Real Time Autonomous Video Image Registration for Endomicroscopy: Fighting The Compromises

Tom Vercauteren*[a,b], Alexander Meining[c], François Lacombe[a], Aymeric Perchant[a]

[a]Mauna Kea Technolgies, 9 rue d'Enghien, Paris, France;
[b]Asclepios Reasearch Group, INRIA, Sophia Antipolis, France;
[c]Klinikum Rechts der Isar, TUM, Munich, Germany

ABSTRACT

Confocal endomicroscopy provides tools for *in vivo* imaging of human cell architecture endoscopically. These technologies are a tough challenge since multiple trade-offs have to be overcome: resolution versus field of view, dynamic versus stability, contrast versus low laser power or low contrast agent doses. Many difficult clinical applications, such as lung, bile duct, urethral imaging and NOTES applications, need to optimize miniaturization, resolution, frame rate and contrast agent dose simultaneously. We propose one solution based on real-time video image processing to efficiently address these trade-offs.

Dynamic imaging provides a flow of images that we process in real time. Images are aligned using efficient algorithms specifically adapted to confocal devices. From the displacement that we find across the images, instantaneous velocities are computed and used to compensate for motion distortions. All images are stitched together onto the same reference space and displayed in real-time to reconstruct an image of the entire surface explored during the clinical procedure. This representation brings both stability and an increased field of view. Moreover, because a given area can be imaged by several frames, the contrast can be improved using temporal adaptive averaging. Such processing enhances the visualization of the video sequence, overcoming most classical trade-offs. The stability and increased field of view help the clinician better focus his attention on his practice which improves the patient benefit. Our tools are currently evaluated in a multicenter clinical trial to assess the improvement of the clinical practice.

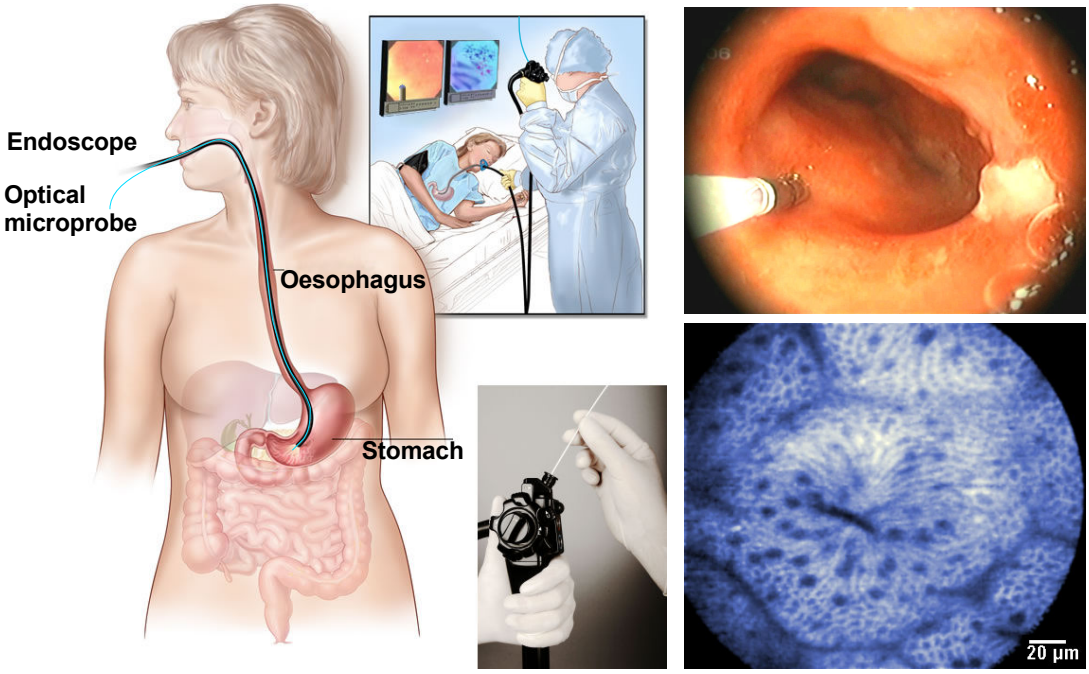
Keywords: Cellvizio®, microscopy, mosaicing, real-time

1. INTRODUCTION

Current *in vivo* imaging technologies such as MRI, CT, PET, cytometrics, bioluminescence, fluorescence tomography, high-resolution ultrasound or SPECT are only capable of producing images at resolutions between 30 µm and 3 mm. This range of resolution, while largely acceptable for a wide scope of applications, is insufficient for cellular level imaging. On the other side of the resolution range, we find several types of microscopy. The vast majority of microscopes, whether conventional or confocal, are limited for use with cell cultures or *ex vivo* tissue samples. The tiny fraction of microscopes that are dedicated to *in vivo* use, and that can function inside the living organism, are called intravital microscopes. These apparatus are cumbersome, difficult to put into use and restricted to research use on small animals.[1] They also require a very delicate and specific preparation of the animal which includes installing a window on the animal through which the microscope can look into the body.

A promising tool to fill this gap is given by fibered confocal microscopy[2–4] (FCM). Classical confocal microscopy is an established optical imaging technique that can be used to obtain high-resolution images of cells on tissue samples or cell cultures. Translation of this technology for *in vivo* applications can be achieved by using optical fibers, miniature optics and robust laser scanning approaches. Fibered confocal microscopy, and especially Cellvizio®, developed by Mauna Kea Technologies, Paris, allows clinicians and biologists to easily get a real-time view of cellular structures without removal of biological tissue. Ultimately, these tools should enable the practitioner to perform what can be referred to as an *optical biopsy*: a histological examination of biological tissues *in vivo* and *in situ*, i.e., in the living organism directly onto the tissue of interest.

*tom.vercauteren@maunakeatech.com



(a) (b)

Figure 1: Illustration of a typical Cellvizio®-GI setting. (a) The miniaturized flexible optical microprobe is small enough to fit into the accessory channel of any flexible endoscope. The optical microprobe is made to be used like biopsy forceps, its flexibility allows most anatomical configuration of the endoscope. (b) Wide field-of-view imaging remains available when the optical microprobe is in contact with the tissue and microscopic imaging comes in.

As interesting as dynamic sequences may be during the time of the medical procedure or biological experiment, there is a need for the expert to get an efficient and complete representation of the entire imaged region. Since fibered confocal microscopy is a contact imaging modality, there is also an inevitable hardware trade-off between resolution, field-of-view and invasiveness.

In previous work,[5,6] we proposed a post-processing robust mosaicing algorithm. This offline scheme can automatically combine successive frames of the acquired video sequence, cancel motion artifacts, and reconstitute wide FOV images of the tissues. To image and explore a region of interest, the confocal microprobe needs simply be glided along the tissue, either hand-held or through the endoscope bending. With the aid of this mosaicing software, we have shown that it is possible to get at the same time a microscopic resolution and a wide field-of-view without having to increase the size of the optical microprobe and thus the invasiveness.[6]

What makes the interest of Cellvizio® is not only its capability of acquiring microscopic images of tissues *in vivo* but mainly the fact that it allows to do so in real time with a direct visualization by the clinician. It is exactly this that makes it possible to unify, within a single procedure, the disease suspicion made during endoscopy, the actual diagnosis and the treatment.

The mosaics we create with our post-processing scheme are very useful for a careful inspection of the Cellvizio® data after the endoscopy and for inclusion in the patient record. An ideal situation would however be to have a mosaic constructed in real time, on the fly during endoscopy as shown in Fig. 2. Such a tool would also help the investigator evaluate whether the area intended for examination was adequately targeted.

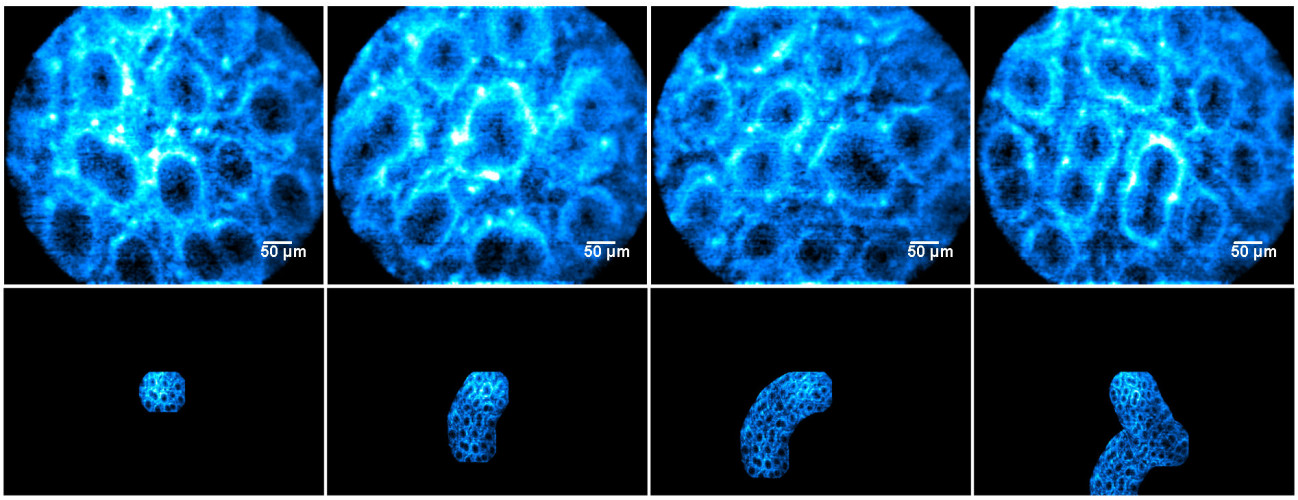


Figure 2: The concept of live mosaicing on a sequence of lymphocytic colitis. As time goes by (left to right), the current image (top row) is roughly register and stitched to the current mosaic (bottom row). We therefore have a growing mosaic. Courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich.

2. REAL-TIME USER FEEDBACK TO IMPROVE THE ACQUISITION

2.1 Our Approach to Live Mosaicing

The computational burden imposed by running our full post-processing mosaicing pipeline makes it impossible to get such a detailed mosaic reconstruction in real time. In a recent work,[7] Loewke *et al.* proposed to use a robot to hold the probe and a set of sensors to get an estimate of the global positioning of the frames. While such an external information allows to ease the computational cost, using a robot will definitely not fit into the current endoscopic procedure. It will also be cumbersome and more invasive as sensor needs to be placed at the tip of the optical microprobe.

Our approach to real-time mosaicing will thus focus on keeping the ease of use of Cellvizio® at least at its current state. This implies that we have to resort to using a purely algorithmic solution to the problem of real-time mosaicing. To let this be feasible, we will fallback to fast but less accurate models than the one presented in our previous work[6] and also fast but less precise reconstruction schemes.

In addition to the *real-time field-of-view enhancement* target, a major goal of our live mosaicing tool lies in the training of the practitioner. Thanks to the direct visual feedback that the live mosaicing provides, clinicians can assess in real time whether their acquisition is smooth and of good image quality. As shown in Fig. 3, the live mosaicing will indeed provide nice-looking mosaics as long as the motion of the optical microprobe is smooth and as long as the image quality is sufficient to register pairs of consecutive images. When it is not the case, the clinician will either see no mosaic being constructed or a rather random stitching of the input images. This is of major interest for the Cellvizio® system because without such feedback, it is often difficult for the user to assess the quality of his acquisition. Early qualitative evaluation has shown that the live mosaicing tool has the effect of reducing the learning curve needed to harness Cellvizio®. It can even improves the quality of the acquisition for experienced users.

2.2 Real-time Mosaicing Algorithm

Let us now focus on how our live mosaicing algorithm manages to register the images and real-time and compensate for some motion distortion.

2.2.1 Spatial Transformations and Algorithm Overview

An interesting point of scanning imaging devices such as Cellvizio® is that the output image is not a representation of a given instant, but a juxtaposition of points acquired at different times. Instead of motion blur, we get geometric distortions, e.g. a circle is distorted into an ellipse. During a rigid motion translation of a given

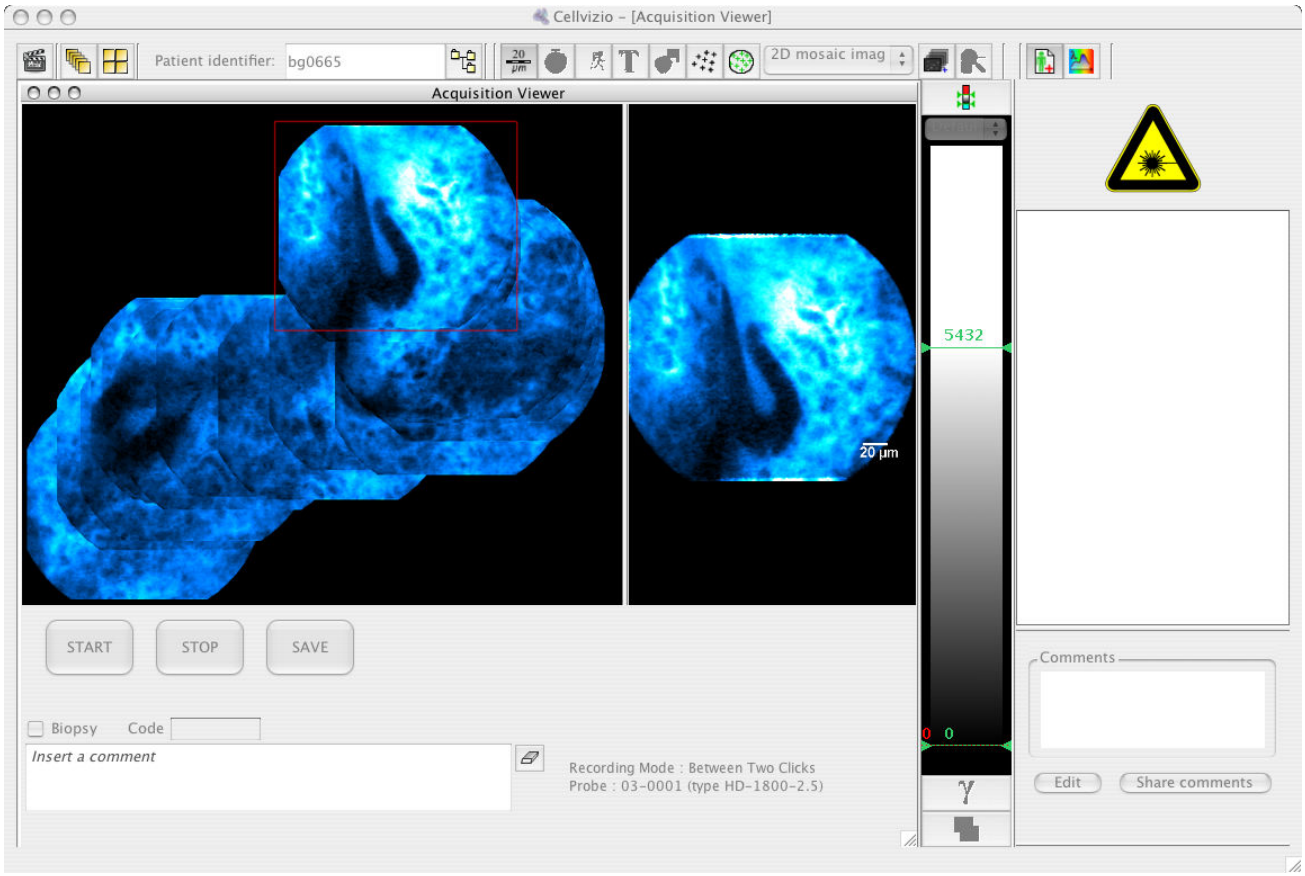


Figure 3: The graphical user interface of the live mosaicing. The clinician can see on the same screen, the regular acquisition view on the right and the simple mosaic, stitched on-the-fly, on the left. Courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich.

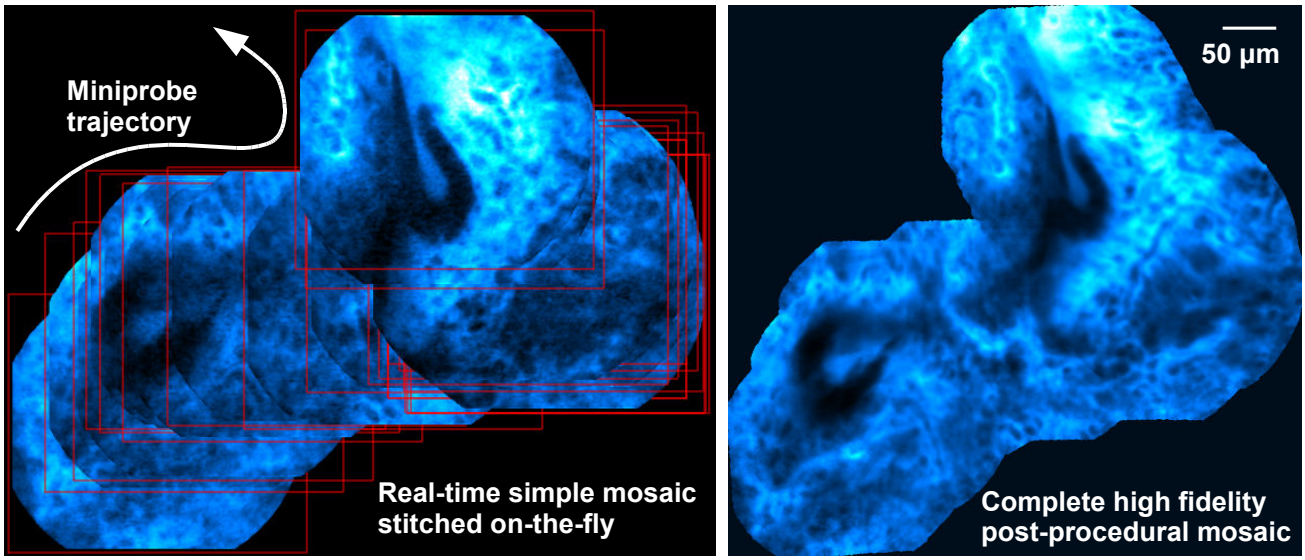


Figure 4: Comparison of the real-time mosaic and the post-processing mosaic on a Cellvizio® sequence of a severe ulcerative colitis. Courtesy of PD. Dr. A. Meining, Klinikum rechts der Isar, Munich.

object in the field of view, the transformation in image coordinates is a translation composed with a skew transformation.[8] This feature has been successfully used for red blood cell velocimetry on single images.[8]

The scanning device we use, Cellvizio®, has a fast horizontal scanning, and a slow vertical scanning. We can assume[8] that the horizontal scanning has an infinite speed, and that the vertical scanning is described by the vertical position of the scanned point: $y_s(t) = v_y.t$, where $v_y =$ is the vertical scanning velocity.

Let us suppose that the motion between two contiguous frames is a translation at speed $\eta = [\eta^x, \eta^y]$. A scanned line with vertical position y, will be sampled at the time $t(y) = t(0) + \frac{y}{v_y}$. During the scanning, a point $p = [x, y]$ in the image reference plane will be sampled at position $p_d = [x + \eta^x y, (1 + \eta^y)y]$ in the object reference plane. This linear transformation is noted v_k. Each point p of a frame k is mapped to a reference space coordinate system by the transformation $f_k(p) = r_k \circ v_k(p)$. Between two frames j and k, the transformation is then given by:

$$f_{j,k}(p) = v_j^{-1} \circ r_j^{-1} \circ r_k \circ v_k = v_j^{-1} \circ r_{j,k} \circ v_k. \tag{1}$$

The estimation of the velocities is done using only the translation part of the $r_{j,k}$. This velocity is used in the following algorithm to compensate for the scanning distorsions. For each contiguous frames the following steps are performed:

1. Estimation of the translation using a 2D normalized cross correlation
2. Estimation of the velocity from the translation
3. Computation of the distortion transformation
4. Estimation of the rigid transformation

2.2.2 Image Registration

The fast normalized correlation matching algorithm[9] allows us to estimate in real time the translation component of (1). The main idea of the fast normalized correlation, is to evaluate, in one pass, the correlation coefficient between the fixed image F and the translated moving image $M \circ \tau$ for every translation τ with integer components.

This algorithm has been designed for template matching. As such it is theoretically correct when the support of $M \circ \tau$ is included in the support of F. It is not completely accurate to use it when F and M have the same size. In practice however it works also well in this case and shows manageable border effects. Here is a very brief overview of this scheme. The similarity criterion can be written as:

$$\mathrm{Sim}\,(F, M \circ \tau)\, \frac{\sum_p (F(p) - \bar{F})(M(p+\tau) - \bar{M})}{\sqrt{\sum_p (F(p) - \bar{F})^2 \sum_p (M(p+\tau) - \bar{M})^2}} = \gamma(-\tau), \tag{2}$$

where $\bar{F}$ is the mean of F and $\bar{M}$ is the mean of M. Let us look at the numerator of (2). Let $F'(p) = F(p) - \bar{F}$, $M'(p) = M(p) - \bar{M}$ and $M'_{rev}(p) = M'(-p)$. We see that

$$\gamma^{num}(\tau) = \sum_p F'(p) M'_{rev}(\tau - p) \tag{3}$$

is the convolution of the normalized fixed image F' with the reversed normalized moving image M'_{rev}. This can efficiently be computed with the Fourier transform $\mathcal{F}$:

$$\gamma^{num}(\tau) = \mathcal{F}^{-1}\big(\mathcal{F}(F')\,\mathcal{F}(M'_{rev})\big) = \mathcal{F}^{-1}\big(\mathcal{F}(F')\,\mathcal{F}^*(M')\big),$$

where we used the convolution theorem and the fact that for a real signal, time reversal of the signal is accomplished through complex conjugate of the Fourier transform. Furthermore running sums that compute the integral of the image intensity and the integral of the squared image intensity are used to compute the denominator.[9] Once the full correlation coefficient map has been computed, we only need to find its maximum to get the optimal translation.

Besides its fast computation time and global optimality properties, this algorithm has the very nice property of not requiring any gradient-descent like loop. This is a very important property for real-time algorithms that needs to terminate their computations before the deadline of the next event. If a gradient-descent like scheme were to be used, the only way we could enforce the worst-case execution time would be to limit the number of iterations. This implies that it would become possible not to reach convergence.

2.2.3 Mosaic Visualization

Thanks to the fast normalized correlation matching and to the relationship between velocity and motion distortion, we have a way to register in real time the consecutive frames of the acquired Cellvizio® sequence and compensate for the motion distortion. What we now need to do is to display a mosaic based on these alignments. Depending on the available computational power and on the specific applications, several options are possible.

The first possibility is to use an advanced mosaic reconstruction scheme such as the one used for post-processing mosaicing.[6] Because a given area is in general imaged by several frames, temporal averaging leads to an improvement of signal-to-noise ratio, and image contrast. In several cases we even achieve some super-resolution in addition to extending the field of view.

In some cases, we need to keep the computational requirement as low as possible. An alternative to the advanced reconstruction schemes, is to use a simple *dead leaves model*. The current frame is simply overlaid on top of the previous frames at a position dictated by the registration. In addition to requiring very little computational time, this approach has the advantage on not being too sensitive to registration errors. As we do not mix the information of several images, a small misregistration will not lead to a blurred image but rather to an image where seams are still visible at the edge of the images, as shown in Fig. 3. If an undetected gross registration error appears, the good thing is that we still visualize correctly the information of the current frame.

For both approaches, it is advantageous to detect gross registration errors and wipe-out the history when such a discontinuity is detected. The scheme we use needs to run in real time and should use as little processing as necessary. From the fast normalized correlation matching, we trivially have access to the correlation coefficient between the images for the optimum translation. We found that, for the real-time algorithm, a simple threshold on this correlation coefficient was sufficient to provide decent visual comfort to the clinician.

3. CLINICAL EVALUATION OF LIVE MOSAICING

Our goal in this section is to measure the value of our tools on a well-defined medical problem, using standard real-life clinical procedures. In this work, we chose to focus on the surveillance of a pre-cancerous condition known as Barrett's esophagus. The strongest evidence one can get to support a clinical hypothesis is provided by a systematic review of a randomized multicenter clinical trial. We have thus integrated our mosaicing tool within such a trial whose main goal is to measure the performance of Cellvizio® with respect to the current gold standard for Barrett's esophagus. This trial allows us to get a fair evaluation of our mosaicing tools for a specific medical application.

3.1 A Multicenter Clinical Trial for the Diagnosis of Barrett's Esophagus

Barrett's esophagus refers to an abnormal change in the cells of the lower end of the esophagus. It is considered to be a pre-malignant condition and is associated with an increased risk of esophageal cancer. As shown in Fig. 5, Barrett's esophagus is visible grossly through an endoscope, but the tissue must be examined at the microscopic level to confirm the diagnosis and determine the malignancy of the cells.

It is widely accepted that the current gold standard for the surveillance of Barrett's esophagus, which is based on somewhat random biopsies,[12] is, at the same time, not specific enough, not reproducible enough and uncomfortable for the patient.[13]

We have mentioned that fibered confocal microscopy allows the physician to get a microscopic view of the tissues during endoscopy. In this case, we are thus not limited by the number of biopsies and can expect a diagnosis that is more specific, more reproducible and more comfortable for the patient. The main goal of the multicenter clinical trial on Barrett's esophagus is to assess the performance of fibered confocal microscopy with respect to the current gold standard. The trial is sponsored by Mauna Kea Technologies. The investigators are

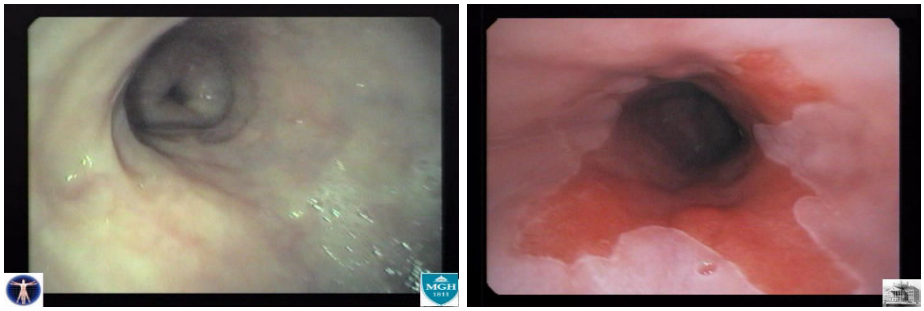


Figure 5: Endoscopic images of the esophagus. Left: healthy esophagus. Right: Barrett's epithelium is recognizable by its salmon color whereas the normal mucosa has a pearly white appearance. Images taken from the DAVE project.[10,11]

PD. Dr. Alexander Meining, Technical University Munich (lead investigator), Pr. Dr. Thomas Rösch Charite University, Berlin and Pr. Dr. Stephan Miehlke, Dresden University of Technology. This trial also serves to validate our mosaicing algorithm from an applicative point of view.

3.2 Results

Table 1: A qualitative evaluation of the usefulness of the live mosaicing during GI endoscopy. Three different investigators have been using Cellvizio® for some time and have then been given the opportunity to test live mosaicing. They answered the following questions after at least several tens of Cellvizio® procedures with live mosaicing. The answers written in the table are direct quotes from the investigators.

	A	B	C
How often do you use the live mosaicing mode with respect to the classical acquisition mode?	Every patient	Now 100%	Every patient
When using the live mosaicing mode, how often do you find the real-time mosaic to be informative?	80%	100%	75%
Are there some protocols for which the live mosaicing does not bring any added value?	No	No	I think it is always an advantage. It provides an idea, how stable you are and how far you slip with the probe. It also gives you an idea on how to cover more than just one spot in case the area is very stable.
When using the live mosaicing mode, how often do you look at the mosaic window with respect to the movie only window?	Stable picture: no need to look at mosaicing. Movement: look at mosaicing window only.	80–90%	50%

As shown in Fig. 4, even in simple mosaicing problems, the output of the live mosaicing naturally does not matches the output of the post-processing mosaicing algorithm in terms of image quality, image details and signal to noise ratio. It however still does a good job for a real-time algorithm.

In order to judge the clinical relevance of our live mosaicing tool, we have decided to perform a qualitative evaluation based on clinical expertise. The three investigators involved in the clinical trial had been working with Cellvizio® for some time before we gave them the live mosaicing software shown in Fig. 3. After a short training period, the investigators were given the opportunity to use the live mosaicing or not. After a few tens of patients each, we asked them to rate the usefulness of our tool based on a few questions that are listed in Table 1. It can be seen from the results that all the investigators chose to keep using the live mosaicing during the acquisition.

This fact only should support the clinical relevance of the scheme. With a subjective perspective, we also found that the data we received from the investigators became better suited for the post-processing mosaicing solution.

4. CONCLUSION

Fibered confocal microscopy offers, in particular, unprecedented characterization capabilities of the GI mucosa in real time during endoscopy. This can potentially improve the specificity and accuracy of the diagnosis while preserving patient comfort. This may in turn limit the cost of the current Barrett's esophagus surveillance protocol.

However, multiple trade-offs have to be overcome to meet the clinical requirements. For example, the quest for non-invasiveness (organs should not be damaged) often leads to tissue or imaging device motions that result in motion artifacts and small field of view.

In this paper, we proposed a real-time video mosaicing solution that allows us to move beyond most of the hardware trade-offs. This representation brings both stability and an increased field of view. By integrating our tools within a multicenter clinical trial, we have been able to measure their clinical relevance. Thanks to the direct visual feedback that the investigators get from the live mosaicing, their attention is more focused on their practice and this helps them improve patient benefit.

REFERENCES

1. D. M. McDonald and P. L. Choyke, "Imaging of angiogenesis: From microscope to clinic," *Nat. Med.* **9**, pp. 713–725, June 2003.
2. G. Le Goualher, A. Perchant, M. Genet, C. Cavé, B. Viellerobe, F. Berier, B. Abrat, and N. Ayache, "Towards optical biopsies with an integrated fibered confocal fluorescence microscope," in *Proc. MICCAI'04*, C. Barillot, D. R. Haynor, and P. Hellier, eds., *Lect. Notes Comput. Sci.* **3217**, pp. 761–768, Springer-Verlag, 2004.
3. A. Meining, M. Bajbouj, S. Delius, and C. Prinz, "Confocal laser scanning microscopy for in vivo histopathology of the gastrointestinal tract," *Arab J. Gastroenterol.* **8**, pp. 1–4, Mar. 2007.
4. L. Thiberville, S. Moreno-Swirc, T. Vercauteren, E. Peltier, C. Cavé, and G. Bourg Heckly, "*In vivo* imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy," *Am. J. Respir. Crit. Care Med.* **175**, pp. 22–31, Jan. 2007. Chosen for the cover of the AJRCCM paper issue.
5. T. Vercauteren, A. Perchant, X. Pennec, and N. Ayache, "Mosaicing of confocal microscopic *in vivo* soft tissue video sequences," in *Proc. MICCAI'05*, J. S. Duncan and G. Gerig, eds., *Lect. Notes Comput. Sci.* **3749**, pp. 753–760, Springer-Verlag, 2005.
6. T. Vercauteren, A. Perchant, G. Malandain, X. Pennec, and N. Ayache, "Robust mosaicing with correction of motion distortions and tissue deformation for *in vivo* fibered microscopy," *Medical Image Analysis* **10**(5), pp. 673–692, 2006. Annual MedIA/MICCAI Best Paper Award 2006.
7. K. E. Loewke, D. B. Camarillo, C. A. Jobst, and J. K. Salisbury, "Real-time image mosaicing for medical applications," in *Proc. MMVR'07*, pp. 304–309, Feb. 2007.
8. N. Savoire, G. Le Goualher, A. Perchant, F. Lacombe, G. Malandain, and N. Ayache, "Measuring blood cells velocity in microvessels from a single image: Application to *in vivo* and *in situ* confocal microscopy," in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07)*, pp. 456–459, Apr. 2004.
9. J. P. Lewis, "Fast template matching," in *Proc. VI'95*, pp. 120–123, 1995.
10. P. B. Kelsey, "Esophagus - Barrett's esophagus." The DAVE Project, May 2004.
11. P. B. Kelsey, "Stomach - Normal upper endoscopy." The DAVE Project, Jan. 2005.
12. R. E. Sampliner and The Practice Parameters Committee of the American College of Gastroenterology, "Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus," *Am. J. Gastroenterol.* **97**, pp. 1888–1895, Aug. 2002.
13. P. Sharma, S. Wani, and A. Bansal, "The quest for intestinal metaplasia - Is it worth the effort?," *Am. J. Gastroenterol.* **102**, pp. 1162–1165, June 2007.