

# 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

24 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.

Date: 24 May 2017

---

Ayush Baid  
(12D100002)





# Abstract

Laparoscopy images exhibit artifacts like occlusion from surgical smoke, specular highlights, and noise. These artifacts hinder visibility, and degrade post processing (e.g. segmentation). There is a lack of literature on smoke removal in laparoscopy images, and to the best of our knowledge, no prior work on jointly tackling these three artifacts. We tackle these degradations as a novel *unified Bayesian inference problem*. We use *probabilistic graphical models* and employ *sparse dictionary models* and novel *image intensity distribution model* as image priors. We obtain maximum-a-priori probability (MAP) estimate by **expectation maximization** and use variational Bayesian factorization to overcome the encountered intractabilities. Results on simulated and real-world laparoscopy images show that our joint optimization strategy outperforms the state-of-the-art.

***Index terms*** — Laparoscopy, smoke removal, specular highlights removal, denoising, variational Bayes, EM, graphical models



# Table of Contents

<b>Abstract</b>	<b>ix</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Survey</b>	<b>3</b>
2.1 Speckle removal in laparoscopy images . . . . .	3
2.2 Dehazing . . . . .	4
2.3 Simultaneous dehazing and denoising . . . . .	5
2.4 Desmoking . . . . .	5
<b>3 Formulation</b>	<b>7</b>
3.1 Image formation . . . . .	7
3.2 Sparse coding under a dictionary . . . . .	8
3.3 Image intensity distribution . . . . .	9
3.4 Kernel density estimation . . . . .	10
3.5 Prior on image $\mathbf{X}$ . . . . .	11
3.6 Prior on transmission map $\mathbf{T}$ . . . . .	12
<b>4 Estimation</b>	<b>13</b>
4.1 Variational Bayesian Expectation Maximization . . . . .	13
4.1.1 Expectation step . . . . .	14
4.1.2 M step . . . . .	15

<b>5 Results</b>	<b>17</b>
5.1 Experiment details . . . . .	17
5.2 Synthetic corruption on simulated data . . . . .	18
5.3 High quality laparoscopy data and synthetic corruption . . . . .	19
5.4 Clinical validation . . . . .	20
<b>A Optimal factors for transmission map <math>T</math></b>	<b>23</b>
<b>B Optimal factors for codes <math>S</math></b>	<b>25</b>
<b>List of Publications</b>	<b>31</b>
<b>Acknowledgements</b>	<b>33</b>

# List of Figures

3.1	<b>Learning Prior PDFs on Color.</b> Empirical histograms (bar plots) and fitted parametric PDFs (solid curves) in uncorrupted laparoscopic images, for 3 channel components: <b>(a)</b> gamma $\Gamma_1$ , <b>(b)</b> Gaussian $G_2$ , <b>(c)</b> Gaussian $G_3$ . . . . .	10
5.1	<b>Qualitative Validation on Simulated Data.</b> <b>(a)</b> Phantom (color component values $\in [0, 255]$ ). <b>(b)</b> Corrupted phantom with smoke, specularities, and low noise ( $\sigma = 5$ ). Results of processing image (b), using: <b>(c)</b> adaptive filtering [1] followed by inpainting; <b>(d)</b> <i>proposed method VBEM1</i> ; <b>(e)</b> denoising and desmoking [2] followed by inpainting; <b>(f)</b> bilateral filter for denoising followed by dehazing [3] followed by inpainting. Zoomed sections: <b>(g)</b> of (d); <b>(h)</b> of (e). . . . .	18
5.2	<b>Quantitative Validation on Simulated Data.</b> Box plots for RRMSE <b>(a, b)</b> and SSIM <b>(c, d)</b> for different combinations of smoke levels and noise levels. Each combination is run 50 times. <b>(a), (c)</b> grouped by noise level $\in [0, 255]$ ; <b>(b), (d)</b> grouped by smoke level. . . . .	19
5.3	<b>Qualitative Validation on Simulated Data.</b> <b>(a)</b> Ground truth (color component values $\in [0, 255]$ ). <b>(b)</b> Corrupted phantom with smoke, specular highlights, and low noise ( $\sigma = 5$ ). Results of processing image (b), using: <b>(c)</b> adaptive filtering [1] followed by inpainting; <b>(d)</b> <i>proposed method VBEM1</i> ; <b>(e)</b> denoising and desmoking [2] followed by inpainting; <b>(f)</b> bilateral filter for denoising followed by dehazing [3] followed by inpainting. . . . .	20
5.4	<b>Qualitative Validation on Simulated Data.</b> The images <b>(a)</b> to <b>(f)</b> are zoomed in subparts of the corresponding images in Figure 5.3 . . . . .	21

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. . . . . 22

# List of Tables

<b>5.1 Clinical ratings</b> . . . . .	21
---------------------------------------	----





# Chapter 1

## Introduction

Laparoscopy is a popular *minimally invasive surgery* technique in which operations are performed by inserting equipments through small incisions. Laparoscopy surgery has 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.

Chapter 2 covers the related work in this area. Recent work in specular highlight removal in laparoscopy use anisotropic diffusion and does not preserve texture and edges. There is lack of literature for smoke removal in laparoscopy images, and hence we look into haze removal in outdoor images. These methods lack some properties of laparoscopy images which can be imposed due to the smaller variation in laparoscopy images than outdoor images.

In Chapter 3, we will introduce the image formation and model the variables in the system using priors on probabilistic graphical models. We formulate a unified Bayesian inference problem in Chapter 4, which is solved using expectation-maximization (EM) algorithm. We introduce variational Bayesian approximation to overcome the analytical intractability in the optimization scheme. We will present two versions of our algorithm, VBEM1 and VBEM2, where the second version is to make our method more sound in mathematical sense.

We present the validation metric and results of our algorithm compared to the state-of-the-art in Chapter 5. We conclude and discuss some future directions in Chapter 6.

# 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 ubiquitous in literature. Equation (2.1) captures the effect of haze.

$$X(i) = T(i)J(i) + (1 - T(i))A \quad (2.1)$$

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

In this chapter, we will define the variables in the system and present the image formation model in Section 3.1. We use Markov random fields (MRFs) to model the variables and will introduce the prior functions in the following sections. We will then conclude this chapter by discussing the MRF structure and the final prior distribution for the variables in Section ?? and Section ??.

Our notation scheme is to use upper case for random variables and lower case for their specific instance. Bold face will denote a set of random variables over the whole image, one for each pixel.  $\mathbf{X}$  is an uncorrupted image, which we want to estimate.  $\mathbf{Y}$  is the observed image,  $\mathbf{T}$  is the smoke transmission map, and  $\mathbf{R}$  is the speckle map.  $K_{spec}$  and  $K_{smoke}$  are speckle and smoke color respectively.

We have two variables to model,  $\mathbf{X}$  and  $\mathbf{T}$ . For  $I$  pixels in the image, the MRFs are denoted by  $\mathbf{X} := \{X_i\}_{i=1}^I$  and  $\mathbf{T} := \{T_i\}_{i=1}^I$ . Each  $X_i$  is a vector denoting value of pixel  $i$  in  $RGB$  space and  $T_i$  is a scalar in  $[0, 1]$ . We want our estimate of the underlying true image to have sharp texture and edges, and natural colors sans the smoke. We also want the transmission map to have spatial regularity.

### 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  $\mathbf{R}$ , where  $R_i$  having a value 1 denotes the presence of speckle. Equation (3.2) captures the smoke using smoke map  $\mathbf{T}$ ,  $T_i \in [0, 1]$  and an i.i.d white Gaussian noise  $n_i$  is added at each pixel.  $\mathbf{Y}$  is captured and available for processing.

$$Z_i = (1 - R_i)X_i + R_iK_{spec} \quad (3.1)$$

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

### 3.2 Sparse coding under a dictionary

Given a data vector  $X$ , and a dictionary  $\mathbf{D}$ , a linear decomposition in terms of codes  $S$  is of the form

$$X \approx \mathbf{D}S \quad (3.3)$$

In sparse coding [17], we want to represent an input vector to be using as few atoms of the dictionary as possible. [18] demonstrated that the dictionary atoms learnt via sparse coding bears similarity to preprocessing by visual cortex by learning features which are localized, oriented, and bandpass. A common way to enforce sparsity is to regularize the codes [ Equation (3.4) ], where  $\lambda$  controls the trade-off between sparsity and reconstruction accuracy.

$$J(\mathbf{D}, S) = \frac{1}{2} \|X - \mathbf{D}S\|_2^2 + \lambda \sum_j f(S_j) \quad (3.4)$$

The sparsity function  $f$  typically is strictly increasing function of the absolute value of its argument.

Sparse coding using L1 regularization, i.e.  $f(S_j) = |S_j|$ , was used to perform digit recognition in [19]. Non-negative sparse coding [20] constrains the dictionary and codes to be non-negative. The choice of sparsity function is  $f(S_j) = |S_j| = S_j$ . [21] proposed an online algorithm for dictionary learning in sparse coding framework using L1 regularization. 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 datasets.

Given a dictionary  $\mathbf{D}$  and input vector  $X$ , the weighted least square optimization problem in Equation (3.5), where  $\Gamma$  is a diagonal matrix containing the weight for each coefficient. Given



a fixed  $\Gamma$ , the optimum code is given by Equation (3.6).

$$S^* = \arg \min_S \frac{1}{2} \|X - \mathbf{D}S\|_2^2 + \lambda \|\Gamma S\|_2^2 \quad (3.5)$$

$$= (\mathbf{D}^\top \mathbf{D} + 2\lambda \Gamma^\top \Gamma)^{-1} \mathbf{D}^\top X \quad (3.6)$$

Iterative reweighted least squares algorithm is an iterative algorithm, where in each step the weighted L2 regularized optimization problem is solved. It can be used to solve L1 regularized least square problem where the weights are chosen to be inverse of the absolute value of the current coefficients. If  $S^n$  is the estimate before iteration  $n$ , then the optimization objective will be

$$S^{n+1} = \arg \min_S \frac{1}{2} \|X - \mathbf{D}S\|_2^2 + \lambda \sum_j \frac{1}{|S_j^n|} (S_j)^2 \quad (3.7)$$

This iterative method of fixing the weights and solving a weighted L2 problem runs until convergence.

### 3.3 Image intensity distribution

Surgical smoke, generally gray, perturbs the color in the image. We need some model distribution of colors in uncorrupted high quality laparoscopy 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 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

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.3568 & 0.8413 & 0.5304 \\ 0.0760 & -0.2006 & 0.1239 \\ 0.2267 & 0.3574 & -0.6512 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.8)$$

The correlation between channels in *RGB* space is  $[0.96, 0.99, .94]$  and that in  $l\alpha\beta$  space is  $[0.62, 0.63, 0.81]$ . The new space has a clear advantage when we model the intensity distribution

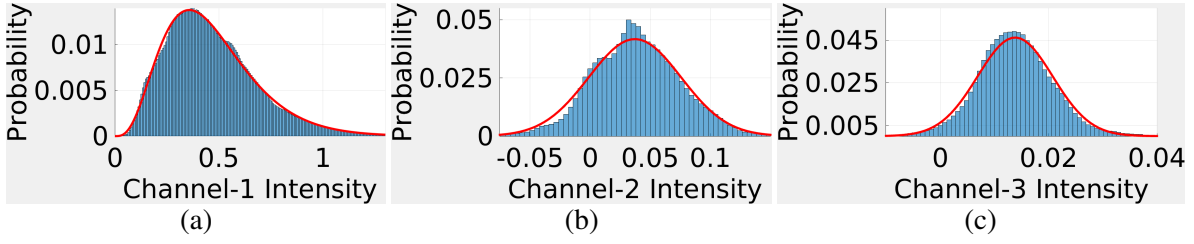


Figure 3.1: **Learning Prior PDFs on Color.** Empirical histograms (bar plots) and fitted parametric PDFs (solid curves) in uncorrupted laparoscopic images, for 3 channel components: (a) gamma  $\Gamma_1$ , (b) Gaussian  $G_2$ , (c) Gaussian  $G_3$ .

channel-wise and independent of other channels. We use a Gamma distribution to fit the  $l$  channel and Gaussian distributions to fit  $\alpha$  and  $\beta$  channels respectively. The empirical distributions and the fits are illustrated in Figure 3.1

### 3.4 Kernel density estimation

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

$$f^{\mathbf{x}}(b) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{b - x_i}{h}\right) \quad (3.9)$$

where  $K(\cdot)$  is a kernel which takes non-negative values, integrates to one, and has mean zero.  $h > 0$  is the bandwidth and controls the trade-off between bias and variance of the estimator. We will use the Gaussian kernel due to its mathematical properties like differentiability. Due to our choice of Gaussian kernel, we can use the rule of thumb estimate [22] for bandwidth using the sample standard deviation  $\hat{\sigma}$ . Equation (3.10) is used to tune the bandwidth using training data.

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \quad (3.10)$$

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

$$f^{\mathbf{x}}(b) = \frac{1}{nh\sqrt{2\pi\sigma^2}} \sum_{i=1}^n \exp\left(-\frac{(b-x_i)^2}{2\sigma^2h^2}\right) \quad (3.11)$$

$$F^{\mathbf{x}}(b) = \sum_{s=x_{min}}^b f^{\mathbf{x}}(s) \quad (3.12)$$

$$\frac{\partial f^{\mathbf{x}}(b)}{\partial x_i} = \frac{1}{nh\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(b-x_i)^2}{2\sigma^2h^2}\right) \frac{(b-x_i)}{\sigma^2h^2} \quad (3.13)$$

$$\frac{\partial F^{\mathbf{x}}(b)}{\partial x_i} = \sum_{s=x_{min}}^b \frac{\partial f_h^{\mathbf{x}}(s)}{\partial x_i} \quad (3.14)$$

We will now introduce the penalty function  $J_{dist}$  in Equation (3.15) and its gradient, where  $H(\cdot)$  evaluates the transformation of a point under CDF matching of the empirical distribution with the reference [Equation (3.17)]. The choice of the penalty function was made as it exhibits better convergence when gradient descent is used for optimization.

$$J_{dist}(\mathbf{x}) = \sum_{b \in B} \{b - H(b; F^{\mathbf{x}}, F^{ref})\}^2 \quad (3.15)$$

$$\frac{\partial J_{dist}(\mathbf{x})}{\partial x_i} = 2 \sum_{b \in B} \{b - H(b; F^{\mathbf{x}}, F^{ref})\} \frac{\partial H(b; F^{\mathbf{x}}, F^{ref})}{\partial x_i} \quad (3.16)$$

$$H(b; F^{\mathbf{x}}, F^{ref}) = F_{ref}^{-1}(F^{\mathbf{x}}(b)) \quad (3.17)$$

$$\frac{\partial H(b; F^{\mathbf{x}}, F^{ref})}{\partial x_i} = F_{ref}^{-1}(F^{\mathbf{x}}(b)) \frac{\partial F^{\mathbf{x}}(b; x_1^n)}{\partial x_i} \quad (3.18)$$

For use in laparoscopy data, we will add up the penalty for each channel  $c_i$  independently.

$$J_{KDE}(\mathbf{x}) = \sum_{i=1}^3 \sum_{b \in B_{c_i}} \{b - H(b; F_{c_i}^{\mathbf{x}}, F_{c_i}^{ref})\}^2 \quad (3.19)$$

## 3.5 Prior on image $\mathbf{X}$

We will define the prior on  $\mathbf{X}$  conditional on the sparse code  $\mathbf{S}$ . The MRF  $\mathbf{X} := \{X_i\}_{i=1}^I$  has a fully connected neighborhood system. We use the sparse coding prior to preserve texture and remove noise, and KDE prior to preserve image intensity statistics. We define potentials on each square clique of size  $m \times m$ , where  $m$  is odd. For KDE prior, we define potential on clique

containing all the pixel locations. The probability distribution for  $\mathbf{X}$  is

$$P(\mathbf{X}|\mathbf{S}) = P(\mathbf{S}|\mathbf{X}) P(\mathbf{X}) = \frac{1}{Z} \exp \left( -\gamma_1 \sum_{i=1}^I \left( \|\mathbf{X}_i^m - \mathbf{D}S_i\|_2^2 + \lambda \sum_j w_{ij} S_{ij}^2 \right) - \gamma_2 J_{KDE}(\mathbf{X}) \right) \quad (3.20)$$

where  $\mathbf{X}_i^m$  is an  $m \times m$  patch centered at pixel  $i$ , and its code under the prelearned dictionary is  $S_i$ .  $\gamma_1, \gamma_2$  are free parameters to adjust the weights.  $Z$  is the normalization constant.

### 3.6 Prior on transmission map $\mathbf{T}$

To enforce spatial smoothness, 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 pixels in  $5 \times 5$  patch centered at pixel  $i$ . The prior distribution is defined using potentials on all the cliques as follows

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

where the outer sum is performed over all pixels,  $\gamma_3 \in \mathbb{R}^+$  is a free parameter, and  $Z$  is the normalization constant.

# Chapter 4

## Estimation

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

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

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 Variational Bayesian Expectation Maximization

We will introduce smoke transmission map  $\mathbf{T}$  and dictionary codes  $\mathbf{S}$  as latent variables in the system. The aposterior probability is

$$P(\mathbf{y}|\mathbf{x}, \mathbf{r}) P(\mathbf{x}) = \int_{\mathbf{t}, \mathbf{s}} P(\mathbf{y}, \mathbf{t}, \mathbf{x}, \mathbf{s}|\mathbf{r}) d\mathbf{t} d\mathbf{s} \quad (4.2)$$

$$= \int_{\mathbf{t}, \mathbf{s}} P(\mathbf{y}|\mathbf{t}, \mathbf{x}, \mathbf{s}, \mathbf{r}) P(\mathbf{x}, \mathbf{s}) P(\mathbf{t}) d\mathbf{t} d\mathbf{s} \quad (4.3)$$

$$= \int_{\mathbf{t}, \mathbf{s}} P(\mathbf{y}|\mathbf{t}, \mathbf{x}, \mathbf{r}) P(\mathbf{x}, \mathbf{s}) P(\mathbf{t}) d\mathbf{t} d\mathbf{s} \quad (4.4)$$

Equation (4.4) follows as  $\mathbf{Y}$  is independent of  $\mathbf{S}$ , given  $\mathbf{T}$  and  $\mathbf{X}$  due to our image formation model. The final term in Equation (4.4) has three components. The first one is the likelihood of the output being observed. The second and third component are priors on  $\mathbf{X}$ ,  $\mathbf{S}$  and  $\mathbf{T}$ .

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

$$\log P(\mathbf{y}|\mathbf{x}, \mathbf{t}, \mathbf{r}) = - \sum_{i|r_i=0} (y_i - t_i x_i - (1 - t_i) K_{smoke})^2 \quad (4.5)$$

#### 4.1.1 Expectation step

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

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

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

$$P(\mathbf{T}, \mathbf{S}|\mathbf{y}, \mathbf{x}^n, \mathbf{r}) = P(\mathbf{T}|\mathbf{y}, \mathbf{x}^n, \mathbf{r}) P(\mathbf{S}|\mathbf{x}^n) \quad (4.7)$$

$$P(\mathbf{S}|\mathbf{x}^n) = \prod_i P(S_i|\mathbf{x}^n) \quad (4.8)$$

$$P(\mathbf{y}, \mathbf{x}, \mathbf{T}, \mathbf{S}|\mathbf{r}) = P(\mathbf{y}|\mathbf{x}, \mathbf{T}, \mathbf{r}) P(\mathbf{S}|\mathbf{x}) P(\mathbf{X}) P(\mathbf{T}) \quad (4.9)$$

where the factorization in Equation (4.8) is evident from Equation (3.20).

The expectation is analytically intractable to compute. We use variational factorization for each component in Equation (4.7). We assume pixel by pixel factorization for  $\mathbf{T}$  and over each atom for  $\{S_i\}$ .

$$P(\mathbf{T}|\mathbf{y}, \mathbf{x}^n) \approx \prod_{i=1}^I F_i(T_i|\mathbf{y}, \mathbf{x}^n) \quad (4.10)$$

$$P(S_i|\mathbf{x}^n) \approx \prod_j G_{ij}(S_{ij}|\mathbf{x}^n) \quad (4.11)$$

where  $F_i$  is the factor of  $\mathbf{T}$  at pixel  $i$ , and  $G_{ij}$  is the factor corresponding to  $j^{th}$  coefficient of code vector  $S_i$ .  $\{F_i\}$  are truncated Gaussians with support  $[0, 1]$ , means  $\{\mu_i^T\}$ , standard deviations  $\{\sigma_i^T\}$ .  $\{G_{ij}\}$  are Gaussian distributions with means  $\{\mu_{ij}^S\}$  and standard deviation  $\{\sigma_{ij}^S\}$ . We have to optimize the factorization before solving for the  $Q$  function. Equation (4.12) and Equation (4.13) are solved iteratively for optimum factors  $\{F_i^*\}$  and  $\{G_{ij}^*\}$  until convergence.

$c_1$  and  $c_2$  are additive constants which will be absorbed in normalization of factor distributions.

$$\log F_i^*(T_i) = \mathbb{E}_{\prod_{k \neq i} F_k(T_k|\mathbf{y}, \mathbf{x}^n)} [\log P(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r})] + c_1 \quad (4.12)$$

$$\log G_{ij}^*(S_{ij}) = \mathbb{E}_{\prod_{k \neq j} G_{ik}(S_{ik}|\mathbf{x}^n)} [\log P(S_i|\mathbf{x}^n)] + c_2 \quad (4.13)$$

The parameters of optimum factors are derived in Appendix 5.4 and Appendix ???. Once we have the optimum factorization, solving for the  $Q$  function becomes simple. The simplified function is

$$Q(\mathbf{x}; \mathbf{x}^n) = \mathbb{E}_{P(\mathbf{T}|\mathbf{x}^n, \mathbf{y}, \mathbf{r})} [\log P(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r})] + \sum_i \mathbb{E}_{P(S_i|\mathbf{x}^n)} [\log P(S_i|\mathbf{X})] + \log P(\mathbf{X}) \quad (4.14)$$

We will substitute the factorization assumed on the posterior distribution of latent variables  $\mathbf{T}$  and  $\mathbf{S}$ .

$$\mathbb{E}_{P(\mathbf{T}|\mathbf{y}, \mathbf{x}^n)} [\log P(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r})] = \mathbb{E}_{\prod_{i=1}^I F_i^*(T_i|\mathbf{y}, \mathbf{x}^n)} [\log P(\mathbf{y}, \mathbf{T}|\mathbf{x}, \mathbf{r})] \quad (4.15)$$

$$\mathbb{E}_{P(S_i|\mathbf{x}^n)} [\log P(S_i|\mathbf{X})] = \mathbb{E}_{\prod_j G_{ij}^*(S_{ij}|\mathbf{x}^n)} [\log P(S_i|\mathbf{X})] \quad (4.16)$$

Equation (4.15) is solved by substituting  $\mu_i^{\mathbf{T}}$  for  $T_i$ ,  $(\mu_i^{\mathbf{T}})^2 + (\sigma_i^{\mathbf{T}})^2$  for  $T_i^2$  and  $\mu_i^{\mathbf{T}} \mu_j^{\mathbf{T}}$  for  $T_i T_j$ . Similarly Equation (4.16) is solved by substituting  $\mu_{ij}^{\mathbf{S}}$  for  $S_{ij}$ .

### 4.1.2 M step

In this step, we will maximize  $Q(\mathbf{x}; \mathbf{x}^n)$  with respect to  $\mathbf{x}$  to obtain a new estimate  $\mathbf{x}^{n+1}$ . The optimization in Equation (4.17) is solved using adaptive gradient descent.

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





# 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 isotropic diffusion used in [5, 10]. The methods used for comparison 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 specular 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.

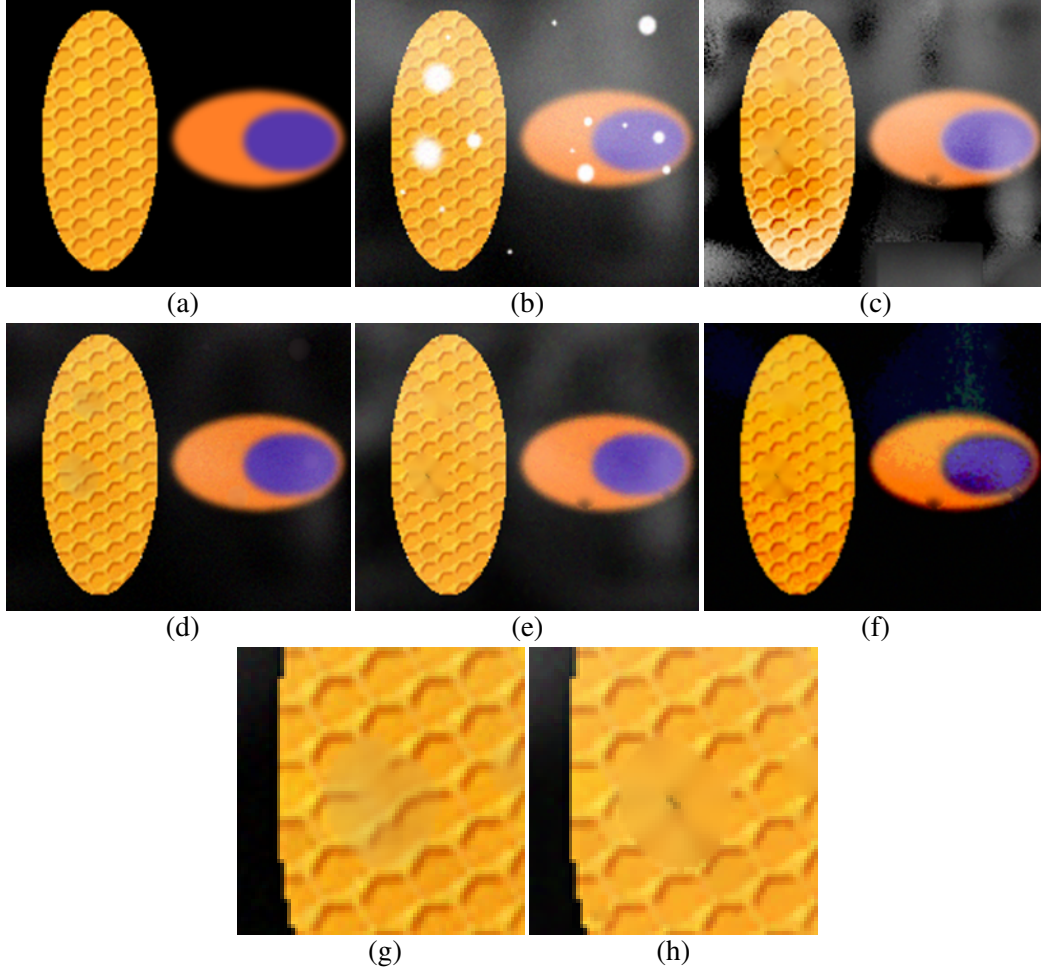


Figure 5.1: **Qualitative Validation on Simulated Data.** (a) Phantom (color component values  $\in [0, 255]$ ). (b) Corrupted phantom with smoke, specularities, and low noise ( $\sigma = 5$ ). Results of processing image (b), using: (c) adaptive filtering [1] followed by inpainting; (d) *proposed method 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 inpainting 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.

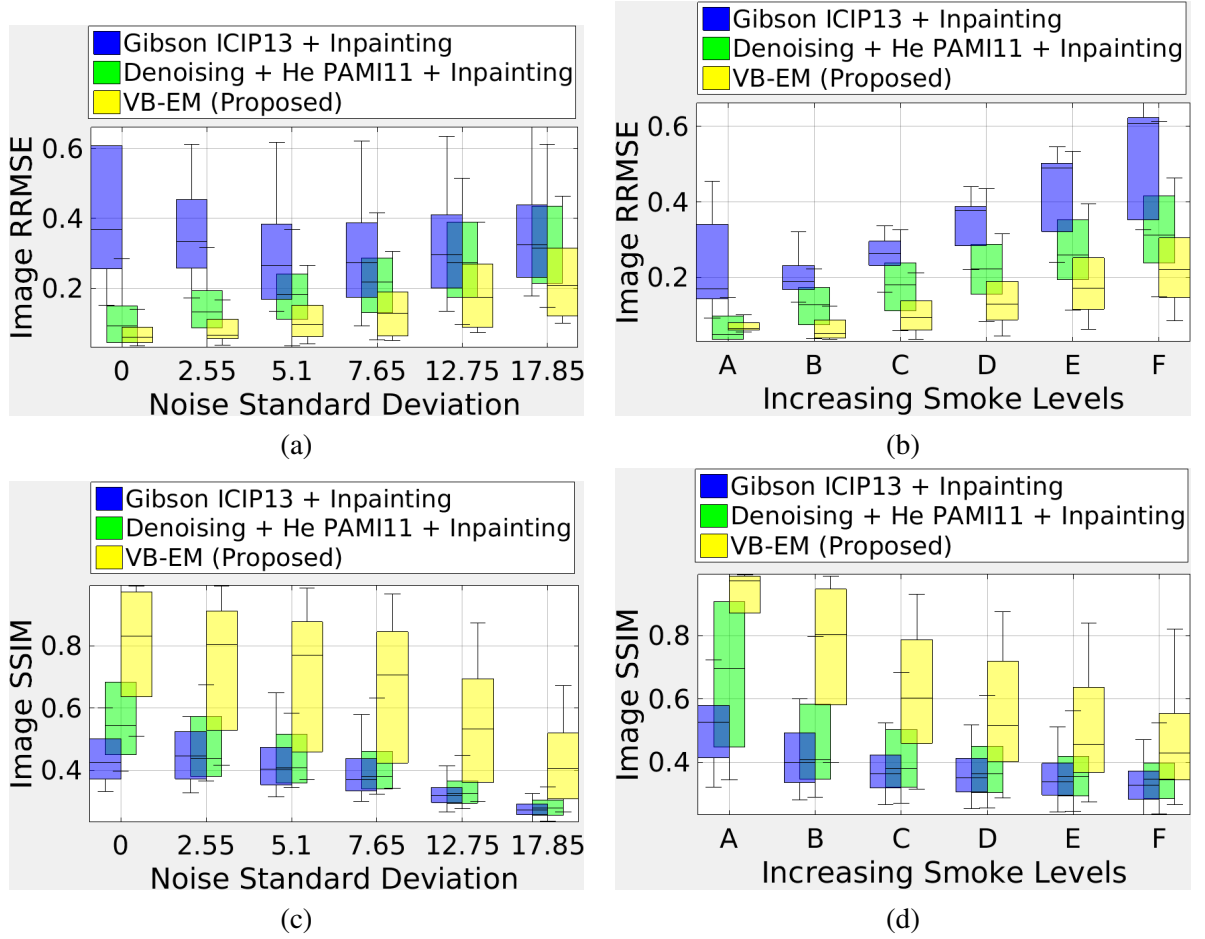


Figure 5.2: **Quantitative Validation on Simulated Data.** Box plots for RRMSE (a, b) and SSIM (c, d) for different combinations of smoke levels and noise levels. Each combination is run 50 times. (a), (c) grouped by noise level  $\in [0, 255]$ ; (b), (d) grouped by smoke level.

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 demonstrated 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 shows the results of processing on laparoscopy data. The observations are similar to those for simulated data. Gibson ICIP13 [1] plus inpainting does a poor job at removing

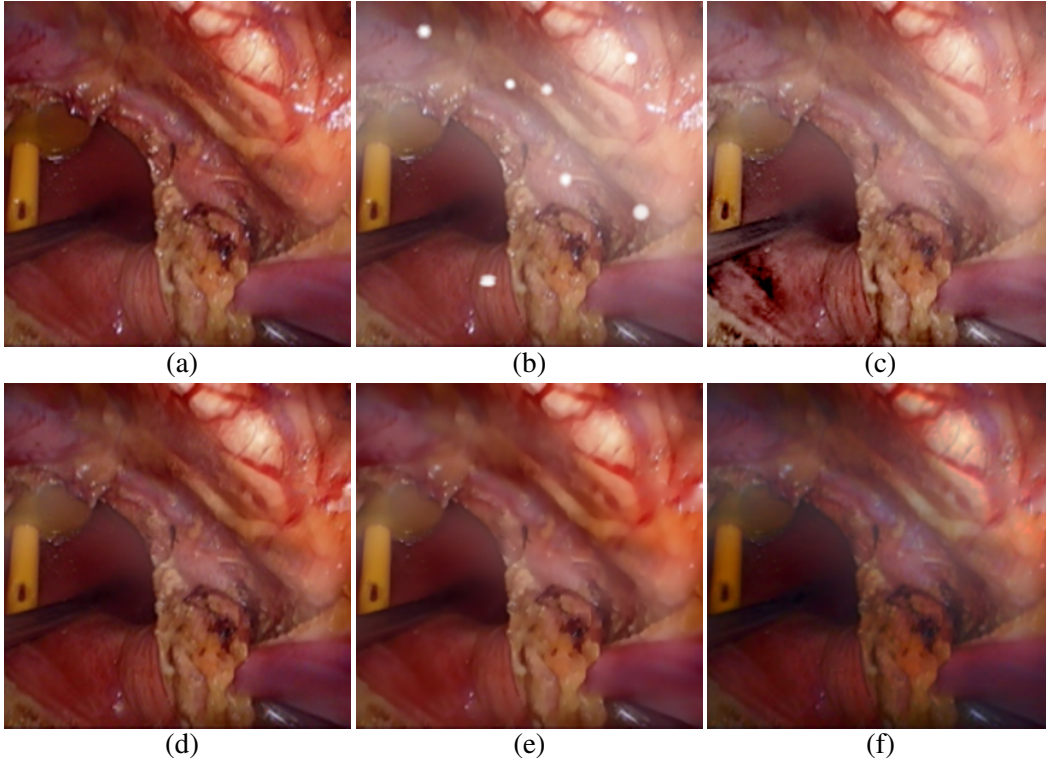


Figure 5.3: **Qualitative Validation on Simulated Data.** (a) Ground truth (color component values  $\in [0, 255]$ ). (b) Corrupted phantom with smoke, specular highlights, and low noise ( $\sigma = 5$ ). Results of processing image (b), using: (c) adaptive filtering [1] followed by inpainting; (d) *proposed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

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.

## 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.

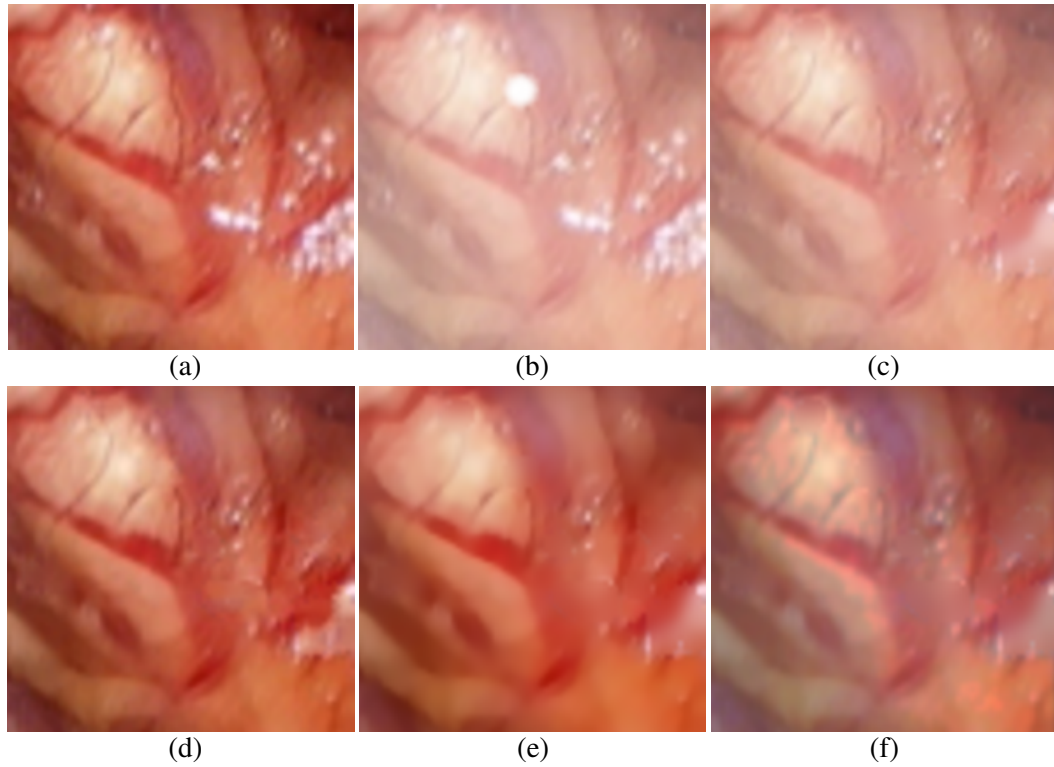
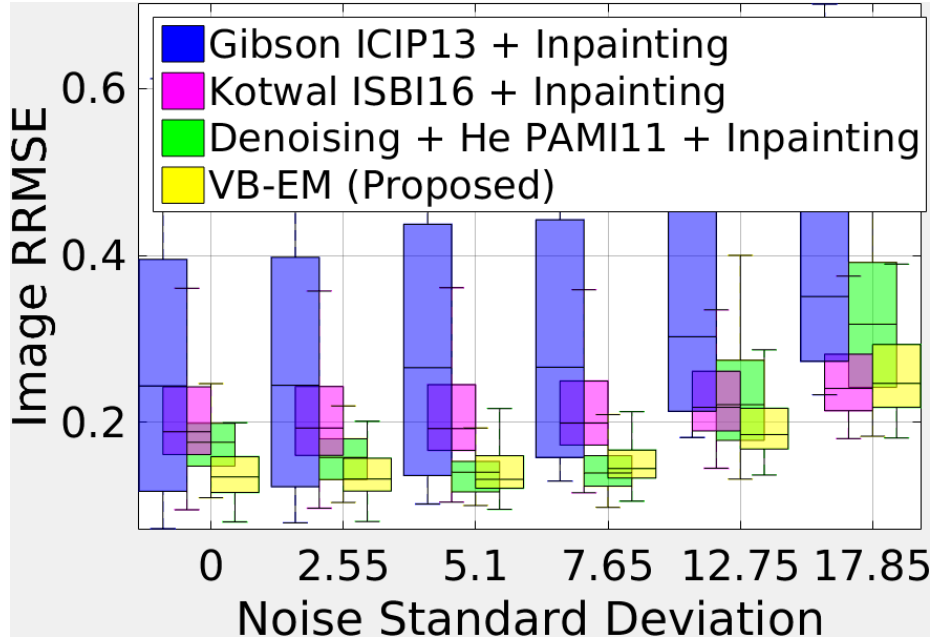


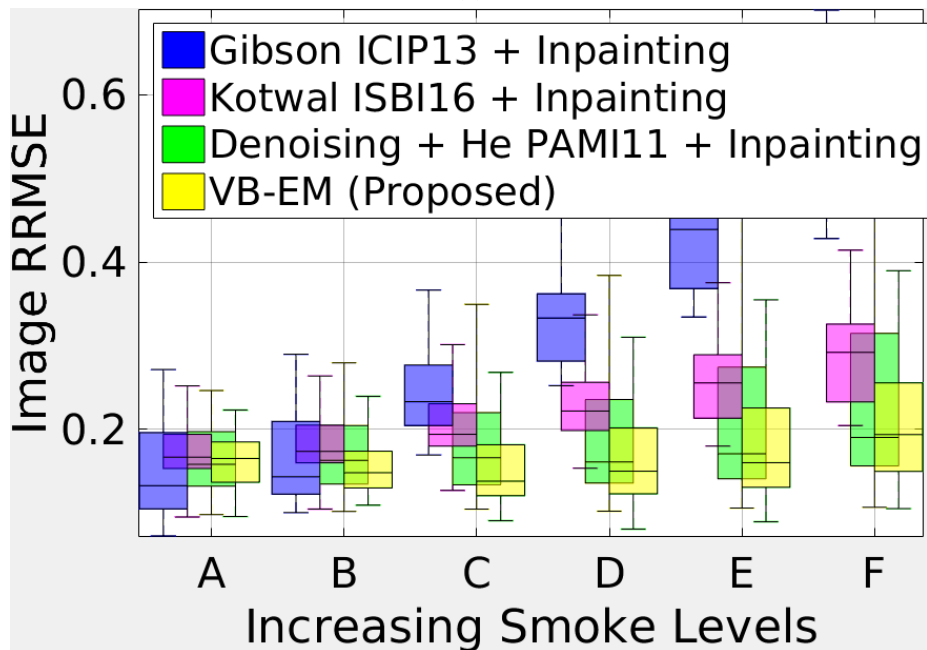
Figure 5.4: **Qualitative Validation on Simulated Data.** The images (a) to (f) are zoomed in subparts of the corresponding images in Figure 5.3

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

## Optimal factors for tranmission map $\mathbf{T}$

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.

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

To solve Equation (??), 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 (A.1). This will result in a quadratic equation in  $T_i$ , from which the parameters  $(\mu_i, \sigma_i)$  of Gaussian factor  $F_i$  can be solved for. The final solution is derived as

$$\bar{\mu}_i = \frac{(y_i - K_{smoke})(z_i - K_{smoke}) + 2\gamma_3 \sum_{j \in \mathcal{N}_i^T} w_{ij}^T \mu_j}{(z_i - K_{smoke})^2 + 2\gamma_3 \sum_{j \in \mathcal{N}_i^T} w_{ij}^T} \quad (\text{A.2})$$

$$\bar{\sigma}_i = \frac{1}{\sqrt{2}} \left( (z_i - K_{smoke})^2 + 2\gamma_3 \sum_{j \in \mathcal{N}_i^T} w_{ij}^T \right)^{-\frac{1}{2}} \quad (\text{A.3})$$

$$\alpha_i = -\frac{\bar{\mu}_i}{\bar{\sigma}_i} \quad (\text{A.4})$$

$$\beta_i = \frac{1 - \bar{\mu}_i}{\bar{\sigma}_i} \quad (\text{A.5})$$

$$\mu_i = \bar{\mu}_i + \bar{\sigma}_i \frac{\phi(\alpha_i) - \phi(\beta_i)}{\Phi(\beta_i) - \Phi(\alpha_i)} \quad (\text{A.6})$$

$$\sigma_i = \bar{\sigma}_i^2 \left[ 1 + \frac{\alpha_i \phi(\alpha_i) - \beta_i \phi(\beta_i)}{\Phi(\beta_i) - \Phi(\alpha_i)} - \left( \frac{\phi(\alpha_i) - \phi(\beta_i)}{\Phi(\beta_i) - \Phi(\alpha_i)} \right)^2 \right] \quad (\text{A.7})$$

where  $\phi$  and  $\Phi$  are PDF and CDF of standard normal distribution.





## Appendix B

### Optimal factors for codes S

$$\mu_{ij} = \frac{D_j^\top \left( (\mathbf{x}^n)_i^m - \sum_{k \neq j} D_k S_{ik} \right)}{D_j^\top D_j + 2\lambda} \quad (\text{B.1})$$

$$\sigma_{ij} = \frac{1}{\sqrt{D_j^\top D_j + 2\lambda}} \quad (\text{B.2})$$



# 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] B. A. Olshausen and D. J. Field, "Sparse coding of natural images produces localized, oriented, bandpass receptive fields," *Nature*, vol. 381, no. 60, p. 609, 1996.
- [19] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," in *Advances in neural information processing systems*, 2009, pp. 1033–1040.

- [20] 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.
- [21] 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.
- [22] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.
- [23] A. Baid, A. Kotwal, R. Bhalodia, S. Merchant, and S. P. Awate, “Joint desmoking, specular 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, specular 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

24 May 2017