]\begin{bmatrix}\theta_1(i)\\ \theta_{2(i)}\end{bmatrix} + \frac{\partial I(x+i,y)}{\partial x}\right]^2, (9)$$

where

$$\theta_1(i) = \frac{k-i}{k} \quad , \quad \theta_2(i) = \frac{i}{k} \qquad (10)$$

and i indicates the distance from the previous node (0 to k). The displacement α_x appears in the error calculations for the range α_{x-k} to α_x as well as α_x to α_{x+k}. The solution for the set of displacements mapping one line of pixels to the adjacent line is identified as the set with the lowest cumulative error over the full line.

The nodal framework of 1DCGI mitigates the impact of noise and allows for the detection of gradual edges. Furthermore, interpolation is performed by inserting new lines of data as:

$$I\left(x+\frac{\alpha}{2}, y+\frac{1}{2}\right) = \frac{1}{2}[I(x,y) + I(x+\alpha, y+1)] \quad , (11)$$

making the 1DCGI approach suitable for deinterlacing applications. A drawback of 1DCGI for this application is the use of an iterative, conjugate gradient approach for determining the optimal displacements. More recently we have developed a method called, segment adaptive gradient angle (SAGA) interpolation that is based on a similar parameterized definition of image isophotes and nodal structure [11]. Unlike 1DCGI however, the cost function used to determine the optimal displacements for a given line of data is formulated as a least squares problem with a tri-diagonal coefficient matrix. This allows for a numerical solution with $O(N)$ complexity.

Use of the SAGA interpolator as an intra-frame deinterlacing method is presented as an alternative to ELA. SAGA is also extended as an intra-frame deinterlacer SAGA+TF by applying the temporal component of the VTF filter and is defined by:

$$\widehat{F_n}(i,j) = \begin{cases} F_n(i,j)\,, \ (j\ mod 2 = n\ mod 2) \\ \frac{F_{n-1}(i,j) + F_{n+1}(i,j)}{2}\,, d_n(i,j) < t \\ SAGA(i,j) + \sum_m \sum_k F_{n+m}(i, j+k)\, h_m(k)\,, (else) \end{cases} \qquad (12a)$$

where

$$h_m(k) = \begin{cases} 0{,}0\ , (k = -1{,}1\ \&\ \ m = 0) \\ \frac{-1}{16}, \frac{1}{8}, \frac{-1}{16}\ , (k = -2{,}0{,}2\ \&\ m = -1{,}1) \end{cases}. \qquad (12b)$$

The output of the proposed algorithm thus depends on the saliency map, the difference map, and the SAGA estimate as indicated by Equation (4).

3. RESULTS

To facilitate comparisons to published results, we introduce a statistical relevance factor '*r*'. If MSE_{VTF} is the mean squared error from VTF and MSE_{new} is the mean squared error of a new algorithm, then the statistical relevance '*r*' is

$$r = 100 * \left[1 - \frac{MSE_{VTF}}{MSE_{new}}\right]. \qquad (13)$$

Comparison of the proposed approach against CAVTF [6] in terms of *'r'* is given Table 1.

The PSNR results of SAGA, SAGA+TF and the proposed algorithm are provided in Table 2 along with results from other state-of-the-art methods. The performance of the proposed method and other methods is shown in Figure 3.

Table 2 shows that SAGA performs better than ELA and that SAGA+TF outperforms its inter-frame counterparts VTF and STELA. Furthermore, SAGA provides an analytical solution avoiding the expensive sorting involved in ELA and STELA. The proposed algorithm also outperforms the recently published approach CAVTF as shown in Table 1.

4. CONCLUSION

Advantages of the proposed algorithm are particularly noticeable in videos like 'mother', where there are distinct salient and background regions as indicated in Figure 2. SAGA interpolates better than any temporal method along the edges in the salient regions, while the background regions are estimated using either VTF or temporal average (depending on temporal motion). It is clear that this selection process improves the performance of SAGA alone as a deinterlacer and makes the overall approach better than the prior art methods.

a) Original b) Saliency Map c) Deinterlaced

Figure 2. Original frame, saliency map and deinterlaced output

Table 1. Statistical relevance '*r*' values for the proposed algorithm and the CAVTF algorithm proposed in [6]. VTF is used as the reference algorithm.

Video	CAVTF[6]	Proposed Algorithm
Akiyo	70.620	**74.941**
Container	93.200	**97.090**
Foreman	41.920	**66.547**
Hall Monitor	76.770	**89.694**
Mother	57.040	**82.529**

Table 2. PSNR (dB) values for the proposed algorithm and other deinterlacing algorithms.

Video	Weave	LA	ELA	STELA	VTF	SAGA	SAGA+TF	Proposed Algorithm
Akiyo	33.302	38.359	36.758	41.237	41.117	38.720	46.024	**47.301**
Bowing	31.617	36.850	34.617	37.013	40.962	36.228	42.881	**46.122**
Bridge Far	29.854	32.244	32.135	38.788	33.689	32.225	39.138	**42.423**
Container	24.436	28.017	27.795	35.479	31.055	28.769	46.278	**46.417**
Deadline	27.178	30.443	28.519	35.662	33.152	30.603	42.954	**42.814**
Foreman	28.162	31.519	32.149	31.467	32.202	34.539	35.709	**36.957**
Galleon	21.411	24.333	23.344	31.609	27.058	24.290	31.780	**42.048**
Hall Monitor	25.801	29.945	30.435	36.942	32.023	31.566	37.948	**41.892**
Mother	33.103	36.637	36.016	42.599	38.058	36.690	44.351	**45.635**
News	28.469	34.202	31.765	36.855	39.088	34.068	40.053	**44.597**
Students	28.124	31.906	30.994	37.086	33.436	32.323	38.479	**45.173**
Paris	23.541	26.717	25.370	30.943	28.934	26.764	32.114	**33.799**
Sign Irene	32.667	36.468	36.268	36.181	36.401	37.462	37.489	**40.108**

a) Original b) ELA c) STELA d) VTF e) SAGA f) Proposed

Figure 3. Visual comparison among deinterlacing algorithms.

4. REFERNCES

[1] G.D. Haan, and E.B. Bellers, "Deinterlacing – an overview," *in Proc. of the IEEE*, vol. 86, no.9, pp. 1839-1857, Sep 1998.

[2] C. J. Kuo, C. Liao, and C.C. Lin, "Adaptive interpolation technique for scanning rate conversion," *in IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no.3, pp. 317-320. June 1996.

[3] Y. L. Chang, S. F. Lin, and L. G. Chen, "Extended intelligent edge based line average with its implementation and test method," *in Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 341–344. Vancouver, Canada, May 2004.

[4] Hong et, al. "Method switching algorithm for intra-field deinterlacing", *in IEEE 15th International Symposium on Consumer Electronics*, pp. 518-523, Aug 2011.

[5] M. Weston, "Interpolating lines of video signals", US-patent 4,789,893, Dec. 1988.

[6] K. Lee, C. Lee, "High quality deinterlacing using content adaptive vertical temporal filtering", *in IEEE Transactions on Consumer Electronics*, pp. 2469 – 2474, Jan 2011.

[7] H. S. Oh, Y. Kim, Y. Y. Jung, A. W. Morales and S. J. Ko, "Spatio- Temporal Edge-Based Median Filtering for Deinterlacing," *in Proc. of International Conference on Consumer Electronics*, pp. 52- 53, June 2000.

[8] C. M. Zwart and D. H. Frakes, "Biaxial control grid interpolation: Reducing isophote preservation to optical flow," *in Proc. IEEE Digital Signal Processing Workshop*, Jan 2011, pp. 140-145.

[9] C. Guo et al., "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," *in Computer Vision and Pattern Recognition*, pp. 518-523, June 2008.

[10] S.J. Sangwine and T.A. Ell, "Hypercomplex Fourier transforms of color images," *in IEEE Transactions on Image Processing*, pp. 22-35, 2001.

[11] C. M. Zwart, D. H. Frakes, "Soft Adaptive Gradient Angle Interpolation of Grayscale Images", *in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2012.