Abstract

This paper introduces motion focus as a complement to conventional camera focusing. We show how a relative shift between the image sensor and lens of a camera offers the ability to focus on specific motion in a scene. Focusing on motion allows the use of a longer exposure time while preventing motion blur of the subject of interest. As a result, the subject is captured with a significantly improved image quality. We derive a theoretical performance gain and compare it to existing imaging methods. Furthermore, we demonstrate an experimental motion focus implementation using modified off-the-shelf optical image stabilization hardware, which can obtain an effective exposure time extension of a factor of 94.

1. Introduction

A camera requires a longer exposure period as the illumination level decreases, so that photography of moving subjects becomes problematic. As a result, subject motion during the exposure leads to motion blur, which is detrimental to system performance in applications such as video surveillance systems, traffic monitoring and machine vision. This paper proposes a method for blur-free imaging of moving objects with a fixed camera. We use modified off-the-shelf Optical Image Stabilization (OIS) hardware to briefly track object motion for the duration of image exposure. As a result, only the moving objects within a predetermined velocity range will appear ‘in focus’, hence the name motion focus.

Previous work on motion blur primarily concentrates on removing blur by video signal post-processing, or preventing blur by using one or more shorter exposure periods. Our approach is different in that it captures blur-free images of a moving object without any post-processing or resorting to short exposure periods. Motion focus seems to be a promising technique for imaging of moving objects at low illumination levels. The exposure time can be significantly extended without blurring of the object in focus. This allows the image sensor to accumulate more light, which improves the signal quality. However, it does trade off blurring of the moving object against blurring of objects with different motion trajectories.

This paper presents three main contributions. First, we introduce motion focus as a new concept in photography and discuss the correspondences between our approach and conventional camera focusing techniques. Second, we derive a theoretical bound to the imaging performance of motion focus and compare it to state-of-the-art motion deblurring methods. Third, we describe an experimental implementation using modified optical image stabilizer hardware as a proof of concept.

The outline of this paper is as follows: Section 2 provides an overview of related work. Next, Section 3 introduces motion focus and we evaluate the theoretical performance in Section 4. Section 5 describes the hardware modifications on a commercially available interchangeable lens. Practical results are presented in Section 6. Section 7 summarizes the main conclusions and provides directions for future research.

2. Related work

Panning photography is a closely related technique where the photographer tracks a moving object by panning the camera. To our knowledge, there has been no prior scientific work on panning photography. On the other hand, the removal of motion blur by modified exposure, post-processing and combinations of the two have been extensively studied in scientific literature.

Deblurring reverses the blurring by applying the inverse blurring operator. The blurring operator is mathematically described as a convolution with a blurring kernel. Blind deblurring [4] jointly estimates the blurring kernel and recovers the deblurred image.

Computational Imaging (CI) aims to improve the imaging performance by optical coding followed by a computational decoding step. The optical coding is designed to capture more light, while also making the blur inversion stage numerically stable. Raskar [11] et al. introduce...
**Coded Exposure** to temporally modulate the shutter with a pseudo-random pattern. Levin [6] *et al.* describe a technique called *Motion-Invariant Photography* (MIP), which performs a parabolic camera movement during exposure.

Alternatively, *Multi-Image Denoising* (MID [15]) rapidly captures multiple blur-free images within a short time frame with short exposure time and high camera sensitivity. After aligning the sequence of images, the noisy and underexposed images are accumulated to create a single blur-free and properly exposed image.

*Optical Image Stabilization* prevents blur due to camera motion by inducing a compensating relative shift between the sensor and lens. Jelinek [3] has recently patented a concept for an iris imaging application where a moving object is tracked using a fixed camera.

McCloskey [8] and Mohan [9] have shown that relative motion between lens and imaging plane can be used to obtain custom blurring effects. McCloskey modified a commercially available lens equipped with OIS to capture motion-invariant images as proposed by Levin [6]. On the other hand, the system of Mohan was designed to achieve a shallow depth of field by moving both the lens and the imaging plane. This enables a virtual decrease of the depth of field, primarily for aesthetic purposes.

Most of the prior work on deblurring is not sufficiently robust for real-world applications. Numerical blur inversion typically amplifies noise and can introduce artifacts. Furthermore, multi-image techniques require a high frame-rate camera, increasing the system cost. For this reason, we have explored an alternative technique as introduced in Section 3.

### 3. Motion focus

We propose motion focus, which aims at focusing on a specific object velocity using camera motion. Figure 1 provides a schematic overview of an implementation where a shift of the lens is used to track the object. The lens motion is chosen such that a point with a specific velocity is projected stationary on the imaging plane. There is a high correspondence between motion focus and conventional camera focusing, of which we discuss the basic concepts of aperture, depth of field and focal distance.

**Aperture.** While conventional lens aperture diameter determines the trade-off between depth of field and the amount of light projected on the image plane, shutter speed determines the range of object velocities appearing in focus and the amount of light accumulated by the image sensor. A faster shutter speed renders a larger range of velocities in focus, while a slower shutter speed allows more light to be accumulated.

**Focal distance.** The focal distance $d$ of a camera is approximately related to the focal length $f$ of the lens and the image distance $v$ via the thin-lens law [10], which is given by:

$$
\frac{1}{f} = \frac{1}{d} + \frac{1}{v}.
$$

Motion focus matches the lens motion to the object motion such that the object is projected stationary on the image plane. The lens velocity $s_l$ relates to the object velocity $s$ as follows:

$$
sl = \frac{v}{v + d} \cdot s.
$$

**Depth of field.** Depth of field refers to the depth range in which defocus blur is below the perception level. The smaller the lens aperture diameter $D$, the larger the depth of field $d_{\beta} - d_{\alpha}$. Given a focus distance $d$, any point that is not at depth $d$ is defocused with a blurring circle of diameter $b$ [5]. Let $c$ define the largest imperceptible defocus blur diameter and using Equation (1), we want to obtain an expression for the maximum lens aperture diameter. If we require that points in the depth range $[d_{\alpha}, d_{\beta}]$ have a defocus blur smaller than $c$, we derive the maximum lens aperture diameter as

$$
D \leq c \frac{d_{\alpha}d_{\beta}}{(d_{\beta} - d_{\alpha})f}.
$$

Similarly, given the same maximum motion blur size $c$ for a range of subject velocities in the interval $[s_{\alpha}, s_{\beta}]$, the maximum exposure time $T$ is given by

$$
T \leq \frac{2cv}{d(s_{\beta} - s_{\alpha})}.
$$

These correspondences inspire to apply conventional camera focusing methods and techniques to the new concept of motion focus.

### 4. Noise performance

We want to obtain the best possible image quality of a moving object by optimizing the image capturing method. The baseline method is *impulse imaging*, where the exposure time $T_I$ is chosen short enough to reduce motion blur to within one pixel [1]. This section compares the theoretical performance of our method to the performance of computational imaging and multi-image denoising. We consider the
**performance gain** $G$ which was introduced by Cossairt et al. [1], which denotes the improvement of signal-to-noise ratio (SNR) over impulse imaging. The performance gain is a function of scene illumination, sensor read noise and total exposure time. The best imaging method is the one that reconstructs an image with the highest SNR within a given time budget.

For this comparison, we assume that the object velocity is a-priori known and that the background is uniformly dark. Furthermore, we assume that the camera has no overhead time between capturing frames.

### 4.1. Computational imaging

Cossairt [1] considers the performance of single-shot computational imaging with a linear image formation model and additive Gaussian noise, such that

$$
g = Hf + \eta, \quad \hat{f} = H^{-1}g, \quad (5)$$

where $g$ denotes the measurements vector of size $N$ (the blurred image), $f$ is the latent signal vector of size $N$ (the sharp image) and $H$ is the measurement matrix. The matrix $H$ is the mathematical description of the optical coding. Light-throughput $C(H)$ is defined as the sum of elements in each row of $H$. The light-throughput $C(H)$ ($C$ for brevity) is a measure for the amount of light captured if $H$ is the measurement matrix. The measurement matrix for impulse imaging is the identity matrix $H = I$ and then its light-throughput equals $C = 1$. We consider methods using a stationary camera and masking-based optical coding (i.e. attenuating light). As a result, the elements of $H$ are between 0 and 1.

The noise vector $\eta$ is sampled from a zero-mean Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Cossairt uses an affine noise model [1] with two noise components; a signal-independent read noise component with variance $\sigma^2_r$ and a signal-dependent photon noise component with variance equal to the actual signal level $J$. The total noise variance sums up to $\sigma^2 = J + \sigma^2_r$.

Cossairt defines the performance gain $G_{CI}$ as the ratio of the SNR of the recovered image $\hat{f}$ (SNR$_{CI}$) to the SNR of the image obtained with impulse imaging (SNR$_I$). The SNR of an image with noise variance $\sigma^2$ is defined as SNR = $J/\sigma$, resulting in the following expression for $G_{CI}$:

$$G_{CI} = \frac{\text{SNR}_I}{\text{SNR}_I} = \frac{\sigma_I}{\sigma_J}. \quad (6)$$

Building on the work of Ratner et al. [12, 13], Cossairt derives an upper bound to the performance gain $G_{CI}$ as a function of the light-throughput $C$ and image size $N$, giving

$$G_{CI} < \sqrt{\frac{(N-C)C^2}{NC - 2C + 1}} \cdot \frac{J + \sigma^2_f}{CJ + \sigma^2_r}, \quad (7)$$

where $J$ is the light intensity (number of photons) captured by the impulse camera. Assuming a large image size ($N \gg 1$ and $N \gg C$), the above bound simplifies to

$$G_{CI} < \sqrt{\frac{JC^2}{NC - 2C + 1}} \cdot \frac{J + \sigma^2_f}{CJ + \sigma^2_r}. \quad (8)$$

### 4.2. Multi-image denoising

Whereas impulse imaging captures a single image with a short exposure time, multi-image denoising captures a sequence of impulse images. Let $K$ denote the integer number of impulse images ($H = I$) in the sequence, $\mathbf{u}_k$ the translation motion of the moving object in the $k$th image and $\delta_{\mathbf{u}_k}$ the corresponding shift operator. The image formation model and its inverse thus become

$$\mathbf{g}_k = \delta_{\mathbf{u}_k} \cdot \mathbf{f} + \eta_k, \quad \hat{\mathbf{f}} = \frac{1}{K} \sum_{k=1}^{K} \delta_{-\mathbf{u}_k} \cdot \mathbf{g}_k, \quad (9)$$

where the noise $\eta_k$ is independently sampled from the zero-mean Gaussian distribution $\mathcal{N}(0, \sigma^2_r)$ with $\sigma^2_r = J + \sigma^2_r$ denoting the noise variance in each of the impulse images.

Assuming a spatially constant but signal-dependent noise level, Zhang et al. [15] derive the following expression for the noise level in $\hat{f}$

$$\sigma^2_{\hat{f}} = \frac{1}{K} \sigma^2_r. \quad (10)$$

Note that the light-throughput $C$ is equal to the number of captures images $K$. Hence, the performance gain $G_m$ of MID compared to impulse imaging ($K = 1$) is upper bounded by

$$G_m = \sqrt{K} \leq \sqrt{|C|}. \quad (11)$$

### 4.3. Motion focus

Motion focus captures a single image with exposure time $T = CT_I$, while the object motion is tracked by camera motion. As a result, the object is not blurred ($H = C I$). This results in the following image formation model and its corresponding inverse model:

$$\mathbf{g} = C I \cdot \mathbf{f} + \eta, \quad \hat{\mathbf{f}} = \frac{1}{C} \mathbf{g}, \quad (12)$$

where the scalar $C$ denotes the factor by which the exposure time is increased compared to the impulse camera. The noise vector $\eta$ is independently sampled from a zero-mean Gaussian distribution with variance $\sigma^2 = CJ + \sigma^2_r$. The corresponding performance gain is then given by

$$G_{mf} = \sqrt{\frac{J + \sigma^2_f}{J + \frac{\sigma^2_r}{C} + \frac{\sigma^2_r}{C^2}}} = \sqrt{\frac{C J + \sigma^2_f}{C J + \frac{\sigma^2_r}{C} + \frac{\sigma^2_r}{C^2}}} \quad (13)$$

We use Equations (8), (11) and (13) to compare the noise performance of computational imaging, multi-image denoising and motion focus.
4.4. Comparison

We express the performance gain of the described methods as a function of (1) the ratio between photon noise and read noise variance \( J/\sigma^2 \) and (2) the ratio between total exposure time and the impulse camera exposure time \( T/T_i \). Furthermore, the light-throughput \( C \) cannot grow larger than the integration factor from a longer exposure time, so that

\[
C \leq \frac{T}{T_i}.
\]

The discussed theoretical performance bounds are depicted in Figure 2. All compared techniques offer significant improvements over impulse imaging at low illumination levels (signal level below read noise variance). The performance advantage of computational imaging quickly diminishes at higher illumination levels, which was noted by Cossairt as well [1]. Multi-image denoising and motion focus both offer substantial image quality improvements, but motion focus has a clear advantage at low-light levels.

5. Experimental set-up with OIS hardware

Motion focus requires camera movement to track a moving object. Levin [6] uses a platform to move the entire camera during exposure. McCloskey [8] shows that modified Optical Image Stabilization hardware can be used to induce camera motion as well. Commercial Optical Image Stabilizers compensate for camera shaking by either shifting the image sensor or by shifting one or more lens elements. The latter method is chosen for our prototype because of the easy availability of photography lenses equipped with image stabilization. This section describes the modification of a EF-S 18-55mm IS lens to allow for direct control of the lens shifting unit.

5.1. Off-the-shelf hardware

The EF-S 18-55mm IS lens \(^1\) contains an integrated image stabilizer assembly. A pair of pitch and yaw gyroscopes senses the camera rotation and an on-board microcontroller computes the compensating \( x \) (horizontal) and \( y \) (vertical) lens shifts. The corrective lens element is shifted by two voice coils, one for each axis. A dual H-bridge driver feeds the two voice coils and is controlled by a total of 4 Pulse-Width Modulated (PWM) signals from the microcontroller. The moving lens element is suspended by a pair of springs and dampers [7].

5.2. Modifications

The 4 PWM signals from the integrated microcontroller are disconnected from the dual H-bridge driver and replaced by an external microcontroller to directly control the lens shift. The external microcontroller receives horizontal and vertical velocity setpoints via a serial interface. Figure 3 depicts a block diagram of the image stabilizer system with the described modifications. The modified lens is mounted on a Nox-20 Internet Protocol camera with a pixel pitch of 6.4 \( \mu \text{m} \) which continuously captures images at a rate of 25 frames per second. The external microcontroller synchronizes lens motion with the exposure period using the camera flash trigger. The microcontroller ramps up the PWM duty cycles during the exposure time such that the lens shift velocity corresponds to the velocity setpoints. After the exposure has completed, the microcontroller returns the lens to its initial position prior to starting the next exposure.

5.3. Calibration

Accurate tracking of moving objects requires calibration of the dynamic response of the shifting lens. We calibrate the steady-state response and step response of our modified image stabilizer.

Steady-state response. The steady-state response characterizes the lens position for a given PWM duty cycle after the transient oscillations have vanished. We characterize the response by gradually increasing the differential PWM duty cycle of a single axis (i.e. \( PWM_{X+} \) minus \( PWM_{X-} \) for the horizontal axis) from -100% to +100% and from +100% to -100%. The camera captures an image and the lens displacement is computed from the image displacement. Both the horizontal and vertical lens displacement in response to both horizontal and vertical PWM duty cycle are calibrated and the results are depicted in Figure 4.

The maximum lens shift is approximately 300 \( \mu \text{m} \) from the center, making the total range of motion approximately 600 \( \mu \text{m} \) for both axes, corresponding to 94 pixels. There is significant crosstalk between the two axes, as illustrated by Figures 4(b) and 4(c). This may be caused by a small skew between the image sensor and the lens shift assembly.

We model the lens shift with the following linear model:

\[
\begin{bmatrix}
x_h \\
x_v
\end{bmatrix} = C \begin{bmatrix}
PWM_{h} \\
PWM_{v}
\end{bmatrix},
\]

\(^2\)Commercially available from Ampeye Ltd.

---

\(^1\)Commercially available from Canon Inc.
Figure 2. Theoretical performance bounds of computational imaging, multi-image denoising and motion focus combined in each subfigure, for three different lighting scenarios. Motion focus offers a significant improvement if the signal level is below the read noise variance.

(a) Low-light scenario ($J/\sigma^2 = 0.1$). (b) Medium-light scenario ($J/\sigma^2 = 1$). (c) High-light scenario ($J/\sigma^2 = 10$).

Figure 4. Optical Image Stabilizer steady-state response. Solid line depicts lens position response to increasing PWM duty cycle, dashed line represents the response to decreasing duty cycle.

\[
\begin{bmatrix}
\text{PWM}_h \\
\text{PWM}_v
\end{bmatrix} = C^{-1}
\begin{bmatrix}
x_h \\
x_v
\end{bmatrix},
\]

where the PWM duty cycle is expressed as a percentage and the lens position in micrometers. The following calibration matrix was obtained by linear least-squares approximation using the data from Figure 4, leading to

\[
C = \begin{bmatrix} 3.38 & -0.21 \\ 0.25 & 3.39 \end{bmatrix}.
\]

The root-mean-square position error between the linear approximation and the real lens position has been found to be 14 µm, which corresponds to a positioning inaccuracy of approximately 2 pixels.

**Step response.** The step response characterizes the lens position as a function of time in response to a sudden increase of the PWM duty cycle. After applying the step input, the lens position trajectory between 0 and 20 ms is depicted in Figure 4. The image exposure may only start after the transient response has finished. We observe that after 5 ms, the lens position stays within 10% (8 µm) from its steady-state position, which is roughly the size of one pixel.

6. Practical results

Our camera system was set up to capture images of traffic scenes with vehicles moving at a speed of 50 km/h. The object distance $d$ is approximately 110 m and the image distance of the camera $v$ is set to 55 mm. Applying Equation (2), we find that a horizontal lens speed of 6.9 mm/s is required to implement motion focus for the moving traffic.

First, we emulate the impulse camera by a short exposure time of 5 ms and without shifting the lens in Figure 6(a). Although there is no motion blur, the image brightness is too low. Second, although the object motion causes significant motion blur, we increase the exposure time to 25 ms to obtain an appropriate brightness as in Figure 6(b). Finally, we 'focus' on the object by shifting the lens as in Figure 6(c).
7. Conclusions and future work

We have introduced motion focus as a new concept in photography and videography which allows for blur-free imaging of moving objects without resorting to shorter exposure times. The use of relatively long exposure times drastically improves the image quality of the object of interest, at the expense of blurring the background or objects with different velocities. We have shown that motion focus using a moving camera lens offers significant image quality improvements in low-light conditions where the signal level is below the read noise variance. Finally, a proof of concept with off-the-shelf image stabilizer hardware inducing the required camera movement, has been demonstrated.

Whereas conventional focusing has a long history in photography, motion focus is a novel but logical next step to enhance the capabilities of cameras. Future research is planned to explore methods for rapidly capturing and merging multiple images focusing at different velocity ranges. The work of this paper can be further extended to the area of focal stacking [2, 5, 14]. Furthermore, research may be performed on how motion focus relates to and compares with Motion-Invariant Photography [6].

References


