

# An Evaluation of Photo-consistency for Intra-operative Registration in an Image Enhanced Surgical Navigation (IESN) System

Gerardo Gonzalez<sup>a</sup> and Rudy Lapeer<sup>a\*</sup>

<sup>a</sup>School of Computing Sciences, University of East Anglia, Norwich, UK

**Abstract.** In this paper, we propose a method to intra-operatively register real and virtual models in an Image-Enhanced Surgical Navigation (IESN) system based on an intensity metric known as photo-consistency. The application is aimed at Ear, Nose and Throat (ENT) procedures using a pair of cameras connected to a stereoscopic surgical microscope. Three optimisation techniques for searching global optima were compared and an evaluation of different photo-consistency metrics available in the literature was performed. The experiments provide an insight of the method based on a human skull used as a phantom patient. The preliminary results obtained demonstrate an improvement in the optimisation algorithms implementation and a final 3D registration error between 0.5 and 1 mm.

## 1 Introduction

In the field of computer graphics, Augmented Reality (AR) aims to add visual information to the natural world by overlaying real and virtual imagery directly on the user's view. In the medical domain, AR-based Surgical Navigation (SN) systems; also known as Image Enhanced Surgical Navigation (IESN); provide the surgeon with an enhanced vision of a patient's anatomy. This is achieved by superimposing pre-operatively scanned models such as X-Ray Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) on top of surgical images captured by optical devices (i.e. endoscope or microscope).

The required stages to produce an enhanced view consists of 1) *Camera calibration*, which involves the derivation of the physical and optical camera parameters. 2) *Registration*, that aligns real images and CT/MRI models, and 3) *Tracking*, which captures the position and orientation of cameras and/or patient involved during the intervention. While the first two stages are performed pre-operatively in order to generate a static overlay, the third step reflects the dynamic pose of the models throughout surgery. A problem found in IESN is that each step produces a certain level of accuracy error on its own. When these stages are combined, the accumulated errors are aggravated by the use of the tracking device over a prolonged time, which can hinder the surgery overlay accuracy [1].

Because of surgical requirements of time and efficiency, it is inappropriate to repeat the initial calibration and registration procedures. For this reason, we propose a registration technique during the intra-operative stage to reduce the alignment mismatch. The method is based on the visual information obtained from a pair of cameras connected to a surgical microscope. We employ a similarity metric called photo-consistency that compares intensity value differences between the captured images and evaluates the accuracy in the alignment between real and virtual models. In this paper we introduce preliminary results about the implementation of photo-consistency as a registration tool employing a phantom skull.

The concept of photo-consistency has been previously used in the reconstruction of 3D shapes from colour or grayscale images [2]. Clarkson et al. [3] used this technique to match the projection of a set of 2D images to a 3D surface model. They employed calibrated cameras and an optimisation algorithm based on gradient ascent search. Jankó and Chetverikov [4] relied on a genetic algorithm in order to extend the registration method by finding the aligning pose and performing the camera calibration concurrently. In both approaches the authors utilised full-sized polygonal models.

In the field of medical applications, Chen et al. [5] implemented a cost function based on photo-consistency to intra-operatively register various endoscopic images taken at different positions to a volumetric model. They implemented Powell's method to perform the optimisation. Figl et al. [6] followed a different approach that consisted of a video sequence of a beating heart model obtained by a calibrated stereo endoscope. Due to the inherent cyclical cardiac motion, the alignment aims to match multiple CT data sets to the real heart artefact.

---

\*Corresponding Author: School of Computing Sciences, University of East Anglia, Norwich, UK. E-mail: R.Lapeer@uea.ac.uk

## 2 Method

### 2.1 Calculation of photo-consistency based cost function

The photo-consistency metric is based on the difference between intensity or colour values in a visible set of a 3D object's points which are projected on two or more images. Therefore, the related pixels on each viewpoint that correspond to the same 3D point should ideally possess the same intensity or colour attributes. The images are said to be photo-consistent if the difference among all related pixel values is null or near zero. It is assumed that the scene complies with a Lambertian model in which the lighting is static and the visible object maintains an equal luminance regardless of the perspective.

In the current implementation, we connect two black and white video cameras to the eye pieces of a surgical microscope for Ear, Nose and Throat (ENT) interventions. The cameras are calibrated in a pre-operative stage using Tsai's algorithm [7]. This generates a projection matrix  $P = KM$  that relates the internal parameters  $K$  and the external parameters  $M$ , in which  $K$  is a 3 x 3 upper triangular matrix and  $M$  is a 3 x 4 matrix that depicts the orientation and position of the camera. The projection of a model's 3D point  $\mathbf{X}$  on each camera viewport is computed as:  $x_l \sim P_l \mathbf{X}$  and  $x_r \sim P_r \mathbf{X}$  where  $P_l$  and  $P_r$  are the 3 x 4 projection matrices for the left and right cameras, respectively; and  $x_l$  and  $x_r$  are the corresponding projected pixels of the same point  $\mathbf{X}$ . The sign  $\sim$  indicates that the projection is defined up to a scale factor.

We determine the value of the photo-consistency function  $PC$  through comparing the pixel intensity levels ( $I_1$  and  $I_2$ ) in the pair of captured images:

$$PC = \frac{1}{N} \sum_{i=1}^N \|I_1(x_{l,i}) - I_2(x_{r,i})\|^2. \quad (1)$$

where  $N$  represents the total number of visible pixels  $i$  in both images. Clarkson et al. [3] provide an alternative similarity measure for the computation of photo-consistency by first determining a mean of pixel values. In the case of two viewpoints this is:  $\bar{x} = (x_{l,i} + x_{r,i})/2$ .

Then, the total sum of squared differences is computed according to the following equation:

$$PC_{squared} = \frac{1}{N} \sum_{i=1}^N \frac{(x_{l,i} - \bar{x})^2 + (x_{r,i} - \bar{x})^2}{2}. \quad (2)$$

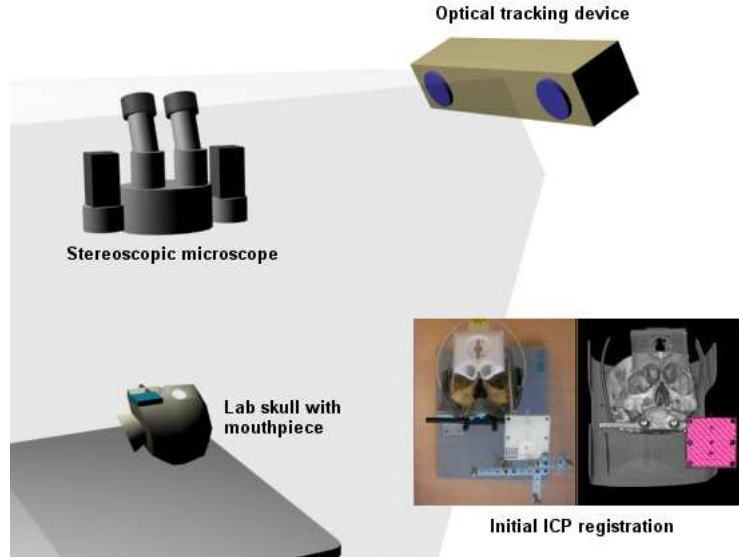
A final cost function also described in [3], aims to reduce the effect of outliers by calculating the inverse of squared differences. This is achieved through a threshold  $\epsilon$  related to the noise level found in intensity images:

$$PC_{inverse} = \frac{1}{N} \sum_{i=1}^N \frac{\epsilon^2}{\epsilon^2 + ((x_{l,i} - \bar{x})^2 + (x_{r,i} - \bar{x})^2)}. \quad (3)$$

### 2.2 Intra-operative registration

For the purpose of aligning both real and virtual models at the beginning of the medical procedure it is necessary to carry out an initial registration. We use the Iterative Closest Point (ICP) algorithm [8] to match a VBH mouthpiece [9] worn by the patient during the pre-operative scan and at the time of surgery (See Figure 1 for a representation of the equipment setup). As we have described earlier, the accumulated errors generated throughout surgery tend to affect the original alignment. Therefore, we apply a cost function based on photo-consistency in order to evaluate the best registration pose that represents the lowest intensity difference between images. Due to the fact that both cameras have been previously calibrated, the cost function only involves the computation of six degrees of freedom (DOF), i.e. three for translation and three for rotation.

As a first step, a set of visible voxels in the volumetric model is selected by backprojecting screen pixels of that model within a user-defined selection window. This technique is similar to a raycasting projection. A voxel is detected for each projected ray that collides on the virtual model. As several screen pixels will map to one voxel because of the magnification levels provided by the microscope, duplicated voxels are neglected. Later, a forward projection ray is created from the selected voxel to each of the cameras. This allows determining the corresponding pixel coordinates in both 2D images.



**Figure 1.** Representation of the setup used for registration, which includes a stereo microscope, an optical tracking device and a human skull. The right-bottom insert shows the initial ICP registration between real and virtual models without viewing magnification.

It is necessary to evaluate any potential occlusion that obstructs visibility of a 3D point on the camera viewports. In [3], a z-buffer technique was implemented in order to render only external visible points on a surface mesh model. Another approach involves calculating surface normals to ignore areas of the surface that are not oriented towards the cameras [4]. In our case, the CT or MRI virtual models are made of voxels with different levels of transparency. Thus, we have decided to use a direct check of possible voxels that can partially obstruct the forward-projected ray from a selected 3D point. If the ray collides with a voxel that contains a higher transparency level than a specific threshold, we determine it as an occlusion and the corresponding projected pixels are not evaluated in the photo-consistency cost function. As mentioned earlier, the projection of a voxel does not relate to a single pixel on the pair of captured images (voxel-to-pixel relation is 1:many) due to magnification. For this purpose it is required to compute the voxel dimension on the virtual model and project the vertices of the voxel face that point towards each camera. From these four vertices it is possible to determine a sub-window that relates to the visible voxel hence define the number of its projected pixels. We then apply a median filter to the pixels inside the convolution window.

### 2.3 Optimisation method

The objective of the optimisation method is to minimise the cost function described in Eq. (1) by iteratively changing the registration pose (over six DOFs) for a maximum number of iterations until a global minimum value is reached. The optimised result must correspond to the best matching pose between real and virtual models. Several optimisation techniques were tested which do not require the calculation of derivatives as the shape of the global function is unknown. Among these, we tried the well-established Powell's and Quasi-Newton methods. Although the latter tends to perform faster, we found that Powell's method provides a greater control in the parameters modification for the overall DOFs during optimisation. A number of genetic programming approaches based on Differential Evolution (DE) [10] were also applied. DE aims to optimise a function by mixing the individuals from a set of potential solutions. This strategy selects the best existing candidates and arithmetically combines them until the best value is found. Nevertheless, its main disadvantage consists on correctly setting the initial control parameters for each particular problem, e.g. in [11] the authors suggest ten different approaches depending on the problem features. Also, a wrong choice of initial parameters can affect the overall performance. For this reason, we selected a technique proposed by Salman et al. called Self-adaptive Differential Evolution (SDE) [12]. SDE eliminates most of the manual selection by constantly changing the control parameters and selecting the ones that produce the best results, which exploits a wider search in the function shape and avoids stagnation in local minima. However, a minimum of two parameters are still required to initialise the optimisation. Another strategy we have tried is a stochastic algorithm called CODEQ [13] that is based on four different optimisation strategies: Chaotic search, Opposition-based Learning, Differential Evolution and Quantum mechanics. The main advantage of this approach relies on being a parameter-free method and self-adaptive to the objective function.

### 3 Experiments

#### 3.1 Intra-operative registration laboratory test

We used a human skull as a dummy patient in order to evaluate the intra-operative registration accuracy. The procedure carried out begins by calibrating the cameras connected to the surgical microscope and performing the initial ICP registration. We then move our region of interest (ROI) towards one of the eye sockets (calibration and ICP registration are performed on the attached VBH mouthpiece). We select a number of voxels on the CT model by back-projecting pixels in a selection window and record their 3D position. An initial intensity evaluation is executed in order to calculate the photo-consistency in the current pose. The reason for the latter operation is because the initial photo-consistency value obtained does not result in a zero value even though the real and virtual models are aligned. This is likely due to the error on the ICP registration (1-2mm) and the fact that the illumination produced by the microscope is not strictly Lambertian. For this reason, the 3D position and corresponding photo-consistency values at the matching pose are used as ground truth to assess the registration results. Then, the CT-based virtual model is offset for the purpose of simulating a misregistration during surgery. The applied offset levels of the “starting pose” were of the order of magnitude of millimetres in object space. The initial registration results obtained were not accurate because the images are not Lambertian and the cost function aims to find the global minimum on intensity differences. This led the photo-consistency procedure to stagnate at incorrect positions that had smooth surfaces or showed low lighting (i.e. areas inside the eye socket). In other words, the lowest photo-consistency value (global minimum) does not always correspond to the best registration pose at the magnification level (6x) that we deal with. Therefore, we adapted the cost function in order to minimise the difference between the current and ground-truth photo-consistency values instead of finding the lowest photo-consistency value. This improved the registration procedure as we are refining the search space within the function shape. We compared the registration accuracy of different cost functions proposed in the literature for photo-consistency as discussed in section 2.1, i.e. Equations 1, 2 and 3 and three optimisation methods, i.e. SDE, CODEQ and Powell’s method - see Table 1. The 3D root mean square distance (RMSD) error was calculated between the voxels located at the ground truth position and at the final registration pose.

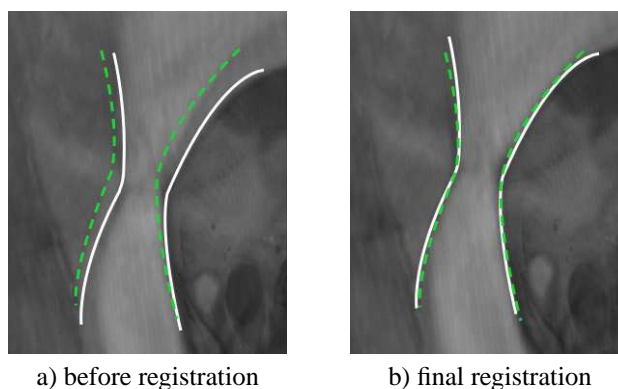
	RMSD at optimum	RMSD at best registration
<i>PC</i>	1.102±1.541	0.8667
<i>PC<sub>squared</sub></i>	1.025±0.321	0.3787
<i>PC<sub>inverse</sub></i>	2.545±1.089	1.0224
SDE	0.6012±0.646	0.5283
CODEQ	0.530±0.785	0.3557
Powell’s method	1.626±0.284	1.4503

**Table 1.** Photo-consistency mean error values ( $\pm$  SD) in mm obtained by each cost function - top three rows; and mean error values ( $\pm$  SD) in mm obtained by three different optimisation methods - bottom three rows.

### 4 Discussion

The experiment presented a laboratory-based test of image enhanced surgical navigation registration when the virtual counterpart (CT data) of a skull was offset after the initial ICP algorithm. The results show that the photo-consistency method is capable of recovering the earlier registration. The first three rows in Table 1 show that the *PC<sub>squared</sub>* cost function provides the best registration accuracy among the described methods. From the results in the bottom three rows of Table 1 we can see that CODEQ generates the lowest RMSD. Figure 2 illustrates the final overlay registration around the skull eye socket. As expected, this method is the most suitable for finding the minimum in a function, followed by SDE. Powell’s method occasionally stagnates at incorrect areas in the image that have low intensity levels. However, a problem we have found using CODEQ is that during registration it takes a longer time to find the optima in comparison to SDE. This is because of its tendency to perform extra cost function evaluations in each generation, leading a search in the opposite direction of the current position. Such behaviour can be problematic as in particular cases the virtual model goes out of scope in the visible image causing the final position to be visually misregistered.

In this first implementation we have used the ground-truth pose as a starting step in order to optimise the registration, which provides an RMSD evaluation. However, further experiments will be based solely on the comparison of pixel intensity levels. A way to perform this could involve comparing the intensities within visible areas on both real images, giving a more realistic ground truth than the arbitrary position. Also, the model will be extended to include specular reflection, which will assume real conditions in a surgical environment. This is likely to generate steeper photo-consistency gradients resulting in less ambiguous extrema hence a likely better result.



**Figure 2.** Registration of real and virtual models around the skull eye socket. Dashed lines show contour features in the real model. Solid lines indicate contour features in the virtual model.

## 5 Conclusion

We have presented an intra-operatively registration method based on photo-consistency for the use of a stereo microscope in IESN. The results obtained demonstrate that a DE-based strategy is suitable for alignment between virtual and real models using intensity information. Although CODEQ generally produces a good match, it can require a long time to find the optimum and in some cases it might produce a visual misregistration. Current processing times are of the order of a couple of minutes on an Intel Core2 Quad processor 2.4Ghz and 2GB RAM. Therefore, the alternative method of using SDE to obtain an approximate solution and subsequently apply a different optimisation technique such as Powell's method to refine the search, is worthwhile of further investigation. Further experiments will involve a comparison of intensity levels as a ground truth based solely on the pair of images and include a non-Lambertian scenario. Also, the implementation of the algorithm on the GPU using CUDA will significantly speed up the registration process so it becomes a viable method for real-time registration in image enhanced surgical navigation.

## Acknowledgements

G. Gonzalez would like to thank the Mexican National Council on Science and Technology (CONACYT) for financial support during his PhD studentship.

## References

1. R. Lapeer, M. Chen, G. Gonzalez et al. "Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking." *IJMRCAS* **4**, pp. 32 – 45, 2008.
2. S. Seitz & C. Dyer. "Photorealistic scene reconstruction by voxel coloring." *Int J Comput Vision* **35(2)**, pp. 151 – 173, 1999.
3. M. Clarkson, D. Rueckert, D. Hill et al. "Using photo-consistency to register 2D optical images of the human face to a 3D surface model." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23(11)**, pp. 1266 – 1281, 2001.
4. Z. Jankó & D. Chetverikov. "Photo-consistency based registration of an uncalibrated image pair to a 3D surface model using genetic algorithm." In *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 616 – 622. 2004.
5. M. Chen, G. Gonzalez & R. Lapeer. "Intra-operative registration for image enhanced endoscopic sinus surgery using photo-consistency." *MMVR15 proceedings, Studies in Health Technology and Informatics* **125**, pp. 67 – 72, 2007.
6. M. Figl, D. Rueckert, D. Hawkes et al. "Registration of a 4D cardiac motion model to endoscopic video for augmented reality image guidance of robotic coronary artery bypass." AMI-ARCS workshop of MICCAI 2008, September 2008.
7. R. Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses." *IEEE J. Robotics and Automation* **3(4)**, pp. 323 – 344, 1987.
8. P. Besl & N. McKay. "A method for registration of 3D shapes." *IEEE Trans. Pattern Analysis and Machine Intelligences* **14(2)**, pp. 239 – 256, 1992.
9. A. Martin, R. Bale, M. Vogele et al. "Vogele-Bale-Hohner mouthpiece: Registration device for frameless stereotactic surgery." *Radiology* **208**, pp. 261 – 265, 1998.
10. R. Storn & K. Price. "Differential evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces." Technical Report TR-95-012, International Computer Science Institute, 1995.
11. K. Price, R. Storn & J. Lampinen. *Differential Evolution - A Practical Approach to Global Optimization*. Springer, 2005.
12. A. Salman, A. Engelbrecht & M. Omran. "Empirical analysis of self-adaptive differential evolution." *Eur J Oper Res* **183(2)**, pp. 785 – 804, 2007.
13. M. Omran. "CODEQ: An efficient meta-heuristic for continuous global optimization." In press.