An optimized algorithm of image stitching in the case of a multi-modal probe for monitoring the evolution of scars

Rami Kassab, Sylvie Treuillet, Franck Marzani, Christian Pieralli, Jean-Christophe Lapayre

An optimized algorithm of image stitching in the case of a multi-modal probe for monitoring the evolution of scars

R. Kassab[a], S. Treuillet[b], F. Marzani[*c], C. Pieralli[a], J.C. Lapayre[a]
[a]FEMTO-ST Institute, UMR CNRS, Franche-Comte University, 16 route de Gray, 25030 Besançon Cedex, France; [b]Laboratoire PRISME, Université d'Orléans, 12 rue de Blois, 45067 Orléans Cedex 2, France; [c]Le2i Lab., UMR CNRS 6306, Bourgogne University, B.P. 47870, 21078 Dijon Cedex, France

ABSTRACT

We propose a new system that makes possible to monitor the evolution of scars after the excision of a tumorous dermatosis. The hardware part of this system is composed of a new optical innovative probe with which two types of images can be acquired simultaneously: an anatomic image acquired under a white light and a functional one based on autofluorescence from the protoporphyrin within the cancer cells. For technical reasons related to the maximum size of the area covered by the probe, acquired images are too small to cover the whole scar. That is why a sequence of overlapping images is taken in order to cover the required area.

The main goal of this paper is to describe the creation of two panoramic images (anatomic and functional). Fluorescence images do not have enough salient information for matching the images; stitching algorithms are applied over each couple of successive white light images to produce an anatomic panorama of the entire scar. The same transformations obtained from this step are used to register and stitch the functional images. Several experiments have been implemented using different stitching algorithms (SIFT, ASIFT and SURF), with various transformation parameters (angles of rotation, projection, scaling, etc…) and different types of skin images. We present the results of these experiments that propose the best solution.

Thus, clinician has two panoramic images superimposed and usable for diagnostic support. A collaborative layer is added to the system to allow sharing panoramas among several practitioners over different places.

Keywords: Scars evolution monitoring, stitching, multi-modal probe, autofluorescence.

1. INTRODUCTION

Skin cancer is the most common form of human cancer. It is estimated that over 1 million new cases occur annually. The annual rates of all forms of skin cancer are increasing each year, representing a growing public concern. It has also been estimated that nearly half of all Americans who live to age 65 will develop skin cancer at least once [L1, L2].

The most common warning sign of skin cancer is a change in the appearance of the skin, such as a new growth or a sore that will not heal. The term "skin cancer" refers to three different conditions. From the least to the most dangerous, they are either basal cell carcinoma (or basal cell carcinoma epithelioma), or squamous cell carcinoma (the first stage of which is called actinic keratosis) or melanoma.

After eradication operations, the risk of the existence of effected cells is big, and the danger is that these cells could redeploy on the skin, this is why an accurate observation of the scar is required to assure that no cells are still affected anymore. Different optical methods for monitoring have been developed over the time starting from clinical methods of detection and characterization where only the personnel experience of practitioners plays the principle role, to the spectroscopic methods of diagnosis based on the fact that the spectral emission of a tissue returns information about its historical, morphological, and biochemical nature. Fluorescence, elastic scattering and Raman scattering are types of spectroscopic methods of diagnosis.

[*] franck.marzani@u-bourgogne.fr; phone +33 380 39 63 33; fax +33 380 39 59 10; le2i.cnrs.fr

In terms of imaging, color and gray scale images provide interesting anatomical information; they can be used to compare scar over time. The use of fluorescence images provides very interesting functional information depending on the reaction behavior of certain molecules in the skin, like protoporphyrin. The emission spectrum of this molecule can be used for characterization of infected/carcinogenic cells. For all these reasons we developed an optical probe system that acquires tow images (gray scale and fluorescence one) at the same time. This system consists of a CCD camera, a blue light laser source to acquire florescence images and a white light source to acquire anatomical images.

Because the sensor covers a small area of skin, it is necessary to make several shots to cover the scar with a slight overlap between each pair of images. Image registration algorithms are used to obtain panoramic images of the scar in both modes (white light and fluorescence) starting from the couples of images. Because fluorescence images do not have enough information to be used by registration algorithms, we apply these methods over the white light images to obtain the transformation parameters and apply these parameters over fluorescence as well as they have been taken in same place at the same time.

Panoramic images obtained by our system, especially the fluorescence ones, are used by practitioners to observe and characterize scars and their evolution over the time. New panoramas are required every two weeks. That is why a multi-view image registration has to be done over the new panoramas to make them comparable with previous one.

Telemedicine can be then used to share these panoramas among several practitioners over different places using different kinds of devices (PCs, PADs, Cell phones, etc…).

In the next section we present some state-of-the-art of stitching methods. Section 3 explains our contribution with some experiments and results. Then the stitching scenario we propose is given in a collaborative way in section 4. Finally the conclusion is followed by further work ideas.

2. METHODS OF STITCHING

Image stitching algorithms are common in computer vision. The domains where they are used vary from satellite maps images to mobile applications to have wide angle photos. They are also used for medical applications. For example, standard X-ray images using conventional screen-film technique have a limited field of view and failed to visualize the entire long bone on a single image. To produce images with whole body parts, digitized images from the films that contain portions of the body parts are assembled using image stitching [1]. The essential step of the most image stitching methods is to extract required pixels information by using the suitable mathematical models. This information is important to relate the corresponding pixels in one image and the other. Before we explain image stitching algorithms, we mention some basic models that are used to determine required information in image stitching methods.

2.1 Homography

Image alignment and stitching between two images obviously require determining the similar parts or pixels in the common area between those images. The general idea is that when the camera takes one photo and turns to take the second one, certain kinds of transformation happens that makes us in need to remap pixels coordinate from one image to the other. The parameters of these transformations can be regrouped in a matrix called homography H [2]. The complexity of this matrix relates to the different kinds of transformation. H is a 3x3 matrix and its general form is:

$$\mathrm{H} = \begin{pmatrix} h1 & h2 & h3 \\ h4 & h5 & h6 \\ h7 & h8 & h9 \end{pmatrix}. \quad (1)$$

So all pixels X in one image can be remapped in the other X' by the general equation:

$$\mathrm{X}' = \mathrm{H}\,\mathrm{X} \quad (2)$$

where

$$\mathrm{X}' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}, \mathrm{X} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

In our case, a simple 2D transformation (Fig. 1) is enough to determine image stitching parameters [3, 4] between each pair of images.

The general form of a homography which is called the projective transform homography also known as a perspective transform is as follows:

$$H = \begin{pmatrix} A & t \\ v^T & 1 \end{pmatrix} \quad (3)$$

with A = R(θ)R(-φ)DR(φ). R(θ) and R(φ) are the rotation matrices for angles θ et φ, and D is a diagonal matrix:

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, t = \begin{pmatrix} t1 \\ t2 \end{pmatrix}.$$

$D = \begin{pmatrix} \lambda\mathbf{1} & 0 \\ 0 & \lambda\mathbf{2} \end{pmatrix}$ where $\lambda\mathbf{1}$ and $\lambda\mathbf{2}$ can be considered as two scaling values. The matrix A is thus a concatenation of a rotation by φ, a scaling by $\lambda\mathbf{1}$ in the x direction, a scaling by $\lambda\mathbf{2}$ in the y direction, a rotation back by –φ, then another rotation by θ and $v = (v1, v2)^T$.

The key difference between the affine and projective transformation is the vector v, which is null in the affine case. This vector is responsible for the non-linear effects of the projectivity.

Depending on the values of homography parameters transformation could varied from simple translation to projective.

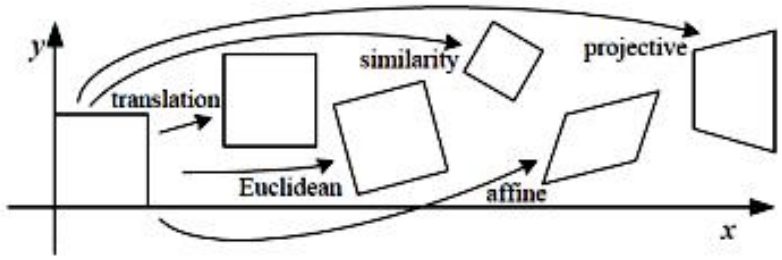

Figure 1. Different transformations.

After we had a glance over different transformation models that can face us in our research for stitching parameters, we review the different image stitching methods that use these models in order to estimate stitching parameters. The most well-known methods could be regrouped into two wide domains, pixel-based alignment (direct) and feature-based alignment [4].

One approach is to shift or warp the images relative to each other and to look at how much the pixels agree. Approaches that use pixel-to-pixel matching are often called direct methods. To use a direct method, a suitable error metric must first be chosen to compare the images. Once this has been established, a suitable search technique must be devised. The simplest technique is to exhaustively try all possible alignments. In practice, this may be too slow, so hierarchical coarse-to-fine techniques based on image pyramids have been developed. Comparing with feature-based registration methods, measuring the contribution of every pixel by direct methods shows better performance over specific images like blurred images where feature-based registration could fail especially. Knowing that this kind of images is rarely occurred in our work and that feature-based methods like SIFT show a robust performance for changing in scale and lighting, we focus more over the feature-based alignment.

2.2 Feature-based registration

While there are technical differences between different feature-based registrations, most of them share the same essential steps depending over the mathematical models explained previously. These common steps are explained in Fig. 2, where

detecting points with characteristic features in images are the two first steps. Then these features are used to find the matching between successive pairs of images. Finally, depending on the matching, the transformation parameters are processed to register images and stitch them together.

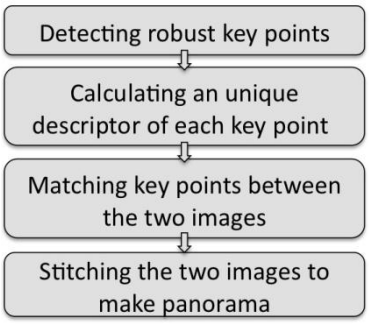

Figure 2. Feature-based stitching steps.

SIFT Algorithm

Detecting key points was primarily designed for robotic applications in order to recognize objects within a captured image using a visual database as a reference. Lowe and Brown [5, 6] also showed a practical application of image registration for automatic panorama recognition. The four major stages of computation used to generate the panorama in the SIFT algorithm [7, 8, 9] are:

Scale-space extrema "key-point" detection: the first stage of computation seeks these points over all scales and image locations. It is efficiently implemented by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)$$

where $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$.

L(x, y, σ) is produced from the convolution of a variable-scale Gaussian G(x, y, σ) with an input image I(x, y). The parameter σ refers to the smooth parameter of Gaussian function and k is the factor that multiplies the smooth between two nearby levels. D(x, y, σ) is the value of at the pixel (x, y) resulted from the subtraction between the two nearby blurred levels.

Key-point localization: at each candidate location, a detailed model is fit to determine location and scale. Key-points are selected based on measures of their stability.

Orientation assignment: one or more orientations are assigned to each key point location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

The scale of the keypoint is used to select the Gaussian smoothed image L with the closest scale, so that all computations are performed in a scale-invariant manner. For each image sample L(x, y) at this scale, the gradient magnitude m(x, y) and orientation θ (x, y) are precomputed using pixel differences [6]:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \quad (5)$$

$$\theta(x, y) = tan^{-1}(L(x, y+1) - L(x, y-1))/(L(x+1, y) - L(x-1, y))). \quad (6)$$

Key-point descriptor: local image gradients are measured at the selected scale in the region around each key-point. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination [4].

SURF Algorithm

The approach for interest point detection uses a very basic Hessian-matrix approximation, because of its reliable accuracy. More precisely, blob-like structures at locations where the determinant is optimum are detected. In contrast to

the Hessian-Laplace detector by Mikolajczyk and Schmid [10], the SURF algorithm also relies on the Hessian determinant for scale selection, as performed by Lindeberg [11].
The major steps of image feature detection of SURF and SIFT are similar, but SURF lends itself to the use of integral images as made popular by Viola and Jones [12], thus reducing the computational time drastically. Integral images fit in the more general framework of boxlets, as proposed by Simard et al [13]. The major goal of using integral images in SURF is to decrease the time of computing key-points [14]. They allow for fast computation of box type convolution filters.
An integral image, denoted i(x, y) at location (x, y) contains the sum of the pixel values above and to the left of (x, y). Formally, we have:

$$\mathrm{i}(\mathrm{x},\mathrm{y}) = \sum_{x',y' \le x,y} (i(x', y')). \quad (7)$$

ASIFT Algorithm

This algorithm depends essentially on SIFT and proposes a crucial parameter for evaluating the performance of affine recognition, the transition tilt [15]. The transition tilt measures the degree of viewpoint change from one view to another, giving a first intuitive approach to absolute tilt and transition tilt. It illustrates why simulating large tilts on both compared images proves necessary to obtain a fully affine invariant recognition. Indeed, transition tilts can be much larger than absolute tilts. In fact they can behave like the square of absolute tilts.

3. CONTRIBUTION

3.1 System overview

Our acquisition system consists of different parts: a CCD camera + lenses, two sources of light with automatic switch on/off (one of white light and one at 406 nm for fluorescence excitation), and a high pass filter at 420 nm to reflect laser light and to let fluorescence emission passes to the camera. Reflected light from skin covers the same area and is captured by the camera. The probe we use has a special size limited to the size of blue light, so it cannot cover the whole area of a scar. Therefore, we acquire a sequence of pairs of images to cover the requested area of skin. Hence we calculate the registration parameters between each pair of successive images acquired under white light because it is much easier than processing with fluorescence data which have less salient points. Registration parameters describe the transformation (rotation, translation, scale …) and are used to achieve the whole panoramic fluorescence image. This registration is possible if a common area of skin always appears between each pair of successive images.

Transformations which have to be determined with our system are mainly based on rotation and scale variance. Because rapidity is important to integrate efficiently our work in a collaborative system, we chose to base our stitching process on feature-based registration. So, we apply several tests over several well-known methods in this domain to achieve best scenario of work depending on the results that we had. Giving a sight over the state-of-the-art of feature-based methods, we finally decided to apply our tests over SIFT, ASIFT, and SURF algorithms, which are the methods that look like to deal with the closest problems to ours.

3.2 Experiments and results

Two types of tests are required: the pretest “objective” that makes us able to predict the situation and the limits between the algorithms, and the “subjective” test at the end of the panoramic process where we check the accuracy of the stitching after transformation. To implement stitching algorithms and related tests, we used C language with the OpenCV libraries.

Depending on elder results in our last researches, we focused over ASIFT algorithms to test the performance of this algorithm over several pre-transformations in different conditions and different images acquired by the probe. These transformations varied from different rotation angles to different scale and projection values.

Fig.3 discusses the error for estimating the rotation angle according to the real rotation angle and to the percentage of overlapping between the two images. It shows that the error of rotation angle in the estimated homography is less with small angles and still accepted to 35 degrees with good overlapping.

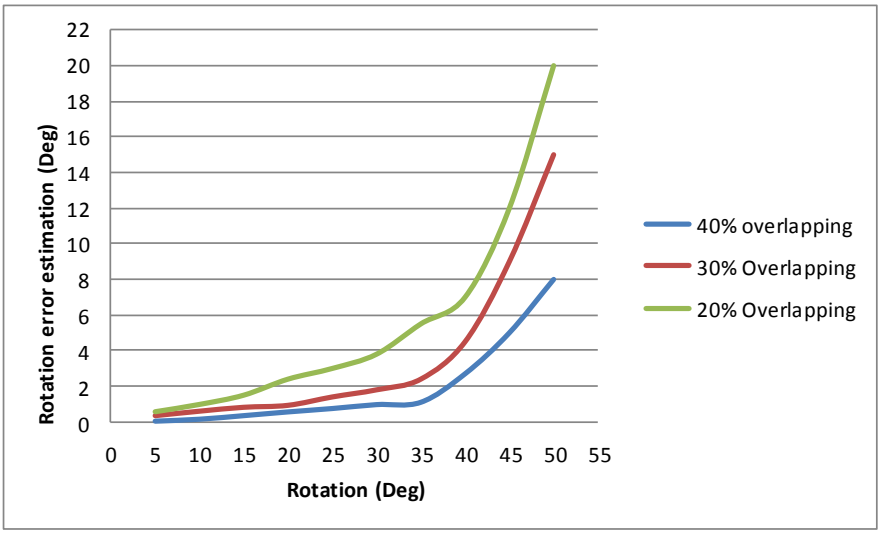

Figure 3. Relation between rotation error estimation and real rotation angle, depending on the overlapping between two images.

Fig.4 discuses about scale transformations. It shows that with scaling between 80% up to 120%, ASIFT has a good performance to estimate it with overlapping between 20% and 40%.

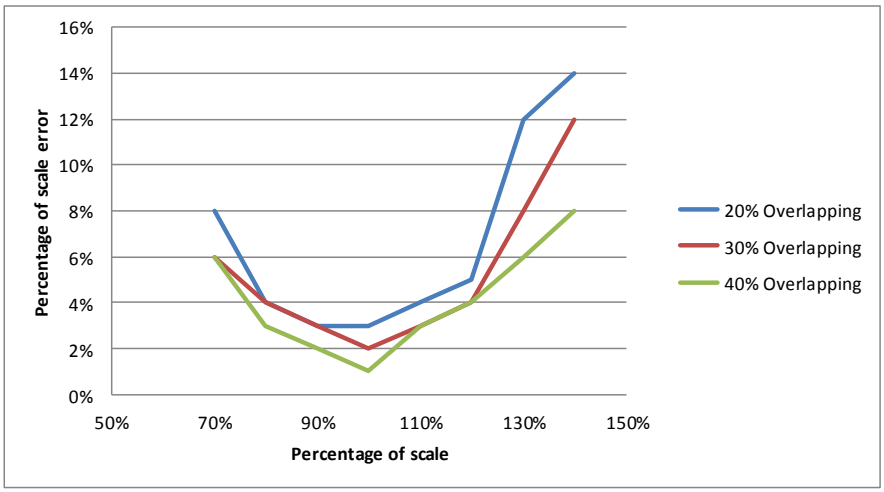

Figure 4. Relation between scale and overlapping with no rotation nor projection.

If we consider that the number of robust matching points detected between two successive images is a proof of good performance to estimate the right homography, Fig. 5 shows that the performance of ASIFT with our probe images is still good (30-140 matching points) with overlapping between 20% and 40%, rotation between 5 to 30 degrees with projection in second image. These conditions could represent all deformations that we could face with real experiments.

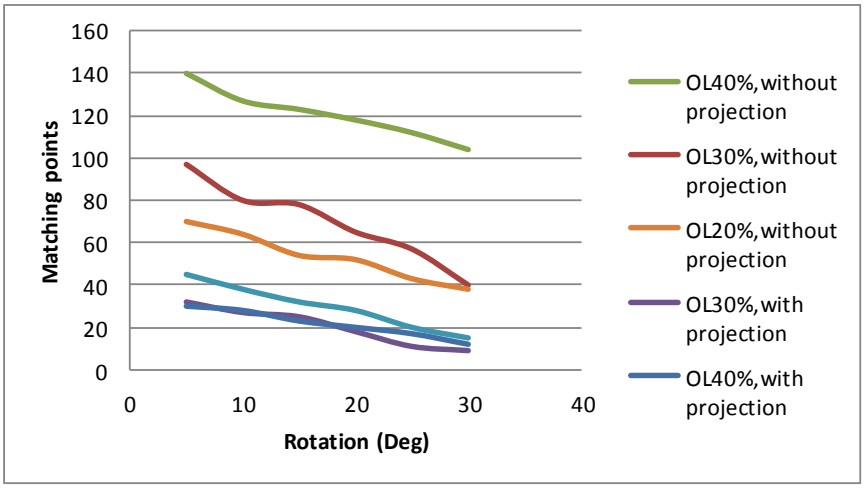

Figure5. ASIFT performance depending on matching points number
with multiple transformation (OL: OverLapping, Rotation, and Projection)

The second part of our experiments was made over images of the probe without knowing in advance the transformations. The results are judged by observing it by human eyes to determine whether that was good or not (subjective test).

In Fig. 6, we can see that with good overlapping, results can be satisfactory, where transformation parameters estimated from white light images where used in fluorescence ones. Fig. 6(a) represents grayscale images from where transformation parameters are really estimated, while in Fig. 6(b) fluorescence images are stitched with the same parameters to get the panorama.

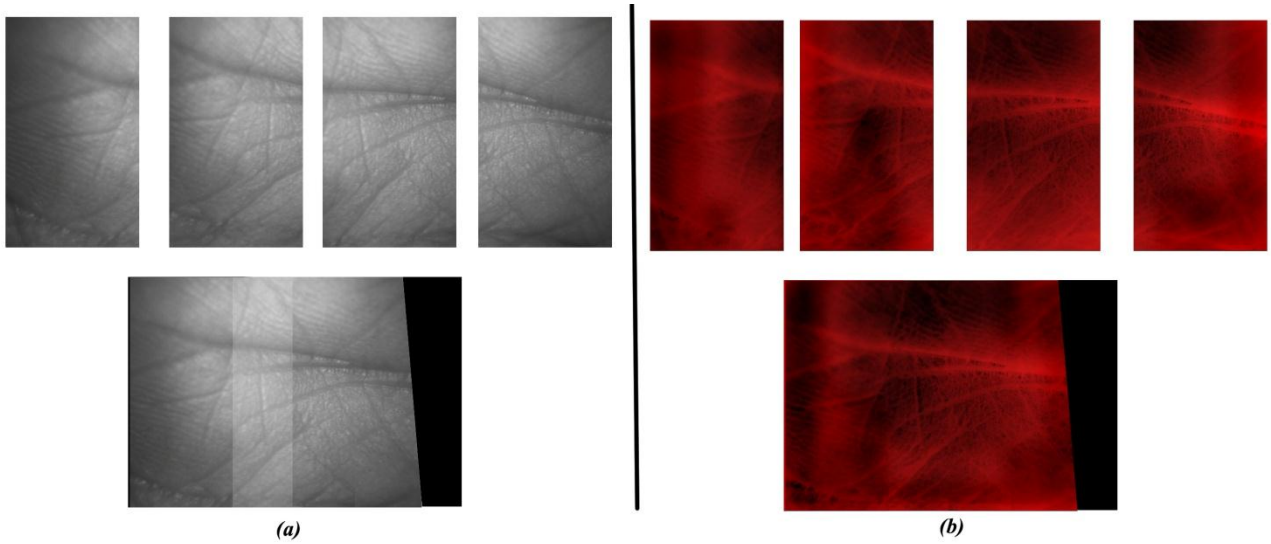

Figure6. (a) white light panorama, (b) fluorescence panorama.

4. STITCHING SCENARIO IN A COLLABORATIVE WAY

Depending on the results of experiments explained in the previous paragraph, we propose the next stitching scenario which is figured in Fig. 7.

Depending on preliminary information and feedback of the session opened between practitioners:

If speed is important but not an essential condition, we start with SIFT. If the result is not good enough, we apply ASIFT. If speed is not important, we start with ASIFT. In such a way, we can guarantee the robustness for most of the examples in our case study. As well as SURF proved to be always the faster but the less robust we use it only if the speed is essential.

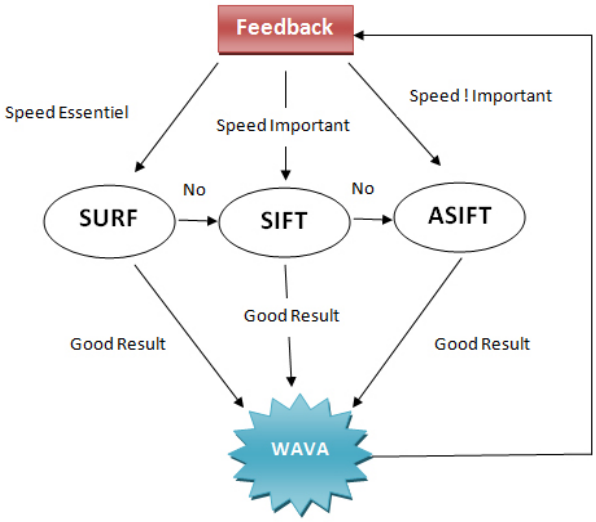

Figure 7. Best method decision.

In our case we suppose that the optical probe is manipulated by hospital staff in one place and practitioners located in other different cities. To make this system works well we need an efficient collaborative system that guarantees data sharing in robust and secure ways. The scenario will be as follows: Hospital staff takes the conventional and fluorescence images, and sends them to a web service by any possible connection (internet, mobile, satellite, etc...). The web service receives these images and manipulates them producing the requested panorama of both conventional and fluorescence modes. Then it distributes these two panoramas to all practitioners who are involved in the telediagnosis. In Fig. 8, we can see that the system sends the same panorama to each practitioner.

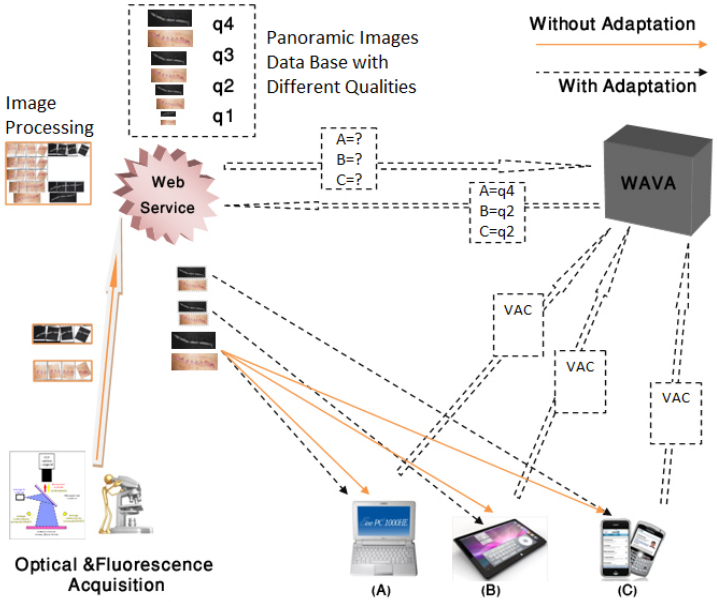

Figure 8. Collaborative system for images co-interpretation.

Our first step in collaborative work could involve problems with the quality of service. That's why we propose a solution for adapting data distribution. We suppose that the practitioners could have different types of connections to the system (for example Internet, Wi-Fi or 3G) and as well different terminals with different capabilities. Because they cannot receive the same panorama, we switch to the adapted solution by adding the WAVA, a new "Web service for Automatic Video data flow Adaptation" module, which adapts the previous scenario like this: Practitioners (A, B, C) connect to WAVA to register themselves, transmitting their capabilities as an XML file called VAC (Virtual Adaptation Card) [16, 17]. WAVA gives each of them a rank. The web service receives the images from the acquisition component. It applies the process to produce both the panoramic conventional and fluorescence images, making different copies with different qualities, saving them in a database. When practitioners connect to the web service searching for panoramic images, the web service sends a quality request to WAVA, which sends back the adapted data. At the end, the web service sends to each practitioner the adapted panorama, from which he is making his diagnosis (doted gray arrows in Fig. 8).

5. CONCLUSION AND FURTHER WORK

In this paper we have tested three different image stitching algorithms applied over our optical probe which provides at the same time two types of panoramic images of the same scar: one image acquired under white light (anatomic information) and another one resulting from tissue autoflorescence (functional image). These two images have been stitched by using parameters extracted from a feature-based algorithm processed on images acquired under white light.
We have tested three stitching algorithms giving a suitable scenario to use them in a collaborative system. While ASIFT shows a good performance in affine transformation, SIFT and SURF are faster. Then we made some experiments applying ASIFT method on the images acquired by our probe. Results presented that ASIT could be satisfactory with several kinds of affine transformations.
In future work we plan to profit of the rapidity of SURF algorithm and the accurateness of ASIFT to build a new algorithm, more robust and quicker. Also we will determine the weak points of our first version of our probe, like the positioning of white light LEDs and the type of CCD and exceed them in new versions.

REFERENCE LINKING

[L1] http://www.skincancer.org.
[L2] http://www.medicinenet.com.

REFERENCES

[1] S. Salbiah, A. Somaya, H. Arof, Z. S. Saleh, F. Ibrahim A new approach to medical image stitching using minimum average correlation energy filter and peak to side-lobe ratio International Journal of Imaging Systems and Technology archive, Volume 22 Issue 3, pp. 166-171, Sept. 2012.
[2] G. Bradski and A. Kaehler. Learning OpenCV. Book published by O'Reilly Media, Inc, sept 2008, ISBN 978-0-596-51613-0, 2008.
[3] Elan Dubrofsky, Homography Estimation, Master's thesis, TheUniversity of British Columbia, 2009.
[4] Richard Szeliski Microsoft Research, Journal Foundations and Trends® in Computer Graphics and Vision, Volume 2 Issue 1, pp. 1 – 104, Jan. 2006.
[5] S. A. J. Winder and M. Brown. Learning local image descriptors. In CVPR, pp 1-8, 2007.
[6] D. G. Lowe. Distinctive image features from scale-invariant keypoints, Int. J. of Comp. Vision, Vol. 60(2), pp. 91-110, 2004.
[7] M. Ozuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code. In Proc. Int. Conf. on Computing Vision and Pattern Recognition, pp. 1-8, June 2007.
[8] Martins, A.L.D., Mascarenhas, N.D.A., and Suazo, C.A.T., Spatio-Temporal Resolution Enhancement of Vocal Tract MRI Sequences Based on Image Registration, Integrated Computer-Aided Engineering, Vol. 18(2), pp. 143-156, 2011.
[9] M. Brown and D.G. Lowe. Automatic panoramic image stitching using invariant features, Int. J. of Computer Vision, Vol. 74(1), pp. 59-73, 2007.

[10] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In ICCV'01, Vol. 1, pp. 525-531, 2001.
[11] T. Lindeberg. Feature detection with automatic scale selection. In Int. J. of Computer Visio, Vol. 30(2), pp. 79-116, 1998.
[12] P.A. Viola and M.J. Jones. Rapid object detection using a boosted cascade of simple features. In CVPR 2001, Vol. 1, pp. 511-518, 2001.
[13] P. Simard, L. Bottou, P. Ha_ner, and Y. LeCun. Boxlets: A fast convolution algorithm for signal processing and neural networks.In NIPS'98, pp. 571-577, 1998.
[14] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding (CVIU), Vol. 110(3), pp. 346-359, 2008.
[15] J.-M. Morel and G. Yu. ASIFT: a new framework for fully affine invariant image comparison, Intern. Int. IEEE Conf. ICASSP'09, pp. 1597-1600, April 2009.
[16] J.-B. Aupet, N. Elmarzouqi, E. Garcia, and J.-C. Lapayre. Virtual awareness card for adaptability in collaborative virtual environments. In IEEE SETIT 2009, pp. 228-234, Hammamet, Tunisia, March 2009.
[17] J.-B. Aupet, R. Kassab, and J.-C. Lapayre. Wava: a new web service for automatic video data flow adaptation in heterogeneous collaborative environments. In CDVE2009, pp. 280-288, Luxembourg City, Springer-Verlag in LNCS, September 2009.