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

2 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: 2 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 [1, 2], surgical smoke [3], 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 [4, 5] and tracking [6]. 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

[7] 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 [2]. They use discrete convolutions with a kernel repeatedly until convergence is reached. [1] 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 [8]. 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.

[9] 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 [10] 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.

[11] 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. [12] use adaptive patch size and replace the soft matting step with guided filtering. [13] calculate the dark channel and then apply adaptive wiener filter to smooth out the transmission map.

124 images are used by [14] 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

[15] 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

$$\mathbf{X} := \{X_i\}_{i=1}^{I} \quad (3.3)$$

$$\mathbf{T} := \{T_i\}_{i=1}^{I} \quad (3.4)$$

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 **X** and **T**. We want **X** to have texture to tackle the speckles, and to have high contrast to counter the smoke. We want **T** to be spatially smooth.

3.2.1 Sparse coding under a dictionary

Given a data matrix **X**, whose columns are i.i.d. random vectors, a linear decomposition is of the form

$$\mathbf{X} \approx \mathbf{DS} \quad (3.5)$$

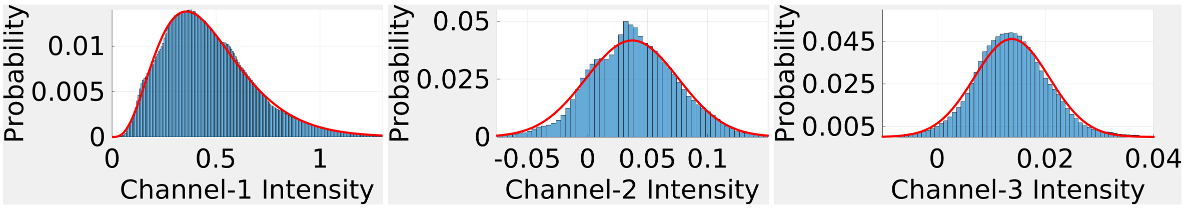
In Equation (3.5), **D** is the dictionary and **S** is a code of **X**.

In sparse coding [16], 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.

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

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 [17] can be applied when **X** is non-negative. It constrains **D** and **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.

[18] 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

$$KS(F_{emp}; F_{ref}) = \max_x |F_{emp}(x) - F_{ref}(x)| \tag{3.8}$$

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

3.2.4 Kernel density estimation

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

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3.9}$$

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 [19] for bandwidth using the standard deviation of samples $\hat{\sigma}$. Equation (**??**)eqn:kdebw) is used to tune the bandwidth using training data.

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

We will now derive the CDF estimator using the Gaussian kernel

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

$$\hat{F}_h(x) = \sum_{s=x_{min}}^{x} \hat{f}_h(x) \tag{3.12}$$

3.2.5 Spatial smoothness prior

3.3 MAP Estimation

3.3.1 EM

3.3.2 VB-EM version 1

3.3.3 VB-EM version 2

Chapter 4

Results

4.1 Simulated data and synthetic corruption

4.2 High quality laparoscopy data and synthetic corruption

4.3 Clinical validation

Appendix A

Supporting Material

Bibliography

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

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

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

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

List of Publications

Put your publications from the thesis here. The packages `multibib` or `bibtopic` or `biblatex` or enumerate environment or thebibliography environment etc. can be used to handle multiple different bibliographies in the document.

Acknowledgements

This section is for the acknowledgments. Please keep this brief and resist the temptation of writing flowery prose! Do include all those who helped you, e.g. other faculty/staff you consulted, colleagues who assisted etc.

Ayush Baid
IIT Bombay
2 May 2017