Video Deinterlacing by Graphic based Visual Saliency

Vayugundla Preethi
Department of Electronics and Communication Engineering
Narayana Engineering College, Nellore, AP, India

Dr. K.S Sagar Reddy
Professor, Department of ECE
Narayana Engineering College, Nellore, AP, India

J.Sunil Kumar
Associate Professor, Department of ECE
Narayana Engineering College, Nellore

Abstract: Video Deinterlacing is a technique in which the interlaced video format is converted into progressive scan format for smooth viewing of the video. The interlacing is generally a trade-off between the frame-rate and bandwidth capacity of the video which is to be processed. As of today's generation all the videos are in digital formats. The available displays require the use of progressive video input. The process of de-interlacing involves converting stream of interlaced frames in a video to progressive frames. A wide variety of de-interlacing techniques are available. Deinterlacing requires the display to buffer one or more fields and recombine them to a full frame. There are various methods to de-interlace a video and each method produces its own artifacts. Here we use to our advantage the visual saliency model, Graphical Based Visual Saliency (GBVS) to determine the salient regions of the frame with the help of feature extraction. Then apply our motion compensated technique to lower heavy computation. Experimental results demonstrate a significant gain in quality reconstruction as compared to other de-interlacing implementations.

Keywords: Progressive Scan, De-interlacing, Feature Extraction.

I.INTRODUCTION

The process of deinterlacing involves transforming a stream of interlaced frames in a video to a progressive frames. An interlaced video is a sequence of fields. A field is half of the scan lines of an entire image. An image consists of even scans (0, 2, 4 ...) or the odd lines (1, 3, 5...) considering a unique and independent image. Another way to think about interlaced video is to imagine a camera that stores 59.94 frames per second [1]. This camera only stores half of the scanlines and switches between storing the even and odd scanlines. Interlacing is a very effective form of compression. The human eye is less sensitive to high frequency motion. To makes this statement simpler to understand. Think about a spinning bicycle, at proper speed the eye can't distinguish the individual spokes. Applying the same concept here with interlacing a video at proper frequency (59.94 fps)if we rapidly switch between even and odd scanlines the image will appear as proper as whole image.

Humans or most of the vertebrates have ability to move their eyes. This common development in animals has helped to draw the most visible features of a scene and spends relatively less processing the background of the video. A human when looks at bird in the sky focuses on to the primary object 'the bird'. Humans fixates on the bird and also fixates-time free viewing the scenario meaning observing the bird fly. When asked to recall the situation, we do best at recalling the way the bird flu or the colors of the bird feathers, but we won't be able to describe the number of clouds. Recalling and remembering situation of the image is a process from the vision community. We engineers have a tendency to learn and develop most of our technology from techniques inherited from the modern biology. This lends to a very important engineering applications. The standard approach to motivated feature selection leads to compression and recognition. Feature selection is the motivation for this process, the distinct features of the scene (or video) are selected and the center surrounding operations are highlight the local gradients [3]. Visual Saliency is a distinct subjective perceptional quality which makes some items in the world stand out from their neighbors and immediately grab our attention.

II.RELATED WORKS

2.1 Visual Saliency-

Various approach with vivid differences all comes to differences as it depends on the piecewise performance. Removing uncertain background details from the origin, thus below are three stages to attain visual saliency.
(vs1) Extraction: draw out feature vectors at locations over the image[5].
(vs2) Activation: with the help of feature vectors draw out activation maps.
(vs3) Normalization/combination: normalize the activation maps. Combine the maps into a single activation maps.

This model is simple and admits combination with other maps. It is also plausible and as of now capable to be naturally paralleled. Classically these algorithms, step(vs1) is accomplished by biological filters step (vs2) is done by subtracting feature maps at different scales (therefore, "c-s" for "centre" and "surround". When it comes to (vs3) the final step is accomplished with the help of intermediate stages.
1. A normalization scheme based on local maxima ("max-avg")
2. An interactive scheme based on convolution with a Difference-of-Gaussians Filter ("DoG")
3. A Nonlinear Integrations ("NL") approach which splits local feature values by weighted averages of surrounding values in a way that is modeled to psychophysics data. These above steps helps us to retrieve the image graph required to further process.

III.OUR PROPOSED METHOD

3.1Graphic-Based Visual Saliency-

To achieve the improved bottom-up visual saliency model, considering two important steps which will lead us to the map required, firstly forming activation maps on a particular feature channel [7]. After that we normalize them. Here we introduce a map to obtain the normalization operator $N(.)$. The normalization operator promotes maps in which a small number of strong peaks of movement in obvious present locations, at the same time smoothening maps which containing numerous peaks responses $N(.)$ showing in detail as follows,

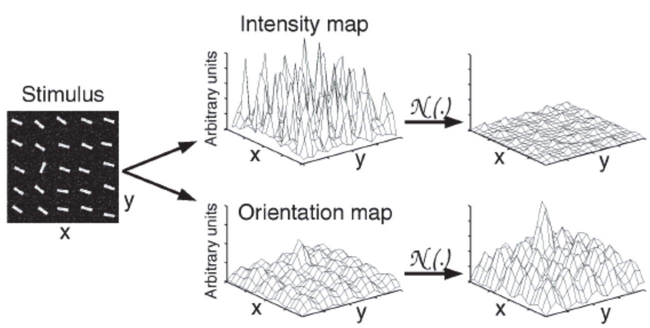

Figure. 1. The normalization operator *N* (.).

This is a simple and easy to understand method which also comes with an advantage, it is biolocially plausible and as naturally parallelized.This architecture, proposed by Koch and Ullman [4].This model has the ability to predict human fixations on 749 variations of 108 natural pictures which gives an appericative percentage of 98 of the ROC area in the human-based control. Whereas in the classical algorithms of Itti& Koch achive only 84%.

To explain in detail the process of extraction, consider an image$I$, draw out the distinct features of the image, highlight the regions establishing vectors and create feature maps.A possible linear filtering is followed by some elementary nonlinearity, followed by activation and normalization and combination.

We implement saliency computations on real images taken from the natural world to achieve optimum results and compare these with resulting maps to predict human fixations. We take a set of natural images and use both the standard techniques and the activated feature maps along with normalization algorithms then we simply sum over the feature channels. The output results in a master saliency map is relative collection of human fixations corresponding to the input image.

3.2 Adaptive Motion Compensated Approach-

The motion compensated (MC) methods are the most frequently used methods to deinterlace a video they presume to provide the best results. We can't say it's the most convenient method either as it comes with its own artifacts. The motion compensated approach requires the largest amounts of storage recourses along with a requirement of large computation. Motion computation must be processed to every pixel of the image which becomes hassle and is not practical approach. Therefore this method isn't real-time applicable. To overcome the burden complex computation, most of the MC methods use a block-based motion-estimation (ME). With the help of blocking artifacts the deinterlaced image looks smooth. There is a possibility that unreliable motion information may limit the usage of MC techniques.

Here with the help of our proposed method with the introduction of motion compensation on high dynamics and smooth areas, they help figure out the "degree of motion" around the edges and decides which type of deinterlacing (MC, spatial or spatial-temporal) is most suitable for this region. This helps us achieve the best quality and smooth deinterlacing. Moreover, since this proposed algorithm is quite adaptive it applies to appropriate interpolation scheme on different parts of the field, the computational complexity of our procedure helps lower the complexcity related to the Motion Computation deinterlacing.

3.3 Deinterlacing-

The primary stage in our deinterlacing process consists in region texture classification based on the countours context. We determine the texture by the number of pixels on the edges with respect to the total number of pixels belonging to that region.

For proper string edge detection we select the optimal Canny Filter. The Canny Edge Detector works in a multi-stage process.

1. To smoothen out the edges of the image we take the help of Gaussian Convolution. Gaussian Filters help also to remove the noise.
2. Then the smooth high gradient fields and afirst directive 2-D operator is used to highlight the regions [2].
3. The pixels which are not part of the local maxima are set to zero, this suppresses all the image information and helps us determine non-maximum suppression.
4. Using hysteresis threshold the edges are finally identified [6]. We achieve a complete suppression of false contouring and the classification of different blocks are improved.

Second stage of deinterlacing comprises of region classification based on the motion of the image. With the help of temporal-deinterlacing we can avoid any abrupt changes in motion. Image is detected for shot-cuts.The first step is based on motion-based area classification. If the blocks are related to the first class, presuming it has an edgy content and dynamic behavior we consider a spatial deinterlacing method is preferred. To avoid any artifacts causedby temporal deinerlacing. The blocks are classified as a highly textured and belonging to a static area are deinterlacing in spatial-temporal manner. If no shots are detected in the current field and its neighbors. The temporal fields are adjusted with the help of forward motion-estimation (ME).

Here k is a weight for the MC interpolation, and xo is obtained by the edge-line minimization technique. In the prior existing algorithm the value of k was determined a value as k=8 an integer. In our proposed algorithm we obtain the k value from the saliency map. The present k value is continuously changing for evidently every block.

Our proposed method is computing model, with a bottom-up saliency graphs and shows remarkable consistency with the attention deployment of human subjects. This method uses a novel ideas and puts together an application of a graph theory to concentrate mass on activation maps and forms these activation maps from the barest (raw) features. We compared our model with other established models and found that ours had a better overall performance.

IV.SIMULATION RESULTS

Figure 2: Input Image Before Deinterlacing

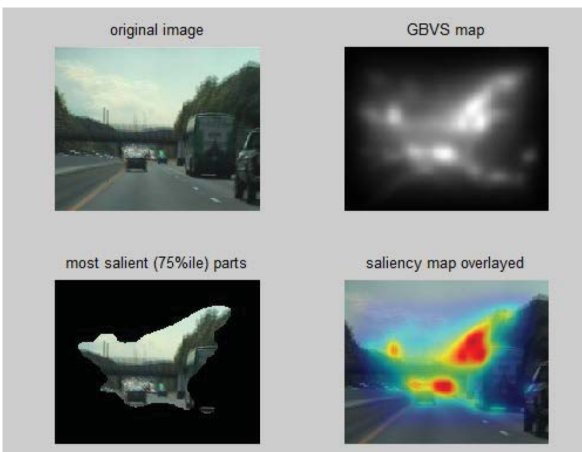

Figure 3: Output of Saliency Map

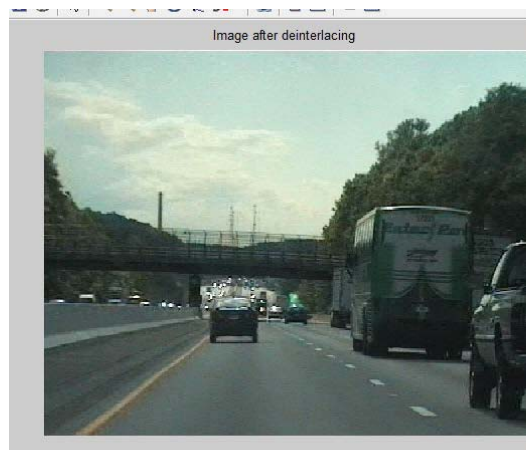

Figure 4: Output Image After Deinterlacing

V.CONCLUSION

Our proposed method is computing model, with a bottom-up saliency graphs and shows remarkable consistency with the attention deployment of human subjects. This method uses a novel ideas and puts together an application of a graph theory to concentrate mass on activation maps and forms these activation maps from the barest (raw) features. We compared our model with other established models and found that ours had a better overall performance.

REFERENCES

[1] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y.H. Lai, N. Davis, and F. Nuflo, "Modelling Visual Attention via Selective Tuning," Artificial Intelligence, vol. 78, no. 1-2, pp. 507–545, Oct. 1995.

[2] M.W. Cannon and S.C. Fullenkamp, "A Model for Inhibitory Lateral Interaction Effects in Perceived Contrast," Vision Res., vol. 36, no. 8, pp. 1,115–1,125, Apr. 1996.

[3] J.M. Wolfe, "Guided Search 2.0: A Revised Model of Visual Search," Psychonomic Bull. Rev., vol. 1, pp. 202–238, 1994.

[4] C. Koch and S. Ullman, "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," Human Neurobiology, vol. 4, pp. 219–227, 1985.

[5] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C.H. Anderson, "Overcomplete Steerable Pyramid Filters and Rotation Invariance," Proc. IEEE Computer Vision and Pattern Recognition, pp. 222-228, Seattle, Wash., June 1994.
[6] The C++ implementation of the model and numerous examples of attentional predictions on natural and synthetic images can be retrieved from http://www.klab.caltech.edu/~itti/attention/.
[7] J.P. Gottlieb, M. Kusunoki, and M.E. Goldberg, "The Representation of Visual Salience in Monkey Parietal Cortex," Nature, vol. 391, no. 6,666, pp. 481-484, Jan. 1998.
[8] M.I. Posner and Y. Cohen, "Components of Visual Orienting," H. Bouma and D.G. Bouwhuis, eds., Attention and Performance, vol. 10, pp. 531–556. Hilldale, N.J.: Erlbaum, 1984.
[9] A.G. Leventhal, The Neural Basis of Visual Function: Vision and Visual Dysfunction, vol. 4. Boca Raton, Fla.: CRC Press, 1991.
[10] R. Milanese, S. Gil, and T. Pun, "Attentive Mechanisms for Dynamic and Static Scene Analysis," Optical Eng., vol. 34, no. 8, pp. 2,428–2,434, Aug. 1995.