Joint Brightness and Tone Stabilization of Capsule Endoscopy Videos

Sibren van Vliet[1], André Sobiecki[1,2] and Alexandru C. Telea[1]
[1]*Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands*
[2]*ZiuZ Visual Intelligence, Gorredijk, The Netherlands*

Keywords: Capsule Endoscopy, Video Processing, Color Stabilization.

Abstract: Pill endoscopy cameras generate hours-long videos that need to be manually inspected by medical specialists. Technical limitations of pill cameras often create large and uninformative color variations between neighboring frames, which make exploration more difficult. To increase the exploration efficiency, we propose an automatic method for joint intensity and hue (tone) stabilization that reduces such artifacts. Our method works in real time, has no free parameters, and is simple to implement. We thoroughly tested our method on several real-world videos and quantitatively and qualitatively assessed its results and optimal parameter values by both image quality metrics and user studies. Both types of comparisons strongly support the effectiveness, ease-of-use, and added value claims for our new method.

1 INTRODUCTION

Endoscopy of the gastrointestinal tract is since long used to screen, diagnose, locate, or treat conditions such as gastrointestinal bleeding, inflammatory bowel disease, celiac disease, polyps, and certain cancer types (Classen and Phillip, 1984). This is traditionally done by using a small camera at the end of a thin flexible tube inserted into the mouth and guided through the tract. However, this method does not reach the many tight bends of the intestines.

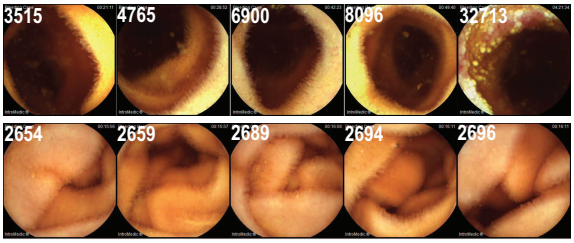

Figure 1: Sample frames from endoscopy pill camera footage illustrating intensity (top row) and hue (bottom row) problems.

A recent disruptive technology is the pill camera, a small capsule holding a camera and lights (Hale et al., 2014). After being swallowed, the camera records 8 to 12 hours of video. While cheaper, less intrusive, and better covering the full gastrointestinal tract, pill cameras have several issues. Figure 1 shows sample frames from a video recorded by the MiroCam pill camera (Hale et al., 2014) at 3 frames per second at 320^2 pixel resolution. Each frame contains a circular picture surrounded by black borders, with the frame number in white. In the top row frames, areas close to the camera are very bright, and far away areas are completely dark, due to the distance from the capsule's lights. Consider frame 4765. All tissue here has in reality the same color, but it is not imaged as such. As the capsule moves onwards from frame 4765, the moderately lit area in the center of frame 4765 becomes too bright, as the light approaches it. Also, the too dark area top-left in frame 4765 becomes moderately lit due to the camera motion. All in all, the same tissue area is shown in differing intensities over time. Figure 1(bottom row) shows a second type of problem: All images are of the same tissue type, so they should have the same color tone (hue). Yet, as the camera automatically adjusts its color balance, tone fluctuates over time. For instance, frame 2654 has a pink tone; frame 2659 has a more orange tone; frame 2689 appears pink again; frame 2694 appears orange; and frame 2696 shifts to pink again.

Medical practitioners viewing endoscopy videos are being distracted by sudden tone and/or intensity fluctuations, which do not contain any information. Color correction (also called stabilization) methods are an effective way to alleviate such problems. However, such methods should not introduce any artifacts which could mislead the physycian. From discussions with gastroenterologists, we found two key

Vliet, S., Sobiecki, A. and Telea, A.
Joint Brightness and Tone Stabilization of Capsule Endoscopy Videos.
DOI: 10.5220/0006552401010112
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 4: VISAPP*, pages 101-112
ISBN: 978-989-758-290-5

requirements for a stabilization method: (i) the relative intensity of pixels in the corrected and original image should be the same (if a pixel a is brighter than another pixel b in the input image I, then a should also be brighter than b in the corrected image I'; and (ii) hue changes should be small enough so that a tissue type can be reliably recognized in stabilized images. While many generic color correction algorithms exist (Anbarjafari, 2014; Vig et al., 2016; Purushothaman et al., 2016; Gautam and Tiwari, 2015; González et al., 2016; Moradi et al., 2015), few have been developed with the specific constraints of endoscopy videos: low resolution, poor lighting of large image areas, relatively low framerate, rapid variation of the light direction, real-time operation, and the avoidance of misleading artifacts in the corrected video. Moreover, such algorithms have various parameters which influence their results. We are not aware of any studies showing how to find optimal parameter values that smooth out intensity and tone changes but do not create significant artifacts.

In this paper we attack the problem of joint intensity-and-tone stabilization in endoscopy videos. We analyze a large set of existing intensity-and-tone stabilization techniques *vs* the video endoscopy constraints (Sec. 2). We select the best candidate, which we next enhance to optimally meet all these constraints (Secs. 3, 4). We evaluate our enhanced algorithm quantitatively (by image similarity metrics) and qualitatively (by an extensive user study), on a set of endoscopy videos showing a wide variation of imaged tissues and lighting conditions (Sec. 5). The evaluation shows that our improved algorithm surpasses the best-so-far algorithm we could find, by performing joint intensity and tone stabilization, being parameter free, guaranteeing good image quality, and working at the same speed as the pill camera. Finally, we conclude with directions for future work (Sec. 6).

2 RELATED WORK

Color correction has a long history in image and video processing (Gijsenij et al., 2011). Early methods include greyscale histogram equalization (GHE) (Kim and Yang, 2006) and dynamic histogram equalization (DHE) (Sun et al., 2005). Few methods were designed for, or tested on, endoscopy videos. Hence, besides considering endoscopy-specific methods, it is useful to study if more generic methods can be used, with suitable modifications, for our problem. We discuss below ten methods which target (partially) our intensity and hue stabilization goal, and are either well-known in image processing or else are designed to handle endoscopy videos. We assess these methods by rating them on a Likert scale (5=very good, 4=good, 3=average, 2=poor, 1=very poor) against the following requirements:

- *Validation* measures how well the claims of a method are defended by results shown in the respective paper. Methods showing stronger validation are more interesting candidates to adapt to our endoscopy use-case.
- *Reproductibility* measures how easy is to (re)implement a method and obtain the results described in that paper. This is essential: without reproducibility, we cannot validate and/or extend a given method.
- Complexity measures the computational complexity of a method for a video of n frames of $w \times h$ pixels. Ideally, we want a (near) linear complexity method in video size so that we can achieve interactive exploration.
- *Usability* tells how easy can a non-technical user run the method. It is measured by the number and intuitiveness of the exposed parameters. A method with many parameters which are not intuitive or easy to set is not very usable. This is a critical requirement for an application that aims to *decrease* the workload for a medical specialist.

(Anbarjafari, 2014) proposed an iterative n^{th} root and n^{th} power color equalization for single generic images. The intensity channel of an image in HSI space is passed through a non-linear transfer function $f(x) = x^{\ln(0.5)/\ln(\overline{x})}$, where $\overline{x}$ is the image's mean intensity. The operation is repeated until the final image achieves a mean 'goal' intensity equal to γ, set typically to $\gamma = 0.5$. The method is good in lighting very dark image areas and darkening too bright areas. However, it does not address our problem of tone stabilization in videos.

(Vig et al., 2016) equalize colors in single images by increasing the intensity of dark areas, but keeps bright areas unchanged, akin to overexposing. Not darkening very bright areas is a limitation in our context. Also, this technique does contrast enhancement; this can create artifacts in endoscopy images which typically contain only low contrast tissue.

(Purushothaman et al., 2016) propose a differential histogram equalization method for color images which increases the contrast of color images so as to make the color information more visible to the human eye. However, as a result, brightly lit areas may become even brighter, losing potentially valuable information in endoscopy imagery.

(Gautam and Tiwari, 2015) propose yet another histogram equalization based method for single ima-

ges which increases contrast in dimly lit areas while not brightening properly lit areas. However, too bright areas are not darkened, which conflicts with our intensity equalization goal.

(González et al., 2016) propose an improvement of the earlier luminance Multi-Scale Retinex method (Funt et al., 1997) that targets hues. The method is very powerful at brightening dark areas and thus revealing rich color information. However, already well lit areas may become too bright.

(Moradi et al., 2015) propose a method specifically targeted at endoscopy images which increases contrast and removes noise. However, intensity normalization is not specifically addressed. Also, the method does not specifically handle tone stabilization.

(Vazquez-Corral and Bertalmio, 2014) propose a so-called video tone stabilization method which equalizes a set of images taken from several cameras or from a single camera where white balance and/or exposure change over time. The method works by making all input images more similar with respect to a so-called reference image. It works in both hue and intensity channels, both which are important for our context. However, an open challenge is how to automatically select a single reference frame.

(Wang et al., 2014) propose yet another video tone stabilization, based on smoothing differences between neighbor frames, much like an average running through time, applied on the trajectory of the color state in color space. A parameter allows turning the smoothing off to keep large tone temporal differences which can encode important information.

(Farbman and Lischinski, 2011) also propose a video tone stabilization method for videos, based on the same reference frame idea as (Vazquez-Corral and Bertalmio, 2014). While the results of this method are impressive, a major drawback is that it appears to be closed-source and patented, which makes its replication and application hard at best.

(Bassiou and Kotropoulos, 2007) present a single-image method based on histogram equalization. The method uses multi-level smoothing correct images in HSI space, using the probability density functions of the saturation and intensity components while keeping hue unchanged. The method can equalize intensity very well. However, it does not directly address the problem of tone stabilization.

Table 1 summarizes our survey. The method of (Anbarjafari, 2014) (referred next to as 'Anbarjafari') gets the best overall rating, with the methods of (Vazquez-Corral and Bertalmio, 2014) and (Bassiou and Kotropoulos, 2007) coming next. As such, we considered extending these three methods for our goal. However, replicating the algorithms in (Vazquez-Corral and Bertalmio, 2014) and (Bassiou and Kotropoulos, 2007) did not succeed in producing the same results as in the respective papers, as several crucial details were omitted in the papers. As such, we settled with extending the method of (Anbarjafari, 2014) to suit our goals, as described next.

3 PROPOSED METHOD

As explained in Sec. 2, the Anbarjafari method brightens dark areas and darkens bright areas in single images. However, we want to equalize intensity *and* smooth out hue fluctuations over time. For this, we extend the Anbarjafari method as follows.

We smooth out fluctuations in an image channel over time by detecting large variances between the channel's histograms (computed over all input image pixels) of all frames within a time window, and next changing the pixel values so that the histogram is suitably compressed. By compressing the histogram, differences between pixel values are made smaller. When applied to all frames within a time window, the compression rate should progress gradually, in order to smoothen out sudden differences. This technique can be applied to any image channel in any color space, *e.g.*, RGB or HSI. As discussed next in Sec. 5, we will apply our technique on both the intensity and saturation channels of a HSI-space image, and combine it with the original Anbarjafari method, which we will also apply on both above channels. The hue channel is left untouched, as changing it easily yields undesired artifacts.

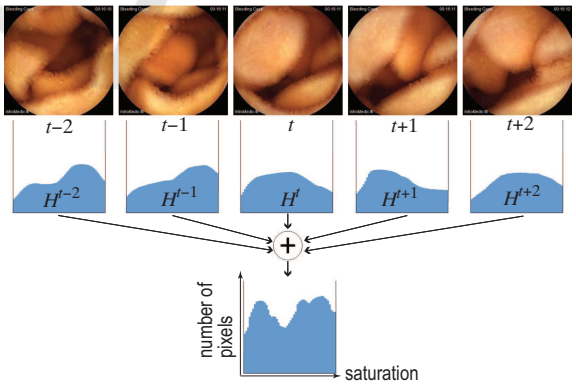

Figure 2: Five successive video frames in which tone fluctuation occurs (from orange to pink). Below each frame, a histogram of saturation values is shown. Summing these histograms results in a cumulative histogram.

The histogram compression works as follows. Consider the current frame t in the video and a time-window of $2k+1$ frames centered at t. Figure 2 shows

Table 1: Brightness and/or tone stabilization methods reviewed in this work.

Method	Validation	Reproductibility	Complexity	Usability
(Anbarjafari, 2014)	**(4)** Very good results for two test-sets	**(4)** MATLAB code provided	**(4)** $O(whnx)$ with $x \approx 10$	**(4)** A single intuitive parameter to set (goal mean).
(Vig et al., 2016)	**(2)** Good results for two test-sets, but only for brightening dark areas	**(2)** No code provided, reproducing is difficult	**(4)** $O(whn)$	**(2)** Four not very intuitive parameters
(Purushothaman et al., 2016)	**(3)** Good results on two test-sets, but mainly for brightening dark areas	**(3)** No code provided, but implementation clear and easy to reproduce)	**(2)** $O((wh)^2nx)$ with $x \approx 128$)	**(3)** A single parameter which is easy to understand
(Gautam and Tiwari, 2015)	**(3)** Good results on five test-sets, but dark areas can become undesirably darker	**(2)** No code provided, reproducing is moderately difficult	**(4)** $O(whn)$	**(5)** No parameters to be set
(González et al., 2016)	**(3)** Good results on six test-sets, but all only show brightening dark areas	**(3)** No code provided, reproducing is moderately difficult	**(4)** $O(whNn)$ where N is the constant size of a small neighborhood around each pixel	**(3)** Three parameters, of which two are not directly intuitive
(Moradi et al., 2015)	**(4)** Good results on four test-sets	**(2)** No code provided, reproducing is difficult due to vague description	**(4)** $O(whn)$	**(2)** Two parameters which do not have an intuitive meaning
(Vazquez-Corral and Bertalmio, 2014)	**(5)** Very good results on 24 test-sets.	**(3)** No code provided, algorithm explanation leaves out some important details	**(4)** $O(whn)$ (authors mention that real-time operation is feasible)	**(3)** Two parameters which do not have an intuitive meaning
(Wang et al., 2014)	**(5)** Good results on seven test-sets	**(2)** No code provided, reproducing seems difficult	**(4)** $O(whn)$	**(1)** Five parameters which do not have an intuitive meaning
(Farbman and Lischinski, 2011)	**(4)** Good results on five test-sets	**(1)** No code provided, algorithm patented by authors	**(4)** $O(whn)$	**(4)** A single parameter with clear usage instructions
(Bassiou and Kotropoulos, 2007)	**(4)** Good results on five test-sets	**(3)** Third party code used in the paper produces undesired results	**(4)** $O(whn)$	**(4)** Parameter(s) of probability smoothing step not explained

this for a window of 5 frames. Below each frame t, the histogram H^t of its saturation channel is shown (in the following, we use saturation as example, though our technique also works on the intensity channel, as already stated). In the frames, we observe an undesired tone shift from orange to pink. We also observe a distinct shape change of the saturation histograms. Hence, the shape change can be used as an indicator of the amount of color variation. For this, we need a way to measure the amount of shape change. To do this, we first compute a cumulative histogram H^C whose bins are given by

$$H_x^C = \sum_{i=-k}^{k} H_x^{t+i}$$

where H_x^{t+i} is the bin for saturation value x of the histogram for frame $t+i$. As our pill camera images are 8 bit per channel, we use histograms of 255 bins. We next compute the mean μ and variance σ^2 of H^C and use the latter as a measure of the shape change of all histograms within the time window. A small variance indicates a small tone fluctuation, meaning that very little histogram compression is needed. A large variance indicates a large tone fluctuation, meaning that more compression is needed to smooth out the fluctuation.

We can now proceed with the actual histogram compression (see also Fig. 3). We start with the computed mean μ and variance σ^2 of the cumulative histogram H^C (Fig. 3a). Secondly, we eliminate the mean by subtracting μ from the saturations of all pixels (Fig. 3b). Thirdly, we compress the histogram by dividing the saturations by $a\sigma^2$ (Fig. 3c). Here, $a \in [1/\sigma^2, 1]$ controls the compression amount: For

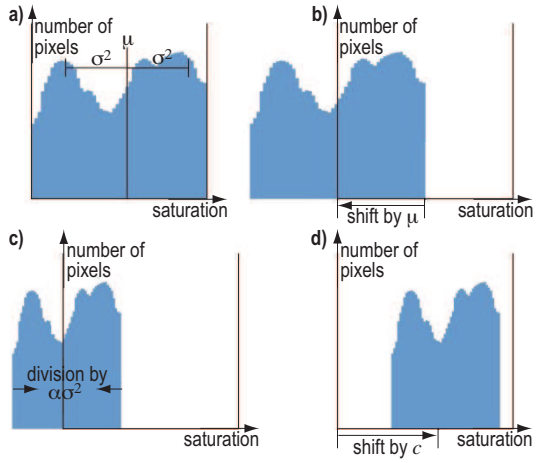

Figure 3: Histogram compression. a) The histogram's mean μ and variance σ^2 are computed. b) The histogram is shifted μ bins to the left so that its mean is zero. c) The histogram is compressed by dividing all saturation values by $a\sigma^2$. d) The histogram is shifted right by c bins.

$a = 1$, all saturations are divided by σ^2, so that the histogram is compressed by an amount proportional to the variance. For $a = 1/\sigma^2$, no compression occurs. After this step, a part of the histogram will correspond to negative saturation values, which of course make no sense. To fix this, it seems natural to shift the histogram back with the same value μ we used in step one. However, we verified that doing so produces unnatural looking tones – pixel saturations appear higher or lower than desired. To solve this issue, we use a shift value $c \in [0,1]$ (Fig. 3), as follows. If $c = 0$, the histogram is shifted so that its leftmost bin corresponds to saturation value 0; if $c = 1$, the histogram is shifted so that its rightmost bin corresponds to saturation value 255. Intermediate values for c produce linearly interpolated shifts between these two extremes.

Several comments are due. The proposed histogram compression extends the relative pixel intensity constraint mentioned in Sec. 1 to pixel saturations. Indeed, the applied transformations are linear, and the shape of the histogram is preserved. Separately, while the histogram compression is computed on the cumulative time-window histogram, the individual pixel intensity or saturation manipulations are done separately on each frame. This ensures that these manipulations will vary smoothly in time, as the cumulative histogram has the effect of a smoothing sliding-window time filter.

4 IMPLEMENTATION

We implemented our method in single-threaded C++ under Linux and Windows. Our tool covers both the original Anbarjafari method and our new method, and allows one to apply them separately, or in sequence, on the saturation and/or intensity channels. The tool loads a pill-camera video in MPEG format, allows changing the parameters k, a, and c of our algorithm and the mean goal γ of Anbarjafari, plays the original and stabilized videos side-by-side, and saves the stabilized video as an MPEG file (Fig. 4).

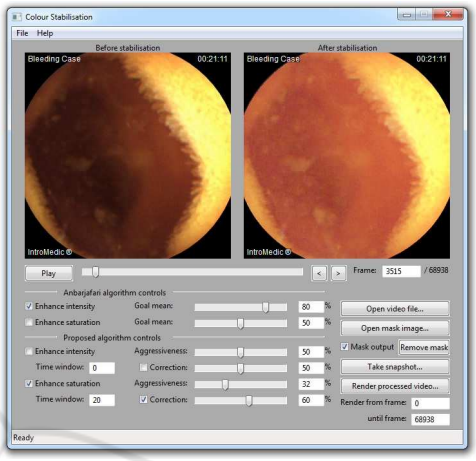

Figure 4: Software tool for color stabilization and video exploration.

For a time window of 41 frames ($k = 20$), computing histograms takes about 3 seconds on a 2.3 GHz laptop with 4GB RAM. The video stabilization runs smoothly at 3 frames/second, which is the recording speed of the pill-camera video (Sec. 1). The computational complexity is $O(whn)$ for processing a video of n frames each of $w \times h$ pixels, *i.e.* linear in input size. After computing the histograms, changing all parameters is, however, instantaneous. This allows a physician to focus on an image of interest and explore it to *e.g.* brighten or darken its various areas in real time.

5 EVALUATION

As already outlined, only very few evaluations of color stabilization for endoscopy videos are present in the literature. Moreover, these take the form of presenting the stabilized images, but come with limited or even no actual evaluation of the *quality* thereof. We improve upon this by presenting next both a qualitative user-study based evaluation (Sec. 5.1) and a quantitative metrics-based evaluation (Sec. 5.2).

5.1 Qualitative Evaluation

The nature of color stabilization is quite application-specific and possibly even user-specific. It is not easy to formally measure how much 'better' a given stabilized image is than another one. Also, note that we have no ground truth, in the sense of an 'optimally' stabilized image. As such, it is definitely important to compare different stabilization methods or parameter settings by means of user studies. To this end, we performed a survey in which users were asked to rank images produced by different stabilization methods and parameter values, as described next.

5.1.1 Evaluation Materials

We acquired several endoscopy videos, each 8 hours long, recorded using the MiroCam pill camera (Medivators, 2017), from medical specialists at a major regional hospital in the Netherlands. The videos were pre-screened by the specialists for suitability – that is, containing no major artifacts due to camera malfunction, and containing a wide range of image intensities and tones that would pose difficulties in manual analysis and for which stabilization would be of added value. Since organizing a study where multiple users examine thousands of images such as present in our videos was infeasible, we first manually grouped the available video frames into five representative classes, depending on the color and intensity distribution, as follows:

- Dark area directly bordering a very bright area (Fig. 1, frame 3515);
- Dark area separated from a very bright area by moderate illumination (Fig. 1, frame 4765);
- Dark area directly surrounded by bright areas on all sides (Fig. 1, frame 6900);
- Dark area surrounded by bright areas on all sides, with a moderate illumination transition zone (Fig. 1, frame 8096);
- Dark area bordering a bright area of varied color and structure (Fig. 1, frame 8096).

Next, we randomly selected a few images in each class for the qualitative study. For each image, we ran several combinations of the Anbarjafari method (A) and our proposed method (P) described in Sec. 3, applied on the intensity (I) and saturation channels (S), as described below. Note that only the first combination (A used solely on I) is covered by existing literature, all other combinations being novel.

1. **A $\rightarrow$ I:** A applied to I only;
2. **A $\rightarrow$ (I,S):** A applied to both I and S channels;
3. **P $\rightarrow$ I:** P applied to I only;
4. **P $\rightarrow$ (I,S):** P applied to both I and S channels;
5. **(A,P) $\rightarrow$ I:** A applied to I, followed by applying P to the resulting I;
6. **(A,P) $\rightarrow$ (I,S):** A applied to I, followed by applying P to the resulting I; and A applied to S, followed by applying P to the resulting S.

For each combination, we ran the involved methods for several parameter values. Specifically, we set the mean goal γ in Anbarjafari to values in $\{0.6, 0.7, 0.8, 0.9, 1\%\}$; and the compression a of our method to values in $\{0.02, 0.04, 0.08, 0.16, 0.32\}$. The latter set of values is chosen as such since c is used as a denominator (Sec. 3), so it affects a function of hyperbolic type $1/x$. For the time window size and correction, we used the fixed values of $k = 20$ frames and $c = 0.4$ respectively, which have been determined by us empirically by testing stabilization on several videos.

Figure 5 shows the stabilization results obtained for frame 8096 (Fig. 1) for several method and parameter combinations. Due to space limitations, we cannot show all the tested results which entail several hundreds of images. The rows in Fig. 1 indicate method combinations; columns indicate parameter-value combinations. Below we discuss the findings we observed ourselves – that is, before using these results in the actual survey, which is described next in Sec. 5.1.2.

A $\rightarrow$ I: We see that, as the parameter γ increases, dark areas are brightened, and colors and details get more easily visible to the human eye. For all five frames in the top row in Fig. 5, we found that $\gamma = 0.7$ yields the greatest intensity increase with acceptable loss of details. When $\gamma > 0.7$, images become too noisy. Moreover, in endoscopy images, detail such as edges is mainly defined by intensity and not hue, so too much brightening erases such detail.

A $\rightarrow$ (I,S): Similar to brightening dark areas, increasing γ now makes the color of low-saturation (gray-like) areas more vivid. Since low saturation areas match very well dark areas in the gastrointestinal tract, this method additionally boosts dark areas by making them not only brighter, but also more colorful. As for the A $\rightarrow$ I method, we found an optimal value around $\gamma = 0.7$. Larger γ values affect color tones too much, which can create undesirable artifacts, like rendering a normal tissue too red, thus suggesting an internal bleeding.

P $\rightarrow$ I: Similar to A $\rightarrow$ I, this method makes dark areas become brighter as c increases. However,

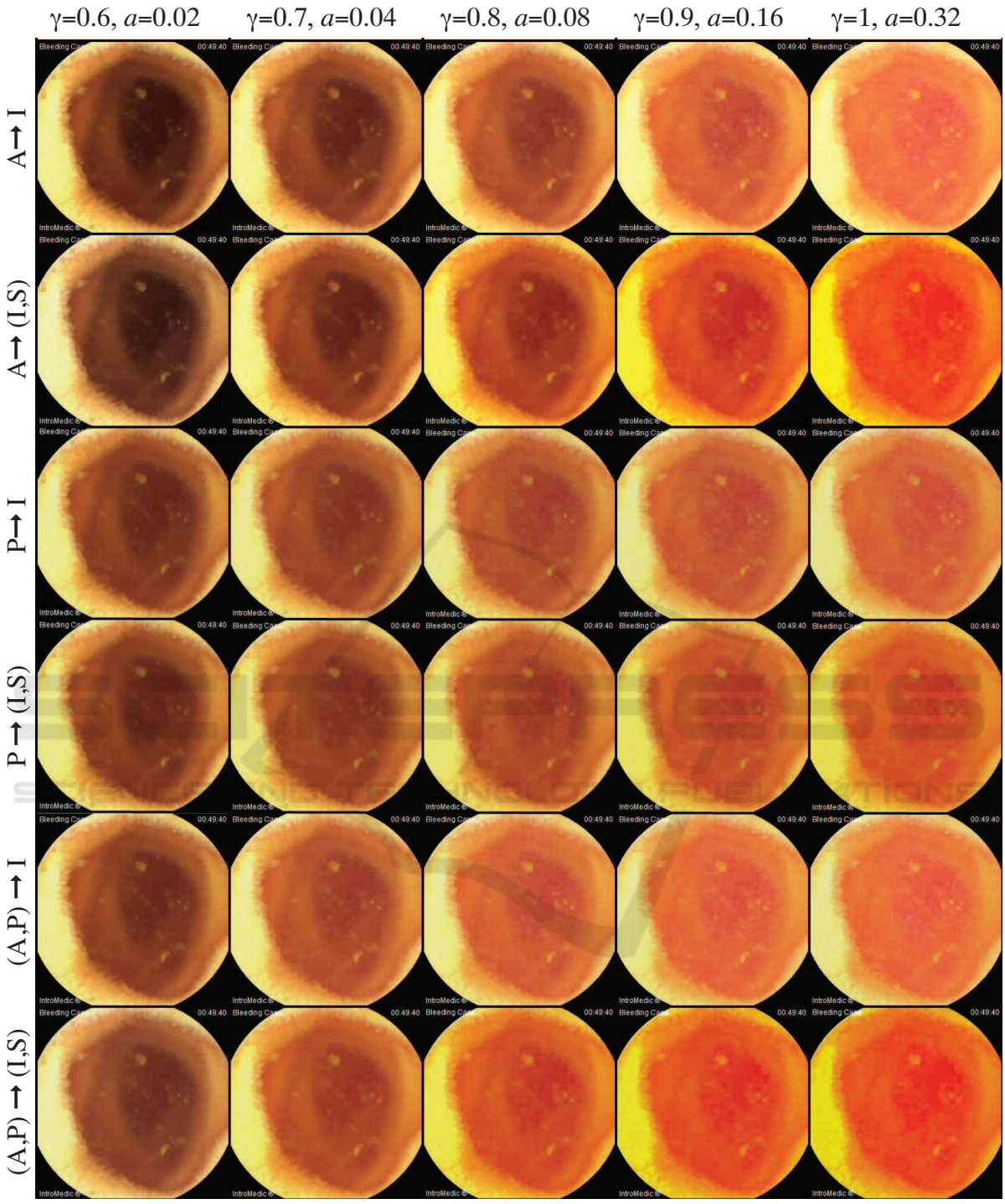

Figure 5: Frame 8096 (shown in Fig. 1) processed with various combinations of algorithms and parameters.

details in dark areas are lost earlier than in the A → I case. We also note that this method yields overall brighter images than A → I (compare rows 1 and 3 in Fig. 5). However, detail shading is slightly less well visible. This is expected, since the goal of our method (P) is not to enhance single images, but to smooth sudden changes in video sequences. Since P → I essentially compresses the intensity channel histogram, edges captured by intensity differences may become less visible.

P → (I,S): In addition to the previously discussed

effect on the intensity levels, this method makes colors more saturated as c increases. Interestingly, saturation is not increased as aggressively as in A $\rightarrow$ (I,S). Again, this is because our algorithm does not try to increase saturation to a certain predefined level γ, but aims to smooth out sudden differences in the saturation histograms of neighboring frames. This is why, as we will discuss later, our method is better for stabilizing saturation in videos rather than single images.

(A,P) $\rightarrow$ I: We observe that the results of this method are nearly identical to those of P $\rightarrow$ I. We explain this by the fact that P compresses the histogram after A enhanced the intensity. This largely undoes the enhancements that the A method made. As a result, the output images suffer from the same problems we observed when using P $\rightarrow$ I, namely loss of details due to the histogram compression.

(A,P) $\rightarrow$ (I,S): We observe that the results of this method are very similar to those of A $\rightarrow$ (I,P). However, the saturation is less dramatically increased. We explain this by the fact that, after the A method has made the saturation very high, the P method compresses the saturation histogram, thus making the color vibrance less extreme.

From all above, we draw the following preliminary qualitative conclusions. The Anbarjafari method (A) with a mean goal value around $\gamma = 0.7$ shows itself to be best for intensity stabilization of single images. However, it is not effective in stabilizing tone fluctuations – when applied to saturation (A $\rightarrow$ S), it may actually *enhance* tone fluctuations. In contrast, our method (P) is effective in smoothing tone fluctuations, but less effective in stabilizing intensity.

5.1.2 User Survey

We refined the qualitative observations presented above, which are drawn from our own study of the computed results, by conducting an online survey that involved a wide group of people, thereby realizing a more representative qualitative evaluation. The survey material consisted of five pages, one page for an image in each image class defined in Sec. 5.1.1. Each page contained all stabilized images for the respective input image, laid out identically to Fig. 5. We also included an additional column representing the actual input image. However, the column was not marked as such, so the participants could not know which is the input and which the outputs of the stabilization. For each image row, the participant was asked to pick the image that they thought was the best in terms of enhancing the information in the brighter and darker areas of the image and without introducing too much noise or losing information. This answers the question 'which parameter values are best for a given method combination?'. Next, at the end of each page, participants were asked to review the six images they picked as best for the six rows and pick the best one among these. This answers the question 'which method combination delivers the best results, given that all methods are run with their optimal parameter values?'.

The survey was conducted using Google Forms. Participants were encouraged to look at each row of images for roughly 10 seconds, so that the survey could be finished in about 5 minutes. However, the participants could spend more time if desired, and were also allowed to go back to previous pages to review or change their answers. Note that the participants did not see any annotations on the survey pages such as the method names and parameter values in Fig. 5. Eighteen people participated in the survey. All are specialists in image processing and computer vision, and are well familiar with endoscopy videos and their issues. The participants were aged between 20 and 50, the majority being male.

Table 2 presents the aggregated results of the survey. Rows indicate method combinations, and columns indicate parameter values, just like in Fig. 5. Each cell contains two numbers, separated by a slash. The first number indicates how many times an image generated by the respective method and parameter-values combination was chosen best in a row of images – thus, best for all tested parameter values. The second number (in bold) indicates how many times an image was chosen as best for an entire survey page – thus, best for all method and parameter values combinations tested.

We get several insights from these figures. First, we see that the parameter values $\gamma = 0.6, a = 0.02$ and $\gamma = 0.7, a = 0.04$ get most votes, the former being seen best when the combined method (A,P) is used, and the latter when the individual Anbarjafari (A) method is used, respectively. These are thus good values for a wide set of images and a wide set of users. Note that the setting $\gamma = 0.7$ matches what we found ourselves in our preliminary qualitative evaluation (Sec. 5.1.1). Hence, we use these values as presets in our tool (Sec. 4). Secondly, we see that very high parameter values are never preferred. This matches our own findings that such values yield too much disappearance of relevant details (Sec. 5.1.1). Thirdly, we see that the Anbarjafari method applied to saturation (A $\rightarrow$ S) with $\gamma = 0.7, a = 0.04$ has the highest number of overall best results. This matches our ear-

lier observations that this method is indeed very good for stabilizing *single* images. Moreover, this is an interesting novel result, as the Anbarjafari method has been originally proposed to work on intensity only. Separately, as explained earlier, this method is not aimed at stabilizing tone fluctuations in *video* sequences – something that our survey could not capture, as participants were shown only individual frames. Finally, we see that the combination (A,P) $\rightarrow$ (I,S) with $\gamma = 0.6, a = 0.02$ scores the best image-in-a-row. As such, this method combination is arguably good for video color stabilization, albeit it scores lower for single frame stabilization.

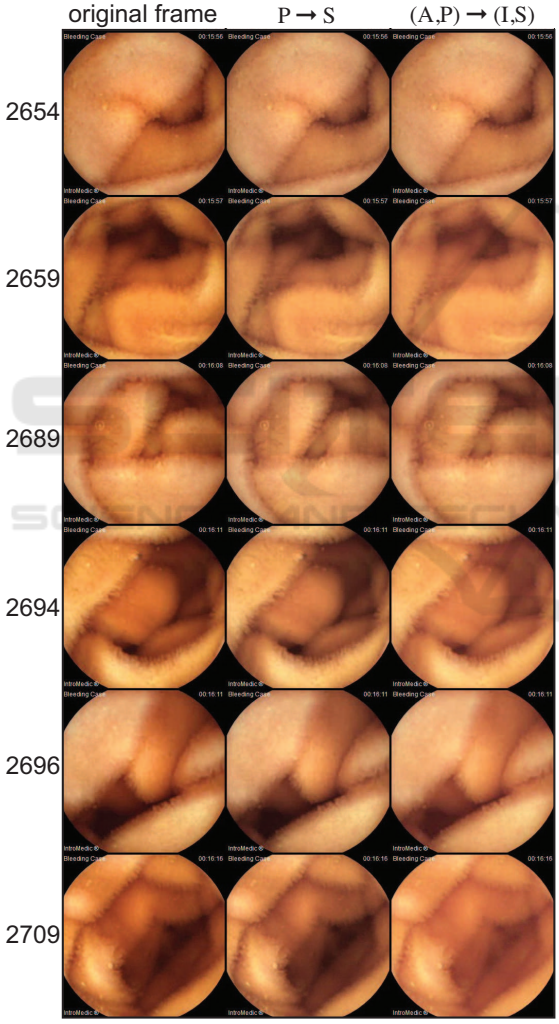

Figure 6: Selected frames from a video fragment demonstrating how the combination (A,P) $\rightarrow$ (I,S) successfully stabilizes both intensity and tone in image sequences.

5.1.3 Video Intensity and Tone Stabilization

Among the studied methods, we found the original Anbarjafari method to be the best for intensity stabilization in single images. Yet, this method does not handle tone stabilization in video sequences. Consider Fig. 6, left column, which shows a selection of frames from a video of a bleeding gastrointestinal tissue. The first five frames are identical to those in Fig. 1, bottom row. As also outlined in Sec. 1, a certain amount of tone fluctuation is visible even in this short sequence.

We next show how the combination of Anbarjafari and our method solves this problem. First, as a baseline, we apply only our method to the saturation channel (P $\rightarrow$ S), see Fig. 6 middle column, with a time window $k = 40$, compression $a = 0.04$, and correction $c = 0.4$, in line with the optimal values found for our method (P) in the survey. We see how the sudden tone changes have now been smoothed out – all frames in Fig. 6, middle column, have a pinkish tone. The tone stabilization is even more evident when watching the actual video. However, the intensity is not stabilized. To solve this, we apply the combination of Anbarjafari and our method to both the intensity and saturation channels ((A,P) $\rightarrow$ (I,S)), see Fig.6, right column. In addition to the previous parameters, we use a mean goal $\gamma = 0.7$, shown to be optimal in our survey (Sec. 5.1.2). As visible, especially for frames 2659 and 2709, the intensity is more uniform now; in addition, the tone fluctuations are low, thanks to our method. All in all, we conclude that the combination (A,P) $\rightarrow$ (I,S) is indeed a good way to stabilize *both* intensity and tone fluctuations.

5.2 Quantitative Evaluation

The qualitative evaluation of the various combinations of methods and parameters in Sec. 5.1 has empirically found good parameter values that yield images perceived by users as stabilized. However, as explained already in Sec. 1, stabilization should not create artifacts which could lead to misinterpretation of the imaged tissue structures. Formally put, stabilization can be thought of a function $\Phi(\gamma, a, I_{input}) = I_{stabilized}$ from images to images which aims to maximize *both* the temporal stability of intensity and tones *and* in the same time minimize the perceptual difference between the original and stabilized images. The behavior of this function is driven by our method's free parameters, of which the most important are the goal mean γ (for Anbarjafari) and the compression a (for our histogram-based compression). To study how Φ affects image similarity, we need a way to compare I_{input} and $I_{stabilized}$. For this, similarly to (Moradi et al., 2015), we use the peak-to-signal noise ratio (PSNR) and structural similarity index (SSIM) metrics, well known in image processing. For 8-bit-per-

Table 2: Image-quality survey results accumulated for all five tested endoscopy image classes.

	original	$\gamma = 0.6, a = 0.02$	$\gamma = 0.7, a = 0.04$	$\gamma = 0.8, a = 0.08$	$\gamma = 0.9, a = 0.16$	$\gamma = 1, a = 0.32$
A → I	6 / **6**	18 / **7**	40 / **12**	24 / **3**	2 / **0**	0 / **0**
A → (I,S)	5 / **5**	14 / **5**	55 / **17**	16 / **3**	0 / **0**	0 / **0**
P → I	11 / **7**	41 / **4**	28 / **6**	7 / **0**	3 / **2**	0 / **0**
P → (I,S)	9 / **7**	41 / **7**	30 / **2**	8 / **1**	2 / **0**	0 / **0**
(A,P) → I	8 / **7**	50 / **6**	21 / **3**	11 / **1**	0 / **0**	0 / **0**
(A,P) → (I,S)	8 / **7**	57 / **7**	23 / **6**	2 / **0**	0 / **0**	0 / **0**

channel images like ours, typical PSNR values for good similarity are between 30 and 50 dB, where higher is better (Huynh-Thu and Ghanbari, 2008). SSIM ranges between -1 and 1 where higher is better (1 denotes identical images) (Wang et al., 2004).

Figure 7 shows the plots of the PSNR and SSIM similarity metrics between the original endoscopy images I_{input} and the stabilized ones $I_{stabilized}$ as function of the key parameters γ (for Anbarjafari) and a (for our method), for the set of images used in our qualitative analysis (see Sec. 5.1.1), and for fixed values of $k = 20$ and $c = 0.4$. As methods, we considered Anbarjafari applied on intensity (A → I) and separately on saturation (A → S), and our method applied on intensity (P → I) and separately on saturation (P → S). From these plots we make the following observations.

Quality: The A → I method peaks for both PSNR and SSIM at γ very close to 0.5, *i.e.*, the mean intensity of I_{input}. This is expected: If the goal mean equals the original mean, no correction needs to be done, as $I_{stabilized}$ is identical to I_{input}. In contrast, A → S peaks at values around $\gamma = 0.7$. This matches very well the optimal γ values found in our qualitative study (Sec. 5.1). Hence, the γ values found best by users to explore the images is *also* the one where the least changes are done by stabilization. Moreover, the maximal PSNR values (over 50 dB) and SSIM values (close to 1) indicate that our stabilization looses very little from the original image features. Separately, we see that both SSIM and PSNR have very good values for a close to 0.04, which was found earlier in our qualitative studies to yield a very good tone stabilization (Secs. 5.1.2 and 5.1.3). This confirms that our preset $a = 0.04$ is indeed a good one.

Intensity *vs* Saturation Stabilization: The plots for A → I and A → S are very similar in shape and magnitude. This matches our earlier qualitative finding that the Anbarjafari method can be used to stabilize both intensity and saturation (Sec. 5.1). In contrast, the plot for P → S is always larger than P → I. This means that our proposed method P is better at stabilizing saturations (tones) than intensities, which again correlates with our qualitative findings.

Parameter Sensitivity: The plots for A → I and A → S have overall quite high derivatives close to the maximum, while the plots for P → I and P → S show a much more stable, and actually monotonic, variation. This tells that setting the compression a for the P method is less sensitive than setting the mean goal γ for the A method. However, this does not mean that tuning γ is sensitive: As explained above, we obtain a very good image quality for values around $\gamma = 0.5$ for the method A → I, and respectively for values around $\gamma = 0.7$ for the method A → P. All in all, we conclude that parameter setting is not a sensitive process.

Consistency and Smoothness: Across the five frames, plots for the same method are similar in shape, position, and peak location. This is desirable, as it tells that optimal parameter values are consistent for quite different inputs. Considering the earlier parameter sensitivity analysis, the parameter *presets* proposed in Sec. 5.1 can be indeed used as default values for entire videos. This makes our method basically parameter-free. Secondly, the plots are smooth, with no jitters, which tells that small parameter-value changes do not massively affect the image similarity. Hence, our method is robust *vs* parameter changing, if users really need to change the preset values.

6 CONCLUSIONS

We propose a new method for jointly stabilizing intensity and tone (hue) in endoscopy videos. For this, we adapt the intensity channel by brightening dark areas and darkening too bright areas, and also minimize tone fluctuations between temporally close frames. Our method is simple to implement, works at the frame-rate of the pill camera, has no free parameters that users should set, delivers consistent results for a wide variety of endoscopy videos, and alters only minimally the input images, thereby reducing the risk of creating misleading artifacts. Summarizing, our main contributions are:

Survey: To our knowledge, our work is the first in which a large set (10) of imaging methods was studied for suitability for the specific case of endoscopy

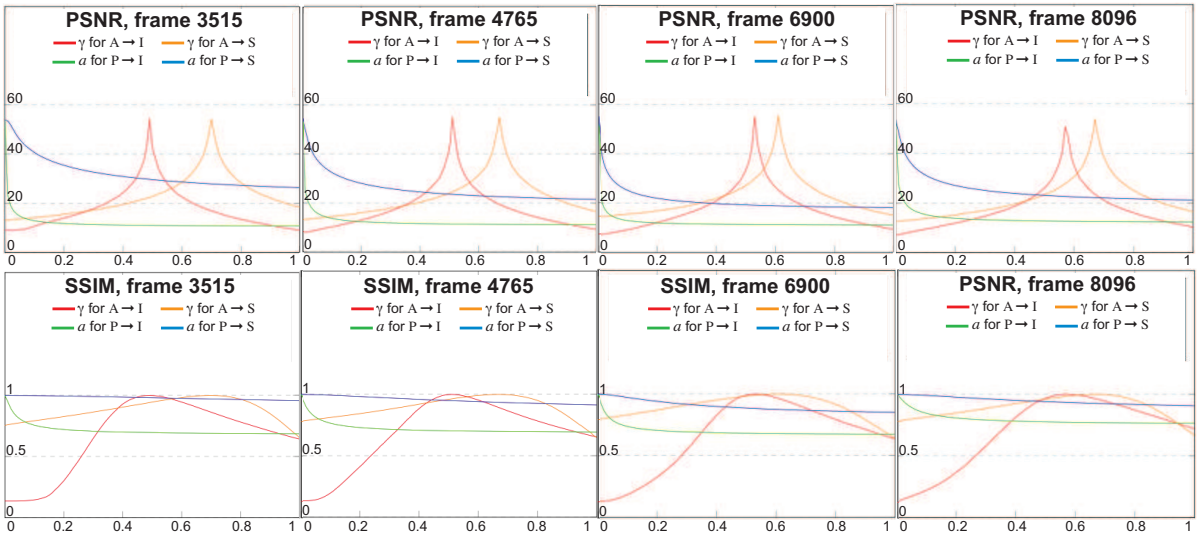

Figure 7: PSNR and SSIM image-similarity plots for several frames from Fig. 1 processed with Anbarjafari and our method. The horizontal axis denotes either the goal mean γ or the compression factor a depending on the graph type.

video stabilization, from a practical perspective including validation, reproductibility, computational complexity, and ease of use.

Joint Stabilization: While several methods perform intensity stabilization, we show how both intensity and tone can be jointly stabilized. For the former, we use an existing method (Anbarjafari, 2014). For the latter, we propose a simple but efficient method based on histogram compression.

Validation: Compared to existing work, we perform a significantly more thorough validation including testing several method types applied on intensity and/or saturation; a detailed user study for finding good method combinations and parameter values; and a quantitative evaluation that shows how to find parameter presets which match the values suggested by our qualitative study and also minimally affect image quality. This makes our method fully parameter-free and guarantees its output quality. Our method can be easily and efficiently implemented.

Limitations: Our search of the algorithm-and-parameter space is, of course, not exhaustive. More methods and parameter values exist which could be assessed. It is also fair to say that our current evaluation already surpasses what one typically encounters in endoscopy video stabilization papers. Separately, one can argue that the differences between the original and stabilized images are quite small, so the entire stabilization process is not worthwhile. Yet, when watching the actual stabilized videos, these differences are well visible, and show that the stabilized material is easier to follow.

Several future work directions exist. More extensive evaluations can be made to compare with additional color stabilization methods, use more videos, or a more users. Machine learning techniques could be used to perform a more fine-grained stabilization based on images or image regions labeled by users as requiring brightening.

ACKNOWLEDGEMENTS

We thank Medisch Centrum Leeuwarden for providing us the capsule endoscopy videos.

REFERENCES

Anbarjafari, G. (2014). HSI based colour image equalization using iterative n^{th} root and n^{th} power. arXiv:1501.00108 [cs.CV].

Bassiou, N. and Kotropoulos, C. (2007). Color image histogram equalization by absolute discounting back-off. *CVIU*, 107(1):108–122.

Classen, M. and Phillip, J. (1984). Electronic endoscopy of the gastrointestinal tract. *Endoscopy*, 16(1):16–19.

Farbman, Z. and Lischinski, D. (2011). Tonal stabilization of video. *ACM Trans Graph*, 30(4):89–101.

Funt, B., Barnard, K., Brockington, M., and Cardei, V. (1997). Luminance-based multi-scale retinex. In *Proc. 8th Congress of the International Colour Association.*

Gautam, C. and Tiwari, N. (2015). Efficient color image contrast enhancement using range limited bi-histogram equalization with adaptive gamma correction. In *Proc. IEEE ICIC.*

Gijsenij, A., Gevers, T., and van der Weijer, J. (2011). Computational color constancy: Survey and experiments. *IEEE Trans Imag Process*, 20(9):2475–2489.

González, D. M., Ponomaryov, V., and Kravchenko, V. (2016). Chromaticity improvement in images with poor lighting using the multiscale-retinex MSR algorithm. In *Proc. IEEE MSSW*.

Hale, M., Sidhu, R., and McAlindon, M. (2014). Capsule endoscopy: Current practice and future directions. *World J Gastroenterol*, 20(24):7752–7759.

Huynh-Thu, Q. and Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment. *Electron Lett*, 44(13).

Kim, T. and Yang, H. (2006). A multidimensional histogram equalization by fitting an isotropic Gaussian mixture to a uniform distribution. In *Proc. IEEE ICIP*, pages 2865–2868.

Medivators (2017). MiroCam capsule endoscope. www.medivators.com/products/gi-physician-products/mirocam-capsule-endoscope.

Moradi, M., Falahati, A., Shahbahrami, A., and Zare-Hassanpour, R. (2015). Improving visual quality in wireless capsule endoscopy images with contrast-limited adaptive histogram equalization. In *Proc. IPRIA*. IEEE.

Purushothaman, J., Kamiyama, M., and Taguchi, A. (2016). Color image enhancement based on hue differential histogram equalization. In *Proc. ISPACS*, pages 322–331. IEEE.

Sun, C. C., Ruan, S. J., Shie, M. C., and T, W. P. (2005). Dynamic contrast enhancement based on histogram specification. *IEEE Trans Consum Electr*, 51(4):1300–1305.

Vazquez-Corral, J. and Bertalmio, M. (2014). Color stabilization along time and across shots of the same scene, for one or several cameras of unknown specifications. *IEEE Trans Imag Process*, 23(10).

Vig, N., Budhiraja, S., and Singh, J. (2016). Hue preserving color image enhancement using guided filter based sub image histogram equalization. In *Proc. 9th Intl. Conf. on Contemporary Computing (IC3)*.

Wang, Y., Tao, D., Li, X., Song, M., Bu, J., and Tan, P. (2014). Video tonal stabilization via color states smoothing. *IEEE Trans Image Process*, 23(11).

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans Imag Process*, 13(4):600–612.