Master Thesis

Multi-Frame Image Restoration by a Variational Bayesian Method for Motion-Blur-Free Multi-Exposure Photography

Supervisor Professor Michihiko Minoh

Department of Intelligence Science and Technology
Graduate School of Informatics
Kyoto University

Motoharu Sonogashira

February 9, 2015

Multi-Frame Image Restoration by a Variational Bayesian Method for Motion-Blur-Free Multi-Exposure Photography

Motoharu Sonogashira

Abstract

In photography, motion blur degrades the quality of images captured in the presence of motion of either a camera or an object. In this work, we aim at photography free from motion blur.

Short exposure reduces motion blur but produces noise. While *multi-exposure* is effective at reducing motion blur without producing noise, it has a problem that inaccurate image registration limits the performance of image restoration.

In this paper, we solve the problem of multi-exposure by *multi-frame image restoration*. To improve registration by restoration, we perform registration and restoration iteratively. Specifically, we enable robust multi-frame image restoration by a *variational Bayesian* method.

The effectiveness of the proposed method was evaluated for synthetic image sequences in terms of multiple image quality metrics. As the result, both for translational and for rotational motion, the proposed method achieved higher image quality than a previous method of multi-exposure. In addition, the proposed method was applied to real images to show its performance in the real world, and successfully reduced motion blur without producing noise.

One of the future directions is to improve approximation of large sparse covariance matrices, which contributes to the robustness of the variational Bayesian method. another direction is to further evaluate the effectiveness of the proposed methods in the presence of more complex motion.

動きぶれのない多重露光画像撮影のための
変分ベイズ法による多フレーム画像復元

薗頭元春

内容梗概

画像撮影においては，カメラや被写体の動きがある場合に撮影された画像は動きぶれによって劣化する．本稿では動きぶれのない画像撮影を目的とする．

短露光によって動きぶれは抑えられるが，雑音が生じてしまう．**多重露光**は雑音を生じることなく動きぶれを抑えるのに有効であるが，不正確な画像位置合わせによって画像復元の性能が限られてしまうという問題がある．

本稿では**多フレーム**画像復元によって多重露光の問題を解決する．復元によって位置合わせを改善するために，位置合わせと復元を反復的に行う．具体的には，変分ベイズ法によって頑健な多フレーム画像復元を実現する．

提案手法の有効性を，合成画像列に対して，複数の画質尺度について評価した．結果として，回転・並進のいずれの動きに対しても，提案手法によって多重露光の従来手法よりも高い画質を達成できた．さらに実世界での性能を示すため，提案手法を実画像に対しても適用し，雑音を生じることなく動きぶれを抑えることに成功した．

今後の課題として，変分ベイズ法の頑健性に影響する，大規模疎共分散行列の近似の改良が挙げられる．より複雑な動きに対して，提案手法の有効性をさらに評価することも課題である．

Multi-Frame Image Restoration by a Variational Bayesian Method for Motion-Blur-Free Multi-Exposure Photography

Contents

Chapter 1 Introduction

Photography in dynamic scenes has always suffered from motion blur. Motion blur is image degradation that appears when either a camera or an object moves during exposure. It is an undesirable effect that corrupts important image details, as is clear by comparing the ideal image in Figure 1a and the blurry image in Figure 1b. Our goal is to realize high-quality photography free from motion blur.

Physically, motion blur results from integration of an image of light varying over an exposure time. As the exposure time becomes long, the image of light is likely to vary a lot, producing severe motion blur, as shown in Figure 1b. Once image details are lost in motion blur, it is difficult to recover them by removing the blur afterward [18]. To suppress motion blur before it occurs, the most naïve approach is *short exposure*, i.e., to shorten the exposure time [13]. Today, we can readily capture an image in short exposure even with a consumer camera, and we can make it much shorter with a dedicated high-speed camera, if available. While short exposure does reduce motion blur, however, it also reduces the amount of light recorded by a camera. Consequently, a short-exposure image is too dark to see, and when the lost of brightness is compensated for by post-amplification, it becomes prohibitively noisy, as shown in Figure 1c. *Amplifier noise* is a typical kind of noise produced by short exposure, which is generated electrically and electronically in a camera [13]. While amplifier noise can be reduced to some extent by improving the noise immunity of the camera, *shot noise*, i.e., fluctuation in the number of photons sensed by the camera [13], still matters. In reality, short-exposure noise is a mixture of them, and often too much to remove, given only a single image [18]. The trade-off between motion blur and noise is a serious problem of photography, especially in scenes with low illumination and large motion, e.g., night or indoor scenes with fast moving objects, where long exposure leads to blur and short exposure to noise.

Recently, in the emerging field of *computational photography*, a new approach to motion-blur-free photography has been developed [18], to which we refer as *multi-exposure*. The concept of multi-exposure for motion-blur-free photogra-

(a) (b) (c)

Figure 1: (a) Ideal image. (b) Blurry image captured in long exposure. (c) Noisy image captured in short exposure.

phy [14] is depicted in Figure 2: First, with a camera, we observe a sequence of multiple images, each of which is exposed shortly, and thus not blurry but noisy. Then, with a computer, we restore a clean image, which is free from both blur and noise, by removing short-exposure noise from one of the noisy images using the others. Owing to the use of the multiple images, we can successfully remove the noise increased by short exposure. Regardless of kinds of motion, e.g., large ego-motion of a camera and motion of an object, multi-exposure is effective at reducing blur without producing noise. In addition, it is a software-driven approach, and thus requires no special hardware but only multi-shot functionality, which many consumer digital cameras have today. Thereby, multi-exposure has been already implemented in several cameras in market [13].

Despite reported successes, the performance of multi-exposure has been limited by underlying *image registration*, i.e., alignment of observed noisy images to compensate for motion among them, which is required prior to image restoration [15], as depicted in Figure 3a. On one hand, to remove short-exposure noise by making full use of multiple images, restoration needs registration to tell precise correspondences among the images. On the other hand, the registration is inherently sensitive to the noise [15]. Thus, the registration becomes inaccurate in the presence of the short-exposure noise, and the restoration based on it cannot remove the noise effectively, or sometimes may produce artifacts, such as oversmoothing of image details, as we see later in Chapter 6.

In this work, we go beyond the limitation of multi-exposure by incorporating registration into a single framework of *multi-frame image restoration*. The idea is as follows: Even if we cannot perform registration of noisy images ac-

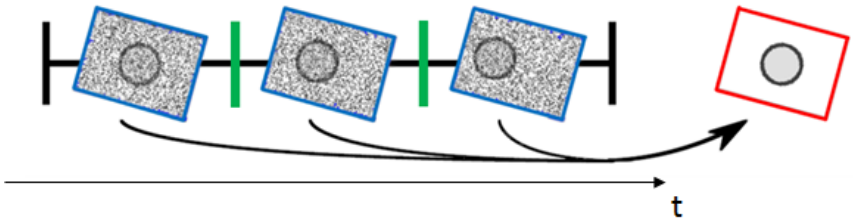


Figure 2: Multi-exposure photography. First, a camera observes a sequence of multiple noisy images. Then, a computer restores a single clean image from the noisy image sequence.

curately enough to perform perfect restoration, we can still obtain a cleaner image than the original ones by restoration. Then, using this less noisy image, we can perform registration more accurately, which further enables better restoration. Thus, we iterate registration and restoration to improve one by the other, as depicted in Figure 3b. Hereby, we enable accurate registration and thus effective restoration even in the presence of severe short-exposure noise. Specifically, we formulate the problem of multi-frame image restoration in terms of Bayesian inference, and overcome mathematical difficulties by a *variational Bayesinan method* [3]. The resultant algorithm reduces to iterative registration and restoration, where we update approximate posterior distributions of latent random variables one by one. In addition, the covariances of the posterior distributions enable us to take account of the uncertainty of registration and restoration, which contributes to the robustness of the variational Bayesian inference.

Through an experiment using synthetic images, we evaluated the effectiveness of the propose method of multi-exposure for motion-blur-free photography, not only qualitatively, but also quantitatively in terms of multiple image quality metrics. In addition, we applied the proposed method to real images to show its performance in the real world.

The rest of this paper is organized as follows: First, we review related works in Chapter 2. Next, we construct a Bayesian model for multi-exposure in Chapter 3, and based on that model, we derive an algorithm of variational Bayesian

(a)

(b)

Figure 3: Flow diagrams of (a) the previous [14] and (b) the proposed method of multi-exposure. While the previous method first performs registration and then restoration, the proposed method does them iteratively to improve one by the other.

inference for multi-frame image restoration in Chapter 4. Details of implementation are discussed in Chapter 5. Then, we demonstrate the effectiveness of the proposed method for motion-blur-free photography through experiments in Chapter 6. Finally, we conclude with discussions on future directions of this work in Chapter 7.

Chapter 2 Related Works on Multi-Exposure and Variational Bayesian Methods

In this chapter, we first review existing works on multi-exposure photography, which directly relate to this work. Then, we introduce several works on image processing by variational Bayesian methods, which shares the same methodology as this work.

2.1 Multi-Exposure for Motion-Blur-Free Photography

Multi-exposure refers to techniques of computational photography, whose motivation is to make use of multiple images to capture a single high-quality images, with the help of digital image processing. Applied to motion-blur-free photography, multi-exposure has been reported to outperform traditional single-exposure approaches [18], e.g., noise removal from a single short-exposure image and blur removal from a single long-exposure one. In addition, it has a notable advantage over *optical image stabilization* [18], which compensates for motion by moving the lens or the sensor of the camera during exposure, in that multi-exposure can deal with not only ego-motion of cameras but also motion of objects.

Tico [14] proposed a multi-exposure method that uses multiple short-exposure images, as with our proposed method. After observing noisy short-exposure images, it first performs local block matching for registration, and then restores each block of a clean image by averaging similar blocks found in the noisy images. To remove noise while preserving image structures like edges, the sizes of blocks are automatically adapted. As exposure gets shorter and noise increases, their method tends to fail due to inaccurate registration, as we demonstrate in Chapter 6. This is because registration alone is not so robust to noise, since it heavily depends on high-frequency structures of natural images, i.e., edges, which are easily corrupted by noise [15]. In this work, we deal with short-exposure noise by variational Bayesian multi-frame image restoration, where we iterate registration and restoration. In this manner, our proposed method performs registration using not only noisy images, but also cleaner ones obtained from restoration, and thus is more effective in the presence of severe

short-exposure noise than the previous method.

Tico *et al.* [17] proposed a different kind of method, which uses a long-exposure image to remove noise from a short-exposure image. Their method combines blur-free regions of the short-exposure one with noise-free regions of the long-exposure one. It also has potential to be extended to another application of multi-exposure, i.e., *high-dynamic-range* imaging [16]. While their method does improve the quality of the short-exposure image to some extent, it is still subjected to the local trade-off between motion blur and noise, since in each region it must choose one of the blurry and the noisy image. In contrast, our proposed method can suppress both motion blur and noise everywhere, since it uses only short-exposure images and removes noise by global restoration.

Apart from noise removal from a short-exposure image, the use of multiple images has also been known to benefit motion blur removal from a long-exposure image. Ben-Ezra *et al.* [2] used multiple short-exposure images to estimate a blur kernel of a long-exposure image, which were then used to remove blur from the long-exposure one by deconvolution. Yuan *et al.* [22] enabled to use only a single short-exposure image. While they performed blur kernel estimation and deconvolution separately, Zhang *et al.* [23] jointly estimated blur kernels, registration parameters, and a blur-free image from multiple long-exposure images. Although multi-exposure does improve the accuracy of blur kernel estimation, blur removal by deconvolution inherently limits tractable ranges of motion blur, since kernels are required to be shift-invariant, i.e., spatially not varying [17]. Our multi-exposure method can deal with a wider range of motion blur, e.g., blur on an object moving independently of others, since it reduces motion blur not by deconvolution but by short exposure.

2.2 Variational Bayesian Methods for Image Processing

Variational Bayesian methods are techniques to approximately deal with intractable marginalization of latent variables in Bayesian inference [3]. They generalize traditional Bayesian techniques, such as *maximum-a-posteriori* (MAP) estimation and *expectation-maximization* (EM) algorithms, and enable Bayesian inference in the presence of multiple variables that distribute differently but

depend on each other. Recently, they have been applied to several image processing problems, including image restoration and registration, and reported to outperform traditional, non-variational-Bayesian methods [6, 7].

Chantas *et al.* [6] proposed one of the state-of-the-art methods of variational-Bayesian single-frame image restoration. They employed a variational Bayesian method to use multiple image priors, each of which effectively removes image degradation such as noise by smoothing an image while preserving image structures. The parameters of these priors, which controls structure adaptation and relative importance of the priors, are estimated jointly with a clean image through variational Bayesian inference. Their method was shown to perform better at preserving detailed structures than traditional methods. However, such a method of single-frame restoration is not directly applicable to multi-exposure, where we need to make full use of multiple images to remove severe short-exposure noise. Meanwhile, Chantas *et al.* [7] proposed a method of variational-Bayesian *optical flow* estimation, which can be used for stand-alone image registration. In this work, we employ the variational Bayesian methodology as in these works on image processing, and introduce it to multi-exposure for motion-blur-free photography. Besides, we make full use of the methodology to enable multi-frame image restoration, which performs restoration and registration jointly rather than separately.

Variational Bayesian methods have also been applied to *superresolution*, which involves registration of images observed by multiple cameras at different viewpoints. Some methods of superresolution employed variational Bayesian methods to jointly perform registration and restoration, e.g., [1, 19]. Our method is different from them, in that we aims at noise removal for motion-blur-free multi-exposure photography rather than resolution enhancement, and that we consider registration for a wide range of motion arising in multi-exposure, rather than for displacements between cameras. Specifically, we parameterize the registration by pixel-wise warping based on an optical flow, rather than by global affine transformation.

Chapter 3 Bayesian Model for Multi-Exposure

In this chapter, we construct a Bayesian model for multi-exposure, aiming at motion-blur-free photography. This model enables us to solve the multi-frame image restoration problem in a Bayesian manner, as described later in Chapter 4. First, we make several basic assumptions, which is needed for both restoration and registration. Then, we define probability distributions of parameters, i.e., a noisy image sequence to be observed, a clean image to be restored, and an optical flow as a registration parameters. After that, we show our complete model graphically.

3.1 Basic Assumptions

Suppose that we first observe a sequence of n_{t} images, each of which has n_{s} pixels, i.e., with $n_{\mathrm{t}}n_{\mathrm{s}}$ pixels in total. We assume that each of these images is exposed in a sufficiently short time, and thus blur-free but noisy. Then, we restore a single clean image with n_{s} pixels from the noisy image sequence. We choose one of the noisy images as the reference of the clean image, i.e., we set the temporal sample point of the clean image to the same as the reference. From now on, we assume that the reference is the first noisy image in the sequence, since in photography we usually want an image that reflects the scene at the beginning of shooting.

In order to register each noisy image with respect to the latent clean image, we assume that each noisy image is basically a warped version of the clean image, i.e., there exists one-to-one correspondences from each grid point in each noisy image to a point in the clean one. Then, we can parameterize registration by an optical flow, e.g., a pixel-wise velocity field, which tells the corresponding point in the clean image given each grid point in the noisy images. This parameterization covers a wide range of motion, including ego-motion of cameras and motion of objects. While the warping assumption may be violated by occlusion due to motion, we deal with it in our noise model, as described shortly.

Let $\boldsymbol{y} \in \mathbb{R}^{n_{\boldsymbol{y}}}$ be a random variable as the noisy image sequence, whose pixels

are flattened into a column vector, $\boldsymbol{x} \in \mathbb{R}^{n_{\boldsymbol{x}}}$ as the clean image, and $\boldsymbol{w} \in \mathbb{R}^{n_{\boldsymbol{w}}}$ as the optical flow, where $n_{\boldsymbol{y}} \equiv n_{\mathrm{t}} n_{\mathrm{s}}$, $n_{\boldsymbol{x}} \equiv n_{\mathrm{s}}$ and $n_{\boldsymbol{w}} \equiv 2 n_{\mathrm{t}} n_{\mathrm{s}}$. While $\boldsymbol{y}$ has one element for each pixel, $\boldsymbol{w}$ has two as the horizontal and vertical components of the velocity at the pixel. The elements of $\boldsymbol{w}$ at the pixels in the first noisy image are fixed to zero to make it the reference. Let $\boldsymbol{W} \in \mathbb{R}^{n_{\boldsymbol{y}} \times n_{\boldsymbol{x}}}$ be the warping matrix with respect to $\boldsymbol{w}$, which transforms the clean image $\boldsymbol{x}$ into a sequence of n_t images. Then, the warping assumption can be expressed as follows:

$$\boldsymbol{y} \simeq \boldsymbol{W}\boldsymbol{x}, \tag{1}$$

where the approximate equality indicates the presence of noise and occlusion.

3.2 Noisy Image Sequence

We assume that the noisy image sequence is mainly degraded by additive zero-mean Gaussian noise, which has been known to well approximates short-exposure noise [17], i.e., amplifier and shot noise. In reality, however, occlusion due to motion produces non-Gaussian differences between each warped clean image and the corresponding noisy image, which can be regarded as impulsive noise. Thus, instead of assuming a simple Gaussian distribution, we assume that the noise in observation follows the t distribution with a precision and a degree of freedom (DoF). When the DoF goes to infinity, the distribution reduces to the Gaussian distribution with the precision [7]. In this sense, the t noise model generalizes the traditional Gaussian noise assumption. On the other hand, when the DoF is not so large, the t distribution is heavy-tailed and thus allows us robust estimation in the presence of outliers [3], i.e., in our case, missing correspondences due to occlusion.

Let $\beta, \xi \in \mathbb{R}$ be the precision and the DoF parameter of the t distribution, respectively. We define a conditional probability distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ and $\boldsymbol{w}$ as follows:

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}) = t(\boldsymbol{y}|\boldsymbol{W}\boldsymbol{x}, \beta \boldsymbol{I}_{\boldsymbol{y}}, \xi), \tag{2}$$

where $\boldsymbol{I}_{\boldsymbol{y}} \in \mathbb{R}^{n_{\boldsymbol{y}} \times n_{\boldsymbol{y}}}$ is an identity matrix, and t is the probability density func-

tion of the t distribution with a mean, a precision, and a DoF such that

$$t(\boldsymbol{y}|\boldsymbol{W}\boldsymbol{x}, \beta\boldsymbol{I}_{\boldsymbol{y}}, \xi) = \prod_{i=1}^{n_{\boldsymbol{y}}} \frac{\Gamma\left(\frac{\xi+1}{2}\right)}{\Gamma\left(\frac{\xi}{2}\right)} \sqrt{\frac{\beta}{\pi\xi}} \left(1 + \frac{\beta}{\xi}\left(\boldsymbol{y}_i - [\boldsymbol{W}\boldsymbol{x}]_i\right)^2\right)^{-\frac{\xi+1}{2}}, \qquad (3)$$

where $[\cdot]_i$ is the ith element of a vector or a matrix, and Γ is the gamma function.

3.3 Clean Image

Next, we define the prior model of $\boldsymbol{x}$ to make it a noise-free version of the reference image. Following standard assumptions in image restoration, we assume a natural image is locally smooth, i.e., it has small local variations, but the smoothness depends on image structures, e.g., edges and textures. We can capture image variations by spatial high-pass filters. Let $\boldsymbol{G} \in \mathbb{R}^{m_{\boldsymbol{G}} n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ be a high-pass filter bank matrix, which itself is composed of $m_{\boldsymbol{G}}$ high-pass filter matrices $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_{m_{\boldsymbol{G}}} \in \mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$. Each ith row of $\boldsymbol{G}_k\boldsymbol{x}$ is a high frequency component, i.e., image variation, captured by $\boldsymbol{G}_k$ at the ith pixel of $\boldsymbol{x}$. Then, the magnitudes of the elements of $\boldsymbol{G}\boldsymbol{x}$ should be small, but in some structural parts of $\boldsymbol{x}$, e.g., around edges, it is allowed to be large. This assumption can be modeled by a t distribution again, where the heavy-tailed property preserves large variations due to structures [6].

Let $\alpha, \nu \in \mathbb{R}$ be the precision and the DoF parameter of the t distribution, respectively. We define the prior probability distribution of $\boldsymbol{x}$ as follows:

$$p(\boldsymbol{x}) \simeq t(\boldsymbol{G}\boldsymbol{x}|\boldsymbol{o}_{\boldsymbol{G}\boldsymbol{x}}, \alpha\boldsymbol{I}_{\boldsymbol{G}\boldsymbol{x}}, \nu), \qquad (4)$$

where $\boldsymbol{I}_{\boldsymbol{G}\boldsymbol{x}} \in \mathbb{R}^{m_{\boldsymbol{G}} n_{\boldsymbol{x}} \times m_{\boldsymbol{G}} n_{\boldsymbol{x}}}$ is an identity matrix, and $\boldsymbol{o}_{\boldsymbol{G}\boldsymbol{x}} \in \mathbb{R}^{m_{\boldsymbol{G}} n_{\boldsymbol{x}}}$ is a vector of zeros.

We note that the right hand side is not exactly a proper distribution of $\boldsymbol{x}$, but of $\boldsymbol{G}\boldsymbol{x}$, hence the approximate equality. Still, such improper priors are known to produce sensible results in Bayesian inference [3]. Some of previous works, e.g., [6], used variable transformation tricks to make priors proper, although they ignored the fact that general high-pass filter matrices such as differentiation operators are rank-deficient. In this work, we leave the prior improper but the resulting algorithm is almost equivalent to that derived by using the tricks for properness.

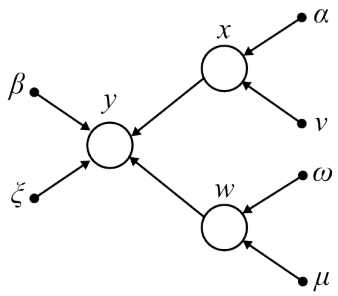


Figure 4: Graphical model for multi-frame image restoration. $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}$ are assumed to be drawn from t distributions.

3.4 Optical Flow

For $\boldsymbol{w}$, which is also assumed to be smooth, depending on structures, we use a prior similar to that of $\boldsymbol{x}$. Let $\boldsymbol{F} \in \mathbb{R}^{m_{\boldsymbol{F}} n_{\boldsymbol{w}} \times n_{\boldsymbol{w}}}$ be another high-pass filter bank matrix, made of $m_{\boldsymbol{F}}$ filter matrices.

Let $\omega, \mu \in \mathbb{R}$ be the be the precision and the DoF parameter of the t distribution, respectively. We define the prior probability distribution of $\boldsymbol{w}$ as follows:

$$p(\boldsymbol{w}) \simeq t(\boldsymbol{F}\boldsymbol{w} | \boldsymbol{o}_{\boldsymbol{F}\boldsymbol{w}}, \omega \boldsymbol{I}_{\boldsymbol{F}\boldsymbol{w}}, \mu). \tag{5}$$

where $\boldsymbol{I}_{\boldsymbol{F}\boldsymbol{w}} \in \mathbb{R}^{m_{\boldsymbol{F}} n_{\boldsymbol{w}} \times m_{\boldsymbol{F}} n_{\boldsymbol{w}}}$ is an identity matrix, and $\boldsymbol{o}_{\boldsymbol{F}\boldsymbol{w}} \in \mathbb{R}^{m_{\boldsymbol{F}} n_{\boldsymbol{w}}}$ is a vector of zeros.

3.5 Complete Model

At this point, we have our whole Bayesian model for multi-frame image restoration as shown by the graphical model in the Figure 4. This model will be further modified to enable variational Bayesian inference in Chapter 4.

Chapter 4 Variational Bayesian Inference for Multi-Frame Image Restoration

In this chapter, we derive an algorithm of variational-Bayesian multi-frame image restoration, based on the model defined in Chapter 3. First, we formulate our problem in terms of Bayesian inference, and introduce a variational Bayesian method to deal with mathematical difficulties. Next, to enable the variational Bayesian inference, we perform further approximation, i.e., we decompose the t distributions in our model, and then linearize the warping with respect to the optical flow. After that, we obtain update formulas of parameters, which constitute the variational Bayesian inference. Finally, we show our complete algorithm, where we employ a coarse-to-fine, iterative update scheme.

4.1 Bayesian Inference with Variational Approximation

We seek for the most probable clean image after observing the noisy image sequence, denoted by $\hat{\boldsymbol{x}}$, by maximizing the posterior probability of $\boldsymbol{x}$ given $\boldsymbol{y}$:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}). \tag{6}$$

To obtain the posterior distribution of $\boldsymbol{x}$, we need to marginalize out the other latent variables, i.e., variables other than the observed $\boldsymbol{y}$, from the joint posterior distribution of the latent variables:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x}, \boldsymbol{w}|\boldsymbol{y}) dw. \tag{7}$$

By the *Bayes' theorem*, the joint posterior distribution can be obtained from our model

$$p(\boldsymbol{x}, \boldsymbol{w}|\boldsymbol{y}) \propto p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}), \tag{8}$$

where $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w})$ is the joint distribution of all the variables, including $\boldsymbol{y}$:

$$p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}) p(\boldsymbol{x}) p(\boldsymbol{w}). \tag{9}$$

However, the exact marginalization with respect to $\boldsymbol{w}$ is analytically intractable.

To deal with the intractable marginalization, we employ a variational Bayesian

method [3], which approximates the joint posterior distribution by factorization. In the following, we denote exact and approximate posterior distributions by p and q, respectively. Then, the exact joint posterior distribution is factorized into approximate posterior distributions of the individual latent variables:

$$p(\boldsymbol{x}, \boldsymbol{w}|\boldsymbol{y}) \simeq q(\boldsymbol{x}, \boldsymbol{w}) \equiv q(\boldsymbol{x})q(\boldsymbol{w}). \tag{10}$$

After this approximation, we can easily marginalize out each latent variable, since its approximate posterior probability is integrated into 1 independently of other variables, and we obtain the approximate posterior distribution of $\boldsymbol{x}$:

$$p(\boldsymbol{x}|\boldsymbol{y}) \simeq \int q(\boldsymbol{x})q(\boldsymbol{w})dw = q(\boldsymbol{x}) \tag{11}$$

We seek for the optimal approximation that makes $q(\boldsymbol{x}, \boldsymbol{w})$ closest to $p(\boldsymbol{x}, \boldsymbol{w}|\boldsymbol{y})$, in terms of minimization of the *Kullback-Leibler* (KL) *divergence* of $p(\boldsymbol{x}, \boldsymbol{w}|\boldsymbol{y})$ from $q(\boldsymbol{x}, \boldsymbol{w})$. Then, for each latent variable, the logarithm of its approximate posterior distribution equals to the logarithmic expectation of the joint distribution of all the variables, which we have in Equation (9), with respect to the other latent variables, up to a constant [3]:

$$\ln q(\boldsymbol{x}) = \mathrm{E}_{q(\boldsymbol{w})} \left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w})\right] + \text{const.}, \tag{12}$$

$$\ln q(\boldsymbol{w}) = \mathrm{E}_{q(\boldsymbol{x})} \left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w})\right] + \text{const.}, \tag{13}$$

where $\mathrm{E}_{q(\cdot)}$ is the expectation with respect to the approximate distribution $q(\cdot)$. Since the approximate posterior distributions , i.e., $q(\boldsymbol{x})$ and $q(\boldsymbol{w})$ depend on each other, we iteratively update the approximate posterior distributions one by one, i.e., given some initial estimate of $q(\boldsymbol{w})$, first we estimate $q(\boldsymbol{x})$ fixing $q(\boldsymbol{w})$, then $q(\boldsymbol{w})$ fixing $q(\boldsymbol{x})$, iterating until convergence. Since the optimization of $q(\boldsymbol{x})$ and $q(\boldsymbol{w})$ corresponds to restoration and registration, respectively, we can interpret this inference as multi-frame image restoration by iterative restoration and registration. Theoretically, it is guaranteed that this iterative procedure converges [3].

4.2 Decomposition of t Distributions

We cannot apply the variational-Bayes approximation directly to our model, since the nonlinearity due to the t distributions prevents us from taking analyt-

ical expectation with respect to $q(\boldsymbol{x})$ and $q(\boldsymbol{w})$. Following the previous works on variational Bayesian image processing [6, 7], we decompose the t distributions into Gaussian and gamma distributions, introducing auxiliary variables.

First, we decompose $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w})$. Let $\boldsymbol{b} \in \mathbb{R}^{n_{\boldsymbol{b}}}$ be the auxiliary random variables for $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w})$, where $n_{\boldsymbol{b}} \equiv n_{\boldsymbol{y}}$. We replace each t distribution by a product of a Gaussian and a gamma:

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}) \to p(\boldsymbol{y}, \boldsymbol{b}|\boldsymbol{x}, \boldsymbol{w}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b})p(\boldsymbol{b}), \tag{14}$$

where

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}) = \mathcal{N}\left(\boldsymbol{y}|\boldsymbol{W}\boldsymbol{x}, \boldsymbol{B}\right), \tag{15}$$

$$p(\boldsymbol{b}) = \mathcal{G}\left(\boldsymbol{b}\middle|\frac{\xi}{2}, \frac{\xi}{2}\boldsymbol{I}_{\boldsymbol{y}}\right), \tag{16}$$

$\boldsymbol{B} \equiv \beta[\boldsymbol{b}]$, $[\cdot]$ is the diagonalization of a vector into a matrix; $\mathcal{N}$ is the probability density functions of the Gaussian distribution with a mean and a precision, and $\mathcal{G}$ is that of the gamma distribution with a shape and a rate, such that

$$\begin{aligned}
\mathcal{N}\left(\boldsymbol{y}|\boldsymbol{W}\boldsymbol{x}, \boldsymbol{B}\right) &= (2\pi)^{-\frac{n_{\boldsymbol{y}}}{2}} |\boldsymbol{B}|^{\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x})^\top \boldsymbol{B}(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x})} \\
&= \prod_{i=1}^{n_{\boldsymbol{y}}} \sqrt{\frac{\beta \boldsymbol{b}_i}{2\pi}} e^{-\frac{\beta \boldsymbol{b}_i}{2}(\boldsymbol{y}_i - [\boldsymbol{W}\boldsymbol{x}]_i)^2}, \qquad (17) \\
\mathcal{G}\left(\boldsymbol{b}\middle|\frac{\xi}{2}, \frac{\xi}{2}\boldsymbol{I}_{\boldsymbol{y}}\right) &= \frac{\left|\frac{\xi}{2}\boldsymbol{I}_{\boldsymbol{y}}\right|^{\frac{\xi}{2}}}{\left(\Gamma\left(\frac{\xi}{2}\right)\right)^{n_{\boldsymbol{b}}}} |[\boldsymbol{b}]|^{\frac{\xi}{2}-1} e^{-\operatorname{tr}\left(\frac{\xi}{2}\boldsymbol{I}_{\boldsymbol{y}}[\boldsymbol{b}]\right)} \\
&= \prod_{i=1}^{n_{\boldsymbol{b}}} \frac{\left(\frac{\xi}{2}\right)^{\frac{\xi}{2}}}{\Gamma\left(\frac{\xi}{2}\right)} \boldsymbol{b}_i{}^{\frac{\xi}{2}-1} e^{-\frac{\xi}{2}\boldsymbol{b}_i}. \qquad (18)
\end{aligned}$$

We can recover the original t distribution $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w})$ by marginalizing out $\boldsymbol{b}$ from $p(\boldsymbol{y}, \boldsymbol{b}|\boldsymbol{x}, \boldsymbol{w})$ [3]. While the distribution of $\boldsymbol{y}$ has become Gaussian, by making each $\boldsymbol{b}_i$ small when $\boldsymbol{y}_i$ is an outlier, e.g., due to occlusion, we can reduce the precision of $p(\boldsymbol{y}_i|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b})$, effectively ignoring the ith pixel of the noisy image sequence. In this sense, the Gaussian distribution is highly *adaptive* owing to the auxiliary variable, and thus preserves the robustness property of the original t distribution. This adaptation is automatically done by treating $\boldsymbol{b}$ as an additional latent variable in the variational Bayesian inference.

In the same manner, we decompose $p(\boldsymbol{x})$ and $p(\boldsymbol{w})$. Let $a \in \mathbb{R}^{n_{\boldsymbol{a}}}$ and $\boldsymbol{z} \in \mathbb{R}^{n_{\boldsymbol{z}}}$ be the auxiliary variable for $p(\boldsymbol{x})$ and $p(\boldsymbol{w})$, respectively, where $n_{\boldsymbol{a}} \equiv m_{\boldsymbol{G}} n_{\boldsymbol{x}}$, and $n_{\boldsymbol{z}} \equiv m_{\boldsymbol{F}} n_{\boldsymbol{w}}$. We decompose $p(\boldsymbol{x})$:

$$p(\boldsymbol{x}) \to p(\boldsymbol{x}, \boldsymbol{a}) = p(\boldsymbol{x}|\boldsymbol{a})p(\boldsymbol{a}), \tag{19}$$

where

$$p(\boldsymbol{x}|\boldsymbol{a}) = \mathcal{N}\left(\boldsymbol{G}\boldsymbol{x}|\boldsymbol{o}_{\boldsymbol{G}\boldsymbol{x}}, \boldsymbol{A}\right), \tag{20}$$

$$p(\boldsymbol{a}) = \mathcal{G}\left(\boldsymbol{a}\middle|\frac{\nu}{2}, \frac{\nu}{2}\boldsymbol{I}_{\boldsymbol{G}\boldsymbol{x}}\right), \tag{21}$$

and $\boldsymbol{A} \equiv \alpha[\boldsymbol{a}]$. We also decompose $p(\boldsymbol{w})$:

$$p(\boldsymbol{w}) \to p(\boldsymbol{w}, \boldsymbol{z}) = p(\boldsymbol{w}|\boldsymbol{z})p(\boldsymbol{z}), \tag{22}$$

where

$$p(\boldsymbol{w}|\boldsymbol{z}) = \mathcal{N}\left(\boldsymbol{F}\boldsymbol{w}|\boldsymbol{o}_{\boldsymbol{F}\boldsymbol{w}}, \boldsymbol{Z}\right), \tag{23}$$

$$p(\boldsymbol{z}) = \mathcal{G}\left(\boldsymbol{z}\middle|\frac{\mu}{2}, \frac{\mu}{2}\boldsymbol{I}_{\boldsymbol{F}\boldsymbol{w}}\right), \tag{24}$$

and $\boldsymbol{Z} \equiv \omega[\boldsymbol{z}]$.

Now, we have the final version of our model, modified with $\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}$, as shown in Figure 5. From this model, we obtain the joint posterior distribution of all the latent variables, including the auxiliary ones $\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}$:

$$p(\boldsymbol{w}, \boldsymbol{x}|\boldsymbol{y}) \to p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}|\boldsymbol{y}) \propto p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}), \tag{25}$$

where

$$p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b})p(\boldsymbol{x}|\boldsymbol{a})(\boldsymbol{w}|\boldsymbol{z})p(\boldsymbol{b})p(\boldsymbol{a})p(\boldsymbol{z}), \tag{26}$$

and the joint approximate posterior distribution of all the latent variables:

$$p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}|\boldsymbol{y}) \simeq q(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) \equiv q(\boldsymbol{x})q(\boldsymbol{w})q(\boldsymbol{b})q(\boldsymbol{a})q(\boldsymbol{z}), \tag{27}$$

where $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b})$, $p(\boldsymbol{x}|\boldsymbol{a})$, $p(\boldsymbol{w}|\boldsymbol{z})$, $p(\boldsymbol{b})$, $p(\boldsymbol{a})$, and $p(\boldsymbol{z})$ are given by Equation (15), (20), (23), (16), (21), (24), respectively.

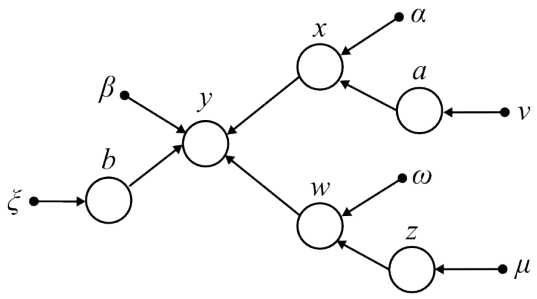


Figure 5: Modified graphical model for multi-frame image restoration. The t distributions of $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}$ are decomposed into a Gaussian and a gamma distribution by introducing auxiliary variables $\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}$, respectively.

4.3 Linearization of Warping

The warping matrix $\boldsymbol{W}$ in our model needs to be constructed in a nonlinear manner given $\boldsymbol{w}$. This nonlinearity still prevents us from taking expectation with respect to $q(\boldsymbol{w})$, as well as from obtaining a explicit solution for $q(\boldsymbol{w})$. Meanwhile, at each update, we have the current estimate of $q(\boldsymbol{w})$, from which we can obtain the current most probable flow, i.e., $\hat{\boldsymbol{w}} = \arg\max q(\boldsymbol{w})$. We make use of this information to linearize $\boldsymbol{W}$.

Let $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{R}^{n_{\boldsymbol{w}}/2}$ be the horizontal and vertical component of $\boldsymbol{w}$, respectively, $\hat{\boldsymbol{w}}_1, \hat{\boldsymbol{w}}_2 \in \mathbb{R}^{n_{\boldsymbol{w}}/2}$ be those of $\hat{\boldsymbol{w}}$, and $\hat{\boldsymbol{W}} \in \mathbb{R}^{n_{\boldsymbol{y}} \times n_{\boldsymbol{x}}}$ be the warping matrix with respect to $\hat{\boldsymbol{w}}$. At each iteration, we approximate $\boldsymbol{W}\boldsymbol{x}$ with respect to $\boldsymbol{w}$ by first-order *Taylor expansion* at $\hat{\boldsymbol{w}}$:

$$\boldsymbol{W}\boldsymbol{x} \simeq \hat{\boldsymbol{W}}\boldsymbol{x} + (\boldsymbol{w}_1 - \hat{\boldsymbol{w}}_1) \circ \left(\hat{\boldsymbol{W}}\boldsymbol{D}_1\boldsymbol{x}\right) + (\boldsymbol{w}_2 - \hat{\boldsymbol{w}}_2) \circ \left(\hat{\boldsymbol{W}}\boldsymbol{D}_2\boldsymbol{x}\right) = \boldsymbol{I}\left(\boldsymbol{w}' \circ \boldsymbol{x}'\right), \tag{28}$$

where $\circ$ is Hadamard (element-wise) multiplication; $\boldsymbol{I} \equiv [\boldsymbol{I}_0\ \boldsymbol{I}_1\ \boldsymbol{I}_2]$, and $\boldsymbol{I}_0, \boldsymbol{I}_1, \boldsymbol{I}_2 \in \mathbb{R}^{n_{\boldsymbol{y}} \times n_{\boldsymbol{y}}}$ are identity matrices; $\boldsymbol{D}_1, \boldsymbol{D}_2 \in \mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ are horizontal and vertical differentiation matrices, respectively; $\boldsymbol{x}' \equiv \boldsymbol{D}\boldsymbol{x}$, and $\boldsymbol{D} \equiv [(\hat{\boldsymbol{W}})^\top\ (\hat{\boldsymbol{W}}\boldsymbol{D}_1)^\top\ (\hat{\boldsymbol{W}}\boldsymbol{D}_2)^\top]^\top$; $\boldsymbol{w}' = \boldsymbol{J}(\boldsymbol{w} - \hat{\boldsymbol{w}}) + \boldsymbol{j}$, $\boldsymbol{J}$ and $\boldsymbol{j}$ are the matrix and the vector, respectively, such that $\boldsymbol{J}(\boldsymbol{w} - \hat{\boldsymbol{w}}) + \boldsymbol{j} = [(\boldsymbol{i})^\top\ (\boldsymbol{w}_1 - \hat{\boldsymbol{w}}_1)^\top\ (\boldsymbol{w}_2 - \hat{\boldsymbol{w}}_2)^\top]^\top$, and $\boldsymbol{i} \in \mathbb{R}^{n_{\boldsymbol{w}}/2}$ is a vector of ones. After this approximation, $\boldsymbol{W}\boldsymbol{x}$ becomes linear with respect to $\boldsymbol{w}$:

$$\boldsymbol{W}\boldsymbol{x} \simeq \boldsymbol{I}[\boldsymbol{x}']\left(\boldsymbol{J}\left(\boldsymbol{w} - \hat{\boldsymbol{w}}\right) + \boldsymbol{j}\right), \tag{29}$$

as well as to $\boldsymbol{x}$:

$$\boldsymbol{W}\boldsymbol{x} \simeq \boldsymbol{I}[\boldsymbol{w}']\boldsymbol{D}\boldsymbol{x}. \tag{30}$$

The linearization at $\boldsymbol{w} = \hat{\boldsymbol{w}}$ is more accurate than naïve linearization at $\boldsymbol{w} = \boldsymbol{0}$, as known in traditional, non-variational-Bayesian optical flow estimation [4].

4.4 Update of Parameters

Finally, we can obtain the approximate posterior distribution for each latent variable by taking an expectation of the joint distribution of all the variables, as described in Appendix. Consequently, we obtain $q(\boldsymbol{x}), q(\boldsymbol{w})$ as Gaussian distributions, and $q(\boldsymbol{b}), q(\boldsymbol{a}), q(\boldsymbol{z})$ as gamma distributions. Since these approximate posterior distributions are well parameterized, update of these posterior distributions reduces to update of their parameters.

Let $\boldsymbol{\mu}_{\boldsymbol{x}} \in \mathbb{R}^{n_{\boldsymbol{x}}}, \boldsymbol{\Sigma}_{\boldsymbol{x}} \in \mathbb{R}^{n_{\boldsymbol{x}} \times n_{\boldsymbol{x}}}$ be the mean and the covariance of the Gaussian $q(\boldsymbol{x})$, respectively, $\boldsymbol{\mu}_{\boldsymbol{w}} \in \mathbb{R}^{n_{\boldsymbol{w}}}, \boldsymbol{\Sigma}_{\boldsymbol{w}} \in \mathbb{R}^{n_{\boldsymbol{w}} \times n_{\boldsymbol{w}}}$ be those of $q(\boldsymbol{w})$, $k_{\boldsymbol{b}} \in \mathbb{R}, \boldsymbol{\theta}_{\boldsymbol{b}} \in \mathbb{R}^{n_{\boldsymbol{b}}}$ be the shape and the scale of the Gamma $q(\boldsymbol{b})$, $k_{\boldsymbol{a}} \in \mathbb{R}, \boldsymbol{\theta}_{\boldsymbol{a}} \in \mathbb{R}^{n_{\boldsymbol{a}}}$ be those of $q(\boldsymbol{a})$, and $k_{\boldsymbol{z}} \in \mathbb{R}, \boldsymbol{\theta}_{\boldsymbol{z}} \in \mathbb{R}^{n_{\boldsymbol{z}}}$ be those of $q(\boldsymbol{z})$. The explicit solution of $q(\boldsymbol{x})$ is as follows:

$$q(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{\boldsymbol{x}}, {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1}), \tag{31}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{D}^{\top}[\overline{\boldsymbol{w}'}]\boldsymbol{I}^{\top}\overline{\boldsymbol{B}}\boldsymbol{y}, \tag{32}$$

$${\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1} = \boldsymbol{D}^{\top}\left(\boldsymbol{I}^{\top}\overline{\boldsymbol{B}}\boldsymbol{I} \circ \overline{\boldsymbol{w}'\boldsymbol{w}'^{\top}}\right)\boldsymbol{D} + \boldsymbol{G}^{\top}\overline{\boldsymbol{A}}\boldsymbol{G}, \tag{33}$$

$\bar{\cdot}$ is expectation of a random variable, $\overline{\boldsymbol{B}} = \beta[\overline{\boldsymbol{b}}] = \beta k_{\boldsymbol{b}}[\boldsymbol{\theta}_{\boldsymbol{b}}]$, $\overline{\boldsymbol{A}} = \alpha[\overline{\boldsymbol{a}}] = \alpha k_{\boldsymbol{a}}[\boldsymbol{\theta}_{\boldsymbol{a}}]$, and

$$\overline{\boldsymbol{w}'\boldsymbol{w}'^{\top}} = \boldsymbol{\mu}_{\boldsymbol{w}}{}'{\boldsymbol{\mu}_{\boldsymbol{w}}}'^{\top} + \boldsymbol{\Sigma}'_{\boldsymbol{w}}, \tag{34}$$

$$\boldsymbol{\mu}_{\boldsymbol{w}}{}' \equiv \boldsymbol{J}(\boldsymbol{\mu}_{\boldsymbol{w}} - \hat{\boldsymbol{w}}) + \boldsymbol{j}, \tag{35}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}}{}' \equiv \boldsymbol{J}\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{J}^{\top}. \tag{36}$$

The explicit solution of $q(\boldsymbol{w})$ is as follows:

$$q(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_{\boldsymbol{w}}, {\boldsymbol{\Sigma}_{\boldsymbol{w}}}^{-1}), \tag{37}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{w}} = \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{J}^{\top} \left([\overline{\boldsymbol{x}'}] \boldsymbol{I}^{\top} \overline{\boldsymbol{B}} \boldsymbol{y} + \left(\boldsymbol{I}^{\top} \overline{\boldsymbol{B}} \boldsymbol{I} \circ \overline{\boldsymbol{x}' \boldsymbol{x}'^{\top}} \right) (\boldsymbol{J} \hat{\boldsymbol{w}} - \boldsymbol{j}) \right), \tag{38}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}}{}^{-1} = \boldsymbol{J}^{\top} \left(\boldsymbol{I}^{\top} \overline{\boldsymbol{B}} \boldsymbol{I} \circ \overline{\boldsymbol{x}' \boldsymbol{x}'^{\top}} \right) \boldsymbol{J} + \boldsymbol{F}^{\top} \overline{\boldsymbol{Z}} \boldsymbol{F}, \tag{39}$$

where $\overline{\boldsymbol{Z}} = \omega[\overline{\boldsymbol{z}}] = \omega k_{\boldsymbol{z}}[\boldsymbol{\theta}_{\boldsymbol{z}}]$, and

$$\overline{\boldsymbol{x}' \boldsymbol{x}'^{\top}} = \boldsymbol{\mu}_{\boldsymbol{x}}{}' \boldsymbol{\mu}_{\boldsymbol{x}}{}'^{\top} + \boldsymbol{\Sigma}'_{\boldsymbol{x}}, \tag{40}$$

$$\boldsymbol{\mu}_{\boldsymbol{x}}{}' \equiv \boldsymbol{D} \boldsymbol{\mu}_{\boldsymbol{x}}, \tag{41}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{x}}{}' \equiv \boldsymbol{D} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{D}^{\top}. \tag{42}$$

We note that $q(\boldsymbol{x})$ is dependent on the parameters of $q(\boldsymbol{w})$, and vise versa. Thus, we have to iteratively update them, which reduces to iterative restoration and registration. We also note that $q(\boldsymbol{x})$ is dependent not only on the mean but also on the covariance of $q(\boldsymbol{w})$. This is due to the variational Bayesian approximation, which does not give only point estimates of parameters but gives full posterior distributions approximately. Consequently, when we estimate $q(\boldsymbol{x})$, we do not consider only the most probable $\boldsymbol{w}$, i.e., for the Gaussian $q(\boldsymbol{x})$, the mean $\boldsymbol{\mu}_{\boldsymbol{w}}$, but also how $\boldsymbol{w}$ distributes around $\boldsymbol{\mu}_{\boldsymbol{w}}$, i.e., the covariance $\boldsymbol{\Sigma}_{\boldsymbol{w}}$. Hereby, we can interpret $\boldsymbol{\Sigma}_{\boldsymbol{w}}, \boldsymbol{\Sigma}_{\boldsymbol{x}}$ as the uncertainty of registration and restoration, respectively, each of which is considered when updating the other. This uncertainty enables us to avoid overfitting to data in the inference [3], i.e., in our case, it makes the inference robust to noise.

The explicit solution of $q(\boldsymbol{b})$ is as follows:

$$q(\boldsymbol{b}) = \mathcal{G}(\boldsymbol{b} | k_{\boldsymbol{b}}, [\boldsymbol{\theta}_{\boldsymbol{b}}{}^{(-1)}]), \tag{43}$$

where $\cdot^{(-1)}$ is the Hadamard (element-wise) inverse, and

$$k_{\boldsymbol{b}} = \frac{\xi + 1}{2}, \tag{44}$$

$$\left[\boldsymbol{\theta}_{\boldsymbol{b}}{}^{(-1)} \right]_i = \frac{\xi + \beta \left[\boldsymbol{y}' \boldsymbol{y}'^{\top} \right]_{i,i}}{2}, \tag{45}$$

for $i \in \{1, \ldots, n_{\boldsymbol{b}}\}$, where $\boldsymbol{y}' \equiv \boldsymbol{y} - \boldsymbol{I}(\boldsymbol{w}' \circ \boldsymbol{x}')$, and

$$\begin{aligned}\overline{\boldsymbol{y}'\boldsymbol{y}'^{\top}} = & \quad (\boldsymbol{I}\,(\boldsymbol{\mu}_{\boldsymbol{x}}{}' \circ \boldsymbol{\mu}_{\boldsymbol{w}}{}') - \boldsymbol{y})\,(\boldsymbol{I}\,(\boldsymbol{\mu}_{\boldsymbol{x}}{}' \circ \boldsymbol{\mu}_{\boldsymbol{w}}{}') - \boldsymbol{y})^{\top} \\ & + \quad \boldsymbol{I}[\boldsymbol{\mu}_{\boldsymbol{x}}{}']\boldsymbol{\Sigma}_{\boldsymbol{w}}{}'[\boldsymbol{\mu}_{\boldsymbol{x}}{}']\boldsymbol{I}^{\top} \\ & + \quad \boldsymbol{I}[\boldsymbol{\mu}_{\boldsymbol{w}}{}']\boldsymbol{\Sigma}_{\boldsymbol{x}}{}'[\boldsymbol{\mu}_{\boldsymbol{w}}{}']\boldsymbol{I}^{\top} \\ & + \quad \boldsymbol{I}\,(\boldsymbol{\Sigma}_{\boldsymbol{x}}{}' \circ \boldsymbol{\Sigma}_{\boldsymbol{y}}{}')\,\boldsymbol{I}^{\top}. \end{aligned} \tag{46}$$

The explicit solution of $q(\boldsymbol{a})$ is as follows:

$$q(\boldsymbol{a}) = \mathcal{G}(\boldsymbol{a}|k_{\boldsymbol{a}}, [\boldsymbol{\theta}_{\boldsymbol{a}}{}^{(-1)}]), \tag{47}$$

where

$$k_{\boldsymbol{a}} = \frac{\nu + 1}{2}, \tag{48}$$

$$\left[\boldsymbol{\theta}_{\boldsymbol{a}}{}^{(-1)}\right]_i = \frac{\nu + \alpha\left[\boldsymbol{G}\overline{\boldsymbol{x}\boldsymbol{x}^{\top}}\boldsymbol{G}^{\top}\right]_{i,i}}{2}, \tag{49}$$

for $i \in \{1, \ldots, n_{\boldsymbol{a}}\}$, where

$$\overline{\boldsymbol{x}\boldsymbol{x}^{\top}} = \boldsymbol{\mu}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{x}}{}^{\top} + \boldsymbol{\Sigma}_{\boldsymbol{x}}. \tag{50}$$

The explicit solution of $q(\boldsymbol{z})$ is as follows:

$$q(\boldsymbol{z}) = \mathcal{G}(\boldsymbol{z}|k_{\boldsymbol{z}}, [\boldsymbol{\theta}_{\boldsymbol{z}}{}^{(-1)}]), \tag{51}$$

where

$$k_{\boldsymbol{z}} = \frac{\mu + 1}{2}, \tag{52}$$

$$\left[\boldsymbol{\theta}_{\boldsymbol{z}}{}^{(-1)}\right]_i = \frac{\mu + \omega\left[\boldsymbol{F}\overline{\boldsymbol{w}\boldsymbol{w}^{\top}}\boldsymbol{F}^{\top}\right]_{i,i}}{2}, \tag{53}$$

for $i \in \{1, \ldots, n_{\boldsymbol{z}}\}$, where

$$\overline{\boldsymbol{w}\boldsymbol{w}^{\top}} = \boldsymbol{\mu}_{\boldsymbol{w}}\boldsymbol{\mu}_{\boldsymbol{w}}{}^{\top} + \boldsymbol{\Sigma}_{\boldsymbol{w}}. \tag{54}$$

We can also obtain the parameters of the original t distributions, by taking the logarithmic expectation of $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$ with respect to $q(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$, taking the derivative of it with respect to each of the parameters, and then

setting it to zero [6, 7]. The precisions β, α, ω are obtained as follows:

$$\beta = \frac{n_{\boldsymbol{b}}}{\sum_{i=1}^{n_{\boldsymbol{b}}} [\overline{\boldsymbol{b}}]_i \left[\overline{\boldsymbol{y}'\boldsymbol{y}'^{\top}}\right]_{i,i}}, \tag{55}$$

$$\alpha = \frac{n_{\boldsymbol{a}}}{\sum_{i=1}^{n_{\boldsymbol{a}}} [\overline{\boldsymbol{a}}]_i \left[\boldsymbol{G}\overline{\boldsymbol{x}\boldsymbol{x}^{\top}}\boldsymbol{G}^{\top}\right]_{i,i}}, \tag{56}$$

$$\omega = \frac{n_{\boldsymbol{z}}}{\sum_{i=1}^{n_{\boldsymbol{z}}} [\overline{\boldsymbol{z}}]_i \left[\boldsymbol{F}\overline{\boldsymbol{w}\boldsymbol{w}^{\top}}\boldsymbol{F}^{\top}\right]_{i,i}}. \tag{57}$$

The conditions that the DoFs ξ, ν, μ should satisfy are as follows:

$$\psi\left(\frac{\xi}{2}\right) - \psi\left(k_{\boldsymbol{b}}\right) + \frac{1}{n_{\boldsymbol{b}}} \sum_{i=1}^{n_{\boldsymbol{b}}} \chi\left(k_{\boldsymbol{b}} [\boldsymbol{\theta}_{\boldsymbol{b}}]_i\right) + 1 = 0, \tag{58}$$

$$\psi\left(\frac{\nu}{2}\right) - \psi\left(k_{\boldsymbol{a}}\right) + \frac{1}{n_{\boldsymbol{a}}} \sum_{i=1}^{n_{\boldsymbol{a}}} \chi\left(k_{\boldsymbol{a}} [\boldsymbol{\theta}_{\boldsymbol{a}}]_i\right) + 1 = 0, \tag{59}$$

$$\psi\left(\frac{\mu}{2}\right) - \psi\left(k_{\boldsymbol{z}}\right) + \frac{1}{n_{\boldsymbol{z}}} \sum_{i=1}^{n_{\boldsymbol{z}}} \chi\left(k_{\boldsymbol{z}} [\boldsymbol{\theta}_{\boldsymbol{z}}]_i\right) + 1 = 0. \tag{60}$$

where

$$\psi(u) \equiv \ln(u) - \digamma(u), \tag{61}$$

$$\chi(u) \equiv \ln(u) - u, \tag{62}$$

for $u \in \mathrm{R}$, and $\digamma$ is the digamma function. As these parameters of the t's are mutually dependent, we update them iteratively, along with the parameters of the approximate posterior distributions.

4.5 Coarse-to-Fine Iterative Algorithm

Before we start iterative update, we have to initialize several parameters. For the clean image $\boldsymbol{x}$, we have the reference noisy image as a reasonable initial estimate. On the other hand, it is well known that optical flow estimation is strongly subjected to initial estimates. Thus, we employ the coarse-to-fine scheme to initialize the optical flow $\boldsymbol{w}$, which is crucial to deal with large motion [4]. That is, we first estimate $q(\boldsymbol{w})$ on coarser grids, where we use downsampled versions of y as input, ignoring the parameters related to $\boldsymbol{x}$. Then, on the finest grids, initializing $q(\boldsymbol{w})$ with a upsampled version of the coarse estimate, we estimate all the parameters iteratively. We can expect that the coarse-grid estimation is relatively less affected by noise, since noise is reduced in

coarse versions of observed images owing to blurring and downsampling. The estimation of the parameters of the t distributions are less sensitive to initial estimates, and thus we simply initialize them with a constant, i.e., we set their elements to 1.

We have three groups of parameters according to the original three t distributions:

$$G_{\boldsymbol{y}} = \{q(\boldsymbol{b}), \beta, \xi\}, \tag{63}$$

$$G_{\boldsymbol{x}} = \{q(\boldsymbol{x}), q(\boldsymbol{a}), \alpha, \nu\}, \tag{64}$$

$$G_{\boldsymbol{w}} = \{q(\boldsymbol{w}), q(\boldsymbol{z}), \omega, \mu\}. \tag{65}$$

The parameters in each of the groups are strongly correlated with each other. While the update of $G_{\boldsymbol{x}}$ and $G_{\boldsymbol{w}}$ correspond to restoration and registration, respectively, that of $G_{\boldsymbol{y}}$ corresponds to measurement of noise and occlusion occurred in observation, which is required by both the restoration and the registration. After initialization, we first perform registration by updating the parameters in $G_{\boldsymbol{y}}$ and then $G_{\boldsymbol{w}}$ iteratively. In each iteration, the sub-groups $\{q(\boldsymbol{b}), \beta, \xi\}$ and $\{q(\boldsymbol{z}), \alpha, \mu\}$ are updated further iteratively, since the parameters in each of them have yet strong correlation. Next, we update those in $G_{\boldsymbol{y}}$ and then $G_{\boldsymbol{x}}$ to perform restoration, iteratively again, with internal iterations for sub-groups $\{q(\boldsymbol{b}), \beta, \xi\}$ and $\{q(\boldsymbol{a}), \alpha, \nu\}$. Then, we iterate the registration and the restoration until both of them converge.

The resultant algorithm is shown in Figure 6.

1: Input $\boldsymbol{y}$.
2: Initialize $q(\boldsymbol{w})$ with an upsampled coarse estimate.
3: Initialize $\boldsymbol{\mu}_x$ of $q(\boldsymbol{x})$ with the reference image part of $\boldsymbol{y}$.
4: Initialize $\beta, \alpha, \omega, \xi, \nu, \mu$ with 1.
5: **repeat**
6: **repeat** ▷ Registration
7: Update $q(\boldsymbol{b}), \beta, \xi$ by Equation (43), (55), and (59) iteratively.
8: Update $q(\boldsymbol{z}), \omega, \mu$ by Equation (51), (57), and (60) iteratively.
9: Update $q(\boldsymbol{w})$ by Equation (37).
10: **until** convergence.
11: **repeat** ▷ Restoration
12: Update $q(\boldsymbol{b}), \beta, \xi$ by Equation (43), (55), and (59) iteratively.
13: Update $q(\boldsymbol{a}), \alpha, \nu$ by Equation (47), (56), and (60) iteratively.
14: Update $q(\boldsymbol{x})$ by Equation (31).
15: **until** convergence.
16: **until** convergence.
17: Output $\hat{\boldsymbol{x}} = \arg\max q(\boldsymbol{x})$.

Figure 6: Algorithm of variational-Bayesian multi-frame image restoration.

Chapter 5 Implementation of Variational Bayesian Multi-Frame Image Restoration

In this chapter, we discuss several points to be considered in implementing the algorithm of variational Bayesian multi-frame image restoration derived in Chapter 4.

5.1 Choice of Operators

In construction of all operator matrices used in the proposed method, the Neumann boundary condition with zero derivatives is assumed, i.e., values at boundaries were repeated outside, in order to avoid oscillation of solutions near the boundaries.

We construct the warping matrix $\boldsymbol{W}$ from $\boldsymbol{w}$ by means of bilinear interpolation. Differentiation matrices $\boldsymbol{D}_1, \boldsymbol{D}_2$, resulting from the linearization of warping, is discretized by central finite difference.

For the high-pass filter bank matrix $\boldsymbol{G}$ in the prior of the clean image, we use horizontal and vertical differentiation operators, which are discretized by forward finite difference. For the filter bank matrix $\boldsymbol{F}$ in the prior of the optical flow, we use a single Laplacian filter, which is discretized using 4 neighbors. Such choices of filters for image and flow priors were known to be successful in the previous works of variational Bayesian image restoration [6] and registration [7].

5.2 Approximation of Large Sparse Covariances

To update $\boldsymbol{\mu}_{\boldsymbol{x}}$ of $q(\boldsymbol{x})$ and $\boldsymbol{\mu}_{\boldsymbol{w}}$ of $q(\boldsymbol{w})$ using Equation (32) and (38), respectively, we need to solve linear systems, which involves inversion of precision matrices ${\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1}, {\boldsymbol{\Sigma}_{\boldsymbol{w}}}^{-1}$ into covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{w}}$, respectively. However, it is computationally intractable to exactly invert these large sparse matrices. Thus, we employ the *conjugate gradient* method [12] to solve the systems without explicit inversion of ${\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1}, {\boldsymbol{\Sigma}_{\boldsymbol{w}}}^{-1}$.

Still, $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ are explicitly required in the update of $q(\boldsymbol{x})$, $q(\boldsymbol{w})$, $q(\boldsymbol{b})$, $q(\boldsymbol{a})$, $q(\boldsymbol{z})$, β, α, and ω, using Equation (31), (37), (43), (47), (51), (55),

(56), and (57), respectively. Hence, we perform low-rank approximation of large sparse inverse matrices to obtain $\boldsymbol{\Sigma}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{w}}$ from $\boldsymbol{\Sigma}_{\boldsymbol{x}}{}^{-1}, \boldsymbol{\Sigma}_{\boldsymbol{w}}{}^{-1}$ approximately. Specifically, we employ the approximation method used in [6], which makes use of conjugate vectors as a byproduct of the conjugate gradient. Then, each approximate inverse is obtained as the sum of the outer products of the conjugate vectors.

5.3 Solution of Non-Linear Equations

The update of the DoF parameters ξ, ν, μ requires to solve nonlinear equations, i.e., Equation (43), (55), and (59), and we cannot obtain explicit solutions from them. Since these parameters are scalar-valued, we use the bisection method to numerically solve the equations, following the previous works on the variational Bayesian method [6, 7].

5.4 Value Clamping for Numerical Stability

At the beginning of the algorithm, estimates of several parameters can take very large values, which makes the algorithm numerically instable. For example, according to Equation (57), ω goes to infinity when w is almost zero and thus very smooth, typically on the coarser grids. Hence, we clamp the values of these parameters to keep the algorithm stable, i.e., we limit their value ranges to $[0.001, 1000]$.

5.5 Numbers and Iterations

In the coarse-to-fine algorithm, we use 4 levels of grids constructed by full-octave coarsening, i.e., given the noisy image sequence as the 1st level, we apply 3 times of spatial downsampling to it, and then we iterate over these 4 levels to obtain an initial estimate of $q(\boldsymbol{w})$ on the finest grid. We note that, if the number of levels is too large, the algorithm becomes numerically instable, due to extremely large or small values of estimates on very coarse grids.

Convergence of the variational Bayesian inference, where we iteratively update the approximate posterior distributions, can be assessed in several different ways, e.g., by examining the differences of estimates between iterations, or that

of the KL divergence to be minimized [3], at the cost of additional computation. Meanwhile, we have empirically found that only a few iterations are required for each of the three types of iteration in our algorithm, i.e., the update of each sub-group, the update for the registration or the restoration, and the overall update that alternates between the registration and the restoration. Hence, we simply iterate 4 times for each of them.

5.6 Programming Language and Libraries

To implement the multi-frame image restoration algorithm, we employ the *C++* programming language, for its performance and wealth of available libraries.

For operations on vectors and matrices, we use the *Eigen* linear algebra library. While we store the parameters as dense vectors, which are efficient for random element access, we store the operators as sparse matrices, since they are too large to treat in a dense form. For basic operations on images, e.g., data type conversion and I/O, we use the *OpenCV* library.

Chapter 6 Experiments on Motion-Blur-Free Photography

In this chapter, we show the results of two experiments to demonstrate the effectiveness of the proposed multi-exposure method for motion-blur-free photography. While synthetic images were used in the first experiment to show how effective the proposed method is, real images were used in the second one to see how it works in the real world.

6.1 Experiment for Synthetic Images

We evaluated the effectiveness of the proposed multi-exposure method for motion-blur-free photography. To evaluate it not only qualitatively but also quantitatively, we used synthetic images, for which we had access to the ground truth. We compared the proposed method with short and long exposure, and with the previous method of multi-exposure [14], which uses short-exposure images only, as with the proposed one, and performs registration by local block matching, followed by block-wise restoration.

We virtually set up two scenes with different types of global motion, i.e., translation and rotation, each of which imitates ego-motion of a camera. For each scene, we synthesized a sequence of noisy images from a real image, which is shown in Figure 7, by transforming and adding Gaussian noise to it. The number of the images in each sequence was 16, the global noise level was 16 in standard derivation, and the displacements due to motion between each adjacent images were 1 pixel and 0.5 degrees for translation and rotation, respectively. The first, middle, and the last images in these sequences are shown in Figure 8. Using these sequences, we captured images by long exposure, short exposure, the previous method [14], and our proposed method by simulation. For the long exposure, we averaged all the images in each sequence, and for the short exposure, we extracted the first image in it. The parameter for the noise level required by the previous method was set to the ground truth.

The resultant images are shown in Figure 9. As for the previous and the proposed methods, close-up versions of the results are shown in Figure 12 for com-

Figure 7: Ground-truth image from which synthetic images are generated.

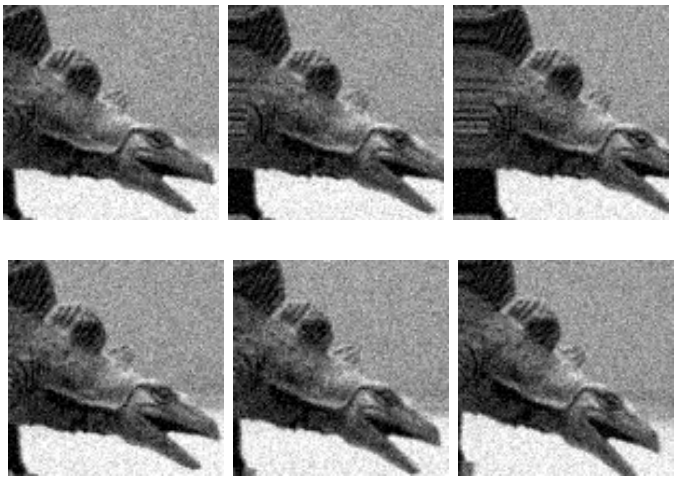
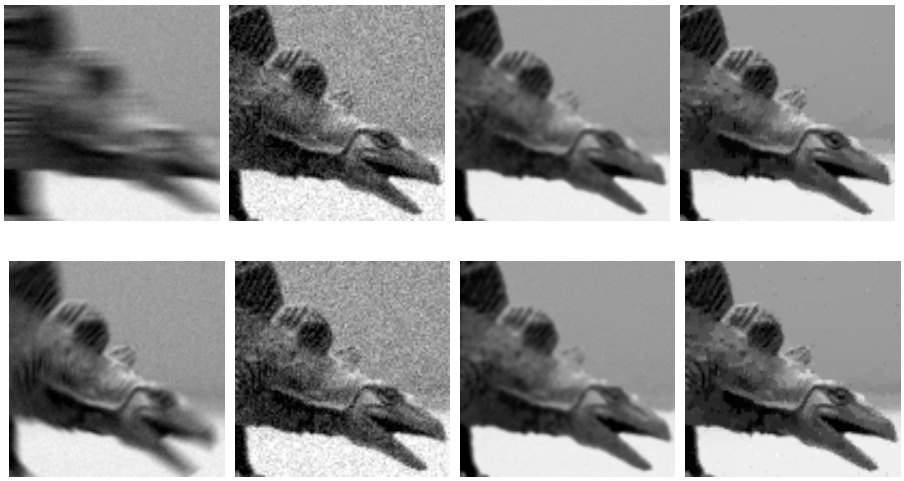


Figure 8: Synthesized image sequences for (1st row) translation and (2nd row) rotation.

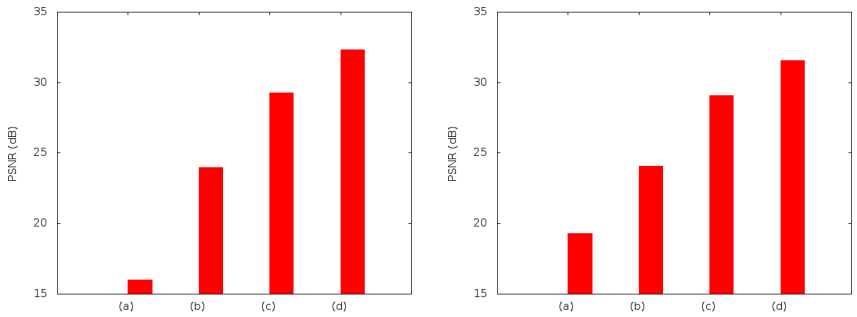
parison. Furthermore, their quantitative image quality was assessed in terms of peak signal-to-noise ratio (PSNR), which is one of standard image quality metrics in the field of image processing [9], and in terms of structural similarity (SSIM) [20], which is known to better reflect image quality perceived by the human visual system [9], as shown in Figure 10 and Figure 11, respectively.

We can see that both of the multi-exposure methods suppressed motion blur and noise, and achieved higher visual image quality than the blurry long-exposure and the noisy short-exposure image. This is the power of multi-exposure photography, which enables us to suppress motion blur without increasing noise. However, the previous method smoothed out the image details, e.g., textures of the object in the scene, resulting in lower image quality, as



(a) (b) (c) (d)

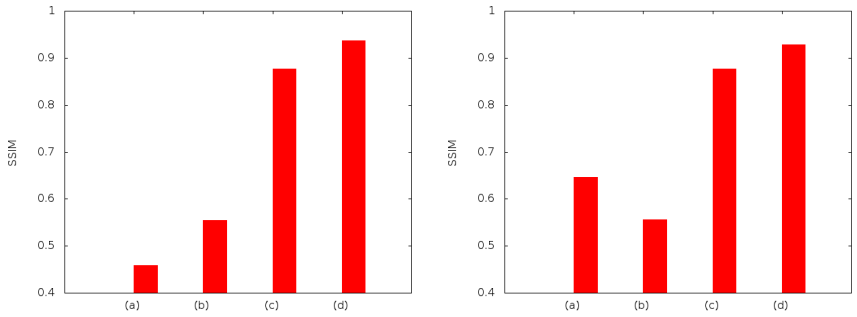
Figure 9: Images captured in the presence of (1st row) translation and (2nd row) rotation, (a) in long exposure, (b) in short exposure, (c) by the previous method of multi-exposure [14], and (d) by the proposed method of multi-exposure.



Translation Rotation

Figure 10: Peak signal-to-noise ratios (PSNRs) of the images captured (a) in long exposure, (b) in short exposure, (c) by the previous method of multi-exposure [14], and (d) by the proposed method of multi-exposure.

we can see from the close-ups in Figure 12. This oversmoothing would occur because the block matching registration between the noisy images made wrong or ambiguous correspondences, leading to integration of unrelated blocks in restoration. On the other hand, the proposed method could preserve the details for both of the scenes, as in Figure 12. Hereby, we can see the effectiveness



Translation Rotation

Figure 11: Structural similarities (SSIMs) of the images captured (a) in long exposure, (b) in short exposure, (c) by the previous method of multi-exposure [14], and (d) by the proposed method of multi-exposure.

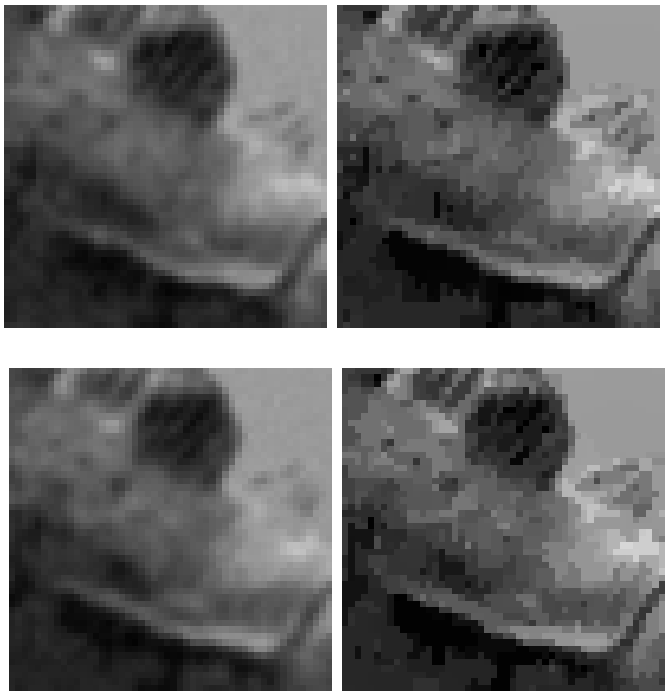
of the proposed method, which can perform registration more accurately and consequently restoration more effectively than the previous one, owing to the iterative registration and restoration by the variational Bayesian method. This qualitative argument was further confirmed quantitatively by PSNR and SSIM, for both of which the proposed method achieved higher image quality than the previous one in the two scenes.

We note that the previous method [14] required the noise parameter to be given manually. Owing to the variational Bayesian method, our algorithm could automatically estimate the parameters that determine the noise distribution, i.e., $q(\boldsymbol{b}), \beta, \xi$.

6.2 Experiment for Real Images

To demonstrate the performance of the proposed method in the real world, we applied the proposed method to real images, i.e., images observed by a real camera.

We set up a scene with relatively low illumination and large motion, where we would face either motion blur from long exposure or noise from short exposure, if we had took the single-exposure approach. In addition, the motion was complex in that both a camera and an object were moving during exposure, and thus it was difficult to deal with it by traditional approaches, such as optical



(c) (d)

Figure 12: Close-ups of images captured in the presence of (1st row) translation and (2nd row) rotation, (c) by the previous method of multi-exposure [14], and (d) by the proposed method of multi-exposure.

image stabilization. First, we observed a sequence of 10 short-exposure images with a real camera in the scene. We compensated for the loss of brightness due to short exposure by the gain of the factor 10. The original image sequence and the one after the brightness compensation were shown in Figure 13. Then, we applied the proposed method to the image sequence, as well as the previous method [14]. As the noise level was unknown, the parameter of the previous method was manually tuned to 10 in standard deviation. We also obtained images that would be captured in short and long exposure, in the same manner as the experiment for synthetic images.

The resultant images are shown, along with their close-ups, in Figure 14. Here, we can compare the results only qualitatively, as in the real world we have no reliable ground truth for quantitative evaluation.

Compared with the long and the short exposure, the proposed method suc-

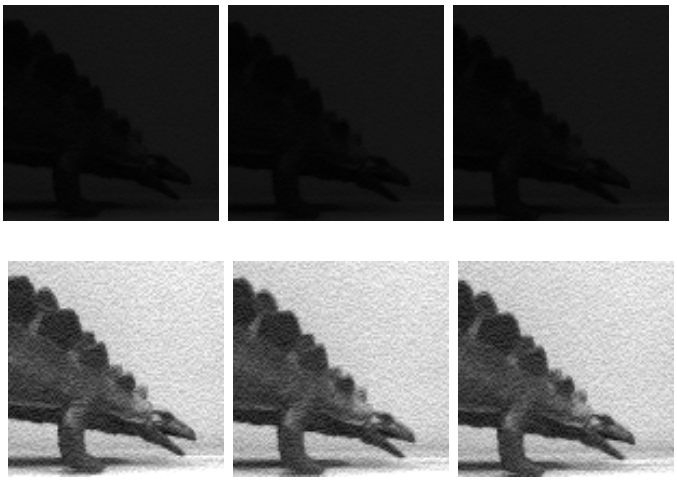


Figure 13: Real image sequences, (1st row) the original and (2nd row) the one after brightness compensation.

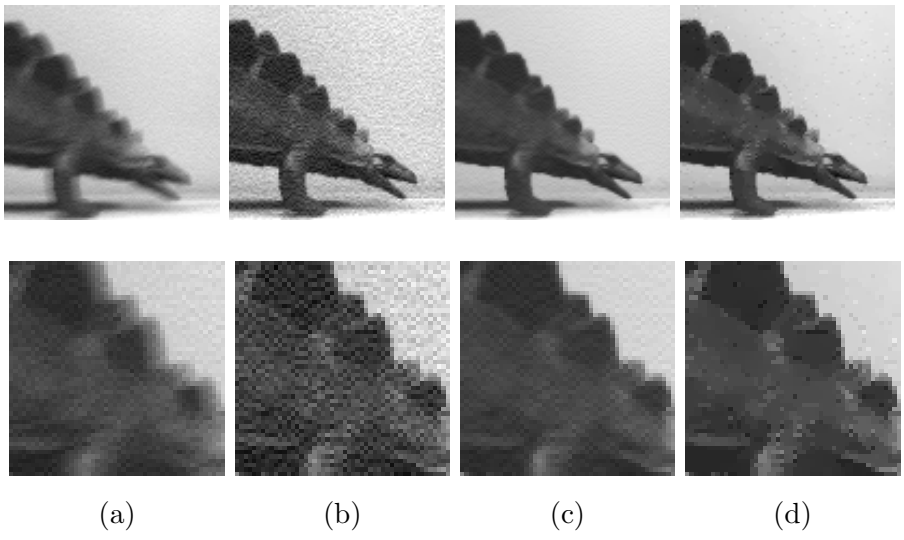


Figure 14: (1st row) Images captured with a real camera, (a) in long exposure, (b) in short exposure, (c) by the previous method of multi-exposure [14], and (d) by the proposed method of multi-exposure, and (2nd row) their close-ups.

cessfully reduced motion blur, without producing noise in the result, albeit in the presence of the complex motion and the real noise. Thereby, it was demonstrated that the proposed method is effective not only in simulation but also in

the real world. In this scene, possibly because the noise was less severe than in the scenes of the experiment with synthetic images, the performances of the previous and the proposed method were comparable, although the previous one required the manual parameter tuning.

Chapter 7 Conclusion

In this work, we propose the method of multi-exposure for motion-blur-free photography. It is build upon the novel variational Bayesian multi-frame image restoration, which iterates registration and restoration to improve one by the other. Thereby, the proposed method works well even in the presence of short-exposure noise. Through the experiments, we demonstrated the effectiveness of the proposed method both for synthetic and for real images. The proposed method extends the applicability of multi-exposure photography to challenging scenes, where we have to deal with severe short-exposure noise to successfully suppress motion blur.

One of the remaining issues of this work is that we could not observe significant contribution of the posterior covariances to the robustness of the proposed method. We suspect that, in our current implementation, the approximation method for the large sparse covariance matrices is not so accurate. Thus, we will investigate the effect of the approximated covariances in more detail, and possibly introduce a better approximation method to fully benefit from the variational Bayesian method. Besides, we will further evaluate the effectiveness of the proposed method by applying it to more challenging scenes with more complex motion, e.g., scenes with multiple moving objects.

Acknowledgments

I would like to express gratitude to my supervisor, Professor Michihiko Minoh for his continuous support and guidance for my master's thesis. My gratitude also goes to the rest of the examiners of the thesis, Professor Atsushi Nakazawa and Professor Masayuki Mukunoki for their commitment and advice on the thesis. Besides, I would like to thank to Professor Masaaki Iiyama and Professor Takuya Funatomi for their help throughout my master's course. Finally, I thank the people in Minoh Laboratory for providing me a great atmosphere for research.

Appendix

In this appendix, we give the details of the derivation of the variational Bayesian inference algorithm described in Chapter 4.

Given our final Bayesian model, which has been modified with the auxiliary variables, we approximate the joint posterior distribution of all the latent variables $p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}|\boldsymbol{y})$ with $q(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$ in Equation (27). Then, the approximate posterior distribution of each latent variable is obtained by taking the logarithmic expectation of the joint distribution of all the variables $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$, which is given by Equation (26), with respect to the approximate posterior distributions of all the other latent variables [3]. Since these distributions are dependent on each other, we update them one by one in the algorithm of the variational Bayesian inference. In the following, for each latent variable, we show how to obtain its approximate posterior distribution explicitly.

A.1 Clean Image

We obtain the approximate posterior distribution of the clean image $q(\boldsymbol{x})$:

$$\ln q(\boldsymbol{x}) = \mathrm{E}_{q(\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})} \left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})\right] + \text{const.}, \tag{A.1}$$

where the constant terms with respect to x are ignored, which belong to the normalizer of $q(\boldsymbol{x})$. To obtain the right hand side, we take the logarithm of $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$, ignoring constant terms with respect to $\boldsymbol{x}$:

$$\begin{aligned} \ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) = &\ \ln p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}) + \ln p(\boldsymbol{x}|\boldsymbol{a}) + \text{const.} \\ = &\ -\frac{1}{2} (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x})^\top \boldsymbol{B} (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}) \\ &\ -\frac{1}{2} (\boldsymbol{G}\boldsymbol{x})^\top \boldsymbol{A} (\boldsymbol{G}\boldsymbol{x}) \\ &\ + \text{const.} \end{aligned} \tag{A.2}$$

By substituting Equation (30), we linearize $\boldsymbol{W}\boldsymbol{x}$ with respect to $\boldsymbol{w}$ at $\hat{\boldsymbol{w}}$:

$$\begin{aligned}
\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) = & -\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{I}[\boldsymbol{w}']\boldsymbol{D}\boldsymbol{x}\right)^\top \boldsymbol{B}\left(\boldsymbol{y} - \boldsymbol{I}[\boldsymbol{w}']\boldsymbol{D}\boldsymbol{x}\right) \\
& -\frac{1}{2}\left(\boldsymbol{G}\boldsymbol{x}\right)^\top \boldsymbol{A}\left(\boldsymbol{G}\boldsymbol{x}\right) \\
& + \text{const.} \\
= & -\frac{1}{2}\boldsymbol{x}^\top \left(\boldsymbol{D}^\top \left(\boldsymbol{I}^\top \boldsymbol{B}\boldsymbol{I} \circ \boldsymbol{w}'\boldsymbol{w}'^\top\right)\boldsymbol{D} + \boldsymbol{G}^\top \boldsymbol{A}\boldsymbol{G}\right)\boldsymbol{x} \\
& + \boldsymbol{x}^\top \boldsymbol{D}^\top [\boldsymbol{w}'] \boldsymbol{I}^\top \boldsymbol{B}\boldsymbol{y} \\
& + \text{const.},
\end{aligned} \tag{A.3}$$

where $\boldsymbol{w}' = \boldsymbol{J}(\boldsymbol{w} - \hat{\boldsymbol{w}}) + \boldsymbol{j}$. Then, we take the expectation of the logarithm with respect to $q(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$:

$$\begin{aligned}
\mathrm{E}_{q(\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})}\left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})\right] = & -\frac{1}{2}\boldsymbol{x}^\top \left(\boldsymbol{D}^\top \left(\boldsymbol{I}^\top \overline{\boldsymbol{B}}\boldsymbol{I} \circ \overline{\boldsymbol{w}'\boldsymbol{w}'^\top}\right)\boldsymbol{D} + \boldsymbol{G}^\top \overline{\boldsymbol{A}}\boldsymbol{G}\right)\boldsymbol{x} \\
& + \boldsymbol{x}^\top \boldsymbol{D}^\top [\overline{\boldsymbol{w}'}] \boldsymbol{I}^\top \overline{\boldsymbol{B}}\boldsymbol{y} \\
& + \text{const.} \\
= & -\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}}\right)^\top {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}}\right) + \text{const.}, \\
= & \ln \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{\boldsymbol{x}}, {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1}),
\end{aligned} \tag{A.4}$$

where $\boldsymbol{\mu}_{\boldsymbol{x}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ are given by Equation (32) and (33), respectively. Therefore, $q(x)$ is the Gaussian distribution with the mean $\boldsymbol{\mu}_{\boldsymbol{x}}$ and the covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, which is equivalent to (31).

A.2 Optical Flow

We obtain the approximate posterior distribution of the optical flow $q(\boldsymbol{w})$:

$$\ln q(\boldsymbol{w}) = \mathrm{E}_{q(\boldsymbol{x},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})}\left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})\right] + \text{const.} \tag{A.5}$$

We take the logarithm of $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$, ignoring constant terms with respect to $\boldsymbol{w}$:

$$\begin{aligned}
\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) = & \ln p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}) + \ln p(\boldsymbol{w}|\boldsymbol{z}) + \text{const.} \\
= & -\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\right)^{\top} \boldsymbol{B}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\right) \\
& -\frac{1}{2}\left(\boldsymbol{F}\boldsymbol{w}\right)^{\top} \boldsymbol{Z}\left(\boldsymbol{F}\boldsymbol{w}\right) \\
& + \text{const.}
\end{aligned} \tag{A.6}$$

By substituting Equation (29), we linearize $\boldsymbol{W}\boldsymbol{x}$ with respect to $\boldsymbol{w}$:

$$\begin{aligned}
\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) = & -\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{I}[\boldsymbol{x}']\left(\boldsymbol{J}\left(\boldsymbol{w} - \hat{\boldsymbol{w}}\right) + \boldsymbol{j}\right)\right)^{\top} \boldsymbol{B}\left(\boldsymbol{y} - \boldsymbol{I}[\boldsymbol{x}']\left(\boldsymbol{J}\left(\boldsymbol{w} - \hat{\boldsymbol{w}}\right) + \boldsymbol{j}\right)\right) \\
& -\frac{1}{2}\left(\boldsymbol{F}\boldsymbol{w}\right)^{\top} \boldsymbol{Z}\left(\boldsymbol{F}\boldsymbol{w}\right) \\
& + \text{const.} \\
= & -\frac{1}{2}\boldsymbol{w}^{\top}\left(\boldsymbol{J}^{\top}\left(\boldsymbol{I}^{\top}\boldsymbol{B}\boldsymbol{I} \circ \boldsymbol{x}'\boldsymbol{x}'^{\top}\right)\boldsymbol{J} + \boldsymbol{F}^{\top}\boldsymbol{Z}\boldsymbol{F}\right)\boldsymbol{w} \\
& + \boldsymbol{w}^{\top}\boldsymbol{J}^{\top}\left([\boldsymbol{x}']\boldsymbol{I}^{\top}\boldsymbol{B}\boldsymbol{y} + \left(\boldsymbol{I}^{\top}\boldsymbol{B}\boldsymbol{I} \circ \boldsymbol{x}'\boldsymbol{x}'^{\top}\right)\left(\boldsymbol{J}\hat{\boldsymbol{w}} - \boldsymbol{j}\right)\right) \\
& + \text{const.},
\end{aligned} \tag{A.7}$$

where $\boldsymbol{x}' \equiv \boldsymbol{D}\boldsymbol{x}$. Then, we take the expectation of the logarithm with respect to $q(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$:

$$\begin{aligned}
\mathrm{E}_{q(\boldsymbol{x},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})}\left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})\right] = & -\frac{1}{2}\boldsymbol{w}^{\top}\left(\boldsymbol{J}^{\top}\left(\boldsymbol{I}^{\top}\overline{\boldsymbol{B}\boldsymbol{I}} \circ \overline{\boldsymbol{x}'\boldsymbol{x}'^{\top}}\right)\boldsymbol{J} + \boldsymbol{F}^{\top}\overline{\boldsymbol{Z}}\boldsymbol{F}\right)\boldsymbol{w} \\
& + \boldsymbol{w}^{\top}\boldsymbol{J}^{\top}\left(\overline{[\boldsymbol{x}']}\boldsymbol{I}^{\top}\overline{\boldsymbol{B}}\boldsymbol{y} + \left(\boldsymbol{I}^{\top}\overline{\boldsymbol{B}\boldsymbol{I}} \circ \overline{\boldsymbol{x}'\boldsymbol{x}'^{\top}}\right)\left(\boldsymbol{J}\hat{\boldsymbol{w}} - \boldsymbol{j}\right)\right) \\
& + \text{const.} \\
= & -\frac{1}{2}\left(\boldsymbol{w} - \boldsymbol{\mu}_{\boldsymbol{w}}\right)^{\top}{\boldsymbol{\Sigma}_{\boldsymbol{w}}}^{-1}\left(\boldsymbol{w} - \boldsymbol{\mu}_{\boldsymbol{w}}\right) + \text{const.}, \\
= & \ln \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_{\boldsymbol{w}}, {\boldsymbol{\Sigma}_{\boldsymbol{w}}}^{-1}),
\end{aligned} \tag{A.8}$$

where $\boldsymbol{\mu}_{\boldsymbol{w}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ are given by Equation (38) and (39), respectively. Therefore, $q(w)$ is the Gaussian distribution with the mean $\boldsymbol{\mu}_{\boldsymbol{w}}$ and the covariance $\boldsymbol{\Sigma}_{\boldsymbol{w}}$, which is equivalent to (37).

A.3 Auxiliary Variables

First, we obtain the approximate posterior distribution $q(\boldsymbol{b})$:

$$\ln q(\boldsymbol{b}) = \mathrm{E}_{q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{a},\boldsymbol{z})} \left[\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})\right] + \text{const.} \tag{A.9}$$

We take the logarithm of $p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})$, ignoring constant terms with respect to $\boldsymbol{b}$:

$$\begin{aligned}
\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z}) = \quad & \ln p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{w},\boldsymbol{b}) + \ln p(\boldsymbol{b}) + \text{const.} \\
= \quad & \frac{1}{2}\ln|[\boldsymbol{b}]| - \frac{1}{2}\operatorname{tr}\left(\beta\left(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}\right)\left(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}\right)^{\top}[\boldsymbol{b}]\right) \\
& + \left(\frac{\xi}{2}-1\right)\ln|[\boldsymbol{b}]| - \frac{1}{2}\operatorname{tr}\left(\xi\boldsymbol{I}_{\boldsymbol{y}}[\boldsymbol{b}]\right) \\
& + \text{const.} \\
= \quad & \left(\frac{\xi+1}{2}-1\right)\ln|[\boldsymbol{b}]| \\
& - \frac{1}{2}\operatorname{tr}\left(\left(\xi\boldsymbol{I}_{\boldsymbol{y}} + \beta\left(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}\right)\left(\boldsymbol{y}-\boldsymbol{W}\boldsymbol{x}\right)^{\top}\right)[\boldsymbol{b}]\right) \\
& + \text{const.}
\end{aligned} \tag{A.10}$$

By substituting Equation (28), we linearize $\boldsymbol{W}\boldsymbol{x}$ with respect to $\boldsymbol{w}$:

$$\begin{aligned}
\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z}) = \quad & \left(\frac{\xi+1}{2}-1\right)\ln|[\boldsymbol{b}]| \\
& - \frac{1}{2}\operatorname{tr}\left(\left(\xi\boldsymbol{I}_{\boldsymbol{y}} + \beta\boldsymbol{y}'\boldsymbol{y}'^{\top}\right)[\boldsymbol{b}]\right) \\
& + \text{const.},
\end{aligned} \tag{A.11}$$

where $\boldsymbol{y}' \equiv \boldsymbol{y} - \boldsymbol{I}(\boldsymbol{w}' \circ \boldsymbol{x}')$. Then, we take the expectation of the logarithm with respect to $q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{a},\boldsymbol{z})$:

$$\begin{aligned}
\mathrm{E}_{q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{a},\boldsymbol{z})}\left[\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})\right] = \quad & \left(\frac{\xi+1}{2}-1\right)\ln|[\boldsymbol{b}]| \\
& - \frac{1}{2}\operatorname{tr}\left(\left(\xi\boldsymbol{I}_{\boldsymbol{y}} + \beta\overline{\boldsymbol{y}'\boldsymbol{y}'^{\top}}\right)[\boldsymbol{b}]\right) \\
& + \text{const.} \\
= \quad & (k_{\boldsymbol{b}}-1)\ln|[\boldsymbol{b}]| - \operatorname{tr}\left([\boldsymbol{\theta}_{\boldsymbol{b}}{}^{(-1)}][\boldsymbol{b}]\right) + \text{const.} \\
= \quad & \ln \mathcal{G}(\boldsymbol{b}|k_{\boldsymbol{b}}, [\boldsymbol{\theta}_{\boldsymbol{b}}{}^{(-1)}]),
\end{aligned} \tag{A.12}$$

where $k_{\boldsymbol{b}}$ and ${\boldsymbol{\theta}_{\boldsymbol{b}}}^{(-1)}$ are given by Equation (44) and (45), respectively. Therefore, $q(\boldsymbol{b})$ is the Gamma distribution with the shape $k_{\boldsymbol{b}}$ and the rate $[{\boldsymbol{\theta}_{\boldsymbol{b}}}^{(-1)}]$, which is equivalent to (43).

Second, we obtain the approximate posterior distribution $q(\boldsymbol{a})$:

$$\ln q(\boldsymbol{a}) = \mathrm{E}_{q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{z})} \left[\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})\right] + \text{const.} \tag{A.13}$$

We take the logarithm of $p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})$, ignoring constant terms with respect to $\boldsymbol{a}$:

$$\begin{aligned}
\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z}) = &\quad \ln p(\boldsymbol{x}|\boldsymbol{a}) + \ln p(\boldsymbol{a}) + \text{const.} \\
= &\quad \frac{1}{2}\ln|[\boldsymbol{a}]| - \frac{1}{2}\operatorname{tr}\left(\alpha\left(\boldsymbol{G}\boldsymbol{x}\right)\left(\boldsymbol{G}\boldsymbol{x}\right)^{\top}[\boldsymbol{a}]\right) \\
&+ \left(\frac{\nu}{2} - 1\right)\ln|[\boldsymbol{a}]| - \frac{1}{2}\operatorname{tr}\left(\nu\boldsymbol{I}_{\boldsymbol{G}\boldsymbol{x}}[\boldsymbol{a}]\right) \\
&+ \text{const.} \\
= &\quad \left(\frac{\nu+1}{2} - 1\right)\ln|[\boldsymbol{a}]| \\
&- \frac{1}{2}\operatorname{tr}\left(\left(\nu\boldsymbol{I}_{\boldsymbol{G}\boldsymbol{x}} + \alpha\boldsymbol{G}\boldsymbol{x}\boldsymbol{x}^{\top}\boldsymbol{G}^{\top}\right)[\boldsymbol{a}]\right) \\
&+ \text{const.}
\end{aligned} \tag{A.14}$$

Then, we take the expectation of the logarithm with respect to $q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{z})$:

$$\begin{aligned}
\mathrm{E}_{q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{z})}\left[\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})\right] = &\quad \left(\frac{\nu+1}{2} - 1\right)\ln|[\boldsymbol{a}]| \\
&- \frac{1}{2}\operatorname{tr}\left(\left(\nu\boldsymbol{I}_{\boldsymbol{G}\boldsymbol{x}} + \alpha\boldsymbol{G}\overline{\boldsymbol{x}\boldsymbol{x}^{\top}}\boldsymbol{G}^{\top}\right)[\boldsymbol{a}]\right) \\
&+ \text{const.} \\
= &\quad (k_{\boldsymbol{a}} - 1)\ln|[\boldsymbol{a}]| - \operatorname{tr}\left([{\boldsymbol{\theta}_{\boldsymbol{a}}}^{(-1)}][\boldsymbol{a}]\right) + \text{const.} \\
= &\quad \ln \mathcal{G}(\boldsymbol{a}|k_{\boldsymbol{a}}, [{\boldsymbol{\theta}_{\boldsymbol{a}}}^{(-1)}]),
\end{aligned} \tag{A.15}$$

where $k_{\boldsymbol{a}}$ and ${\boldsymbol{\theta}_{\boldsymbol{a}}}^{(-1)}$ are given by Equation (48) and (49), respectively. Therefore, $q(\boldsymbol{a})$ is the Gamma distribution with the shape $k_{\boldsymbol{a}}$ and the rate $[{\boldsymbol{\theta}_{\boldsymbol{a}}}^{(-1)}]$, which is equivalent to (47).

Third, we obtain the approximate posterior distribution $q(\boldsymbol{z})$:

$$\ln q(\boldsymbol{z}) = \mathrm{E}_{q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a})} \left[\ln p(\boldsymbol{y},\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a},\boldsymbol{z})\right] + \text{const.} \tag{A.16}$$

We take the logarithm of $p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z})$, ignoring constant terms with respect to $\boldsymbol{z}$:

$$\begin{aligned}
\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) = & \ln p(\boldsymbol{w}|\boldsymbol{z}) + \ln p(\boldsymbol{z}) + \text{const.} \\
= & \frac{1}{2} \ln |[\boldsymbol{z}]| - \frac{1}{2} \operatorname{tr} \left(\omega \left(\boldsymbol{F}\boldsymbol{w} \right) \left(\boldsymbol{F}\boldsymbol{w} \right)^{\top} [\boldsymbol{z}] \right) \\
& + \left(\frac{\mu}{2} - 1 \right) \ln |[\boldsymbol{z}]| - \frac{1}{2} \operatorname{tr} \left(\mu \boldsymbol{I}_{\boldsymbol{F}\boldsymbol{w}} [\boldsymbol{z}] \right) \\
& + \text{const.} \\
= & \left(\frac{\mu + 1}{2} - 1 \right) \ln |[\boldsymbol{z}]| \\
& - \frac{1}{2} \operatorname{tr} \left(\left(\mu \boldsymbol{I}_{\boldsymbol{F}\boldsymbol{w}} + \omega \boldsymbol{F}\boldsymbol{w}\boldsymbol{w}^{\top}\boldsymbol{F}^{\top} \right) [\boldsymbol{z}] \right) \\
& + \text{const.}
\end{aligned} \tag{A.17}$$

Then, we take the expectation of the logarithm with respect to $q(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a})$:

$$\begin{aligned}
\mathrm{E}_{q(\boldsymbol{x},\boldsymbol{w},\boldsymbol{b},\boldsymbol{a})} \left[\ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{z}) \right] = & \left(\frac{\mu + 1}{2} - 1 \right) \ln |[\boldsymbol{z}]| \\
& - \frac{1}{2} \operatorname{tr} \left(\left(\mu \boldsymbol{I}_{\boldsymbol{F}\boldsymbol{w}} + \omega \boldsymbol{F} \overline{\boldsymbol{w}\boldsymbol{w}^{\top}} \boldsymbol{F}^{\top} \right) [\boldsymbol{z}] \right) \\
& + \text{const.} \\
= & (k_{\boldsymbol{z}} - 1) \ln |[\boldsymbol{z}]| - \operatorname{tr} \left([{\boldsymbol{\theta}_{\boldsymbol{z}}}^{(-1)}][\boldsymbol{z}] \right) + \text{const.} \\
= & \ln \mathcal{G}(\boldsymbol{z}|k_{\boldsymbol{z}}, [{\boldsymbol{\theta}_{\boldsymbol{z}}}^{(-1)}]),
\end{aligned} \tag{A.18}$$

where $k_{\boldsymbol{z}}$ and ${\boldsymbol{\theta}_{\boldsymbol{z}}}^{(-1)}$ are given by Equation (52) and (53), respectively. Therefore, $q(\boldsymbol{z})$ is the Gamma distribution with the shape $k_{\boldsymbol{z}}$ and the rate $[{\boldsymbol{\theta}_{\boldsymbol{z}}}^{(-1)}]$, which is equivalent to (51).

References

[1] Babacan, S., Molina, R. and Katsaggelos, A.: Variational Bayesian super resolution, *IEEE Trans. Image Process.*, Vol. 20, No. 4, pp. 984–999 (2011).

[2] Ben-Ezra, M. and Nayar, S.: Motion-based motion deblurring, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 26, No. 6, pp. 689–698 (2004).

[3] Bishop, C.: *Pattern recognition and machine learning*, Springer (2006).

[4] Brox, T., Bruhn, A., Papenberg, N. and Weickert, J.: High accuracy optical flow estimation based on a theory for warping, *Proc. ECCV LNCS*, pp. 25–36 (2004).

[5] Brox, T. and Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 33, No. 3, pp. 500–513 (2011).

[6] Chantas, G., Galatsanos, N., Molina, R. and Katsaggelos, A.: Variational Bayesian image restoration with a product of spatially weighted total variation image priors, *IEEE Trans. Image Process.*, Vol. 19, No. 2, pp. 351–362 (2010).

[7] Chantas, G., Gkamas, T. and Nikou, C.: Variational-Bayes optical flow, *J. Math. Imaging Vis.*, Vol. 50, No. 3, pp. 199–213 (2014).

[8] Gkamas, T., Chantas, G. and Nikou, C.: A probabilistic formulation of the optical flow problem, *Proc. ICPR*, pp. 754–757 (2012).

[9] Hore, A. and Ziou, D.: Image Quality Metrics: PSNR vs. SSIM, *Proc. ICPR*, pp. 2366–2369 (2010).

[10] Horn, B. and Schunck, B.: Determining optical flow, *Artifical Intelligence*, Vol. 17, pp. 185–203 (1981).

[11] Lucas, B. and Kanade, T.: An iterative image registration technique with an application to stereo vision, *Proc. IJCAI*, pp. 674–679 (1981).

[12] Saad, Y.: *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics (2003).

[13] Szeliski, R.: *computer vision: algorithms and applications*, Springer (2011).

[14] Tico, M.: Adaptive block-based approach to image stabilization, *Proc. ICIP*, pp. 521–524 (2008).

[15] Tico, M., Alenius, S. and Vehvilainen, M.: Method of motion estimation for image stabilization, *Proc. ICASSP*, pp. 277–280 (2006).

[16] Tico, M., Gelfand, N. and Pulli, K.: Motion-blur-free exposure fusion, *Proc. ICIP*, pp. 3321–3324 (2010).

[17] Tico, M. and Pulli, K.: Image enhancement method via blur and noisy image fusion, *Proc. ICIP* (2009).

[18] Tico, M. and Pulli, K.: Low-light imaging solutions for mobile devices, *Proc. ACSSC*, pp. 851–855 (2009).

[19] Villena, S., Vega, M., Molina, R. and Katsaggelos, A.: A non-stationary image prior combination in super-resolution, *Digital Signal Processing*, Vol. 32, pp. 1–10 (2014).

[20] Wang, Z., Bovik, A. C., Sheikh, H. and Simoncelli, E.: Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.*, Vol. 13, No. 4, pp. 600–612 (2004).

[21] Wei, Z., Cao, Y. and Newton, A.: Digital image restoration by exposure-splitting and registration, *Proc. ICPR*, pp. 657–660 (2004).

[22] Yuan, L., Sun, J., Quan, L. and Shum, H.: Image deblurring with blurred/noisy image pairs, *ACM Trans. Graph.*, Vol. 26, p. 3 (2007).

[23] Zhang, H. and Carin, L.: Multi-shot imaging: joint alignment, deblurring and resolution-enhancement, *Proc. CVPR*, pp. 2925–2932 (2014).