# Gaze-Contingent 3D Control for Focused Energy Ablation in Robotic Assisted Surgery

Danail Stoyanov, George P. Mylonas, and Guang-Zhong Yang

Institute of Biomedical Engineering,
Imperial College London, London SW7 2AZ, UK
{danail.stoyanov, george.mylonas, g.z.yang}@imperial.ac.uk
http://vip.doc.ic.ac.uk

**Abstract.** The use of focused energy delivery in robotic assisted surgery for atrial fibrillation requires accurate prescription of ablation paths. In this paper, an original framework based on fusing human and machine vision for providing gaze-contigent control in robotic assisted surgery is provided. With the proposed method, binocular eye tracking is used to estimate the 3D fixations of the surgeon, which are further refined by considering the camera geometry and the consistency of image features at reprojected fixations. Nonparametric clustering is then used to optimize the point distribution to provide an accurate ablation path. For experimental validation, a study where eight subjects prescribe an ablation path on the right atrium of the heart using only their gaze control is presented. The accuracy of the proposed method is validated using a phantom heart model with known 3D ground truth.

**Keywords:** Robotic Assisted Surgery, Atrial Fibrillation, Gaze-Contingent Control, 3D Depth Recovery, Focused Energy Delivery.

## 1 Introduction

Robotic assisted Minimally Invasive Surgery (MIS) is increasingly being used to perform complex cardiac procedures such as mitral valve repair or replacement via minimal access ports [1]. Although a significant number of patients undergoing mitral valve surgery suffer from Atrial Fibrillation (AF) concomitant treatment is not widely adopted. Surgery for AF involves creating scar tissue that isolates irregular electrical impulses and prevents their conduction to the rest of the heart. With developments in port access techniques, the use of focused energy delivery based on radiofrequency, microware, cryoablation, and bipolar cautery has been preferred over the conventional approach [1,2]. More recently laser sources have also been proposed for controlled heating of the tissue and forming electrical impedance while preserving the tissue's structural coherence [2]. A challenging problem in concomitant mitral valve and AF surgery is the control of the ablation instrument whilst manipulating the conventional robotic manipulators.

Compared to the use of other Human-Machine Interface (HMI) channels, eye gaze is a versatile option that has not been fully explored. It is the only input modality that implicitly carries information about the focus of the user's attention at a specific point

D. Metaxas et al. (Eds.): MICCAI 2008, Part II, LNCS 5242, pp. 347–355, 2008.

in time. It has previously been shown that binocular eye tracking can be used to infer the 3D location and deformation of the epicardial surface at the surgeon's fixation point [3,4]. The method is robust to dynamic effects and performs well even for surfaces with homogenous appearance. It can be used to deliver gaze-contingent Augmented Reality (AR) and reduce computational overheads to the specific region of the surgeon's interest [5]. For its practical use, in addition to environmental factors and distractions, there are a number of factors that may influence the eye movement and fixations. During periods of visual fixation, small eye movements continuously move the projection of the image on the retina. These fixational eye movements include small saccades, slow drifts, and physiological nystagmus [6]. When the head is not immobilized, movements of the head and body combine with eye movements to further amplify the jittering of the images on the retina. To ensure the robustness of the gaze-contingent approach for robotic control, it is therefore necessary to introduce additional depth and motion cues to refine and filter out inherent errors and inadvertent jitters.

Recently, the computation of 3D structural and motion information automatically form stereo-laparoscope image data has made marked progress [7,8]. Such techniques can provide accurate information about the 3D soft-tissue deformation but their robustness is dictated by the presence of prominent structural features and can be affected by view dependent specular highlights. The results, however, are complementary to those derived from gaze-contingent deformation recovery, suggesting the potential value of combining the two techniques for achieving enhanced accuracy, robustness, and clinical applicability. The purpose of this paper is to propose a framework for enhancing robotic control in AF surgery by using eye gaze coupled with adaptive depth refinement using stereo vision to prescribe 3D paths on tissue surfaces for ablation using focused energy delivery. The method integrates initial 3D position of the surgeon's gaze with optical refinement for accurate 3D localization. The method is validated with a phantom model with known geometry and user experiments on a daVinci® surgical robot. To our knowledge, this is the first effort to fuse human and machine vision in robotic surgery.

## 2 Method

### 2.1 Gaze-Contingent 3D Depth Recovery

In human vision, the parallax between two retinal images is a fundamental depth cue. By measuring the ocular vergence, it is possible to extract quantitative information regarding the 3D position of the surgeon's fixation point [4]. This can be obtained non-intrusively through video-oculography where image sensors are used to measure the corneal reflection from a fixed IR light source in relation to the center of the pupil. These two measurements define a vector, which can be mapped to a unique eye gaze direction. The combined tracking of both eyes provides the binocular vergence measure, which in turn determines the 3D fixation point in the coordinate space of the calibrated stereo-laparoscope.

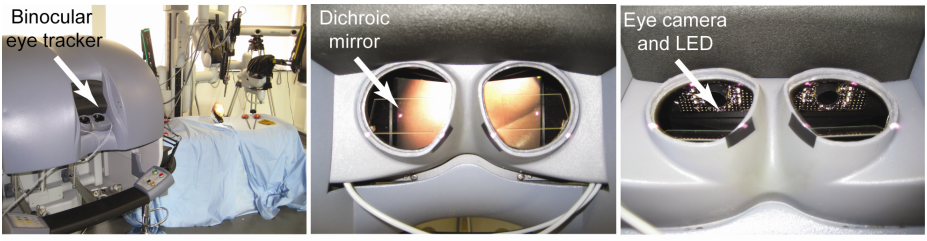



**Fig. 1. (a)** The binocular eye tracking system used in this study when mounted onto the master control console of the daVinci® surgical robot. **(b)** A view of the operating field observed by the surgeon and the dichroic mirrors used to capture the pupil and cornea reflection of the eye. **(c)** The eye tracking cameras and IR-LEDs used to generate the glint vectors.

In order to establish the relationship between pupil-glint vectors and points in 3D space, calibration is required prior to eye tracking. Different methods of binocular eye-tracking calibration exist and in the proposed system, the calibration also takes into consideration the robot help stereo-laparoscope camera characteristics [3,9]. A schematic illustration of the binocular eye tracking framework used in this study is shown in Fig 1. The system allows seamless localization of the surgeon's 3D fixation point relative to the operating field of view. The eye tracking unit is fully integrated with the surgical console and provides video images of the eyes without obstructing the surgeon's view. It consists of a pair of near infrared (IR) sensitive cameras, an array of externally switchable sub-miniature IR emitting diodes at 940nm, and a pair of dichroic beam splitters with their cut-off wavelength set above 750nm. This configuration allows appropriate illumination of the surgeon's eyes and image capture at a rate of 50 Hz. The outputs from the two cameras are digitized and processed in real-time by an eye tracking workstation to determine the 3D fixation of the surgeon.

## 2.2 Optical 3D Fixation Refinement

The binocular eye tracking system provides an estimate of the 3D fixation point, but its value may be compounded by a number of factors as mentioned previously. To ensure accurate tissue targeting, this information needs to be refined automatically using the image data and the known calibration of the stereo-laparoscope. We denote the 3D fixation estimate provided by binocular eye tracking as $\mathbf{X} = [X \quad Y \quad Z]^{\mathrm{T}}$ and its projections onto the left and right image planes as $\mathbf{x}_l = [x \quad y]^{\mathrm{T}}$ and $\mathbf{x}_r = [x' \quad y']^{\mathrm{T}}$. The projection is performed using a perspective camera model combined with non-linear distortion estimation [10]. The optical based fixation correction involves finding the optimal 3D point to evaluate a dissimilarity measure in the image space while maintaining consistency with the original eye fixation estimates, and further considering the reliability of image information alone.

In this study, zero mean normalized cross correlation is used as a dissimilarity measure in the image space. It is calculated centered around the re-projected fixations in each image and denoted as $\upsilon\left(\mathbf{x}_l, \mathbf{x}_r\right)$. A confidence measure is used to provide an indication on the reliability of the reference image region for matching. By assuming

that highly textured image regions are more reliable than those with homogeneous appearance, it is defined as:

$$s(\mathbf{x}) = \max\left\{I(\mathbf{x}+\delta) - I(\mathbf{x})\right\}, \delta \in \left\{(1,0),(-1,0),(0,1),(0,-1)\right\} \tag{1}$$

The confidence measure is normalized in the range $0 \le s(\mathbf{x}) \le 1$ within the template region to maintain consistency. The resulting cost function used to compute the likelihood for a 3D fixation then considers the similarity and confidence measures, as well as the distance away from the original fixation re-projection. The optically corrected image plane fixation $\mathbf{x}_l'$ is found by:

$$\arg\max_{\mathbf{x}_l} \left\{ \left(1 - s(\mathbf{x}_r)\right) \upsilon(\mathbf{x}_l', \mathbf{x}_r) + \frac{s(\mathbf{x}_r)}{2\pi\sigma^2} e^{-\frac{|\mathbf{x}_l - \mathbf{x}_l'|}{2\sigma^2}} \right\} \tag{2}$$

With the above notation, the right eye is assumed as dominant and used as a reference while the reprojection is optimized only in the left image space. However, this is not restrictive as the fixation re-projection in one (or both) of the eyes can be used as a reference and the strategy can accommodate ocular dominance switching in normal vision [11]. Since the search space is constrained by using the information derived from eye tracking, global optimization with exhaustive winner-takes-all can be used to minimise Eq. (2). For all experiments reported in this study the template size was predefined as 11×11 pixels and the respective search space was 70×50 pixels.

### 2.3 Fixation Path Optimization

With the refined estimation of 3D fixation points, the surgeon is able to pinpoint specific locations on the soft-tissue surface. Once the required ablation area is exposed, the gaze-contingent control scheme can be used to prescribe a desired ablation path. Prior to this, a final path optimization is performed before focused energy delivery. The reason for this is that 3D fixations collected during the path prescription process also contain fixational eye movements due to the natural behavior of the eyes.

To fit an optimal path through the collected 3D point set, we use a robust non-parametric clustering technique to reduce the dimensionality of the point cloud and lock onto a suitable path. To avoid the *a priori* selection of a fixed number of fixation clusters, a fast Adaptive Mean Shift (AMS) algorithm was used [12]. To estimate the unknown density function of the data $f(\mathbf{X})$, a kernel estimation function $\hat{f}_K(\mathbf{X})$ is defined as:

$$\hat{f}_K(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^3} K_E\left( \left\| \frac{\mathbf{X} - \mathbf{X}_i}{h_i} \right\|^2 \right) \tag{3}$$

Where $h_i$ is the bandwidth parameter, which has a critical effect on performance but is calculated adaptively in AMS using the $k$ - nearest neighbors algorithm to each data point. In this study, $K_E(\mathbf{X})$ is chosen as the Epinechnikov kernel which has the profile $k_E(x) = 1 - x$ when $0 \le x \le 1$ and $k_E(x) = 0$ when $x$ is outside the

designated range. The resulting spherically symmetric kernel guarantees convergence and is computationally efficient [13]:

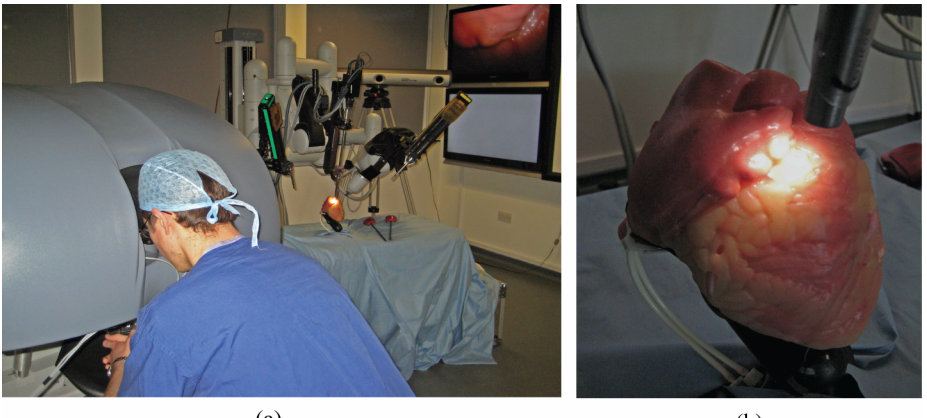
$$K_E(\mathbf{X}) = \begin{cases} \frac{15}{8\pi}(1 - \|\mathbf{X}\|^2) \\ 0 \end{cases} \tag{4}$$

By iteratively choosing sub-samples from the complete 3D fixation point cloud, the procedure converges to produce a robust nonparametric clustering of the data where the output modes represent the cluster centers. The modes are then used as the discrete control points defining the ablation path.

## 3 Experiments and Results

The proposed framework was implemented using master console of the daVinci® surgical robot shown in Fig 2(a). The eye tracking and processing software was optimized to run in real-time on a standard PC. Stereo video streams from the stereo-laparoscope were processed by a separate PC and the two systems were directly connected through a crosswire network cable to enable real-time synchronization. For validation against ground truth data, a phantom heart model (The Chamberlaine Group, MA USA) was scanned using a Siemens Sonata 1.5T MRI Scanner. The ground truth data was registered to the visual scene using fiducial markers and the absolute orientation method developed by Horn [14].

For experimental validation, eight subjects were asked to prescribe an ablation path near the right atrium of the phantom heart using the proposed gaze-contingent framework. Camera and subject-specific eye tracker calibration were performed



(a) (b)

**Fig. 2.** Phantom and user experiment setup where **(a)** shows the binocular eye tracking device used in this study mounted onto the daVinci® surgical console and **(b)** illustrates the stereo-laparoscope observing the phantom heart model

before each experiment. No time constraint was imposed and the subjects were only instructed to follow a smooth ablation path with their eyes. By varying fixation duration, the subject could increase the importance of points to the final result.

Fig 3(a) shows the fixation path for one of the subjects studied (S1). It is overlaid onto images of the phantom heart captured from the stereo-laparoscope. The path consists of 703 fixations, which are accurately reconstructed in metric 3D space as shown in Fig 3(b), and subsequently used to create an ablation control signal using the proposed technique.

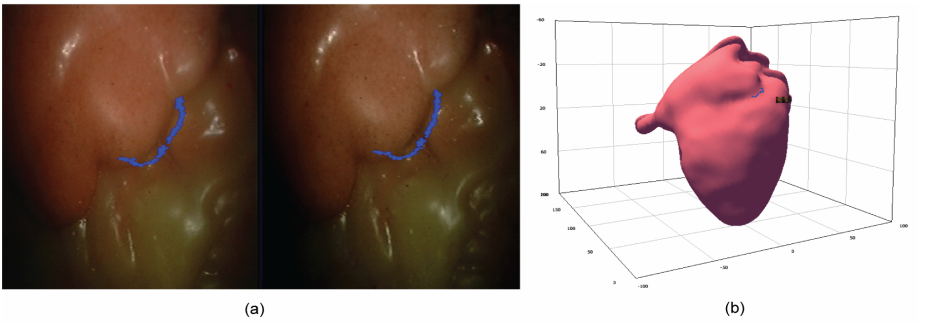



**Fig. 3. (a)** Stereoscopic fixation path superimposed onto the two stereo-laparoscope images of the phantom heart model. **(b)** 3D Rendition of the corresponding 3D fixation path within the 3D scene showing the ground truth of the phantom heart model.

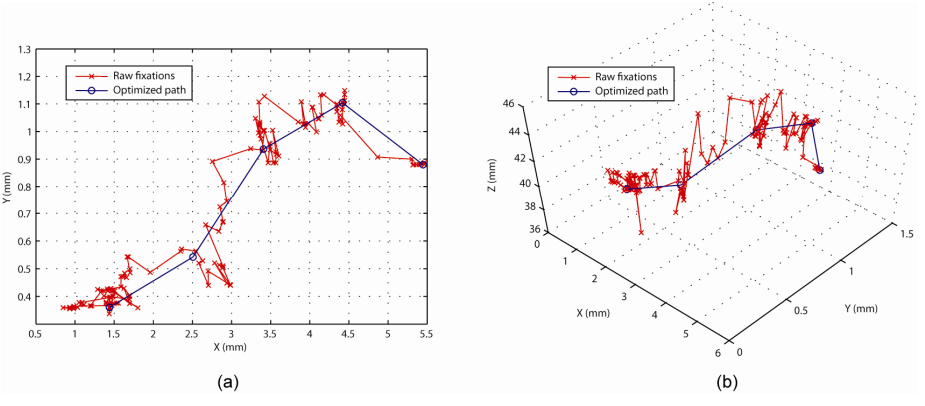



**Fig. 4. (a)** Fixation path along the X and Y axes of the metric camera coordinate system and the corresponding optimized path using data clustering. **(b)** 3D rendition of the same fixation as (a) in the full 3D coordinate space.

The 3D phantom geometry shown in Fig 3(b) is used to measure the error between the 3D fixation reconstruction and the 3D scene geometry. For measurements, we use the Eucledian distance between the recovered 3D fixation point and the point of contact between phantom model and the respective line of sight direction from the eyes. Fig 4(a) shows a subset of the fixation path in Fig 3 in more detail and the

subsequently optimized path for the 150 fixation points along the X and Y axes of the metric coordinate frame. It is evident that the robust path locking provides a jitter free result that is amenable for ablation procedures. Fig 4(b) further illustrates the full 3D metric representation of the path in Fig 4(a).

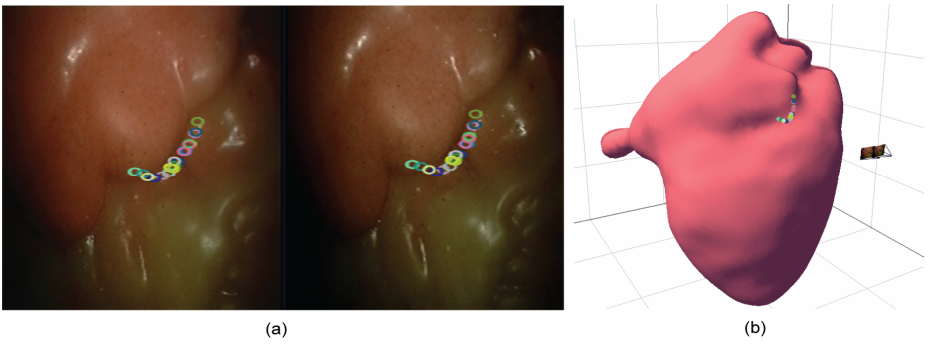



**Fig. 5. (a)** The optimized fixation path from Fig. 3(a) with modes re-projected back into the image space. **(b)** 3D rendition of the corresponding 3D modes in the metric scene with a dense reconstruction of the phantom model.

The re-projection of 3D fixation modes determined using AMS is shown in Fig 5(a). For this subject, the fixation points are reduced to 32 modes that describe the desired ablation path. For the experiments reported in this study, the direct control signal from AMS clustering was used, however, in a final system spline interpolation can be used to generate a smooth path and aviod potential collisions. To help visual assessment of the derived result, Fig 5(b) shows the modes in the 3D metric space with a view of the phantom model mesh.

Table 1 summarizes the results obtained for all eight subjects involved in this study. In this table, the mean and standard deviation of error distance for individual fixations are shown in millimeters. It also provides the final ablation modes derived

**Table 1.** Results obtained for the eight subjects peforming the ablation task on the phantom model. The table provides the mean and standard deviation of error for 3D fixations obtained from the fusion process, as well as for the final modes estimated using AMS. All measurements are reported in metric space with units in millimeters.

| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | **Mean [Std]** |
|---|---|---|---|---|---|---|---|---|---|
| **Fix. Err Mean** | 1.82 | 3.15 | 2.22 | 1.92 | 2.36 | 2.24 | 2.27 | 2.10 | **2.25 [±0.41]** |
| **Fix. Err Std** | 1.83 | 3.00 | 1.18 | 1.19 | 1.50 | 1.63 | 0.95 | 1.36 | **1.58 [±0.63]** |
| **Num Fix.** | 703 | 611 | 499 | 399 | 420 | 340 | 661 | 326 | |
| **Mode Err** | 1.81 | 3.00 | 2.20 | 1.88 | 2.33 | 2.14 | 2.27 | 2.05 | **2.21 [±0.37]** |
| **Mode Std** | 1.80 | 2.72 | 1.14 | 1.12 | 1.47 | 1.12 | 0.93 | 1.31 | **1.44 [±0.58]** |
| **Num Modes** | 32 | 34 | 21 | 16 | 18 | 11 | 28 | 15 | |
| **Time (sec)** | 28.2 | 24.4 | 20.0 | 16.0 | 16.8 | 13.6 | 26.4 | 13.0 | **19.8[±0.59]** |

by using the proposed AMS procedure. Overall, the magnitude of the error is about 2.2mm for all subjects. It should be noted that this error includes the registration error between the 3D model and MR segmentation.

## 4 Discussion and Conclusions

In this paper, we have presented an original framework for fusing human and machine vision for robotic control in MIS. The technique relies on the 3D fixation points of the surgeon derived through binocular eye-tracking and their subsequent refinement with machine vision. The temporal series of 3D fixation points is optimized to define a control path using nonparametric clustering. Results on a phantom model with a group of subjects demonstrate the practical potential of the technique. With further developments in focused energy delivery and cardiac port access surgery, the proposed method represents an attractive way forward for the development of new HMI control for tissue ablation. Future work involves extending the method to explicitly handle temporal variation and accommodate soft-tissue deformation. In addition, an auxiliary interface must be developed to indicate start/stop of control path definition and the final correct path placement before ablation commences. This will assist the system to handle aberrant eye movements and incorrect control prescription.

## References

1. Grossi, E.A., et al.: Minimally invasive mitral valve surgery: a 6-year experience with 714 patients. Ann Thorac Surg. 74, 660–664 (2002)
2. Smith, J.M.: Robots and lasers: The future of cardiac tissue ablation. Int. J. Med. Robotics Comp. Assist Surg. 2, 329–332 (2006)
3. Mylonas, G., Darzi, A., Yang, G.-Z.: Gaze-contingent soft tissue deformation tracking for minimally invasive robotic surgery. Comp. Aided Surg. 11(5), 256–266 (2006)
4. Mylonas, G., Stoyanov, D., Deligianni, F., Darzi, A., Yang, G.-Z.: Gaze-Contingent Soft Tissue Deformation Tracking for Minimally Invasive Robotic Surgery. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 843–850. Springer, Heidelberg (2005)
5. Lerotic, M., Chung, A.J., Mylonas, G., Yang, G.-Z.: pq-space Based Non-Photorealistic Rendering for Augmented Reality. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 102–109. Springer, Heidelberg (2007)
6. Yang, G.-Z., Dempre-Marco, L., Hu, X.-P., Rowe, A.: Visual Search: Psychophysical Models. Practical Applications, Image Vision Comput. 20(4), 291–305 (2002)
7. Stoyanov, D., Darzi, A., Yang, G.-Z.: A practical approach towards accurate dense 3D depth recovery in robotic laparoscopic surgery. Comp. Aid Surg. 10(4), 199–208 (2005)
8. Stoyanov, D., Mylonas, G., Deligianni, F., Yang, G.-Z.: Soft-Tisse Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 139–146. Springer, Heidelberg (2005)
9. Mylonas, G., Darzi, A., Yang, G.-Z.: Gaze Contingent Depth Recovery and Motion Stabilisation for Minimally Invasive Robotic Surgery. In: Yang, G.-Z., Jiang, T. (eds.) MIAR 2004. LNCS, vol. 3150, pp. 311–319. Springer, Heidelberg (2004)
10. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)

11. Khan, A.Z., Crawford, J.D.: Occular Dominance Reverses as a Function of Horizontal Gaze Angle. Vision Research 41, 1743–1748 (2001)
12. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach Towards Feature Space Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603–619 (2002)
13. Georgescu, B., Shimshoni, I., Meer, P.: Mean Shift Based Clustering in High Dimensions: A Texture Classification Example. In: ICCV 2003, vol. 1, pp. 456–463 (2003)
14. Horn, B.: Closed-form Solution of Absolute Orientation Using Unit Quarternions. J. Opt. Soc. Am. 5(7), 629–642 (1987)