

Removal of Specular Reflections in Endoscopic Images

T. Stehle

During an endoscopic examination, pictures from the inside of the human body are displayed on a computer monitor. Disturbing light reflections are often visible in these images. In this paper, we present an approach for removing these reflections and replacing them by an estimate obtained using a spectral deconvolution algorithm.

Keywords: endoscopy, light reflection removal, spectral deconvolution, defect interpolation.

1 Introduction

Endoscopy is a minimally invasive diagnostic procedure that is used to give the physician a realistic impression of almost any part of the gastrointestinal tract or other organs inside the human body. During an endoscopic examination a flexible tube, which contains among other things a video camera and a lamp, is introduced into the human body. It has a steerable tip that enables the physician to change perspective and to ease navigation, e. g. through the colon. Another system for imaging the gastrointestinal tract is a so-called „Camera in a Pill“ or „PillCam“ [1]. This is swallowed by the patient and travels through the gastrointestinal tract in a natural way. The images that are recorded by the camera are sent via a radio frequency transmitter to a data recorder, which is worn on a belt around the patient's waist.

Due to the frontal illumination, there are often light reflections that are disturbing for the physician. In the case of classic endoscopy, the physician may change the perspective by turning the tip of the endoscope to avoid these reflections. In the case of a camera-in-pill examination, however, there is no way to force the pill to move to a better position. Therefore, these reflections need to be removed in a different manner. The algorithm we are going to present may also be applied to classic endoscopy for the physician's convenience.

2 Methods

The algorithm that we apply is used in image communications to conceal image data corrupted by transmission errors [2] or in the restoration of x-ray images with defective detector elements [3]. It models the defective image as a pointwise product of the undisturbed image and a known binary defect map that contains ones where the image is correct and zeros in the case of a defective pixel. This pointwise product in the spatial domain corresponds to a convolution in the spectral domain. To restore the image, a spectral deconvolution is performed.

In case of endoscopy, we first perform a segmentation by thresholding in the YUV color space and then we treat the pixels that belong to light reflections as defective pixels, setting them to zero, and then we apply the defect interpolation algorithm.

2.1 Segmentation of reflections

Before the spectral deconvolution algorithm is applicable, a map has to be generated that contains ones where the image is undisturbed and zeros where specular reflexions occur.

These reflections are very bright, so the affected pixels show high luminance in the YUV color space. Thus, the image is transferred from RGB into the YUV color space by applying the linear transform

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

where the Y component contains luminance and U and V contain chrominance.

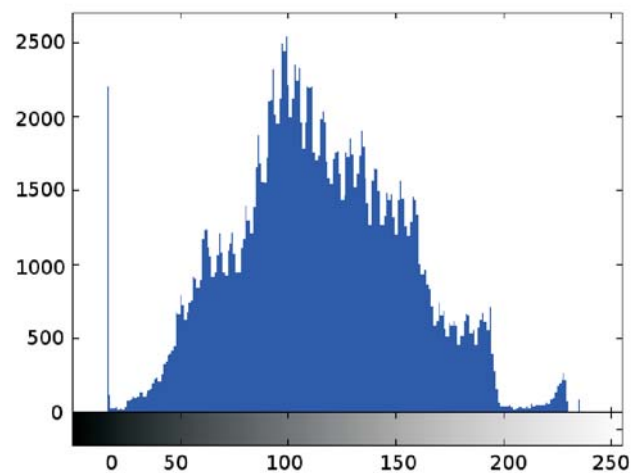


Fig. 1: Histogram of the luminance channel of an endoscopic image containing specular reflections. The x axis shows the luminance value, the y axis shows the number of occurrences

Fig. 1 shows a typical histogram of the luminance channel of an endoscopic image. On the right hand side, there is a small peak that corresponds to pixels that contain specular reflections. We now segment these reflections by thresholding with a value that is just lower than the left edge of this hill. Therefore, we low pass filter the histogram and find the position beginning from the right hand side where the derivative changes from a significant positive value to a value near zero.

Since there are often dark rings around the affected pixels in the reconstructed image, and these are also very disturbing, we enlarge the segmented areas by an erosion operation. Note that we here use erosion instead of dilation to enlarge the segments, since we have a black segment on a white background.

2.2 Spectral deconvolution

For ease of notation, we describe the univariate case. The results can be easily generalized to the bivariate case.

The observed image $g(n)$ is modeled as

$$g(n) = f(n) \cdot w(n)$$



$$\begin{aligned} G(k) &= \frac{1}{N} F(k) * W(k) \\ &= \frac{1}{N} \sum_{l=0}^{N-1} F(l) \cdot W(k-l) \end{aligned}$$

where $0 \leq n, k < N$, $f(n)$ is the undisturbed image, and $w(n)$ a binary window, with $w(n) = 0$ if the corresponding pixel n in $g(n)$ belongs to a reflection and $w(n) = 1$ otherwise. $G(k)$, $F(k)$ and $W(k)$ are the DFT spectra of $g(n)$, $f(n)$ and $w(n)$, respectively. Both $f(n)$ and $g(n)$ are real valued, hence, $G(k) = G^*(N-k)$ and $F(k) = F^*(N-k)$. Our goal is to estimate $F(k) \xleftarrow{\text{DFT}} f(n)$ for $0 \leq n, k < N$. Let us select a spectral line pair $G(s)$ and $G(N-s)$ of $G(k)$. If $F(k)$ consisted only of two lines at s and $N-s$, i.e.

$$F(k) = \hat{F}(k) = \hat{F}(s)\delta(k-s) + \hat{F}(N-s)\delta(k-N+s) \quad (3)$$

convolution with the window spectrum $W(k)$ would yield for the observed pair

$$G(s) = \frac{1}{N} \left(\hat{F}(s) \cdot W(0) + \hat{F}^*(s) \cdot W(2s) \right) \quad (4)$$

$$G^*(s) = \frac{1}{N} \left(\hat{F}^*(s) \cdot W^*(0) + \hat{F}(s) \cdot W^*(2s) \right),$$

where $\hat{F}(s)$ and $\hat{F}(N-s) = \hat{F}^*(s)$ are the estimated coefficients. From (4), $\hat{F}(s)$ can be found to

$$\hat{F}(s) = N \frac{G(s)W(0) - G^*(s)W(2s)}{|W(0)|^2 - |W(2s)|^2}. \quad (5)$$

The estimated spectrum $\hat{F}(k)$ would then be given by (3). Generally, $F(k)$ consists of more than two spectral lines. The error after deconvolution (5) at s and $N-s$ in the spectral domain is given by

$$\begin{aligned} G^{(1)}(k) &= G(k) - \frac{1}{N} \cdot \hat{F}(k) * W(k) \\ &= G(k) - \frac{1}{N} \left(\hat{F}(s) \cdot W(k-s) + \hat{F}^*(s) \cdot W(k+s) \right) \end{aligned} \quad (6)$$

where $G^{(1)}(k)$ is the spectrum of the window internal difference image

$$g^{(1)}(n) = g(n) - \hat{f}(n) \cdot w(n) = w(n) \left(f(n) - \hat{f}(n) \right). \quad (7)$$

Clearly, the spectral error $G^{(1)}(k) = 0$ at the selected line pair $k = s$ and $k = N-s$.

Given s and $N-s$, the estimates $\hat{F}(s)$ and $\hat{F}(N-s)$ are optimal in the minimum mean square error (MMSE) sense if the energy E_g of $g^{(1)}(n)$ is minimized. Using Parseval's theorem, E_g can directly be evaluated in the spectral domain according to

$$\begin{aligned} E_g &= \sum_{n=0}^{N-1} \left(g^{(1)}(n) \right)^2 = \frac{1}{N} \sum_{k=0}^{N-1} |G^{(1)}(k)|^2 \\ &= \frac{1}{N} \cdot E_G. \end{aligned} \quad (8)$$

This provides MMSE estimates for the sought frequency coefficients.

If the selected line pair $\hat{F}(s)$ and $\hat{F}(N-s)$ is dominant in the sense specified below, its convolution with $W(k)$ tends to "hide" other, less dominant spectral coefficients of $F(k)$. This influence is removed by (6), so that another line pair can be selected from $G^{(1)}(k)$, estimated and subtracted. This leads to the following iteration for spectral deconvolution:

- Initialization: $\hat{F}^{(0)}(k) = 0$, $G^{(0)}(k) = G(k)$, $i = 1$.
- i -th iteration step: Select a pair of spectral coefficients $G^{(i-1)}(s^{(i)})$, $G^{(i-1)}(N-s^{(i)})$ out of $G^{(i-1)}(k)$.
- Estimate $\hat{F}(s^{(i)})$, $\hat{F}(N-s^{(i)})$ according to (5) such that $G^{(i)}(s^{(i)}) = G^{(i)}(N-s^{(i)}) = 0$, i.e.

$$G^{(i-1)}(s^{(i)}) = \frac{1}{N} \left(\hat{F}(s^{(i)})W(0) + \hat{F}^*(s^{(i)})W(2s^{(i)}) \right) \quad (9)$$

$$G^*(s^{(i)}) = \frac{1}{N} \left(\hat{F}^*(s^{(i)})W(0) + \hat{F}(s^{(i)})W^*(2s^{(i)}) \right)$$

\Rightarrow

$$\begin{aligned} \hat{F}(s^{(i)}) &= \\ N \frac{G^{(i-1)}(s^{(i)})W(0) - \left(G^{(i-1)}(s^{(i)}) \right)^* W(2s^{(i)})}{|W(0)|^2 - |W(2s^{(i)})|^2}. \end{aligned} \quad (10)$$

Since we seek to minimize the error (8), the line pair we should select in the i th iteration is that which maximizes the energy reduction $\Delta_E(\hat{F}(s^{(i)}))$, which can be calculated to

$$\begin{aligned} \Delta_E(\hat{F}(s^{(i)})) &= \\ \frac{1}{N^2} \sum_{k=0}^{N-1} \left| \hat{F}(s^{(i)})W(k-s^{(i)}) + \hat{F}^*(s^{(i)})W(k+s^{(i)}) \right|^2 \end{aligned} \quad (11)$$

which can be rewritten to

$$\begin{aligned} \Delta_E &= \frac{1}{N^2} \hat{F}^*(s^{(i)}) \\ &\sum_{k=0}^{N-1} \left[\hat{F}(s^{(i)}) |W(k+s^{(i)})|^2 + \hat{F}^*(s^{(i)})W(k+s^{(i)})W^*(k-s^{(i)}) \right] \\ &+ \frac{1}{N^2} \hat{F}(s^{(i)}) \\ &\sum_{k=0}^{N-1} \left[\hat{F}^*(s^{(i)}) |W(k-s^{(i)})|^2 + \hat{F}(s^{(i)})W(k-s^{(i)})W^*(k+s^{(i)}) \right]. \end{aligned} \quad (12)$$

For a binary window, we have $w^2(n) = w(n)$. Inserting this into Parseval's theorem, we obtain

$$\begin{aligned} \sum_{k=0}^{N-1} |W(k+s^{(i)})|^2 &= \sum_{k=0}^{N-1} |W(k-s^{(i)})|^2 \\ &= \sum_{k=0}^{N-1} |W(k)|^2 = N \sum_{n=0}^{N-1} w(n) = N \cdot W(0) \end{aligned} \quad (13)$$

and

$$\sum_{k=0}^{N-1} W(k+s^{(i)})W^*(k-s^{(i)}) \quad (14)$$

$$= N \sum_{n=0}^{N-1} w(n) \exp\left\{-j2p \frac{2s^{(i)}}{N} n\right\} = N \cdot W(2s^{(i)}).$$

Inserting (13) and (14) into (12) yields

$$\Delta_E = \hat{F}^*(s^{(i)}) \cdot G^{(i-1)}(s^{(i)}) + \hat{F}(s^{(i)}) \cdot (G^{*(i-1)}(s^{(i)})). \quad (15)$$

Since we know how to estimate $\hat{F}(s^{(i)})$ optimally (10), $\hat{F}^*(s^{(i)})$ can be eliminated, thus expressing the error energy reduction depending on the available error spectrum line $G^{(i-1)}(k_s^{(i)})$ as

$$\Delta_E = \frac{N}{|W(0)|^2 - |W(2s^{(i)})|^2} \cdot \left(2|G^{(i-1)}(s^{(i)})|^2 |W(0) - 2\operatorname{Re}\left\{ (G^{(i-1)}(s^{(i)}))^2 \cdot W(2s^{(i)}) \right\}| \right). \quad (16)$$

Selecting the best line pair in each iteration hence implies finding s such that Δ_E according to (16) is maximum. Because of the symmetry of the DFT spectra of real valued signals, it suffices to search over only half the coefficients of $G^{(i-1)}(k)$, i.e. from $k=0$ to $k=N/2$, to find $s^{(i)}$.

$$\hat{F}(s^{(i)}) = \frac{N}{W(0)} G^{(i-1)}(s^{(i)}). \quad (17)$$

Similarly, the calculation of the error energy reduction Δ_E according to (16) modifies to

$$\Delta_E = \frac{2N}{W(0)} |G^{(i-1)}(s^{(i)})|^2. \quad (18)$$

Clearly, when selecting only a single line, the error energy reduction depends only on the modulus spectrum $|G^{(i-1)}(s^{(i)})|$, apart from a constant factor. To save computational expense in the selection step, a simplified approach is to always select $s^{(i)}$ such that $|G^{(i-1)}(s^{(i)})|$ is maximum, regardless of whether a single line is tested, that is, $s^{(i)}=0$ or $s^{(i)}=N/2$, or a line pair. Practically, the outcome of the iteration remains almost unchanged.

When the estimated spectrum $\hat{F}^*(s^{(i)})$ contains as many lines as there are samples of $f(n)$ inside the window $w(n)$, the remaining error energy E_G vanishes. The backtransformed estimate $\hat{f}^{(i)}(n)$ is then identical to $g(n)$ inside the observation window, and contains extrapolated information outside. In practice, the iteration is stopped when Δ_E falls below a pre-specified level ε , or when a maximum number of iterations is reached.

To achieve high spectral resolution for the interpolated signal, it is often reasonable to apply zero-padding to $g(n)$ and $w(n)$ before transforming and starting the iteration. An illustration of the entire recursion is given in Fig. 2.

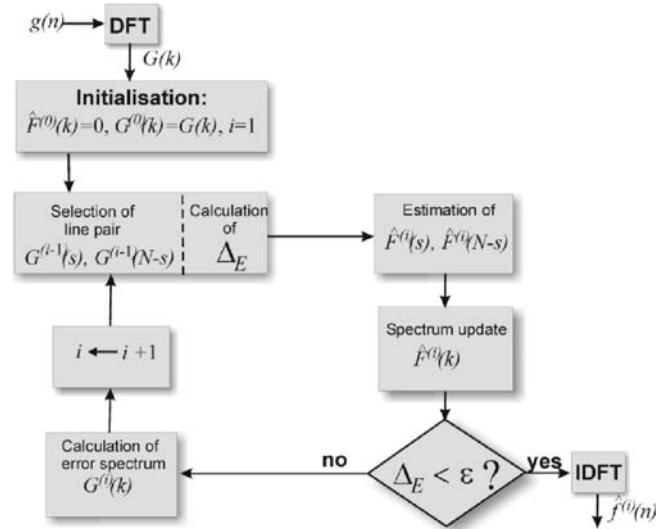


Fig. 2: Illustration of the recursive spectral deconvolution algorithm. From the final spectral estimate $\hat{F}^{(i)}(k)$, the interpolated signal estimate is obtained by an inverse DFT.

3 Results

The algorithm as described above was applied to several endoscopic images. In our experiments, we used a 9×9 circular structure element for erosion of the binary mask image, and we independently applied 100 iterations of the deconvolution algorithm to each channel of the YUV color space. Fig. 3 and Fig. 4 show two examples with the original image and a processed version of the image. The specular reflections are removed and the holes are filled with texture that does not disturb the overall impression of the image.

However, the results are not perfect. In Fig. 3, several vessels are affected by specular reflections in the lower right corner, and in the processed image some of these vessels appear interrupted. In Fig. 4, block artifacts were added to the image in areas where a large number of specular reflections occurred. In some cases, the specular reflections themselves were surrounded by green, pink and yellow circles. These circles were correctly identified not being specular reflections and therefore they were not removed completely. This is not severe if the physician knows that they are caused by reflections. In the processed image these colors are visible, too, but the reflections have been removed, so the physician might think it is the real color of the tissue.

The processed images have not yet been presented to a physician because of the drawbacks described above. However, the results achieved so far are already very promising.

4 Discussion

We have presented an approach for removing specular reflections from endoscopic images by segmentation of the affected pixels and interpolation via a spectral deconvolution algorithm.

The algorithm, however, needs further improvement. We are going to modify the segmentation of the specular reflections such that the typical elliptical shape is incorporated into

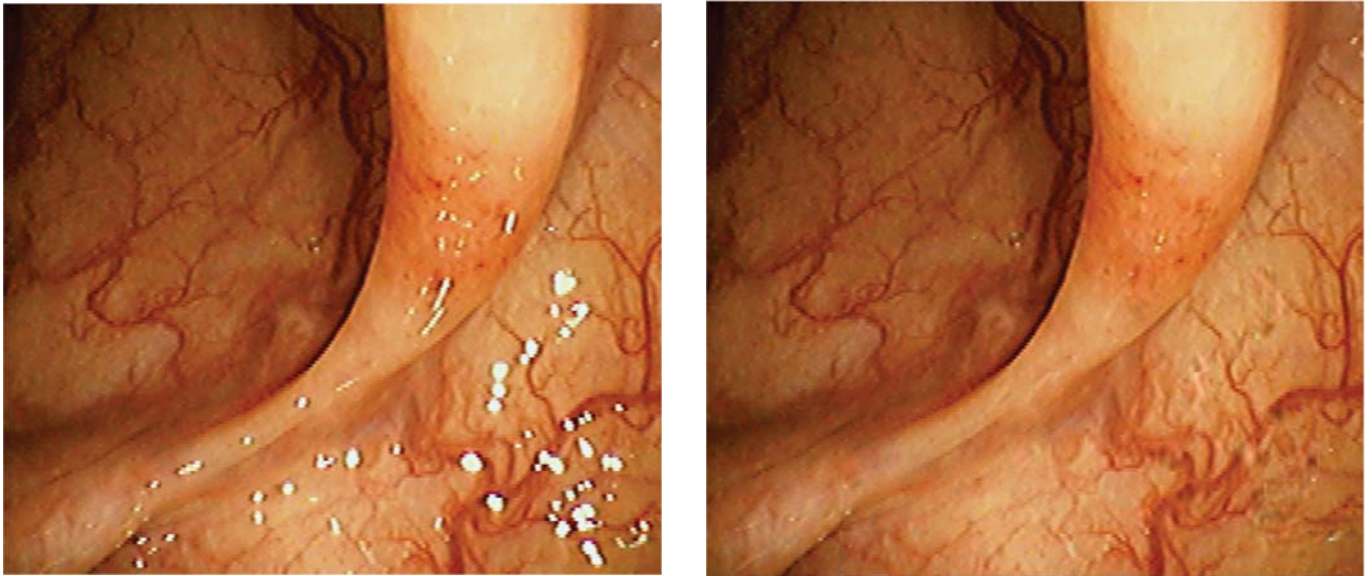


Fig. 3: Endoscopic image of the colon with specular reflections (left) and processed image (right). The overall impression of the processed image is better, but some artifacts have been added to the image. The vessels in the lower right corner appear blurred, and some of them interrupted

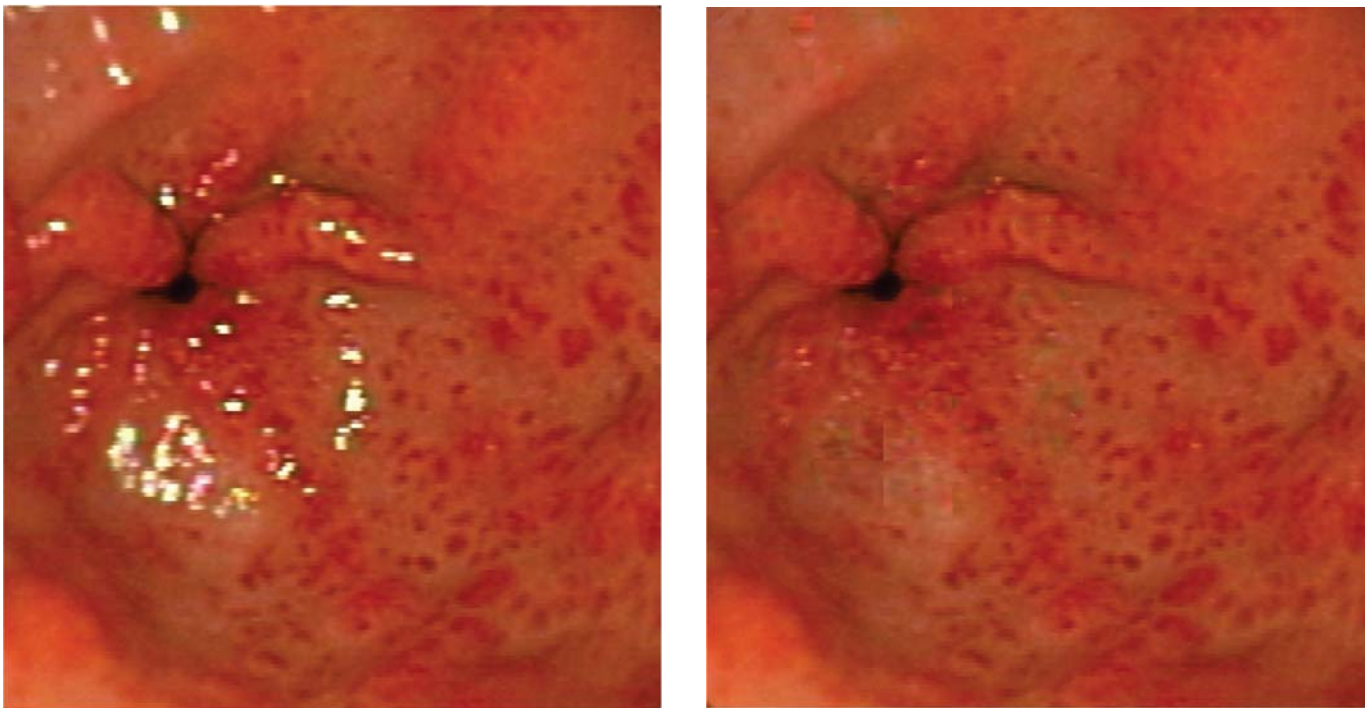


Fig. 4: Endoscopic image of the esophagus with specular reflections (left) and the processed image (right). In areas with a large number of specular reflections block artifacts have been added to the image.

the segmentation process. A rejection criterion will be integrated so that the quantity of false positive segmentation results is reduced. In addition, we will address the color disturbances in the processed images.

5 Acknowledgments

I would like to thank Prof. Dr. T. Aach, Institute of Imaging and Computer Vision, RWTH Aachen University, Germany, who supervised this project. I would also like to

thank Prof. Dr. med. C. Trautwein, University Hospital Aachen, Germany, who kindly provided the endoscopic images used in this paper.

References

- [1] Iddan, G., Meron, G., Glukhovsky, A. Swain P.: Wireless Capsule Endoscopy, *Nature*, 2000, p. 405–417.
- [2] Kaup, A., Meisinger, K., Aach, T.: Frequency Selective Signal Extrapolation with Applications to Error

Concealment in Image Communications. *International Journal of Electronic Communication (AE)*, Vol. **59** (2005), p. 147–156.

- [3] Aach, T., Metzler, V.: Defect Interpolation in Digital Radiography – How Object-Oriented Transform Coding helps. *SPIE Medical Imaging*, Proceedings of SPIE 4322, 2001, p. 824–835.

Thomas Stehle

e-mail: Thomas.Stehle@lfb.rwth-aachen.de

Institute of Imaging & Computer Vision

RWTH Aachen University

D-52056 Aachen, Germany