Robust Visual Enhancement of Moving Contents in Projected Imagery

Xiaodan Hu*, Mohamed A. Naiel*, Zohreh Azimifar§, Mark Lamm†, and Paul Fieguth*

*University of Waterloo, Waterloo, ON, Canada

§Shiraz University, Shiraz, Iran

†Christie Digital Systems Canada Inc., Kitchener, ON, Canada

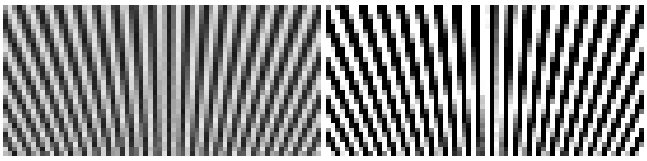
Email: {x226hu, mohamed.naiel, azimifar, pfieguth}@uwaterloo.ca, mark.lamm@christiedigital.com

Abstract

For any projection system, one goal will surely be to maximize the quality of projected imagery at a minimized hardware cost, which is considered a challenging engineering problem. Experience in applying different image filters and enhancements to projected video suggests quite clearly that the quality of a projected enhanced video is very much a function of the content of the video itself; that is, to first order, whether the video contains content which is moving as opposed to still, since the human visual system tolerates much more blur in moving imagery. We would therefore assert that the moving and non-moving pixels of a given video stream should be enhanced differently, using class-dependent video enhancement filters to achieve a maximum visual quality. In this paper, we introduce such a novel motion–dependent content enhancement scheme, based on a pixel-wise moving / non-moving classification, with the actual enhancement obtained via class–dependent Wiener deconvolution filtering. Experimental results on four challenging videos show that the proposed scheme offers improved visual quality.

Author Keywords

Motion Enhancement; Projector Resolution Enhancement; Motion Detection; Motion Artifacts;



Original Image Sharpened Image

Fig. 1. Enhancement of high contrast imagery (left) can produce Moire artifacts (right), which can become dizzying / spinning distractions when moving.

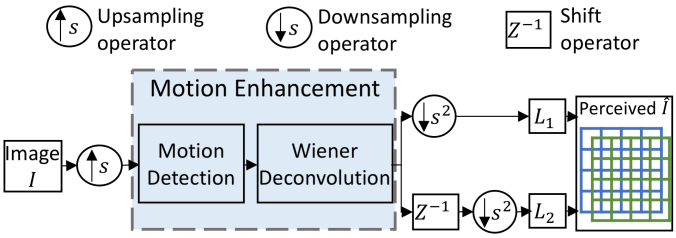


Fig. 2. Proposed moving content enhancement scheme, where s is an upsampling factor in both x and y directions, and L_1 and L_2 are two downsampled sub-images generated with and without one pixel shift, respectively.

1. Introduction

Due to the excessive cost of producing high definition projectors, in many, cases projectors cannot achieve the same high–resolution projection as that of the projected contents [1]. In response, Allen and Ulichney [2] proposed using a low-resolution projection system with an opto-mechanical image shifter to reproduce enhanced high-resolution content, so-called Wobulation, by superimposing two low resolution images in rapid succession to project a higher resolution image. Later, Barshan *et al.* [3] proposed a learning-based resolution enhancement method called shifted superposition, which uses a learning procedure to compute a set of optimized sub-frames. A consequence of projection without enhancement as in [2, 3] is that very unusual time-aliasing can take place in moving imagery, seriously frustrating attempts to enhance video contents. Recently, Ma *et al.* [4] proposed a spatial resolution enhancement kernel to pre-distort the image prior to the projection based on the Wiener deconvolution technique. However, this method sharpens the moving and non-moving contents with the same strength, which may lead to motion artifacts, especially when there is rapid motion inside the video.

In Figure 1, the right image shows the sorts of motion artifacts associated with over–sharpening by using the Wiener deconvolution-based method in [4]. Tests on super-imposed projection display have made very clear that effective filters for static contents, particularly high-contrast contents (such as text), tended to create badly-aliased artifacts in moving regions; similarly filters effective for motion tended to under-enhance / blur other high-contrast contents. Since the human visual system is relatively insensitive to the blur of moving objects, and the super-imposed projection creates the challenge of potential aliasing primarily for moving content, clearly the appropriate response is some sort of motion-dependent enhancement, as shown in the shaded block in Figure 2, which is the focus of this paper.

In this paper, we propose a novel motion-dependent visual enhancement scheme in projector-based systems. In this scheme, a robust motion detection is used to segment the moving regions from the background ones. Then, a less sharpened and a more sharpened Wiener deconvolution filters are applied to moving and non-moving regions, respectively, resulting in a better visual quality. In order to find the best sharpening levels to enhance the image while avoiding severe motion artifacts, an optical flow-based parameter selection technique is proposed. Both quantitative and qualitative results show that the proposed scheme is able to provide a robust solution for videos enhancement including moving objects more than projection without enhancement as well as the recent state-of-the-art enhancement method in [4].

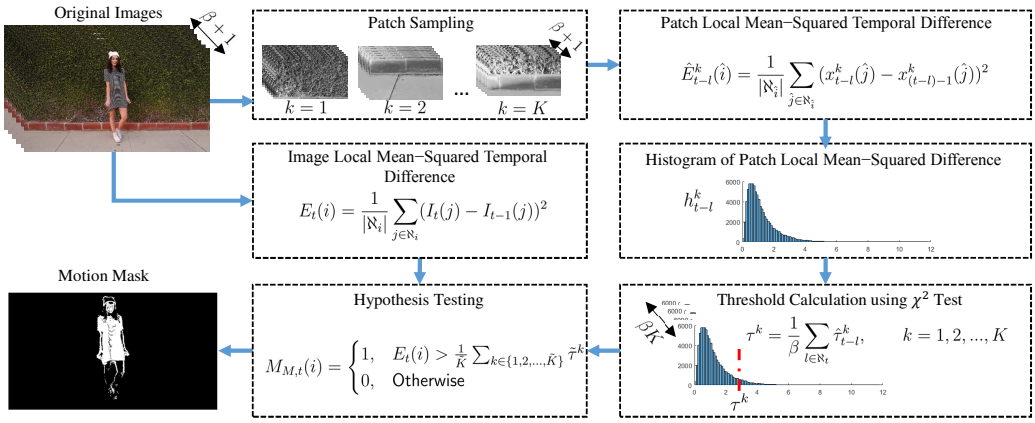


Fig. 3. Block diagram of the proposed motion detection scheme.

2. Methodology

For an effective video enhancement, a robust motion detection method is needed to classify pixels into moving and non-moving classes. Block matching [5] and optical flow [6] are two fundamental motion estimation techniques that have been widely used. However, these methods usually require high computational costs, as well as they do not provide directly a specific class for every pixel. In this section, we introduce a novel hypothesis-testing-based motion detection method to classify each pixel into moving or non-moving class by using statistics of the local mean-squared temporal difference.

2.1. Motion Detection

In order to distinguish moving pixels from non-moving ones, we formulate this classification problem into a statistical hypothesis testing one. In the proposed scheme, a given pixel is assumed to be stationary unless an enough evidence is obtained to argue that this pixel is moving, therefore we define the null and alternative hypotheses as

H_0: A pixel is *stationary*;
H_1: A pixel is *moving*.

Figure 3 shows the block diagram of the proposed motion detection scheme. Let I_t denote the input image at a given time t. The local mean-squared temporal difference, $E_t(i)$, at time t and location i can be obtained as

$$E_t(i) = \frac{1}{j@ij}\sum_{j\,2@_i}(I_t(j) \quad I_{t\ 1}(j))^2 \tag{1}$$

for a spatial neighbourhood $@_i$ around the i^{th} pixel, where $j.j$ denote the cardinality of the set. Since even still video is not perfectly constant (due to camera vibration, sampling error and pixel noise) we essentially having a χ^2 problem. Thus, in order to decide whether a pixel is stationary or not we need to have a threshold on E, where the threshold will need to be dynamic, as this threshold may be content–dependent.

Let the histogram of the $(k,l)^{th}$ patch local mean-squared temporal difference be denoted by $h^k_{t\ l}$, where $k = 1,2,...,K$, $l\ 2\ @_t = \frac{}{2}, \frac{}{2}+1,..., \frac{}{2}$, and K and β are the number of spatial and temporal sampled patches, respectively. Then, the motion threshold of the k^{th} volume location is obtained as:

$$\tau^k = \frac{1}{\beta}\sum_{l2@_t} \hat{h}^k_{t\ l}, \qquad k = 1,2,...,K \tag{2}$$

where $\hat{h}^k_{t\ l} = \underset{e}{\text{argmin}}\{ejCDF(h^k_{t\ l}) \quad p,e\ 2\ E_tg$ (3)

and $CDF(h^k_{t\ l})$ is the cumulative distribution function of $h^k_{t\ l}$, p is chosen to be 0.95, and the total number of thresholds, $\hat{h}^k_{t\ l}$, is $K \quad \beta$. To minimize the effect of moving regions on computing the motion threshold, the lowest $K \quad K$ motion thresholds, τ^k, are selected. Finally, the mask of moving pixels, $M_{M;t}$, is obtained by thresholding $E_t(i)$ as:

$$M_{M;t}(i) = \begin{cases}1, & E_t(i) > \frac{1}{K}\sum_{k2\{1;2;:::;K\}}\tau^k \\ 0, & \text{Otherwise}\end{cases} \tag{4}$$

2.2. Band limited Wiener Deconvolution

Wiener deconvolution filtering has been widely used to enhance the image spatial resolution [7]. In general, such filtering processes operate in the Fourier transform domain and involves domain transformations, convenient conceptually, but resulting in a relatively high computational complexity (for the context of real-time high-resolution data projectors). Recently, Ma *et al.* [4] introduced a spatial kernel derived from the Wiener deconvolution filter, simplifying the filtering operation to be a relatively local spatial convolution. In this section, we give a brief overview of the method in [4] and how we have generalized it to our proposed scheme. The projector's estimated point spread function (PSF) in the Fourier domain is denoted as $H(u,v)$, where u and v represent the spatial frequency indices. The Wiener deconvolution filter $G(u,v)$ is computed as

$$G(u,v) = \frac{1}{H(u,v)}\left[\frac{jH(u,v)j^2}{jH(u,v)j^2 + \frac{1}{SNR}}\right] \tag{5}$$

where G and H are of size $r \quad r$ and SNR is the signal-to-noise ratio. It is shown in [4] that obtaining the associated spatial kernel $g(n,m)$ by directly applying the inverse 2D-DFT on (5) causes oversharpening artifacts for the filtered image. To avoid this problem, a low-pass filter $B(u,v)$ of cutoff frequencies f_{c_1} and f_{c_2} was used in [4] to suppress the high frequency components in $G(u,v)$ as follows:

$$\hat{G}(u,v) = G(u,v)B(u,v) \tag{6}$$

Since $\hat{G}(u,v)$ satisfies the symmetric property conditions of 2D-DFT, the spatial kernel $\hat{g}(n,m)$ can be obtained by employing the inverse 2D-DFT, $F^{\ 1}$, on $\hat{G}(u,v)$ as

$$\hat{g}(n,m) = F^{\ 1}\ \hat{G}(u,v)] \tag{7}$$

Table 1. Quantitative results of the proposed method, Ma *et al.* [4] with different settings and projection without enhancement on the *Spinning* sequence.

Method	SSIM 10 ^{2}	MSE 10 ^{3}	PSNR
Proposed Method (f_M = 28, f_B = 32)	99.79	26.17	15.82
Proposed Method (f_M = 32, f_B = 34)	<u>99.84</u>	<u>19.84</u>	17.02
Ma *et al.* [4] (f_c = 32)	99.82	19.84	<u>17.03</u>
Ma *et al.* [4] (f_c = 34)	**99.86**	**17.98**	**17.45**
Without Enhancement	99.76	31.66	15.00

Note: The best and the second best results on each sequence are shown in boldface and underscore, respectively.

In order to fit the memory of a given hardware, the final normalized spatial enhancement kernel is obtained by cropping $\hat{g}(n,m)$ with a desired size of r r as follows:

$$g \quad n \quad (\frac{r}{2}), m \quad (\frac{r}{2}) \quad ! \quad = \frac{\hat{g}(n,m)}{\sum_{\frac{r-r}{2} \quad n;m < \frac{r+r}{2}} \hat{g}(n,m)} \qquad (8)$$

where $\frac{r}{2}$ $n, m < \frac{r+r}{2}$.

In the proposed scheme, unlike the work in [4], we use two Wiener deconvolution kernels to allow for content–dependent input image enhancement, instead of using only one kernel as in [4]. Let g_M and g_Ω denote the Wiener deconvolution kernels corresponding to the moving and non-moving regions, respectively. We design the two kernels using two different cutoff frequencies f_M and f_Ω, where f_M f_Ω, in order to enhance the input image I according to the motion mask obtained in (4).

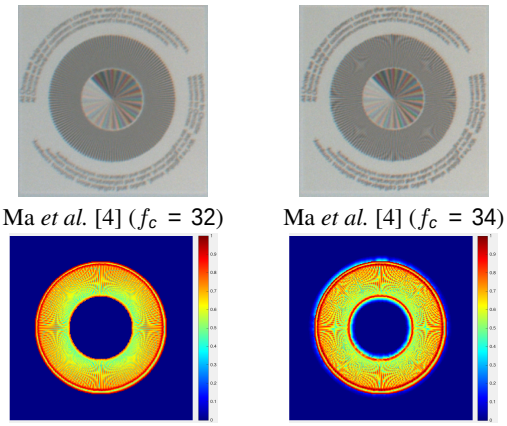
3. Experimental Results

The proposed motion enhancement method has been tested on a 120Hz Christie Digital projector,[1] which includes a piezo-electric actuator introducing a diagonal half-pixel shift. A software–triggered RGB camera was positioned to capture the superimposed projection results. The proposed scheme was evaluated on four videos, namely, *Spinning* (360 frames at 276 276), *RaceCar* (120 frames at 1920 1080), *Girl* (642 frames at 1920 1080) and *ToyTrain* (348 frames at 1280 720). Sample images are shown in the first column of Figure 5. These videos include multiple representations of moving and background regions.

Quantitative Evaluation: To evaluate the performance of the proposed scheme in enhancing projected imagery, we compare the results of our method with that of projection without enhancement and the recent method presented in [4]. For this purpose, the Structural SIMilarity (SSIM) [8], Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) are used.

Table 1 shows the values of these metrics for the proposed method, projection without enhancement and the method in [4] on the *Spinning* video. Although all the three evaluation metrics indicate that the Wiener Deconvolution approach with cut-off frequency f_c = 34 is the best method, on the contrary, it produces severe aliasing shown in the upper right image in Figure 4. Besides, it is seen that the SSIM score of the proposed method is very similar to the method in [4]. Thus, we conclude that SSIM, MSE and PSNR are doing poorly in

[1] Christie Matrix StIM WQ simulation projector



Ma *et al.* [4] (f_c = 32) Ma *et al.* [4] (f_c = 34)

V of the *proposed method* (f_M = 28, f_Ω = 36) V of Ma *et al.* [4] (f_c = 36)

Fig. 4. Comparing the motion artifacts in projected images enhanced by different levels and comparing the magnitude of optical flow, $j\nabla j$, as an assessment of temporal aliasing for two-parameter settings.

evaluating the video quality with motion artifacts and thus, the comparisons using SSIM, MSE and PSNR are all not very meaningful, we suggest to find other metrics for assessing aliasing.

Motion Artifacts Measurement: Since all of the above metrics fail in detecting aliasing introduced in high-frequency patterns and seem to prefer the methods that introduce less blur instead, no matter how severe aliasing in the moving regions is introduced, we propose a new indicator to measure the degree of aliasing artifacts in a given video based on the temporal information.

For this purpose, optical flow [9] has been used to show how close is the enhanced video after projection to the original one. Since the optical flow calculates the apparent content velocities within two successive frames, we believe that the aliasing artifacts will result in additional velocities, which can be measured numerically. To develop the new metrics, let the optical flow error V_t^q between the true velocity $V_t^q(i)$ and estimated velocity $\hat{V}_t^q(i)$ at time t, direction q and location i be defined as:

$$V_t^q(i) = V_t^q(i) \quad V_t^q(i) \qquad (9)$$

Then, the spatial and temporal variances of the optical flow errors are, respectively, obtained as:

$$\sigma_{spa} = \frac{1}{QT} \sum_{q;t} \text{var}(\{\Delta V_t^q(i), i = (1,1),(1,2),...,(N_1,N_2)\}) \qquad (10)$$

$$\sigma_{tmp} = \frac{1}{N_1 N_2 Q} \sum_{i;q} \text{var}(\{\Delta V_t^q(i), t = 1,2,...,T\}) \qquad (11)$$

where var(.) computes the statistical variance, N_1 and N_2 denote the numbers of pixels in the x and y directions, respectively, Q is the number of motion directions, *i.e.*, two directions, and T is the number of frames. The second row of Figure 4 shows the magnitude of the velocities calculated using the optical flow calculation between the 9^{th} and the 10^{th} frames of the *Spinning* video. It is noticed from this figure that the average j Vj of pixels in moving regions increases by increasing the sharpening level of the Wiener deconvolution kernel corresponding to these regions. Table 2 shows the motion artifacts quantified using the spatial and temporal variances, σ_{spa} and σ_{tmp}, calculated for the moving regions of the *Spinning*