Motion Detection in High Resolution Enhancement

Xiaodan Hu [1], Avery Ma [1], Ahmed Gawish [1], Mark Lamm [2], Paul Fieguth [1]

[1]Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada
[2]Christie Digital Systems Canada Inc., Kitchener, Canada

Abstract

Shifted Superposition (SSPOS) is a resolution enhancement method where apparent high-resolution content is displayed using a low-resolution projection system with an opto-mechanical shifter. While SSPOS-enhanced projectors have been showing promising results in still images, they still suffer from motion artifacts in video contents. Motivated by this, we present a novel approach to apparent projector resolution enhancement for videos via motion-based blurring module. We propose the use of a motion detection module and a blurring module to compensate for both SSPOS-resulted and natural motion artifacts in the video content. To accomplish this, we combine both local and global motion estimation algorithms to generate accurate dense flow fields. The detected motion regions are enhanced using directional Gaussian filters. Preliminary results show that the proposed method can produce accurate dense motion vectors and significantly reduce the artifacts in videos.

1 Introduction

High-resolution content projection systems have become increasingly popular in consumer markets. Projectors with ultra-high-resolution have been widely used in theaters and amusement parks. While they have been available for several years, such high-resolution projectors are still extremely costly. As such, displaying high-resolution content using low-resolution projector is highly desired.

The concept of resolution enhancement using low-resolution projectors has been explored in past literature, and a number of approaches have been proposed. In the proposed method, we make use of the SSPOS method proposed by Barshan *et al.* [1]. Given an input high-resolution image, SSPOS generates two optimized low-resolution sub-frames and superimposes them in a rapid succession with a fraction of a pixel shift, resulting a perceived high-resolution content. Input images are pre-distorted using Wiener deconvolution for optical aberration from the projector-lens systems. However, Wiener deconvolution generates motion artifacts due to its inherent sharpening in video content. Flickering artifacts typically appears when moving regions with texture details are sharpened. Therefore, it is crucial to estimate the motion of the regions as humans perceive it and enhance them separately. In our work, we propose a motion-based apparent resolution enhancement method which is accomplished by two separate modules: a motion detection module and a blurring module. This approach allows the detection of regions with specific motion directions in the input video and applies different enhancement on the regions based on their underlying motion characteristics. We propose a novel motion detection method which leverage both local and global motion estimation methods to generate accurate dense motion fields. Directional Gaussian filters are applied to enhance the motion regions based on their magnitudes and directions.

Optical flow has achieved great success in estimating pixel-wise flow. However, optical flow focus on local analysis which leads to errors in motion estimation known as aperture problem. Furthermore, because of the brightness constancy assumption, motion vectors inside the textureless objects cannot be detected. Several methods have been proposed to overcome the shortcomings of previous models [3, 6]. Yang *et al.* [2] proposed several energy functions combining three assumptions: brightness constancy, gradient constancy and smoothness constraint. Bruhn *et al.* [4] proposed a method combining local and global (CLG) optic flow. Liu *et al.* [5] realized CLG method with coarse-to-fine frame to get faster calculation. In this paper, we will apply those method to estimate motion between successive frames in the video in order to attenuate the artifacts appeared during movement.

2 System Model

The pipeline for the proposed resolution enhancement method is visualized in Fig. 1. In the resampling module, each high-resolution frame is reproduced by two low-resolution frames and the two frames are combined with pixel shifting. The pixel shifting is accomplished by shifting the image by one pixel and resampling the image. In the pre-distortion module, a Wiener deconvolution operation is applied to enhance the image. In the motion estimation module, a combined global and local method is applied to detect the motion. And then different Gaussian filters in four directions are applied to furthur enhance the motion regions.
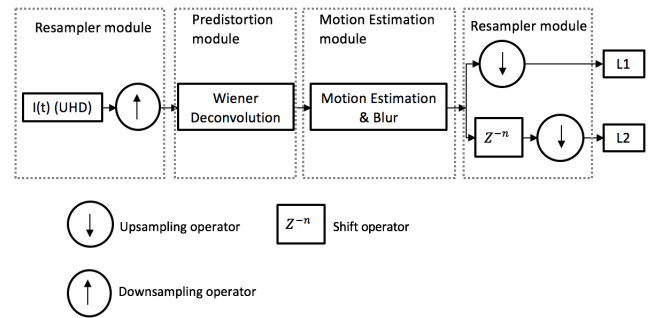

Fig. 1: Pipeline for proposed resolution enhancement method

3 Methods

3.1 SSPOS

The resolution of a video can be enhanced by generating two low resolution frames from the high resolution frame and displaying them within the retinal integration time to achieve higher perceived resolution.

The two sub-frames $I1$ and $I2$ are produced by minimizing the distance between the perceived projector output and the high-resolution target frame. The perceived projection output can be estimated by multiplying the point spread function (PSF) of the projector with the sub-frames independently. The two low-resolution sub-frames are generated with an offset of less than one pixel. The shifting is implemented by re-sampling the sub-frames and shifting diagonally.

Correspondingly, the hardware configuration is set to produce shifted superimposed projection so as to reconstruct the high-resolution frame. The SSPOS is implemented in hardware by optics and mechanics to diagonally move the projected image with a small portion at the target frame rate. The piezoelectric actuator is used to slightly spread the transmissive glass portion to two different positions so that the refraction moves the projection beam between the two positions within one pixel vertically and horizontally. In this case, the original 60Hz ultra-high-definition video is reproduced by a 120Hz low-resolution video shown in two shifted pixel positions.

3.2 Wiener Deconvolution

The projected frames are degraded because of the blurring and noise. The observed image can be modelled as:

$$y(t) = h(t) * x(t) + n(t) \qquad (1)$$

where x is the source image, y is the observed projected image, h is the impulse response (point spread function of the projector/lens system), and n is additive noise.

The objective of deconvolution is to restore a observed degraded image to its original form. The Wiener filter is a linear time-invariant filter whose can compute a statistical estimate of the unknown original signal using the known observed degraded signal as an input and filtering that known signal to produce the estimate as an output which would come as close to the original signal as possible. The Wiener filter can be discribed as follows:

$$G(f) = \frac{1}{H(f)}\left[\frac{|H(f)|^2}{|H(f)|^2 + \frac{N(f)}{S(f)}}\right] \qquad (2)$$

where G(f) and H(f) are the Fourier transforms of g and h at frequency f, S(f)is the mean power spectral density of the original signal x(t), N(f) is the mean power spectral density of the noise n(t).

Thus by pre-distorting the source frames with Wiener filter, the projected frames can be greatly restored.

3.3 Motion Estimation

3.3.1 Grey value constancy assumption

The primary assumption of optical flow estimation is to assume that the grey value of a pixel remains consistent. [10]

$$I(x,y,t) = I(x+u, y+v, t+1) \qquad (3)$$

$$I_x u + I_y v + I_t = 0 \qquad (4)$$

Which involves minimizing the brightness between corresponding pixels. Here the energy function is defined on the two constants u and v at location (x,y,t).

$$E_{data} = \int (|(I_x u + I_y v + I_t)^2|)dx \qquad (5)$$

where the $I_x, I_y = \nabla I_1 = J_1$ and $I_t = e_i$ is the temporal derivative which varies along with the video sequence. By minimizing the energy function, the unknown optic flow vector (u,v) can be determined by solving $\partial_u E_{data} = 0$ and $\partial_v E_{data} = 0$, which gives Eq. 6,

$$\begin{bmatrix} K_\rho * (f_x^2) & K_\rho * (f_x f_y) \\ K_\rho * (f_x f_y) & K_\rho * (f_y^2) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -K_\rho * (f_x f_t) \\ -K_\rho * (f_y f_t) \end{bmatrix} \qquad (6)$$

$$A\Delta u = b \qquad (7)$$

let $A = \begin{bmatrix} K_\rho * (f_x^2) & K_\rho * (f_x f_y) \\ K_\rho * (f_x f_y) & K_\rho * (f_y^2) \end{bmatrix}$ and $b = \begin{bmatrix} -K_\rho * (f_x f_t) \\ -K_\rho * (f_y f_t) \end{bmatrix}$.

However, this is not sufficient to uniquely calculate the two unknowns u and v, which is known as aperture problem. Supposing that a slanted edge is moving horizontally, and if the matching happens in a completely textureless region, it's hard to recover the true motion from Eq. 7.

3.3.2 Gradient constancy assumption

The gradient of the image grey value is assumed not to vary due to the displacement. [6] This gives

$$\nabla I(x,y,t) = \nabla I(x+u, y+v, t+1) \qquad (8)$$

$$E_G = \int |\nabla I(x+u, y+v, t+1) - \nabla I(x,y,t)|^2 dxdy \qquad (9)$$

3.3.3 Smoothness Assumption

Because of the aperture problem, only the flow in normal direction to the gradient can be estimated, which caused ambiguous in flat regions. No neighboring pixels are taken into account. Thus imposing spatial smoothness constraints to the flow field is necessary and useful. The smoothness constraints, shown as squared penalties, assumes that adjacent pixels should move together as much as possible:

$$E_{Smooth} = \sum (u_x^2 + u_y^2 + v_x^2 + v_y^2) \qquad (10)$$

3.3.4 A Combined Local-global (CLG) Method

The total energy is the weighted sum between the data term and the smoothness term

$$E(u,v) = E_{Data} + \gamma E_G + \alpha E_{Smooth} \qquad (11)$$

with some regularization parameter alpha > 0. Now the goal is to find the functions u and v that minimize this energy.

If the local and global estimation are combined together, either the robustness of local approaches or the density of global methods could be gained. we replace $[J_i J_i^T]$ and $J_i e_i$ in Eq. 10 with the A and b in Eq. 7 [4, 10], then local constraints elements are used in global minimization functions.

3.3.5 Coarse to Fine

Coarse-to-fine strategy can incrementally compute the optic flow field for arbitrary video sequence since this strategy is based on an image pyramid model. The resolution is refined from the topmost coarse scale to the lowest fine level step by step. Coarse-to-fine approach can also help us to accelerate the searching process when minimizing the well-defined alignment cost function.

The algorithm starts from the coarser levels at the top of the pyramid fully searching over a smaller number of pixels for the best displacement u that minimize the difference between two frames. For energy functional with a global minimum, the motion vector estimated from this level will be used to correct/warp one of the frames before a search at next finer level. There only small displacement fields are required to computer at the next finer level and only the motion increments du and dv between the first image and the warped second image are estimated.

$$E_{data}^1 = \int g * M_1(x,y)|I_1(x+u_1, y+v_1) - I_2(x,y)| \qquad (12)$$

where g is a Gaussian filter. M_1 and M_2 is the visible mask of a layer at frame H and I. (u_1, v_1) is the flow field from H to I. The data term E_{data}^2 for (u_2, v_2) is defined similarly.

While Coarse-to-fine method might cannot gain the same result as a full search, the calculation for final displacement field obtained by a summarizing all motion increments is much more accurate and faster [5].

3.3.6 Gaussian Filter

In order to eliminate the artifacts appeared when the finest areas are moving, Gaussian filters are applied to the motion area to blur it in corresponding directions. Here we defined a four-direction filter which can divide the estimated motions to horizontal, vertical, main-diagonal and anti-diagonal direction separately. And then all of the four-direction motion vectors will pass through a magnitude filter to reduce the noise. Finally, Gaussian filters in these four directions are applied on the moving regions according to the motion direction of each pixel.

4 Results

In this section, the efficacy of the proposed motion detection method is demonstrated on various test videos. To investigate the performance of the proposed method, motion detection algorithm was performed on videos with different scene characteristics: synthetic and real-world scene.

We made comparison to the results of the other methods. Fig. 3 shows our method compared to the Classic+NL-fast method [7], and Combined Local-Global method [8] as well as Robust Discontinuity-Preserving approach [9]. In general, by inspecting visually, our result for motion estimation is better and gives accurate, dense and smooth motion flow fields.

5 Conclusion

In this paper, we present a motion-based projector apparent resolution enhancement method. Given a high-resolution input image, SSPOS performs resolution enhancement by computing optimized low-resolution sub-frames, de-blurring the images by passing them through Wiener filter while blurring the regions that have motion according to their motion direction to reduce flickering artifacts, and
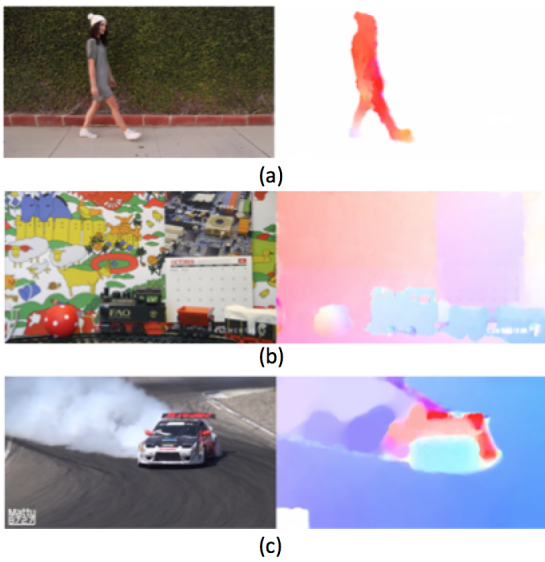
[4] A. Bruhn, J. Weickert. and C. SchnÃűrr. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods *International Journal of Computer Vision (IJCV)* (2005).

[5] C. Liu, Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. *Doctoral Thesis. Massachusetts Institute of Technology* May (2009).

[6] T. Brox, A. Bruhn, N. Papenberg, and J.Weickert. High accuracy optical flow estimation based on a theory for warping. *European Conference on Computer Vision (ECCV)* (2004).

[7] D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision (IJCV)* (2014)

[8] J. Jara-Wilde, M. Cerda, J. Delpiano, and S. HÃďrtel. An Implementation of Combined Local-Global Optical Flow. *Image Processing On Line* (2015)

[9] N. MonzÃşn, A. Salgado, and J. SÃąnchez. Robust Discontinuity Preserving Optical Flow Methods. *Image Processing On Line* (2016)

[10] R. Szeliski. Computer Vision: Algorithms and Applications. (2010)

Fig. 2: Video scene and motion detection result. (a): girl walking from left to right. (b) toy train running from right to left. c) racing car drifting towards bottom left of the frame
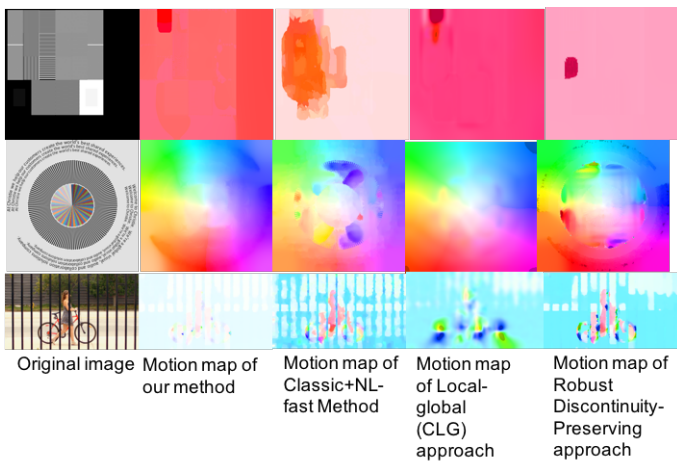

Fig. 3: Comparison with algorithms proposed in [8, 9] on the "shifting lines" sequence, "anti-clockwise spinning" sequence, and "wheeling a bicycle" sequence.

rendering them in rapid succession with a fraction of a pixel shift. We showed that our motion estimation module can produce robust dense motion vectors and our final results weaken the artifacts appeared by movement in the video.

Acknowledgments

References

[1] E. Barshan, M. Lamm, C. Scharfenberger, P. Fieguth. Resolution Enhancement Based on Shifted Superposition. *SID Symposium Digest of Technical Papers* (2015).

[2] J. Yang and H. Li. Dense, Accurate Optical Flow Estimation with Piecewise Parametric Model. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015).

[3] D. Fortun, P. Bouthemy, and C. Kervrann Optical flow modeling and computation: a survey. *Computer Vision and Image Understanding* (2015).