Laparoscopy Image Enhancement

A Dissertation
Submitted in partial fulfillment of
the requirements for the degree of
Master of Technology
by

Ayush Baid
(12D100002)

Supervisors:
Prof. S. N. Merchant
and
Prof. S. Awate

Department of Electrical Engineering
Indian Institute of Technology Bombay

15 May 2017

Dedicated to my parents

Approval Sheet

This dissertation entitled “Laparoscopy Image Enhancement” by Ayush Baid is approved for the degree of Master of Technology.

_____

_____

_____

Examiners

_____

_____

_____

Supervisor (s)

_____

Chairman

Date: _____

Place: _____

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I declare that I have properly and accurately acknowledged all sources used in the production of this report. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Ayush Baid

Date: 15 May 2017 (12D100002)

Abstract

Laparoscopy images exhibit artifacts like occlusion from surgical smoke, specular highlights, and noise. These artifacts hinders visibility, and degrades post processing (e.g. segmentation). We tackle these degradations as a novel *unified Bayesian inference problem*. We propose *probabilistic graphical models* and *sparse dictionary models* as image priors. We obtain maximum-apriori probability (MAP) estimate by *variational Bayesian expectation-maximization*. Results on simulated and real-world laparoscopy images show that our joint optimization strategy outperforms the state-of-the-art.

Index terms — Laparoscopy, desmoking, specularity re-moval, denoising, variational Bayes, EM, graphical models

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Laparoscopy is a popular *minimally invasive surgery* technique in which operations are performed by inserting equipments through small incisions. Laparoscopic surgery offers advantage such as less pain and hemorrhaging, shorter recovery times over open procedures. The key equipment is a **laparoscope**, an optical imaging instrument which relays the visuals on a screen. Another main equipment is a cold light source to illuminate the area of operation.

The closed nature of laparoscopy images presents some challenges. The images can get severely corrupted with specular highlights [4, 5], surgical smoke [6], and noise. Specular highlights result from strong reflection of the light source by body fluids like blood and mucus. Speckles interfere with post-processing like segmentation [7, 8] and tracking [9]. Electrical cauterization of a tissue generates surgical smoke, which hinders visibility for surgeons and robots alike. Noise is present in all optical imaging systems and a laparoscope is no exception.

Our work jointly tackles the mentioned artifacts. We assume that the smoke color, speckle color, and location of speckles is predetermined and available for our use. Probabilistic graphical models are used variables in the system and formulate a unified Bayesian inference problem, which is solved using expectation-maximation (EM) algorithm. We introduce variational Bayesian approximation to overcome the analytical intractability in the optimization scheme.

Chapter 2

Literature Survey

To the best of our knowledge, no existing work tackles smoke, speckles, and noise in a joint setting. We will cover these three and some related problem separately. First, we will look into specular highlights removal in laparoscopy images, which is mostly tackled as an inpainting problems. Inpainting is a process in filling in missing information, usually using true information in the surroundings. Then, we will cover dehazing, both with and without noise removal. Dehazing is haze removal in outdoor images and bears similarity with desmoking laparoscopic images. This will be followed with desmoking. We will not cover denoising as an independent domain.

2.1 Speckle removal in laparoscopy images

[10] use a 2-step inpainting process. In the first step, they fill in the missing data by the centroid of available data within a certain distance and perform strong smoothing using a Gaussian kernel. The smooth image output of the first step and the original image is combined using a weight mask in step 2. The weight mask has high weights near the speckles and decays non-linearly with distance. This results in a gradual transition between original image and the smooth median filtered image. The results however, are smooth and lack texture. This is expected because median filtering is not suitable to interpolate texture.

Isotropic color diffusion is used by [5]. They use discrete convolutions with a kernel repeatedly until convergence is reached. [4] use temporal non-rigid registration to obtain pixel values lost due to speckles. The location of speckles shift with time, and hence missing data can be interpolated by control points obtained after registration with frames captured at different

instances. Both the methods perform averaging for inpainting and hence are unable to fill in texture.

2.2 Dehazing

Outdoor images, particularly of landscapes are often plagued by haze. Haze can be natural (fog) or artificial due to pollution. Haze corrupts the color of image, and when present in large concentration, it can completely obscure the subjects.

The effect of haze is modeled by a linear combination of object's radiance and haze color [11]. The following equation is ubiqutous in literature. Equation (2.1) captures the effect of haze.

$$X(i) = T(i)J(i) + (1 - T(i))A \tag{2.1}$$

where i is pixel location, X is observed image, $T \in [0, 1]$ is the haze transmission coefficient, J is radiance of the scene sans haze, and A is the airlight (considered constant for all pixels). An important property which is exploited quite often is that the haze transmission coefficient T is directly proportional to scene depth, and is hence spatially smooth.

[12] used Markov random field (MRF) to model the transmission map. Squared difference for four nearest-neighbors for each pixel location is penalized to enforce spatial regularity. Spatial regularity of transmission map is also used by [13] as a prior for the MRF model. The image contrast is associated with the number of edges and is optimized for to get haze free high contrast images. Both the methods do not utilize any information about the distribution of colors in the image.

[3] observe a statistical property that most local patches in outdoor haze-free images contain some pixels that have low intensities in at least one color channel. Infact, the lowest intensity in any color channel in a local patch is called *dark channel* and serves as an estimate for the transmission coefficient at that location. Soft matting is used to obtain a smooth final estimate of transmission map. Airlight is estimated by the top 0.1 percent brightest pixel in the dark channel. Laparoscopy image exhibit less variation compared to outdoor images, and hence stronger statistical properties can be derived and used for our problem. [14] use adaptive patch size and replace the soft matting step with guided filtering. [1] calculate the dark channel and then apply adaptive wiener filter to smooth out the transmission map.

124 images are used by [15] to generate a final image performing weighted averaging of transitionally aligned images. Mt. Rainer, the subject of interest in the paper has large white glaciers which do not obey the local dark channel property. They also assume airlight constant for a scan-line and not for the whole image and compute the dark channel value per scan-line. This is then used to compute the transmission map and dehaze the image. This method is impractical for laparoscopy as it requires large number of images for a subject.

2.3 Simultaneous dehazing and denoising

[16] argue that the *dark channel* will be susceptible to outliers resulting from noise. They propose an iterative non-parametric kernel regression. The optimization is performed by alternating between minimization in terms of the transmission map and the uncorrupted image estimate.

2.4 Desmoking

Chapter 3

Formulation

We will now model the system as well as its variables using MRFs. We will introduce priors on the variables. We will then derive the optimization objective.

X is an uncorrupted image instance, which we want to estimate. Y is the observed image, T is the smoke transmission map at hand, and R is the speckle map. K_{spec} and K_{smoke} are speckle and smoke color respectively.

3.1 Image formation

The artifacts are captured in a 3 step fashion. In this section, i denotes the pixel location. Equation (3.1) captures the effect of speckles using a binary speckle map R, where R_i having a value 1 denotes the presence of speckle. Equation (3.2) captures the smoke using smoke map $T, T_i \in [0, 1]$ and an i.i.d white Gaussian noise n_i is added at each pixel.

$$Z_i = (1 - R_i)X_i + R_i K_{spec} \tag{3.1}$$

$$Y_i = T_i Z_i + (1 - T_i)K_{smoke} + \eta_i \tag{3.2}$$

3.2 Variable modeling

We have two variables to model: the original uncorrupted image and the smoke transmission map. They are denoted by Markov random fields (MRFs) $\mathbf{X}$ and $\mathbf{R}$ respectively. If there are I

pixels in the underlying image, then

$$\mathbf{X} := \{X_i\}_{i=1}^{I} \tag{3.3}$$

$$\mathbf{T} := \{T_i\}_{i=1}^{I} \tag{3.4}$$

where $X_i \in [0,1]^3$ is a vector valued random variable denoting a value in RGB color space, and $T_i \in [0,1]$ is a scalar random variable denoting the smoke transmission coefficient.

We will now introduce the priors on $\mathbf{X}$ and $\mathbf{T}$. We want $\mathbf{X}$ to have texture to tackle the speckles, and to have high contrast to counter the smoke. We want $\mathbf{T}$ to be spatially smooth.

3.2.1 Sparse coding under a dictionary

Given a data matrix $\mathbf{X}$, whose columns are i.i.d. random vectors, a linear decomposition is of the form

$$\mathbf{X} \approx \mathbf{DS} \tag{3.5}$$

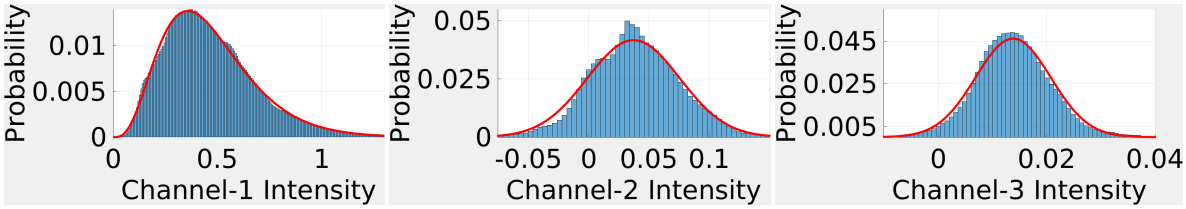
In Equation (3.5), $\mathbf{D}$ is the dictionary and $\mathbf{S}$ is a code of $\mathbf{X}$.

In sparse coding [17], we want an input vector to be using as few atoms of the dictionary as possible. A common way to enforce sparsity is regularization. λ in Equation (3.6) controls the tradeoff between sparseness and reconstruction accuracy. Equation (3.6) is minimized to obtain the best linear sparse representation.

$$J(\mathbf{D},\mathbf{S}) = \frac{1}{2}||\mathbf{X}-\mathbf{DS}||_2^2 + \lambda \sum_{ij} f(S_{ij}) \tag{3.6}$$

The sparsity function f typically is strictly increasing function of the absolute value of its argument. During optimization, there is a decrease in optimization objective by simply scaling up the dictionary atoms and decreasing the value of coefficient in the code commensurately. To prevent this blowup, there is a constraint on the norm of dictionary atoms.

Non-negative sparse coding [18] can be applied when $\mathbf{X}$ is non-negative. It constrains $\mathbf{D}$ and $\mathbf{C}$ to be non-negative. The choice of sparsity function is $f(S_{ij}) = |S_{ij}| = S_{ij}$. The main



(a) (b) (c)

Figure 3.1: **Learning Prior PDFs on Color.** Empirical histograms (bar plots) and fitted parametric PDFs (solid curves) in uncorrupted laparoscopic images, for 3 channel components: **(a)** gamma Γ_1, **(b)** Gaussian G_2, **(c)** Gaussian G_3.

intuition behind introducing non-negative constraint is that in many data classes like images, parts combine additively to form a whole as opposed to canceling each other.

[19] proposed an online algorithm for dictionary learning in the non-negative sparse coding framework. They use least angle regression to solve the sparse coding problem, given a dictionary. Their algorithm and accompanying code provide a significant speedup for learning a dictionary on large dataset.

3.2.2 Image intensity distribution

Smoke, generally gray, damages the color in the image. We need some model distribution of colors in uncorrupted high quality images to compare our image at hand with. A distribution in the combined *RGB* space is the best model but it gives rise to high complexity for its representation as no generic distribution can fit our dataset.

The channels in the *RGB* space exhibit very high correlation and modeling the channels independently is a poor choice. We first transform the data from *RGB* space to *LMS* space, as the latter is more closely related to human perception. We then calculate the eigenvectors and use them as the basis vectors for the new space, which we call $l\alpha\beta$ space. The final transformation is

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.3568 & 0.8413 & 0.5304 \\ 0.0760 & -0.2006 & 0.1239 \\ 0.2267 & 0.3574 & -0.6512 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.7}$$

3.2.3 KS statistic

Kolmogorov-Smirnov (KS) statistic is used to compare a sample with a reference probability distribution. Given an empirical cumulative distribution function (CDF) F_{emp} and reference CDF F_{ref}, the KS distance is

$$KS(F^{emp}; F^{ref}) = \max_k |F^{emp}(k) - F^{ref}(k)| \tag{3.8}$$

KS distance has low computational complexity but it is not differentiable. We can approximate the gradient as $k - \text{cdfmatch}(F^{emp}; F^{ref})(k)$, where cdfmatch is the CDF matching function.

On our data, we will generate the empirical distribution $F_{\mathbf{x}}$, evaluate the KS statistic independently for the three channels and add them up. The total penalty will be

$$J_{KS}(\mathbf{x}) = \sum_{i=1}^{3} KS(F^{\mathbf{x}}_{c_i}; F^{ref}_{c_i}) \tag{3.9}$$

where c_i is channel number i.

3.2.4 Kernel density estimation

Kernel density estimation (KDE) is used to estimate the probability density function in a non-parametric way. Let $\mathbf{x} = (x_1, x_2, ..., x_n)$ be i.i.d. samples from a distribution f. The kernel density estimate is

$$f^{\mathbf{x}}(b) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{b - x_i}{h}\right) \tag{3.10}$$

where $K(.)$ is a kernel which should take non-negative values, integrate to one, and have mean zero. $h > 0$ is the bandwidth and controls the trade-off between bias and variance of the . We will use the Gaussian kernel due to its mathematical property like differentiability. Due to our choice of Gaussian kernel, we can use the rule of thumb estimate [20] for bandwidth using the standard deviation of samples $\hat{\sigma}$. Equation (3.11) is used to tune the bandwidth using training data.

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \tag{3.11}$$

We will now derive the CDF estimator and its gradients using the Gaussian kernel.

$$f^{\mathbf{x}}(b) = \frac{1}{nh\sqrt{2\pi\sigma^2}} \sum_{i=1}^{n} \exp\left(-\frac{(b-x_i)^2}{2\sigma^2 h^2}\right) \tag{3.12}$$

$$F^{\mathbf{x}}(b) = \sum_{s=x_{min}}^{b} f^{\mathbf{x}}(s) \tag{3.13}$$

$$\frac{\partial f^{\mathbf{x}}(b)}{\partial x_i} = \frac{1}{nh\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(b-x_i)^2}{2\sigma^2 h^2}\right) \frac{(b-x_i)}{\sigma^2 h^2} \tag{3.14}$$

$$\frac{\partial F^{\mathbf{x}}(b)}{\partial x_i} = \sum_{s=x_{min}}^{b} \frac{\partial f_h^{\mathbf{x}}(s)}{\partial x_i} \tag{3.15}$$

We will now introduce the cost function J_{dist} in Equation (3.16) and its gradient, where $H(\cdot)$ in evaluates the transformation of a point under CDF matching of the empirical distribution with the reference (Equation (3.18)). The choice of the optimization objective was made as it exhibits better convergence when gradient descent is used.

$$J_{dist}(\mathbf{x}) = \sum_{b \in B} \left\{ b - H\left(b; F^{\mathbf{x}}, F^{ref}\right) \right\}^2 \tag{3.16}$$

$$\frac{\partial J_{dist}(\mathbf{x})}{\partial x_i} = 2 \sum_{b \in B} \left\{ b - H\left(b; F^{\mathbf{x}}, F_{ref}\right) \right\} \frac{\partial H\left(b; F^{\mathbf{x}}, F_{ref}\right)}{\partial x_i} \tag{3.17}$$

$$H\left(b; F^{\mathbf{x}}, F_{ref}\right) = F_{ref}^{-1}\left(F^{\mathbf{x}}(b)\right) \tag{3.18}$$

$$\frac{\partial H\left(b; F^{\mathbf{x}}, F_{ref}\right)}{\partial x_i} = F_{ref}^{-1}\left(F^{\mathbf{x}}(b)\right) \frac{\partial F^{\mathbf{x}}(b; x_1^n)}{\partial x_i} \tag{3.19}$$

On our data, we will add up the penalty for each channel c_i independently.

$$J_{KDE}(\mathbf{x}) = \sum_{i=1}^{3} \sum_{b \in B_{c_i}} \left\{ b - H\left(b; F_{c_i}^{\mathbf{x}}, F_{c_i}^{ref}\right) \right\}^2 \tag{3.20}$$

3.2.5 Prior on image

3.2.6 Spatial smoothness prior

To enforce spatial smoothness prior, we will penalize deviations in a local neighborhood. The MRF $\mathbf{T}$ is defined with a neighborhood system $\mathcal{N}^T := \{\mathcal{N}_i^T\}$, where $\mathcal{N}_i^T$ is the set of 4

nearest neighbors at pixel i. The prior distribution is defined as

$$P(\mathbf{T}) = \frac{1}{Z} \exp\left(-\sum_{i=1}^{I} \sum_{j \in \mathcal{N}_i^T} \gamma_3 (T_i - T_j)^2 \right) \tag{3.21}$$

where the outer sum is performed over all pixels, and $\gamma \in \mathbb{R}^+$ is a free parameter.

Chapter 4

Estimation

For our algorithm, we assume that we are provided with speckle label map R, speckle color K_{spec}, and smoke color K_{smoke}. We will use maximum a posteriori probability (MAP) estimation for image $\mathbf{X}$.

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}, r) = \arg\max_{\mathbf{x}} P(\mathbf{y}|\mathbf{x}, r)P(\mathbf{x}) \tag{4.1}$$

The first part of the final term in Equation (4.1) is the likelihood of observing the output $\mathbf{y}$. The second part is the prior probability of $\mathbf{x}$.

4.1 EM

4.2 VB-EM version 1

We will introduce smoke transmission map $\mathbf{T}$ as a hidden variable, and will treat the dictionary representation codes $\mathbf{S}$ as a part of optimization on $\mathbf{X}$ and not as a variable in our system.

The function to be maximized is

$$\begin{aligned} P(\mathbf{y}|\mathbf{x}, r)P(\mathbf{x}) &= \int_{\mathbf{t}} P(\mathbf{y}, \mathbf{t}|\mathbf{x}, r)P(\mathbf{x})d\mathbf{t} \\ &= \int_{\mathbf{t}} P(\mathbf{y}|\mathbf{x}, \mathbf{t}, r)P(\mathbf{x})P(\mathbf{t})d\mathbf{t} \end{aligned} \tag{4.2}$$

The final term in Equation (4.2) has three components. The first one is the likelihood of the output being observed. The second component is prior on $\mathbf{X}$ which will be defined shortly, and the third one is the smoothness prior on $\mathbf{T}$ [Equation (3.21)].

Due to our modeling of the noise as i.i.d. Gaussian, the likelihood probability distribution turns out to be

$$P(\mathbf{y}|\mathbf{x}, \mathbf{t}, r) = \sum_i (y_i - w_i)^2 \tag{4.3}$$

$$w_i = t_i(1 - r_i)x_i + t_i r_i K_{spec} + (1 - t_i)K_{smoke} \tag{4.4}$$

For $\mathbf{X}$, we use the non-negative sparse coding prior [Equation (3.6)] and KS statistic prior [Equation (3.9)] for color. The prior distribution on $\mathbf{X}$ is

$$P(\mathbf{X}) = \frac{1}{Z} \exp\left(-\gamma_1 \sum_{i=1}^{I} \left(\min_{S_i} ||\mathbf{X}_i^m - \mathbf{D}S_i||_2^2 + \lambda||S_i||_1\right) - \gamma_2 J_{KDE}(\mathbf{X})\right) \tag{4.5}$$

where Z is the normalization constant, and γ_1, γ_2 are weights to adjust the effect of two components of the prior.

4.2.1 E step

Let $\mathbf{x}^n$ be the estimate of the uncorrupted image after n iterations. The Q function in the next iteration is

$$Q(\mathbf{x}; \mathbf{x}^n) = \mathbb{E}_{P(\mathbf{T}|\mathbf{y},\mathbf{x}^n)}[\log P(\mathbf{y}, \mathbf{T}, \mathbf{x}|\mathbf{r})] = \mathbb{E}_{P(\mathbf{T}|\mathbf{y},\mathbf{x}^n)}[\log P(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r})] + \log P(\mathbf{X}) \tag{4.6}$$

The expectation is analytically intractable to compute. Hence, we approximate the posterior probability of latent variable by using a variational factorization over each pixel.

$$P(\mathbf{T}|\mathbf{y}, \mathbf{x}^n) = \prod_{i=1}^{I} F_i(T_i|\mathbf{y}, \mathbf{x}^n) \tag{4.7}$$

where F_i is the factor for pixel i. $\{F_i\}$ are truncated Gaussians with support $[0, 1]$, means $\{mu_i\}$, standard deviations $\{\sigma_i\}$. We have to optimize the factorization before solving for the Q function. Equation (4.8) is solved pixel-wise for optimum factors $\{F_i^*\}$ until convergence

for all the factors. c is an additive constant which will be absorbed in normalization of factor distributions.

$$\log F_i^*(T_i) = \mathbb{E}_{\prod_{j\neq i} F_j(T_j|\mathbf{y},\mathbf{x}^n)} \left[\log P\left(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r}\right)\right] + c \tag{4.8}$$

We will now derive the parameters of optimum factorization. We will use z_i is as defined in Equation (3.1) to simplify the notation.

$$\begin{aligned}\log P\left(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r}\right) = &-\sum_i \left(y_i - T_i z_i - (1-T_i)K_{smoke}\right)^2 \\ &- \gamma_3 \sum_i \sum_{j\in\mathcal{N}_i^T} w_{ij}^T (T_i - T_j)^2\end{aligned} \tag{4.9}$$

To solve Equation (4.8), we replace $\{T_j|j\neq i\}$ with current optimum means $\{\mu_j\}$ and $\{T_j^2|j\neq i\}$ with $\{\mu_j^2+\sigma_j^2\}$ in Equation (4.9). This will result in a quadratic equation in T_i, from which the parameters (μ_i, σ_i) of Gaussian factor F_i can be solved for. The final solution is derived as

$$\bar{\mu}_i = \frac{\left(y_i - K_{smoke}\right)\left(z_i - K_{smoke}\right) + 2\gamma_3 \sum_{j\in\mathcal{N}_i^T} w_{ij}^T \mu_j}{\left(z_i - K_{smoke}\right)^2 + 2\gamma_3 \sum_{j\in\mathcal{N}_i^T} w_{ij}^T} \tag{4.10}$$

$$\bar{\sigma}_i = \frac{1}{\sqrt{2}} \left(\left(z_i - K_{smoke}\right)^2 + 2\gamma_3 \sum_{j\in\mathcal{N}_i^T} w_{ij}^T \right)^{-\frac{1}{2}} \tag{4.11}$$

$$\alpha_i = -\frac{\bar{\mu}_i}{\bar{\sigma}_i} \tag{4.12}$$

$$\beta_i = \frac{1-\bar{\mu}_i}{\bar{\sigma}_i} \tag{4.13}$$

$$\mu_i = \bar{\mu}_i + \bar{\sigma}_i \frac{\phi\left(\alpha_i\right) - \phi\left(\beta_i\right)}{\Phi\left(\beta_i\right) - \Phi\left(\alpha_i\right)} \tag{4.14}$$

$$\sigma_i = \bar{\sigma}_i^2 \left[1 + \frac{\alpha_i \phi\left(\alpha_i\right) - \beta_i \phi\left(\beta_i\right)}{\Phi\left(\beta_i\right) - \Phi\left(\alpha_i\right)} - \left(\frac{\phi\left(\alpha_i\right) - \phi\left(\beta_i\right)}{\Phi\left(\beta_i\right) - \Phi\left(\alpha_i\right)} \right)^2 \right] \tag{4.15}$$

where ϕ and Φ are PDF and CDF of standard normal distribution.

Once we have the optimum factorization, solving for the Q function becomes simple.

$$\mathbb{E}_{P(\mathbf{T}|\mathbf{y},\mathbf{x}^\mathbf{n})} \left[\log P\left(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r}\right)\right] = \mathbb{E}_{\prod_{i=1}^I F_i(T_i|\mathbf{y},\mathbf{x}^n)} \left[\log P\left(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r}\right)\right] \tag{4.16}$$

Equation (4.16) can be solved by making the following substitutions in Equation (4.9):

- μ_i for T_i

- $\mu_i^2 + \sigma_i^2$ for T_i^2
- $\mu_i \mu_j$ for $T_i T_j$

4.2.2 M step

In this step, we will maximize $Q(\mathbf{x}; \mathbf{x}^n)$ with respect to $\mathbf{x}$ to obtain a new estimate $\mathbf{x}^{n+1}$.

$$\mathbf{x}^{n+1} = \arg\max_x Q\left(\mathbf{x}; \mathbf{x}^{n+1}\right) \tag{4.17}$$

There is an inherent minimization with respect to dictionary codes S involved in the prior $P(\mathbf{X})$. Equation (4.17) is optimized alternatively for $\mathbf{X}$ and S until convergence. The algorithm for optimization over a variable is adaptive gradient descent.

4.3 VB-EM version 2

The main change in this version is that we will treat dictionary coefficients S as latent variable along with T.

The function to be maximized is

$$P(\mathbf{y}, \mathbf{x}|r)\, P(x) = \int_{\mathbf{t},S} P(\mathbf{y}, \mathbf{t}, \mathbf{x}, S|r)\, \mathrm{d}\mathbf{t}\, \mathrm{d}S \tag{4.18}$$

$$= \int_{\mathbf{t},S} P(\mathbf{y}|\mathbf{t}, \mathbf{x}, S, r)\, P(\mathbf{x}, S)\, P(\mathbf{t})\, \mathrm{d}\mathbf{t}\, \mathrm{d}S \tag{4.19}$$

$$= \int_{\mathbf{t},S} P(\mathbf{y}|\mathbf{t}, \mathbf{x}, r)\, P(\mathbf{x}, S)\, P(\mathbf{t})\, \mathrm{d}\mathbf{t}\, \mathrm{d}S \tag{4.20}$$

Equation (4.20) follows as $\mathbf{y}$ is independent of S, given $\mathbf{t}$ and $\mathbf{x}$ due to our image formation model.

For our initial attempt, we tried factorizing S keeping the L1 regularization for sparsity and non-negativity constraint. We ran into precision problems leading to numerical inaccuracy. We will use L2 regularization for sparsity and allow the coefficients to have negative values. Kernel density estimate based metric will be used as part of prior on $\mathbf{X}$ to preserve color. The joint

prior distribution on $\mathbf{X}$ and S is

$$\begin{aligned} P(\mathbf{X}, S) &= P(S|\mathbf{X})P(\mathbf{X}) \\ &= \frac{1}{Z}\exp\left(-\gamma_1 \sum_{i=1}^{I}\left(\min_{S_i} ||\mathbf{X}_i^m - \mathbf{D}S_i||_2^2 + \lambda||S_i||_2\right) - \gamma_2 J_{KDE}\left(\mathbf{X}\right)\right) \end{aligned} \tag{4.21}$$

where Z is the normalization constant, and γ_1, γ_2 are weights to adjust the effect of two components of the prior.

4.3.1 E step

Continuing the notation and the derivation process from the previous section, the Q function is

$$Q(\mathbf{x}; \mathbf{x^n}) = \mathbb{E}_{P(\mathbf{T}, S|\mathbf{x}^n, y, r)}\left[\log P\left(\mathbf{y}, \mathbf{T}, \mathbf{x}, \mathbf{S}|r\right)\right] \tag{4.22}$$

Our image formation model leads to the following simplification in the posterior distribution over latent variables and the probability of the joint data

$$P\left(\mathbf{T}, S|\mathbf{x}^n, \mathbf{y}, r\right) = P\left(\mathbf{T}|\mathbf{x}^n, \mathbf{y}, r\right) P\left(S|\mathbf{x}^n\right) \tag{4.23}$$

$$P\left(S|\mathbf{x}^n\right) = \prod_i P\left(S_i|\mathbf{x}^n\right) \tag{4.24}$$

$$P\left(\mathbf{y}, \mathbf{T}, \mathbf{x}, \mathbf{S}|r\right) = P\left(\mathbf{y}|\mathbf{T}, \mathbf{x}, \mathbf{S}, r\right) P\left(\mathbf{S}|\mathbf{x}\right) P\left(\mathbf{X}\right) P\left(\mathbf{T}\right) \tag{4.25}$$

where the factorization in Equation (4.24) is evident from Equation (4.21).

Using these simplifications, the Q function is rewritten in Equation (4.26). The two expectations are analytically intractable. The first term is solved in the exact same manner as VBEM1, using factorization on T per pixel. For the second term, we will assume a factorization on coefficient corresponding to each dictionary atoms for codes $\{S_i\}$.

$$Q(\mathbf{x}; \mathbf{x}^n) = \mathbb{E}_{P(\mathbf{T}|\mathbf{x}^n, \mathbf{y}, r)} \log\left[P\left(\mathbf{y}|\mathbf{T}, \mathbf{x}, \mathbf{S}, r\right) P\left(\mathbf{T}\right)\right] + \sum_i \mathbb{E}_{P(S_i|\mathbf{x}^n)}\left[\log P\left(S_i|\mathbf{X}\right)\right] + \log P\left(\mathbf{X}\right) \tag{4.26}$$

We will assume a Gaussian factorization for S_i. Equation (4.28) is solved sequentially over all j to obtain optimum factorization.

$$P\left(S_i|\mathbf{x}^n\right) \approx \prod_j G_{ij}(S_{ij}|\mathbf{X}) \tag{4.27}$$

$$\log G_{ij}^*\left(S_{ij}\right) = \mathbb{E}_{\prod_{k\neq j} G_{ik}(S_{ij}|\mathbf{x}^n)}\left[\log P\left(S_i|\mathbf{x}^n\right)\right] \tag{4.28}$$

The optimum parameters $\{\mu_{ij}, \sigma_{ij}\}$ of the factors are derived in the following text. We will first write down the probability distribution over S_i.

$$\log P\left(S_i|\mathbf{x}^n\right) = -0.5\gamma_1 || \left(\mathbf{x}^n\right)_i^m - \sum_k D_k S_{ik}||_2^2 - \gamma_1\lambda \sum_k S_{ik}^2 \tag{4.29}$$

To solve Equation (4.28), we substitute $\{S_{ik}|k \neq j\}$ with $\{\mu_{ij}\}$ and $\{S_{ik}^2|k \neq j\}$ with $\{\mu_{ik}^2 + \sigma_{ik}^2\}$ in Equation (4.29). We will obtain the optimum parameters for G_{ij} by the following equations.

$$\mu_{ij} = \frac{D_j^\intercal\left(\left(\mathbf{x}^n\right)_i^m - \sum_{k\neq j} D_k S_{ik}\right)}{D_j^\intercal D_j + 2\lambda} \tag{4.30}$$

$$\sigma_{ij} = \frac{1}{\sqrt{D_j^\intercal D_j + 2\lambda}} \tag{4.31}$$

We will solve for optimum parameters of factors sequentially until convergence. Then 4.26 will be solved by substitution of $\{\mu_{ij}\}$ for S_{ij} in addition to substitutions for $\mathbf{T}$ detailed in VBEM1.

The M step of VBEM2 will be the same as in VBEM1. We will solve an optimization problem over $\mathbf{X}$.

Chapter 5

Results

In this chapter, we will compare our methods with the existing literature. To the best of our knowledge, there is no prior work on joint removal of smoke, speckle, and noise in laparoscopy images. We will combine multiple methods which solves the subproblems and combine them for end-to-end comparison. We will use anisotropic diffusion for inpainting which will preserve texture better than anisotropic diffusion used in [5, 10]. The methods used for comparision are

1. Desmoking and denoising with Kotwal ISBI16 [2], followed by anisotropic diffusion for inpainting.

2. Desmoking and denoising with adaptive wiener filtering by Gibson ICIP13 [1], followed by anisotropic diffusion inpainting.

3. Noise removal with edge preserving bilateral filtering, followed by smoke removal with He PAMI11 [3] and anisotropic diffusion inpainting.

5.1 Experiment details

Proposed as well as competing methods are tuned for best performance at 3 percent noise level. For synthetic corruption, we use 3 transmission coefficient maps and 20 specularity maps. We add i.i.d. Gaussian noise of standard deviation ranging from 0 % to 7%. We get 3 times 6 transmission coefficient maps using 6 scalar multiples or smoke levels. We will first compare relative root mean squared error (RRMSE) by synthetically corrupting simulated data as well as high quality laparoscopy data. After that, we will get the methods evaluated by clinical experts for real world observed laparoscopy images.

5.2 Synthetic corruption on simulated data

This experiment provides proof on concept of our method. The simulated data is designed to provide basis of evaluation in terms of image texture and color.

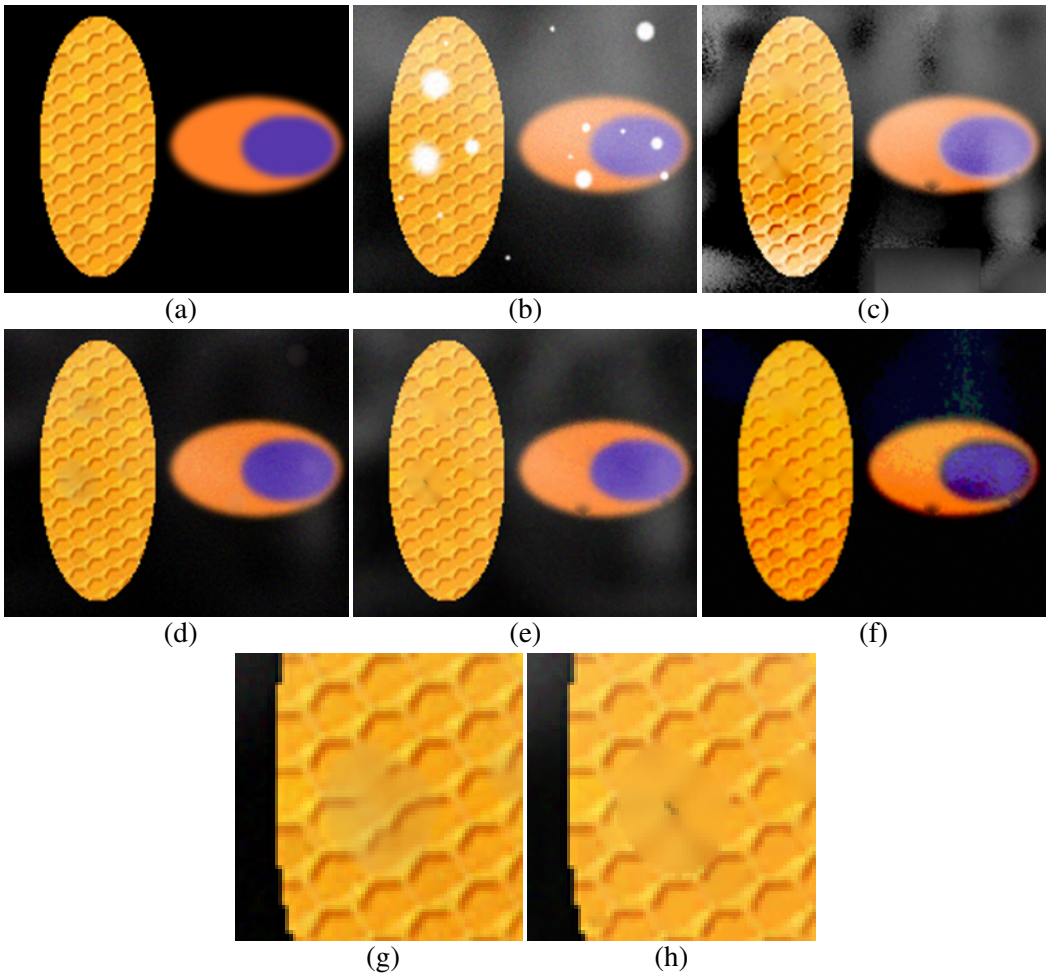
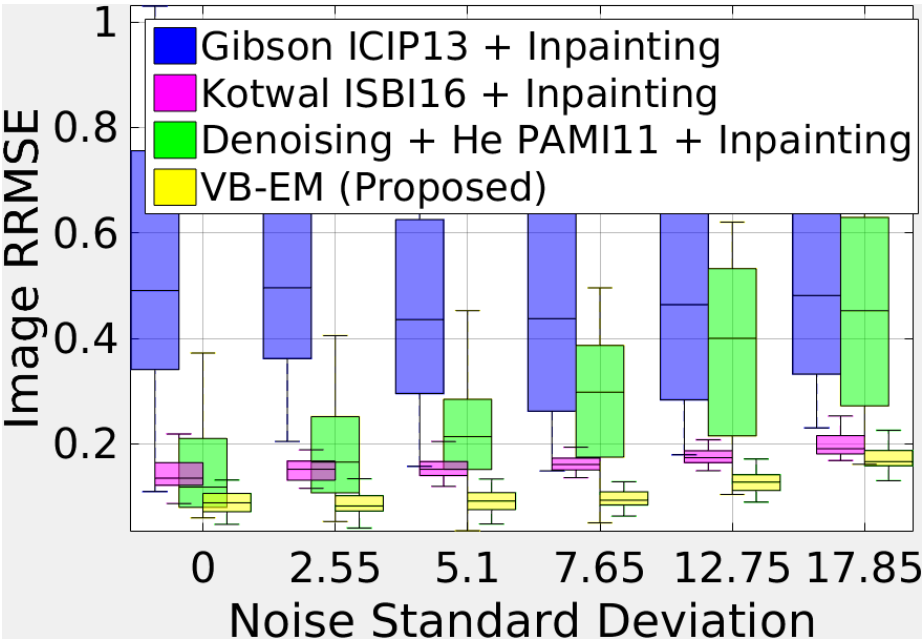
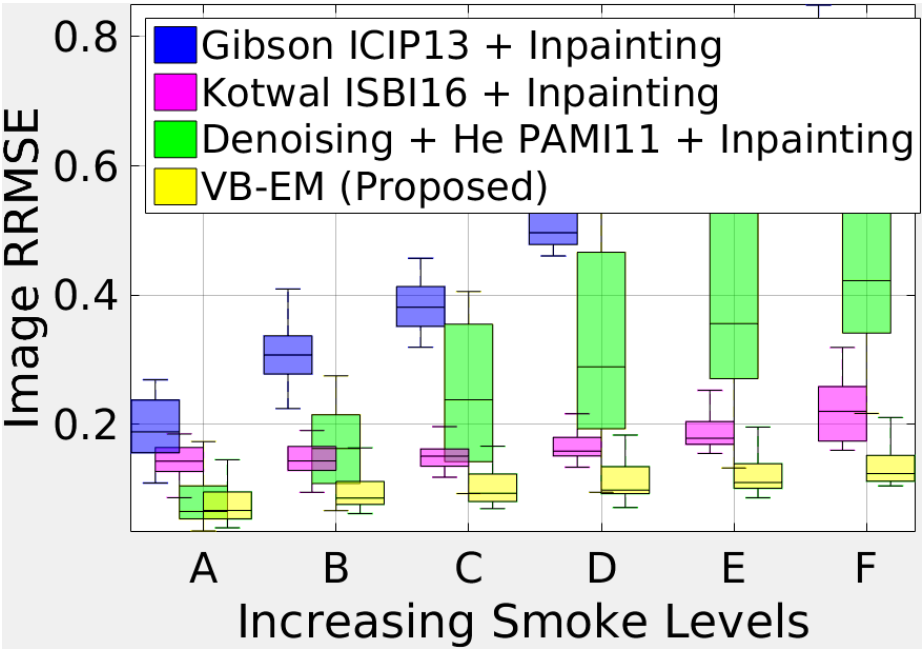


Figure 5.1: **Qualitative Validation on Simulated Data.** **(a)** Phantom (color component values $\in [0, 255]$). **(b)** Corrupted phantom with smoke, specularities, and low noise (σ = 5). Results of processing image (b), using: **(c)** adaptive filtering [1] followed by inpainting; **(d)** *proposed method VBEM1*; **(e)** denoising and desmoking [2] followed by inpainting; **(f)** bilateral filter for denoising followed by dehazing [3] followed by inpainting. Zoomed sections: **(g)** of (d); **(h)** of (e).

An example of processing on synthetically corrupted phantom is Figure 5.1. Gibson ICIP13 [1] plus inpaiting does a poor job at removing smoke. Bilateral filtering, He PAMI11 [3] plus inpainting does a good smoke at desmoking, but the inpainting performance is poor and is displayed in zoomed sections (g) and (h). Our dictionary prior does a better job of filling in texture. Kotwal ISBI16 plus inpainting produce unnatural colors. Proposed method VBEM1 has the best removal of smoke, better texture and colors.



(a)



(b)

Figure 5.2: **Quantitative Validation on Simulated Data.** Box plots for RRMSE for different combinations of smoke levels and noise levels. Each combination is run 50 times. **(a)** grouped by noise level $\in [0, 255]$; **(b)** grouped by smoke level.

We perform quantitative benchmarking using RRMSE values for different methods. The results are presented in Figure 5.2. For the first plot, our method has the small spreads and the lowest means at all noise levels. The robustness of our algorithm is demostrated at higher noise levels. For the second part, we have lower RRMSE values, and the robustness is more pronounced in this plot.

5.3 High quality laparoscopy data and synthetic corruption

We will now perform validation on laparoscopy data. We will take high quality laparoscopy images, corrupt them synthetically, and then process using different algorithms. We will then compare the outputs with the ground truth.

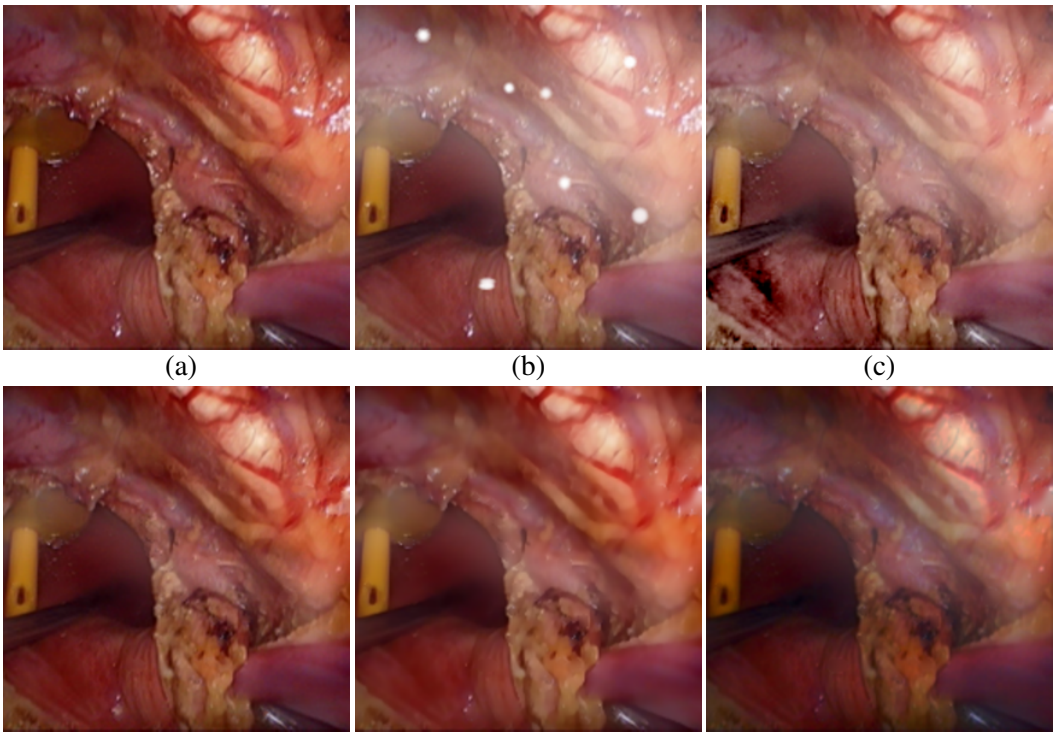


Figure 5.3: **Qualitative Validation on Simulated Data.** **(a)** Ground truth (color component values $\in [0, 255]$). **(b)** Corrupted phantom with smoke, specular highlights, and low noise (σ = 5). Results of processing image (b), using: **(c)** adaptive filtering [1] followed by inpainting; **(d)** *proposed method VBEM1*; **(e)** denoising and desmoking [2] followed by inpainting; **(f)** bilateral filter for denoising followed by dehazing [3] followed by inpainting.

Figure 5.3 shows the results of processing on laparoscopy data. The observations are similar to those for simulated data. Gibson ICIP13 [1] plus inpaiting does a poor job at removing smoke. Bilateral filtering, He PAMI11 [3] plus inpainting does a good smoke at desmoking, but the results has loss of edges and texture. Kotwal ISBI16 plus inpainting produce unnatural colors, particularly in the central regions. Proposed method VBEM1 has the best removal of smoke, better texture and colors. These observations are more clear in the zoomed sections in Figure 5.4

Quantitative evaluation using RRMSE is presented in Figure 5.5. In the first plot, the proposed method has better median RRMSE at all but one noise level. For the second part, we have lower RRMSE values, and robustness at high smoke levels in this plot.

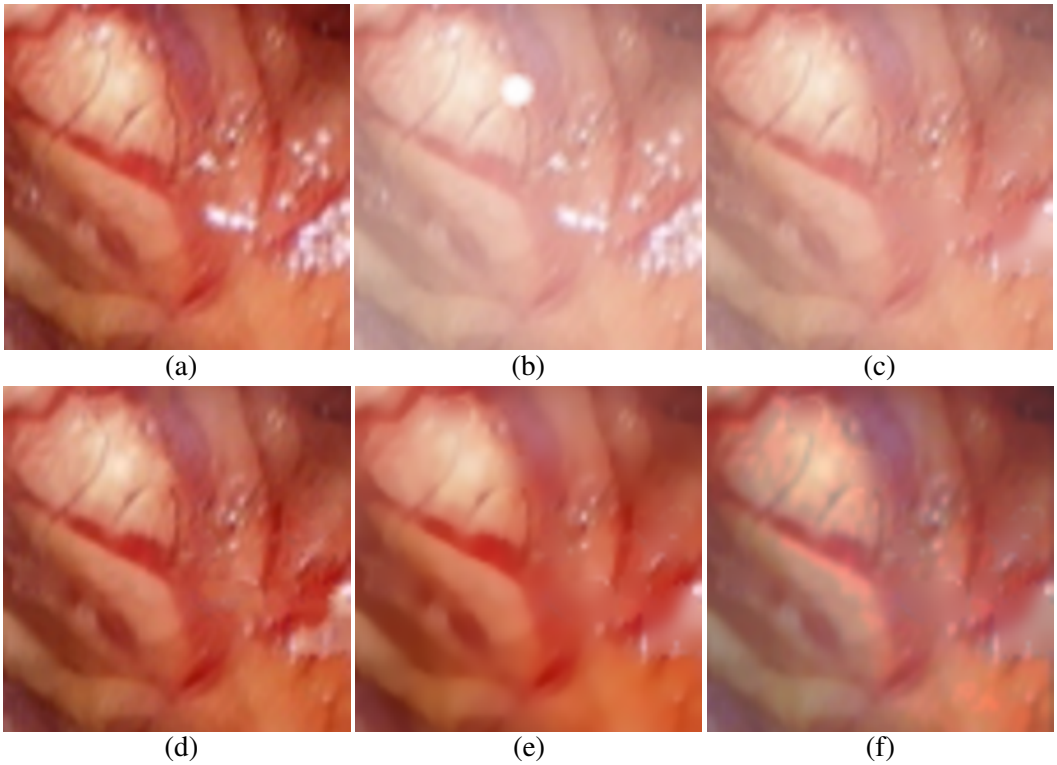


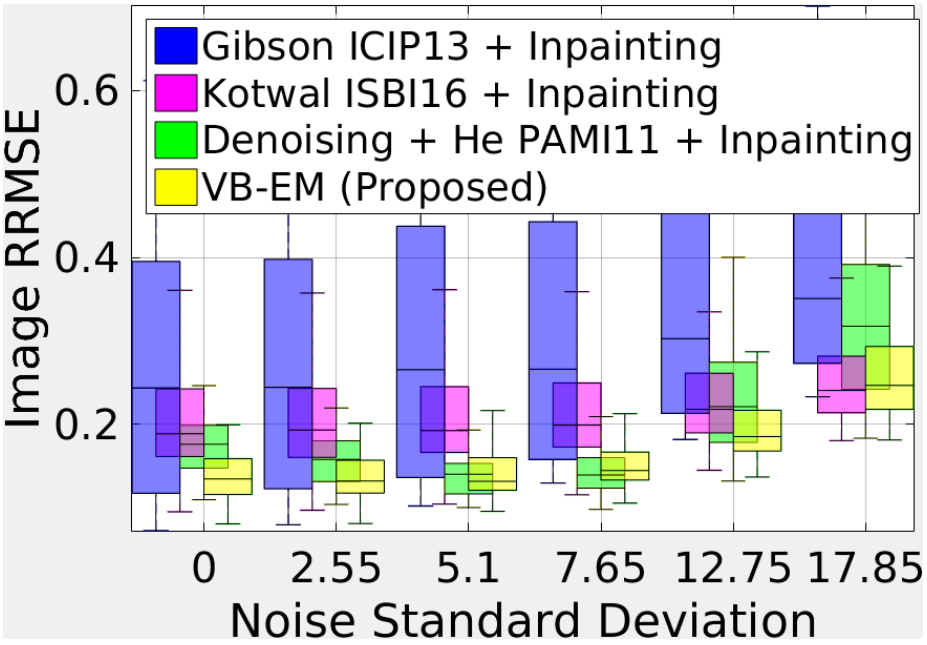
Figure 5.4: **Qualitative Validation on Simulated Data. The images (a) to (f) are zoomed in subparts of the corresponding images in Figure 5.3**

5.4 Clinical validation

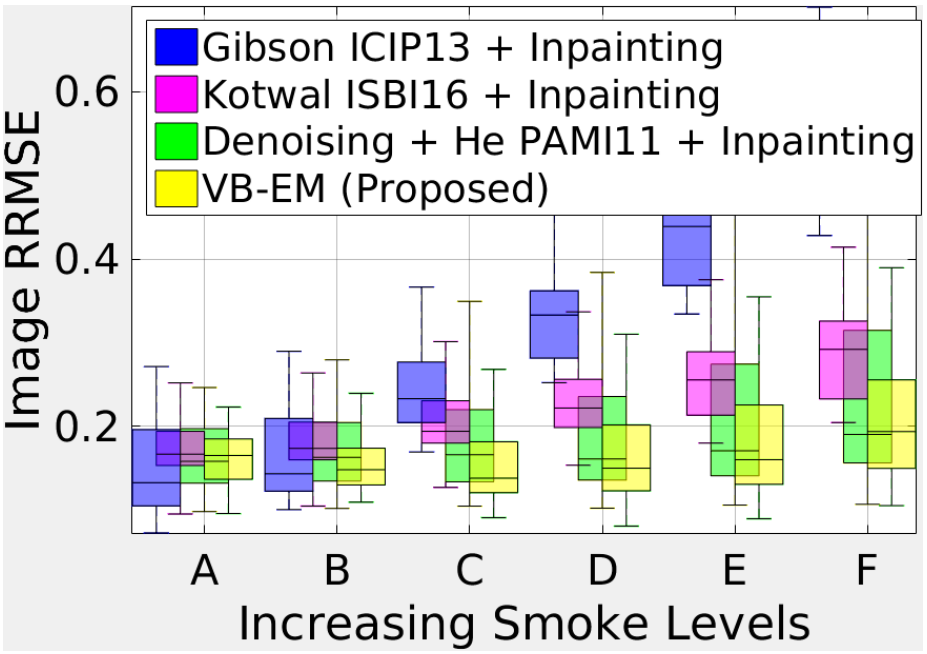
We asked 4 doctors to perform blind evaluation. The rating scale was 1 (bad), 2 (average), 3 (good), 4 (excellent). Same rating can be assigned for multiple methods.

Method	Median	Mean	Standard deviation
Proposed VBEM1	4	3.9	0.3
Gibson ICIP13 + Inpainting	1	1.5	0.8
Kotwal ISBI16 + Inpainting	3	2.9	0.4
Bilat. filtering + He PAMI11 + Inpainting	2	2.2	0.5

Table 5.1: **Clinical ratings**



(a)



(b)

Figure 5.5: **Quantitative Validation on High Quality Laparoscopy Data.** Box plots for RRMSE for different combinations of smoke levels and noise levels. Each combination is run for 24 images **(a)** grouped by noise level $\in [0, 255]$; **(b)** grouped by smoke level.

Appendix A

Expectation Maximization

Bibliography

[1] K. B. Gibson and T. Q. Nguyen, "Fast single image fog removal using the adaptive wiener filter," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 714–718.

[2] A. Kotwal, R. Bhalodia, and S. P. Awate, "Joint desmoking and denoising of laparoscopy images," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 1050–1054.

[3] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[4] D. Stoyanov and G. Z. Yang, "Removing specular reflection components for robotic assisted laparoscopic surgery," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 3. IEEE, 2005, pp. III–632.

[5] C.-A. Saint-Pierre, J. Boisvert, G. Grimard, and F. Cheriet, "Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images," *Machine Vision and Applications*, vol. 22, no. 1, pp. 171–180, 2011.

[6] W. L. Barrett and S. M. Garber, "Surgical smoke: a review of the literature," *Surgical endoscopy*, vol. 17, no. 6, pp. 979–987, 2003.

[7] K. Prokopetc, T. Collins, and A. Bartoli, "Automatic detection of the uterus and fallopian tube junctions in laparoscopic images," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 552–563.

[8] S. Voros, J.-A. Long, and P. Cinquin, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders," *The International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1173–1190, 2007.

[9] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros, "3d tracking of laparoscopic instruments using statistical and geometric modeling," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2011, pp. 203–210.

[10] M. Arnold, A. Ghosh, S. Ameling, and G. Lacey, "Automatic segmentation and inpainting of specular highlights for endoscopic imaging," *EURASIP Journal on Image and Video Processing*, vol. 2010, no. 1, p. 814319, 2010.

[11] H. Koschmieder, *Theorie der horizontalen sichtweite: kontrast und sichtweite*. Keim & Nemnich, 1925.

[12] R. Fattal, "Single image dehazing," *ACM transactions on graphics (TOG)*, vol. 27, no. 3, p. 72, 2008.

[13] R. T. Tan, "Visibility in bad weather from a single image," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[14] J. Pang, O. C. Au, and Z. Guo, "Improved single image dehazing using guided filter," *Proc. APSIPA ASC*, pp. 1–4, 2011.

[15] N. Joshi and M. F. Cohen, "Seeing mt. rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal," in *Computational Photography (ICCP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–8.

[16] E. Matlin and P. Milanfar, "Removal of haze and noise from a single image." in *Computational Imaging*, 2012, p. 82960T.

[17] G. F. Harpur and R. W. Prager, "Development of low entropy coding in a recurrent network," *Network: computation in neural systems*, vol. 7, no. 2, pp. 277–284, 1996.

[18] P. O. Hoyer, "Non-negative sparse coding," in *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*. IEEE, 2002, pp. 557–565.

[19] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 689–696.

[20] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[21] A. Baid, A. Kotwal, R. Bhalodia, S. Merchant, and S. P. Awate, "Joint desmoking, specularity removal, and denoising of laparoscopy images via graphical models and bayesian inference," in *Biomedical Imaging (ISBI), 2017 IEEE 14th International Symposium on*. IEEE, 2017.

List of Publications

- A. Baid, A. Kotwal, R. Bhalodia, S. Merchant, and S. P. Awate, "Joint desmoking, specularity removal, and denoising of laparoscopy images via graphical models and bayesian inference," in *Biomedical Imaging (ISBI), 2017 IEEE 14th International Symposium on*. IEEE, 2017

Acknowledgements

I would like to thank my classmates Alankar Kotwal, Riddhish Bhalodia, and Anand Pathak. Alankar and Riddhish laid down the foundation of this work. Alankar also helped me with debugging and tuning the code, and also actively participated in discussion and planning.

Ayush Baid

IIT Bombay

15 May 2017