A Total Variation Motion Adaptive Deinterlacing Scheme

Sune Keller, François Lauze, and Mads Nielsen

The IT University of Copenhagen
Rued Langgaardsvej 7, 2300 Kbh. S
Denmark
{sunebio,francois,malte}@itu.dk
WWW home page: http://www.itu.dk/image/index.html

Abstract. We propose a new way of deinterlacing using a total variation scheme. Starting by the Bayesian inference formulation of total variation we do MAP by rewriting the problem into PDEs that can be solved by simple numerical schemes. Normally deinterlacing schemes are developed ad hoc with online hardware implementation directly at eye, sometimes with some frequency analysis as only theoretical base. Our belief is that mathematically well based image models are needed to do optimal deinterlacing and by our work presented here, we hope to prove it. Comparing the output of our scheme with those of ten known deinterlacing schemes shows very promising results.

1 Introduction

Interlaced scan has been use since the birth of television in the 1930's and is the scanning format used in the television standards PAL and NTSC. Interlacing is separates a full frame image into two parts called fields, one containing all horizontally odd numbered line and the other containing all the even lines. When recorded in interlaced scan the fields are separated in time and two neighboring fields cannot be merged to one full frame without problems.

Interlacing saves bandwidth and lowers the cost of cameras and CRTs as it is possible to combine a high rate of fields per second (to avoid large area flicker in the image) with a relatively high vertical resolution. This looked fine to the human visual system (HVS) in the early days of television but as screen size grew and television sets produced brighter images, interlacing artefacts started to show.

Interlacing artefacts have many names and are often mixed up when described, as they can be described both from a frequency analysis point of view and by their visual appearance. They are by visual appearance

- Line crawl due to vertical motion in the image and the time difference between the two fields composing one frame.

- Serration of edges due to horizontal motion in the image and the time difference between the two fields composing a frame. It happens to edges at all orientations except those close to and at horizontal orientation.

- Interline flicker due to fine stationary details appearing only in either odd or even fields of the image as they are to small (that is of to high a vertical frequency) to be sampled in both even and odd fields.

A further discussion of frequency analysis and aliasing in interlaced image sequences can be seen in [2] and [11].

One way of reducing the effect of these artifacts in terms of visibility to the human eye is to interpolate new fields and raise the field rate as done in 100 Hz TV sets [7] and [2]. Another way is to convert the interlaced sequence to a progressive sequence by interpolation of image information in the missing lines of the fields to make a full frame of each field. This conversion is called deinterlacing. Progressive scan is used in all PC monitors, in projectors, and in flat panel displays (LCDs and Plasmas). So as many new displays for television are progressive and as PC and television are merged (set top boxes for digital television, DVDs, television tuners for PCs, and video editing on PCs) there is obviously a big need for deinterlacing. Deinterlacing is difficult, as turning e.g. 50 fields per second into 50 frames per second requires a doubling of the amount of image data without introducing new artefacts to annoy the human visual perception. We propose a new scheme for deinterlacing developed from techniques used in image and image sequence inpainting, and we have implemented ten known and widely used deinterlacing schemes to compare it with. Our scheme uses Total Variation (TV) based in a Bayesian framework and do MAP by minimizing an energy functional. This is accomplished by deriving and solving solving corresponding Partial Differential Equations (PDEs) obtained through the calculus of variations. This is in contrast to many known deinterlacers that have been developed ad hoc (and in a heuristic way) with online hardware implementation directly in eye. Therefore they are often simplified to keep hardware costs down. We start with a theoretically well-based offline design that by further development could end up as online hardware. Section 2 will describe the other deinterlacing schemes implemented, section 3 will describe our proposed scheme, section 4 shows the results and in section 5 we draw our conclusions.

2 Standard Deinterlacing

To measure the performance of our deinterlacing scheme, we have implemented ten other schemes known from literature and/or available software and hardware ([2]).

Line Doubling (LDB) is very simple. Every interpolated horizontal line is a repetition of the previous existing line ([15] and [17]). Line Averaging (LAV) is a vertical average of the above and below pixels, since they are both known ([2], [15] and [17]). Field Insertion (FI), a.k.a. merging or weaving, fills in the blanks with neighbouring lines in time and is essentially a temporal version of LDB. The result is very similar to the image seen on an interlaced display ([2] and [17]). Field averaging (FAV) is a temporal version of LAV ([17]), while Vertical Temporal interpolation (VT) is a simple 50/50 combination of LAV and FAV ([17]). Many more advanced but not significantly better VT filters have been suggested, e.g. by BBC Research ([16]). All schemes mentioned so far are fixed, linear filters, whereas the next five are nonlinear and adapt to certain conditions in their local neighborhood and chose one of several possible interpolations depending on the local image content to yield better results.

Median filtering (Med) is a real classic in image processing and is used for deinterlacing in many variations ([2], [3], [8], [14], [13] and [15]). We have chosen a 3-tap vertical temporal version from [2] although we use the forward temporal neighbor instead of the backwards. Motion adaptive deinterlacing (MA) can be done in a countless number of ways and we have chosen a version suggested in [13] and [14]. It does simple Motion detection and takes advantage of the qualities of simpler schemes under different conditions: FAV in presence of no motion, Median filtering when motion is slow and LAV when fast motion is detected. Thresholds classify the motion. Weighted Vertical Temporal deinterlacing (wVT) is a simpler way of doing motion adaptation than the previous mentioned scheme, MA, and gives, instead of a hard switching between schemes, a smooth weighted transition between temporal and vertical interpolation. The scheme is described in detail in [9]. Edge Adaptive deinterlacing (EA) has been suggested in several forms, e.g. in [6], [9] and [15]. We have chosen a scheme that based on Summed Absolute Differences (SAD) selects a direction of interpolation as described in [15], although we have modified it to detect the best of five directions, 0^o, $\pm 26^o$ and $\pm 45^o$ from vertical. Successive Approximation (SA) is the second level of approximation in [9] although the its edge adaptive scheme working on the first deinterlaced approximation has been swapped with the EA scheme that works directly on the interpolated original and thereby taking the successiveness out of the scheme but in the same instance also removing the possibility of error propagation.

Med is a simple adaptive scheme, EA adopts to the orientation of edges while MA, wVT and SA are Motion adaptive.

3 Total Variation Deinterlacing

In this section we introduce a novel deinterlacing scheme based on Total Variation minimization. We first proceed in a Bayesian fashion and deduce a variational formulation through MAP estimation in continuous settings following [10]. We then compute the associated Euler-Lagrange equations and their associated gradient descent formulations. The discretization of the latter will provide our numerical schemes. We first introduce the notations used in the sequel. Ω will denote the spatio-temporal domain of the progressive sequence, $F \subset \Omega$ the domain of the known fields, u_0 will denote the interlaced sequence, and by abuse of notations, it will also denote the known data on F.

3.1 Bayesian Framework

Let u denote a progressive sequence and u_0 the known sequence of interlaced fields. According to Bayes' Theorem

$$p(u|u_0) \propto p(u_0|u)p(u) \ . \tag{1}$$

The term on the left hand side is the a posteriori to be maximized (MAP) and the first term on the right hand the side is a model term and the second is a prior on image sequences. For the model term we choose a simple Dirac distribution $p(u_0|u) = \delta((u - u_0)|_F)$ because we wish to keep the existing pixels unchanged.

We have investigated two distributions for the prior term p(u). First, by viewing the image sequence u as a 3D volume, we set

$$p(u) \propto e^{-\lambda \sum_x |\nabla_3 u(x)|} \tag{2}$$

with x running over all the pixels in the sequence, $\nabla_3 u$ a discrete spatio-temporal gradient and λ a positive constant.

Nevertheless, it is somewhat unnatural to treat an image sequence as a 3D volume. We introduce therefore a simple model that separates spatial and temporal dimensions and we assume independence of the spatial and temporal distributions. Our image prior thus becomes

$$p(u) = p(u_s, u_t) = p(u_s)p(u_t) \tag{3}$$

where $p(u_s)$ refers to the spatial distribution of images and $p(u_t)$ to the temporal correlation between frames. For the spatial prior we use

$$p(u_s) \propto e^{-\lambda \sum_x |\nabla u(x)|} \tag{4}$$

with x again running over all the pixels in the sequence, ∇u a discrete *spatial* gradient and λ a positive constant. This has proven a robust model, well studied in the computer vision community; see for instance [1], [4] or [12]. The temporal prior

$$p(u_t) \propto e^{-\mu \sum_x |\partial_t u(x)|} \tag{5}$$

where $\partial_t u$ denotes the time-derivative of u and introduces the motion adaptive aspect of our algorithm, μ being a positive constant.

3.2 Variational Formulation - Euler-Lagrange Equations

Following [10] in order to compute the Maximum A Posteriori (MAP) solution, u, for our problems, we take the $-\log$ of each term to reformulate it as a minimization problem. Instead of using the $|\cdot|$ function which is non differentiable at the origin, we replace it by the approximation $\psi(s^2) = \sqrt{s^2 + \varepsilon^2}$, with $\varepsilon = 0.1$ or 0.01 in our experiments. From (2), with this modification we obtain u as the solution of

$$\operatorname*{Arg\,min}_u \int_\Omega \psi(|\nabla_3 u|)dx, \quad u = u_0|_F \ . \tag{6}$$

From standard calculus of variations and the fact that $\psi'(s)/s = 1/\psi(s)$, its Euler-Lagrange equation is

$$-\mathrm{div}\left(\frac{\nabla_3 u}{\psi(|\nabla_3 u|)}\right) = 0, \quad u = u_0|_F \tag{7}$$

where div is the divergence operator. The associated gradient descent equation is

$$\partial_\tau u = \operatorname{div}\left(\frac{\nabla_3 u}{\psi(|\nabla_3 u|)}\right) = 0, \quad u = u_0|_F \tag{8}$$

where τ denotes the evolution parameter (in order to not confuse it with the time parameter t of the sequence), which is a 3D *total variation* filter.

From (4) and (5) we obtain the following minimization problem:

$$\operatorname*{Arg\,min}_u \int_\Omega \left(\psi(|\nabla u|) + \alpha\psi(|\partial_t u|)\right) dx, \quad u = u_0|_F \tag{9}$$

the corresponding Euler-Lagrange equation being

$$-\operatorname{div}\left(\frac{\nabla u}{\psi(|\nabla u|)}\right) - \alpha\partial_t\left(\frac{\partial_t u}{\psi(|\partial_t u|)}\right) = 0, \quad u = u_0|_F \tag{10}$$

and its associated gradient descent equation is

$$\partial_\tau u = \operatorname{div}\left(\frac{\nabla u}{\psi(|\nabla u|)}\right) + \alpha\partial_t\left(\frac{\partial_t u}{\psi(|\partial_t u|)}\right), \quad u = u_0|_F \tag{11}$$

which combines a 2D total variation filter for the spatial part and a simple 1D total variation filter for the temporal part. The constant $\alpha = \mu/\lambda$ is a weight between the spatial and the temporal part of the filter. This approach to energy minimization gives convex but not strictly convex, solutions so several global minimums might exist. Therefore the solution can be sensible to initialization.

3.3 Discretizations

The gradient descent equations are solved explicitly, using forward difference for the evolution derivative ∂_τ and central difference for the divergence terms.

For the 3D divergence, we have used a standard discretization on the 6 points spatio-temporal neighborhood (see for instance [5], appendices, for details). For the 2D divergence we have used three different schemes, one using a 4-point neighborhood of the current pixel and two using a full 8-point neighborhood, as described in [1]. The sensibility of the above PDEs to initial values has not given us problems: At $\tau = 0$ to initialize we take the LAV deinterlaced sequence as a rough estimate with good results.

4 Results

We present now the results obtained with four image sequences. The first one, `Person`, is a medium shot of a sitting person, turning the head and talking, the motion can be said to be small. The second sequence, `C&T`, is a shot of a driving car and truck followed by a tracking camera, the motion, which is primarily horizontal, is up to ten pixels between two consecutive frames. The last two sequences are both artificial with high contrast details. The BSNM sequence is stationary while the BS has vertical, horizontal and diagonal motion, both accelerated and constant. Figure 1 shows stills of the two sequences BS and BSNM, whereas stills of the sequences C&T and Person cannot be published due to copyright issues.

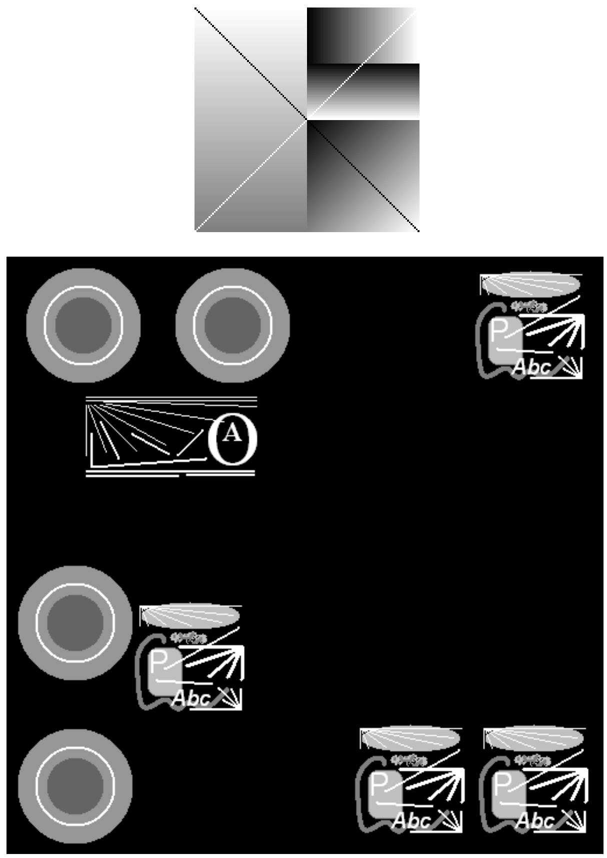

Fig. 1. From the top: A frame from the 128x128 sequence BSNM and frame one of the 512x512 sequence BS.

The four sequences are all progressive, so we have chosen to give the Mean Square Error (MSE) as an objective measure of the performance of the schemes,

$$MSE = \frac{1}{N} \sum_{\Omega \backslash F} (u - u_{org})^2 \tag{12}$$

which measures the square difference between the N interpolated pixels in the output, u, and their removed counterparts in the original progressive sequence, u_{org}.

We also give a subjective evaluation as the final judge of the result is the human visual system. A discussion of how to determine the quality of deinterlacing is given in [2]. Table 1 gives the objective results.

Table 1. MSE from deinterlacing the four sequences `Person`, `C&T`, `BS` and `BSNM`. 3D and 2D+1D are the two versions of our scheme with the number of iterations given after the name. Clearly our schemes give the best results of all on the natural image sequences `Person` and `C&T`

Scheme		`Person`	`C&T`	`BS`	`BSNM`
LDB		17.90	79.72	1623.6	755.8
LAV		5.53	26.31	924.6	678.4
FI		22.26	472.25	3935.1	0
FAV		9.03	284.22	2514.4	0
VT		5.62	94.34	1146.8	169.6
Med		9.27	65.72	1154.1	363.4
MA		5.53	28.18	*840.6*	363.4
wVT		5.07	67.97	1056.0	0
EA		8.77	32.87	957.1	210.1
SA		5.36	48.79	862.4	82.0
3D TV	2	5.02	27.00	1066.2	666.8
3D TV	50	*4.85*	56.29	1078.3	461.8
2D+1D TV	2	4.97	*26.06*	919.2	666.2
2D+1D TV	20	*4.86*	*26.23*	890.9	567.9
2D+1D TV	200	5.11	32.89	*804.4*	242.5

On `C&T` and `BS` it is seen from the MSE's that in presence of large motion, our scheme offers only little improvement, and only for the motion adaptive 2D+1D version, where the spatial and temporal gradients are separated. The 3D version suffers from having a spatio-temporal gradient. Over time (in terms of number of iterations) the 2D+1D improves a lot on the `BS` but not on `C&T`, which contains the larger motions of the two. Although the 3D does not perform to good overall, it actually improves the per-frame MSE in 23 of the 98 frames in `C&T`. In presence of none or only small motion, our scheme wins as it can be seen from the MSE's on the `BSNM` and `Person`. After only 2 iterations a 10 % improvement is seen on the MSE of `Person` and it increases with the number of iterations. Taking SA as initial guess instead of LAV on `BSNM` gave a 9 % improvement in MSE after 20 iterations of 2D+1D.

The results for 2D+1D after 200 iterations show that convergence in MSE stops for the two natural sequences. This is due to the smoothing of the TV prior and noise in the original sequence. Further studies showed that the lowest values in MSE was reached after 40-60 iterations.

Subjectively our scheme produces the best visible results on all four sequences but `BSNM`. `BSNM` is fully stationary, so the temporal schemes give 100 % perfect results on it. On `BS` and `C&T` the improvement is moderate, but on `Person` the results from our scheme are clearly the best. After two iterations we already se a good subjective result for the 2D+1D scheme, but after 20 iterations (and 50 for the 3D) the results are really good, even making us doubt which is the deinterlaced when comparing to the progressive original.

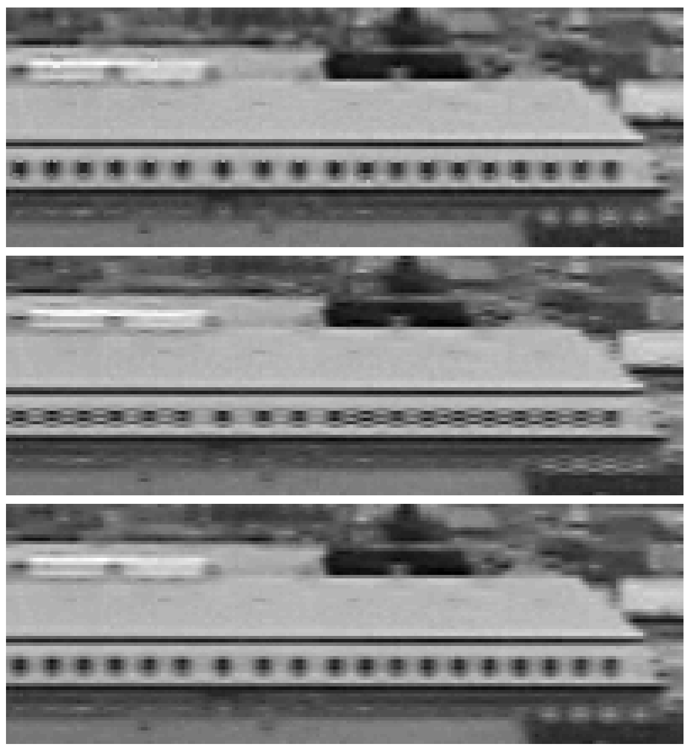

Fig. 2. From top to bottom: deinterlacing with MA, wVT and 2D+1D TV. Only zoom-ins of the full frame is shown here

Figure 2 top, middle and bottom also illustrate the potential of our method. The sequence used for the illustration shows a pan of Christianborg Castle in Copenhagen and as it only exists as interlaced, no MSE's can be calculated. The top picture shows the result of the wVT scheme while the middle one shows the result of the MA scheme. Serious artifacts are visible for both schemes, serration for wVT and erroneous detection and interpolation for MA. The bottom picture shows the result of our 2D+1D scheme after 20 iterations, and clearly asserts the quality improvement obtained with our scheme.

LAV in itself is, given its simplicity, a remarkably well performing deinterlacer as the results in table 1 indicates, but as figure 3 shows, the 2D+1D scheme is able to improve the quality of deinterlacing significantly after 20 iterations. Note in the ornaments how the details have been sharpened and the jagged edges have been removed. The sequence used, `Church`, is stationary, shot with the camera on a camera mounting but rather noisy do to low lighting.

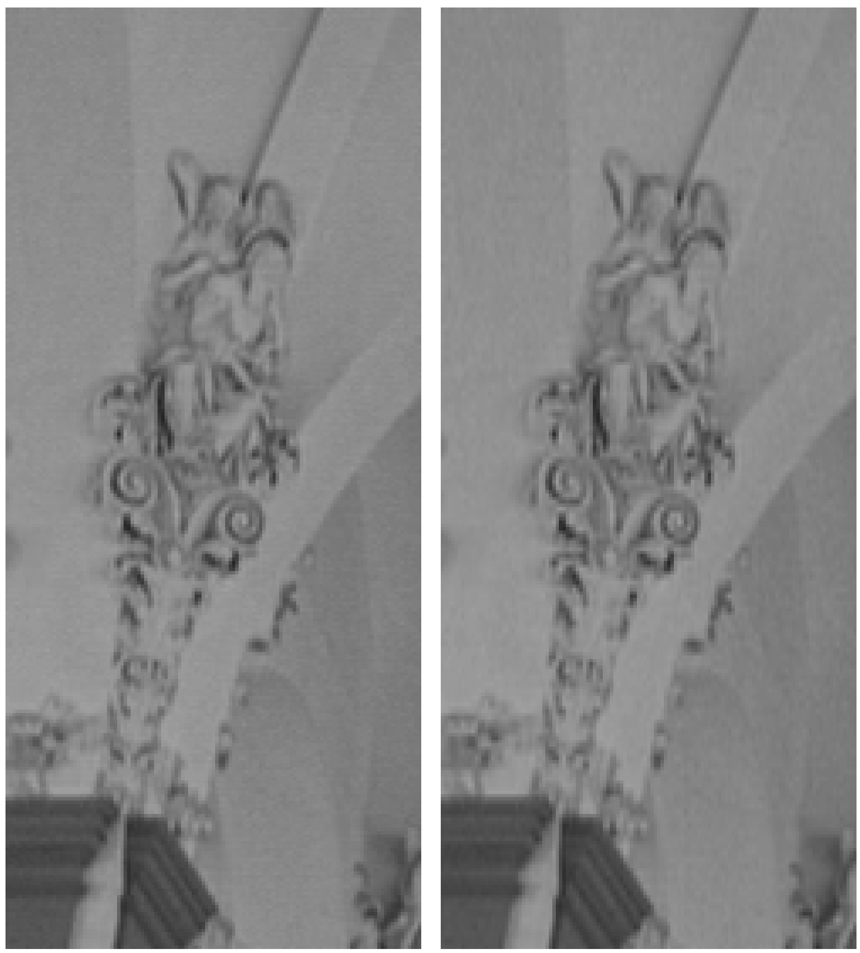

Fig. 3. Deinterlacing of the sequence `Church`. Left: 2D+1D TV after 20 iterations. Right: LAV used as initialization for 2D+1D TV. Clearly 2D+1D TV improves on the LAV initialization. Only zoom-ins of the full frame is shown here. Notice the in particular the arm of the upper angel and the helixes (spirals) in the middle

Further investigations on the 2D+1D schemes has also shown that the number of iterations to obtain a certain quality of the result can be reduced by a factor of three to six by increasing the time step in the gradient descent without the loss of stability. The number of operations and complexity per iteration of the 2D+1D schemes are the same as for the most complex of the ten known deinterlacing schemes, SA. This together with the increase in time step and good results after a few iterations gives rise to our believes that an online hardware implementation of 2D+1D TV MA Deinterlacing is possible.

5 Conclusion

We have shown that a technique so far used for inpainting can be redeveloped to do deinterlacing and further on our Total Variation Motion Adaptive Deinterlacing outperforms ten known fixed or adaptive deinterlacers. Our deinterlacer is still in its youth and its potential not yet fully explored. The quality of results and the computational complexity both indicate that hardware implementation can reach high quality results in realtime.

TV deinterlacing is a novel approach and introduces a whole new theoretical framework for video processing and the results advocate the further exploration of the ideas presented here.

References

1. Aubert, G. Kornprobst, P.: Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations. Applied Mathematical Sciences, Vol. 147. Springer-Verlag, Berlin Heidelberg New York (2002)
2. Bellers, E. B., De Haan, G.: Deinterlacing: A Key Technology for Scan Rate Conversion, Elsevier, Amsterdam (2000)
3. Capodiferro, L.: Interlaced to Progressive Conversion by Median Filtering. In: Chiarglione, L. (ed.): Signal Processing of HDTV, II, Elsevier Science Publishers, Amsterdam (1990)
4. Chan, T. F., Shen, J.: Mathematical Models for Local Nontexture Inpaintings. SIAM Journal on Applied Mathematics, vol. 62, no. 3 (2002) 1019-1043
5. Chanas, L.: Mthodes variationelles pour la restauration de squences d'images fortement dgrades. Application aux images blouies par laser. Ph.D. Thesis, University of Cergy-Pontoise (2001)
6. Doyle, T., Looymans, M.: Progressive Scan Conversion Using Edge Information. In: Chiarglione, L. (ed.): Signal Processing of HDTV, II, Elsevier Science Publishers, Amsterdam (1990)
7. Dreier, H.-J.: Line Flicker Reduction by Adaptive Signal Processing. In: Chiarglione, L. (ed.): Signal Processing of HDTV, II, Elsevier Science Publishers, Amsterdam (1990)
8. Haavisto, P., Juhola, J., Neuvo, Y.: Scan Rate Up-Conversion Using Adaptive Weighted Median Filtering. In: Chiarglione, L. (ed.): Signal Processing of HDTV, II, Elsevier Science Publishers, Amsterdam (1990)
9. Kovacevic, J., Safranek, R. J., Yeh, E. M.: Deinterlacing by Successive Approximation. IEEE Trans. Image Proc., Vol. 6, no. 2 (1997)

10. Mumford, D.: Bayesian Rationale For The Variational Formulation. In: ter Haar Romeny, B. M. (ed.): Geometry-Driven Diffusion In Computer Vision, Kluwer Academic Publishers, Dordrecht (1994) 135–146
11. Pigeon, S., Guillotel, P.: Advantages and Drawbacks of Interlaced and Progressive Scanning Formats. CEC RACE/HAMLET Deliverable no R2110/WP2/DS/R/004/b1 (1995)
12. Rudin, L., Osher, S., Fatemi, E.: Non Linear Total Variation Based Noise Removal Algorithms. Physica D. 60. (1992) 259-286
13. Skarabot, A. Ramponi, G., Buriola, L.: FPGA architecture for a videowall image processor. Proc. SPIE Intern. Symp. Electronic Imaging 2001 (2001)
14. Skarabot, A., Ramponi, G., Toffoli, D.: Image sequence processing for videowall visualization. Proc. IST/SPIE 12th Annual Intern. Symp. Electronic Imaging 2000 (2000)
15. Tekalp, A. M.: Digital Video Processing. Prentice Hall, Upper Saddle River, NJ, (1995)
16. Thomas, G. A.: A Comparison of Motion-Compensated Interlace-to-Progressive Conversion Methods. Signal Process.: Image Commun., vol. 12, no. 3 (1998)
17. Wang, Y. Ostermann, J., Zhang, Y.Q.: Video Processing and Communications. Prentice Hall, Upper Saddle River, NJ (2002)