2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Tracking Intraocular Microdevices Based on Colorspace Evaluation and Statistical Color/Shape Information

Christos Bergeles, Georgios Fagogenis, Jake J. Abbott, and Bradley J. Nelson

Abstract— **Successful ophthalmic surgeries using intraocular untethered microrobots or tethered robotic microtools require methods to robustly track the microdevices in the posterior of the human eye. The dimensions and specularities of the microdevices are major obstacles for accurate tracking. In addition, the optical structure of the human eye makes it challenging to keep the objects of interest constantly in focus, resulting in blurred images. In this paper, the advantages of using different colorspaces for intraocular tracking are examined. After selection of the appropriate colorspace, thresholds that ensure maximum separation of the device from the background are calculated. Based on trained color histograms, level sets are used to track in real time, and the use of statistical shape information is incorporated in the existing tracking framework. The efficacy of the algorithm is demonstrated by tracking a microrobot in a model eye, using a custom made ophthalmoscope and off-the-shelf ophthalmoscopy lenses. With the appropriate colorspace and threshold selection, tracking errors are minimized and are further diminished using shape information.**

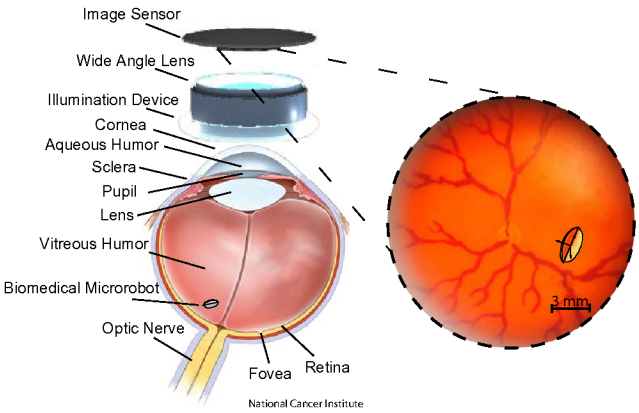


Fig. 1. Illustration of the experimental ophthalmoscopic device, together with the structure of the human eye, and an example of a real microrobot with an overlayed frame in the model eye [7].

I. Introduction

Vitreoretinal surgery requires accuracy and dexterity that is often beyond the limits of human surgeons. A number of robot-assisted solutions have been proposed and developed to improve surgeons' ability to perform these difficult procedures. Our work is motivated by the wireless magnetically steered intraocular microrobot proposed in [1], and can be used with other dextrous microtools [2], [3] as well.

The most sensitive and important structure of the human eye is the retina. In order to control microdevices operating near the retina, knowledge of their position is required. In the case of untethered magnetic devices, knowledge of the position of the device within the magnetic field is needed [4], [5]. The interior of the human eye is externally observable, so vision can be used to perform 3D localization, but for precise localization, robust visual tracking is needed.

The optical elements of the human eye (see Fig. 1) comprising the cornea, the aqueous humor, the lens and the vitreous humor alter the formation of images, and make the use of additional ophthalmic lenses necessary for intraocular imaging [6]. Keeping intraocular objects constantly in focus is challenging, and the captured images are often blurry and noisy. The unstructured illumination that reaches the interior of the eye, either through transpupilary or transscleral means, can deteriorate the images with uneven brightness and backreflections. Moreover, the microtools that operate in the human eye are generally specular, and have no distinctive color features. Due to the noisy nature of the images, edge based algorithms do not operate well.

In this paper, appropriate colorspace selection accompanied by a level-set based tracking method using statistical color/shape information is proposed for biomedical applications. In Sec. II we evaluate the appropriateness of colorspaces, and calculate decision thresholds that ensure object-from-background separation. In Sec. III we enhance a fast probabilistic level-set tracking algorithm to make use of the separation thresholds and statistical information from shape priors. As a result, it can track defocused objects in noisy images. In Sec. IV we demonstrate the efficacy of the method by tracking an intraocular microrobot, and conclude with our observations in Sec. V.

II. Colorspace Selection

The utilization of different colorspaces for image segmentation and visual tracking can significantly enhance algorithmic performance. Each colorspace provides a different representation of color information, and as a result, some are more appropriate for specific applications. In [8] the authors propose a method to choose the best colorspace online, and in [9] the authors create (off-line) a colorspace tailored for specific object tracking. In the field of minimally invasive surgery, [10] proposes using the Hue and Saturation channels of the HSV colorspace to track surgical instruments in the abdominal cavity. In [11] image binarization and

This work was supported by the NCCR Co-Me of the Swiss National Science Foundation.

The authors are with the Institute of Robotics and Intelligent Systems, ETH Zurich, 8092 Zurich, Switzerland {cbergeles, georgefa, bnelson}@ethz.ch. Jake J. Abbott is now with the Department of Mechanical Engineering, University of Utah, Salt Lake City, 84112 Utah, jake.abbott@utah.edu.

Hue/Saturation channels are used in order to segment the lumen, nerves and endoscope in the spinal coord, and in [12], Hue and Homogeneity are used to segment endoscopic images. We are mainly interested in tracking microtools in the posterior of the human eye. In this paper colorspace selection is performed as a preprocessing step based on training data.

A. Colorspace Evaluation

The object of interest Ω_{o_j} is manually segmented from $j = 1 \dots k$ images. The complement of this region, Ω_{b_j}, corresponds to the background in the images. Using the segmented regions, the object's and background's joint-histograms for the channels/colorspace of interest are calculated, resulting in the probabilities $P\Big(\mathbf{u} \mid \mathbf{x} \in \Omega_o\Big)$, and $P\Big(\mathbf{u} \mid \mathbf{x} \in \Omega_b\Big)$, where $\mathbf{x}$ is the pixel of interest, Ω_o is the object region, Ω_b is the background region, and $\mathbf{u}(\mathbf{x}) \in \mathbb{R}^n$ is the observed intensity values vector for the n channels of interest. We define an extended region $\Omega_{e_j} = \Omega_{o_j} \oplus \omega$ where ω is the structural element of preference, such that $\Omega_{o_j} \subset \Omega_{e_j}$. The set $\Omega_{n_j} = \Omega_{e_j} \setminus \Omega_{o_j}$ corresponds to the background in the object neighborhood in image j.

The simplest object-from-background separation criterion is a decision function $F(\mathbf{u}(\mathbf{x})) : \mathbb{R}^n \to \mathbb{R}$, where the pixel $\mathbf{x}$ is classified as belonging to the object ($F > 0$) or as belonging to the background ($F < 0$). The value of this function is based on the previously calculated probabilities.

The desired colorspace is the one that minimizes the misclassifications of background pixels as object pixels, while maximizing the correct classifications. We consider as object pixels the region Ω_{o_j} in each image. The colorspace quality for each image j is calculated by:

$$r_j = c_o + c_n - m_o - m_n \tag{1}$$

where c_o is the ratio of correct object classifications to object size (in pixels), c_n is the ratio of correct neighborhood classifications to neighborhood size, and m_o, m_n the respective misclassification ratios. By comparing the mean and variance of r for different colorspaces, the most appropriate colorspace can be selected.

B. Maximum Separability Thresholding

The classification of a pixel as belonging to the object or the background depends on the value of the decision function F. Instead of using a binary criterion for the classification, we estimate decision thresholds above which a pixel is treated as part of the object, and below which a pixel is considered as part of the background. For values between the two thresholds, no direct decision should be made. However, a decision can be made based on the neighbors, similar to the hysterisis thresholding of Canny's algorithm. The thresholds are estimated by minimizing, for each segmented image and colorspace, the objective function:

$$r_j(t_l, t_h) = (c_o - 1)^2 + (c_n - 1)^2 + m_o^2 + m_n^2 \tag{2}$$

where t_l and t_h are the lower and higher thresholds, respectively. Minimizing this function for each image for different colorspaces ensures the maximum correct classifications and the minimum misclassifications. The thresholds for each colorspace are finally averaged and can be used for more accurate pixel classification, together with the appropriate colorspace representation. Depending on the significance of each term of (2), weights can be added.

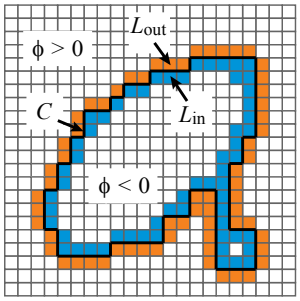


Fig. 2. The sets of pixels where the level-set function is updated. C is the propagating front. This level-set function could represent a tracked microrobot and an erroneously segmented part of the background.

III. Level Set Tracking using Statistical Color/Shape Information

The most common approaches in tracking use active contours, evolving either as parameterized snakes [13], or as higher-order level-set functions [14]. The first approach is computationally efficient, but needs explicit methods to handle reparameterization and topology changes. The second approach is independent of topology and parametrization, but is computationally more complex. Recently, a real-time tracking algorithm based on level sets has been proposed [15], and thus, the computational complexity of tracking using level sets can be overcome.

A. Real-Time Tracking Using Level Sets

The efficacy of the algorithm in [15] lies in the idea that instead of updating the full level-set function, one can update pixels near the propagating front, similarly to the narrow band algorithm presented in [14], or the HERMES algorithm presented in [16]. The algorithm [15] also avoids the computational overhead of solving the curve evolution PDE. The propagating front pixels belong to two sets (see Fig. 2):

$$\begin{aligned} L_{\text{out}} &= \{\mathbf{x} \mid \phi(\mathbf{x}) > 0,\, \exists \mathbf{y} \in N_4(\mathbf{x}) \text{ such that } \phi(\mathbf{y}) < 0\}, \\ L_{\text{in}} &= \{\mathbf{x} \mid \phi(\mathbf{x}) < 0,\, \exists \mathbf{y} \in N_4(\mathbf{x}) \text{ such that } \phi(\mathbf{y}) > 0\} \end{aligned}$$

where $N_4(\mathbf{x})$ are the 4-connectivity neighbors of $\mathbf{x}$, and ϕ is a level-set function:

$$\phi(\mathbf{x}) = \begin{cases} +3 & \text{if } \mathbf{x} \text{ is an exterior pixel,} \\ +1 & \text{if } \mathbf{x} \in L_{\text{out}}, \\ -1 & \text{if } \mathbf{x} \in L_{\text{in}}, \\ -3 & \text{if } \mathbf{x} \text{ is an interior pixel.} \end{cases}$$

Two procedures *switch_in()* and *switch_out()* are defined:

- *switch_in*($\mathbf{x}$) switches the pixel $\mathbf{x} \in L_{\text{out}}$ to L_{in}, and adds its 4-neighboring exterior pixels to L_{out}.

- *switch_out(*$\mathbf{x}$*)* switches the pixel $\mathbf{x} \in L_{\text{in}}$ to L_{out}, and adds its 4-neighboring interior pixels to L_{in}.

The evolution of the sets is carried out in two cycles:

- Cycle One: The decision whether a pixel $\mathbf{x}$ from L_{in} and L_{out} should be switched depends on function F. A decision function can be $F = \log\left(\frac{P(\mathbf{u}(\mathbf{x})|\Omega_o)}{P(\mathbf{u}(\mathbf{x})|\Omega_b)}\right)$, where $\mathbf{u}(\mathbf{x})$ is the feature vector defined at pixel $\mathbf{x}$, Ω_o is the object of interest, and Ω_b is the background. The probabilities are calculated based on the joint-probability histograms that are determined in the training phase, and the features used correspond to the channels of the colorspace that exhibited the maximum quality in Sec. II-A. A pixel $\mathbf{x}$ belonging to L_{out} is passed to *switch_in* if $F(\mathbf{x}) > t_h$, and a pixel $\mathbf{x}$ belonging to L_{in} is passed to *switch_out* if $F(\mathbf{x}) < t_l$. The thresholds t_l and t_h are calculated in the preprocessing step of Sec. II-B. This is a deviation from the original algorithm of [15], and increases the tracking robustness.
- Cycle Two: The switches occur depending on smoothness restrictions. Instead of smoothing the whole level-set function, only the values at L_{out} and L_{in} are smoothed (see [15] for details).

The presented tracking algorithm makes use of statistical color information based on training images, but does not use information about the shapes of the objects of interest. We further increase the tracking robustness by adding a third evolution cycle so that the evolving curve resembles a known shape. In this way, desired objects can be segmented from noisy environments using both color and shape information.

B. Creating a Statistical Shape Model

Algorithms that use known shape information (shape prior) consist of three stages. At the first stage, a training set is created from representations of the desired object. These object representations can be the ones that are used in order to create the joint-probability histogram required for tracking. In order for the images to be used for the shape prior extraction, they should be registered. We register the level-set surfaces corresponding to the binary images of the segmented objects [17], using the method presented in [18] for the 2D translation registration, followed by scale and rotation registration.

The second stage consists of creating a model of the desired object. In [19], the singular value decomposition of a matrix consisting of the vectorized registered level-set functions results in the extraction of eigenshapes. A linear combination of eigenshapes creates a shape prior with the desired precision:

$$\bar{\phi}(\mathbf{a}, \mathbf{p}) = \mu(\mathbf{p}) + \sum_{i=1}^{k} a_i v_i(\mathbf{p}) \tag{3}$$

where $\bar{\phi}$ is the shape prior, μ is the mean level-set function, $\mathbf{p}$ is a vector containing pose parameters, v_i are the k extracted eigenshapes, and $\mathbf{a} = \{a_1, \ldots, a_k\}$ are the weights of the eigenshapes. This representation of the shape prior enables it to be used in a statistical fashion, and in accordance with the level-set tracking algorithm of Sec. III-A.

The third stage consists of estimating the most probable shape prior (i.e. estimate its eigenweights $\mathbf{a}$ and pose parameters $\mathbf{p}$) based on the current evolving level-set function. In [19], this problem is formulated as an *a posteriori* probability maximization problem. We reformulate this approach in order to make it fit the probabilistic evolution of the presented tracking algorithm.

The best estimates for the shape eigenweights $\mathbf{a}$, and pose parameters $\mathbf{p}$ are given as:

$$\langle \mathbf{a}^\star, \mathbf{p}^\star \rangle = \arg\max_{\mathbf{a},\mathbf{p}} P(\mathbf{a}, \mathbf{p} \mid \phi, I) \tag{4}$$

where ϕ is the current level-set function, and I represents image information (e.g. gradient, histogram). Based on Bayes' rule, the previous equation can be rewritten as:

$$P(\mathbf{a}, \mathbf{p} \mid \phi, I) = \frac{P(\phi \mid \mathbf{a}, \mathbf{p}) P(I \mid \mathbf{a}, \mathbf{p}, \phi) P(\mathbf{a}) P(\mathbf{p})}{P(\phi, I)} \tag{5}$$

where for simplification it is assumed that pose and shape are independent (no projective distortion), and thus, $P(\mathbf{a})$ and $P(\mathbf{p})$ can be calculated separately.

Term $P(\phi \mid \mathbf{a}, \mathbf{p})$ is the probability that a given level-set function ϕ is observed based on shape-prior parameters $\mathbf{a}$ and $\mathbf{p}$ (the estimated shape prior is $\bar{\phi}(\mathbf{a}^\star, \mathbf{p}^\star) = \phi^\star$). Contrary to [19], since the shape prior will be used for tracking, we cannot assume that the evolving zero level set of ϕ lies inside the estimated zero level set of $\phi^\star$. The term can be equal to the Laplacian of the difference between the current level-set function, and the estimated shape prior:

$$P(\phi \mid \mathbf{a}, \mathbf{p}) = \exp(- \parallel \phi - \phi^\star \parallel) \tag{6}$$

which takes its maximum value when ϕ and $\phi^\star$ are identical.

Term $P(I \mid \mathbf{a}, \mathbf{p}, \phi)$ is the probability of observing certain image features given the current and estimated level-set functions. In [19], the observed feature in the image is the gradient, whereas here, the features are the intensity values of the channels of the chosen colorspace. The interior of the estimated shape prior $\phi^\star$ should have a color histogram that is similar to the one of the evolving level-set function ϕ. As a result, the value of this term can be the Laplacian of this similarity:

$$P(I \mid \mathbf{a}, \mathbf{p}) = \exp(- \parallel h - h^\star \parallel) \tag{7}$$

where h and $h^\star$ are the joint-probability histograms for the objects segmented by ϕ and $\phi^\star$ respectively.

For the terms $P(\mathbf{a})$ and $P(\mathbf{p})$ in (5), we use [19]:

$$P(\mathbf{a}) = \frac{1}{\sqrt{(2\pi)^k \mid \Sigma_k \mid}} \exp\left(-\frac{1}{2} \mathbf{a}^{\mathbf{T}} \Sigma_k^{-1} \mathbf{a} \right) \tag{8}$$

$$P(\mathbf{p}) = U(-\infty, \infty) \tag{9}$$

where Σ_k is a $k \times k$ diagonal matrix containing the eigenvalues corresponding to the extracted eigenshapes, and U denotes the uniform distribution.

The denominator term of (5) has no dependency on shape or pose, and can be disregarded from the probability maximization.

TABLE I
CYCLE THREE OF THE LEVEL SET TRACKING ALGORITHM

- Compute the shape prior with the maximum *a posteriori* probability $\phi^\star$, based on (5) and the current level-set function ϕ.
- For i = 1:N_s do
  - $\forall \mathbf{x} \in L_{\text{out}}$, if $\phi^\star(\mathbf{x}) < 0$, then *switch_in*($\mathbf{x}$).
  - $\forall \mathbf{x} \in L_{\text{in}}$, if $\forall \mathbf{y} \in N_4(\mathbf{x}), \phi(\mathbf{y}) < 0$, delete $\mathbf{x}$ from L_{in}, and set $\phi(\mathbf{x}) = -3$.
  - $\forall \mathbf{x} \in L_{\text{in}}$, if $\phi^\star(\mathbf{x}) > 0$, then *switch_out*($\mathbf{x}$).
  - $\forall \mathbf{x} \in L_{\text{out}}$, if $\forall \mathbf{y} \in N_4(\mathbf{x}), \phi(\mathbf{y}) > 0$, delete $\mathbf{x}$ from L_{out}, and set $\phi(\mathbf{x}) = +3$.

C. Evolution Using Statistical Shape Information

The estimated shape prior $\phi^\star$ is used as an evolutionary force acting on the propagating front. The estimation can be done using the sets L_{in} and L_{out} to speed up the minimization process. In the framework of the algorithm of Sec. III-A, pixels belonging to the background (object) should be subjected to a negative (positive) force. The predicted shape prior $\phi^\star$ is a level-set function, and as such, if C is its zero-level curve, it is negative in the interior of C, and positive in its exterior. Based on this, a third cycle can be introduced in the level-set tracking algorithm:

- Cycle Three: Pixels $\mathbf{x}$ belonging to L_{out} are passed to *switch_in* if they are inside the estimated shape prior $\phi^\star$, and pixels $\mathbf{x}$ belonging to L_{in} are passed to *switch_out* if they are outside the estimated shape prior.

Cycle three iterates until convergence, or until a maximum number of iterations is reached. Convergence is achieved when the evolving level-set function ϕ, coincides with the estimated shape prior $\phi^\star$. This might not be desirable if ϕ should capture object variability as well. Thus, selecting the appropriate number of iterations requires some experimentation. The exact structure of the third cycle of the algorithm can be seen in Table I.

IV. BIOMEDICAL MICROROBOT TRACKING

We run a Matlab™ implementation of the above algorithm on a Pentium 4 2×3.4 GHz, and conduct three experiments of different complexity. The imaging system is an Edmund Optics 602FC camera, with a FUJINON HF9HA-1B lens.

A. Tracking in Silicone Oil Using Shape Information

In the first experiment, we track a microrobot of the type presented in [1] in silicone oil, but we add an artificial occlusion in the images. With shape information the microrobot is tracked successfully despite the occlusion (see Fig. 3).

B. Tracking in a Model Eye Using Color Information

In the third experiment, we place the biomedical microrobot in a model eye [7] that is commonly used for training by ophthalmologists. The eye is filled with silicone oil in order for the microrobot to "swim" in it.

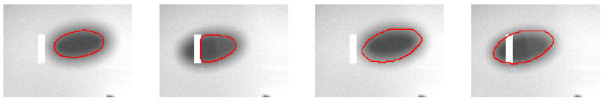


Fig. 3. Microrobot moving from right to left in silicone oil with an artificial occlusion. In (a), (b) the tracking fails, but in (c), (d) the statistical shape information results in successful tracking.

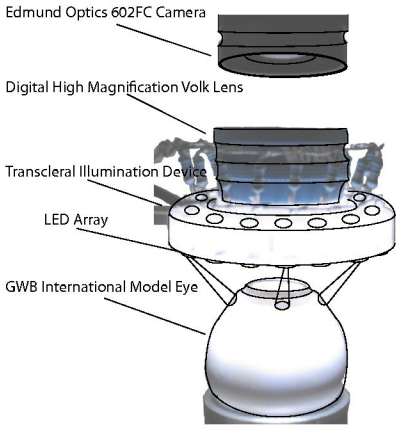


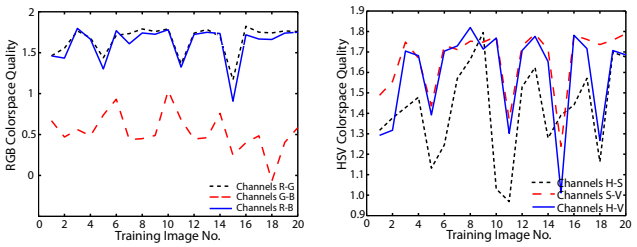
Fig. 4. The experimental imaging setup of Sec. IV-B.

Imaging microdevices in the posterior of an eye is a difficult task in and of itself [6]. Depending on the application, specific ophthalmic lenses must be considered. Here, we require a compromise between high field-of-view and high magnification. We use the Digital High Mag® lens [20] from Volk Optical Inc., which enables a field-of-view of $\sim 60^\circ$, and a magnification of $1.30\times$. This lens is designed for retinal imaging. Capturing the images with the camera ensures a larger depth-of-field. (see Fig. 4).

For illumination inside the model eye we constructed an apparatus based on transscleral illumination principles consisting of multiple LEDs arranged in a circle around the model eye sclera. Transscleral illumination is non-invasive and limits the backreflections from the lens surfaces.

We evaluate the quality of every two-channel combination for some basic colorspaces (see Table II). Using all channels of a colorspace will lead to a large overhead, and we want to examine the efficacy of the tracker with limited information. In Fig. 5(a), and Fig. 5(b), the quality of the channels for the RGB and HSV colorspaces can be seen, respectively. For the RGB colorspace, the highest quality is exhibited by the R-G channel combination. The combination of the H-S channels for the HSV colorspace leads to worst object-from-background separation, even though the H-S channels have been shown to be suitable for tracking tools and endoscopes in the human body [10]–[12]. This indicates that different biomedical applications benefit from the use of specific channels/colorspaces, and thus, their selection is an important preprocessing step.

In Table II, we show the quality mean and variance for



(a) (b)

Fig. 5. Quality measure for the (a) RGB colorspace, and (b) HSV colorspace. The lowest points in the graphs correspond to images where the microrobot is occluding or in the vicinity of a retinal vein.

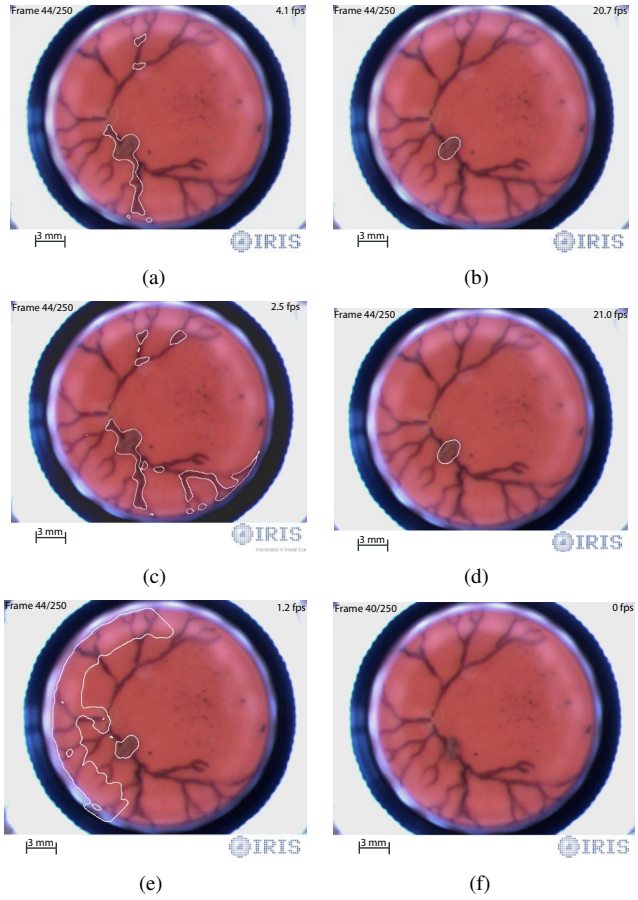
TABLE II
COLORSPACE MEAN QUALITY AND VARIANCE FOR DIFFERENT CHANNEL COMBINATIONS

	$\mathbf{1-2}$	$\mathbf{2-3}$	$\mathbf{1-3}$
RGB	1.66 ± 0.03	0.53 ± 0.06	1.61 ± 0.05
YUV	1.52 ± 0.08	1.63 ± 0.05	1.64 ± 0.04
YIQ	1.63 ± 0.04	1.63 ± 0.03	1.56 ± 0.03
HSV	1.41 ± 0.06	1.66 ± 0.07	1.58 ± 0.06
XYZ	1.59 ± 0.06	1.50 ± 0.07	1.59 ± 0.06
nRGB	1.49 ± 0.05	1.47 ± 0.05	1.49 ± 0.05

channel combinations of typical colorspaces. The channels R-G of the RGB colorspace are the most suitable for tracking, since they have the highest mean quality, and the lowest quality variance. Another good selection is the channels Y-V of the YUV colorspace. We also perform an experiment using the H-S channels of the HSV colorspace, since this selection is appropriate for other biomedical applications. Figure 6(a) shows tracking results for the R-G channels, and Fig. 6(c) shows the results for the Y-V channels; using the best channels/colorspace leads to reduced vein segmentation. Tracking in H-S leads to very poor results (see Fig. 6(e)).

Next, we impose calculated thresholds to the R-G ($t_l = -0.17$, $t_h = 1.71$), the Y-V ($t_l = -0.11\,, t_h = 1.69$), and the H-S ($t_l = -0.77\,, t_h = 0.77$) tracking cases. Successfull tracking occurs at ~ 25 fps, and typical frames can be seen in Fig. 6(b), Fig. 6(d), and Fig. 6(f), respectively. It can be seen that in the R-G, and Y-V cases (i.e. when the chosen channels/colorspace exhibits high quality) thresholding increases the tracking robustness. In the H-S case, the thresholds result in the tracker losing the microrobot.

To accurately estimate the accuracy achieved, we establish ground truth values by manually segmenting the microrobot in 40 equally spaced frames. The microrobot's contour is an ellipse, and the relative errors are calculated with respect to the ellipse's center (X_c, Y_c), the major (A) and minor (B) axis, and the orientation angle (ϕ). Moreover, in order to quantify the importance that thresholding has on a non-optimal colorspace, we track the microrobot using the Y-U channels of the YUV colorspace ($t_l = -0.13\,, t_h = 1.64$). The errors can be seen in Table III, and show that even though thresholding has a great impact on proper tracking, carefully selecting both the colorspace and thresholds leads to the best results.



(a) (b) (c) (d) (e) (f)

Fig. 6. Tracking using (a), (b) the R-G channels of the RGB colorspace without and with thresholds, respectively, (c), (d) the Y-V channels of the YUV colorspace without and with thresholds, respectively, (e), (f) the H-S channels of the HSV colorspace without and with thresholds, respectively.

TABLE III
TRACKING ERRORS FOR DIFFERENT COLORSPACES

	$\delta\mathbf{X_c}$ **(pix)**	$\delta\mathbf{Y_c}$ **(pix)**	$\delta\mathbf{A}$ **(pix)**	$\delta\mathbf{B}$ **(pix)**	$\delta\phi$ **(deg)**
RG	2.15	2.06	2.36	1.65	7.62
YV	2.03	2.20	2.38	1.82	8.26
YU	3.72	9.30	2.96	2.70	18.30

C. Tracking in a Model Eye Using Color/Shape Information

In the last experiment, we demonstrate the effect of tracking in the best colorspace using shape information. Although the method presented in Sec. III-B may not be most appropriate for rigid bodies that exhibit projective distortion, in our case, the shape variability of the microrobot projections can be effectively captured with the extracted eigenshapes (we use 4 eigenshapes). We track the microrobot using the R-G channels of the RGB colorspace, and we impose relatively lesser thresholds ($t_l = 0$, $t_h = 0.6$) since this ensures that the full microrobot is always detected even with some misclassifications. The misclassifications are discarded by the shape information. Figure 7 compares tracking in R-G, and tracking in R-G using shape information; when shape information is incorporated the results are improved.

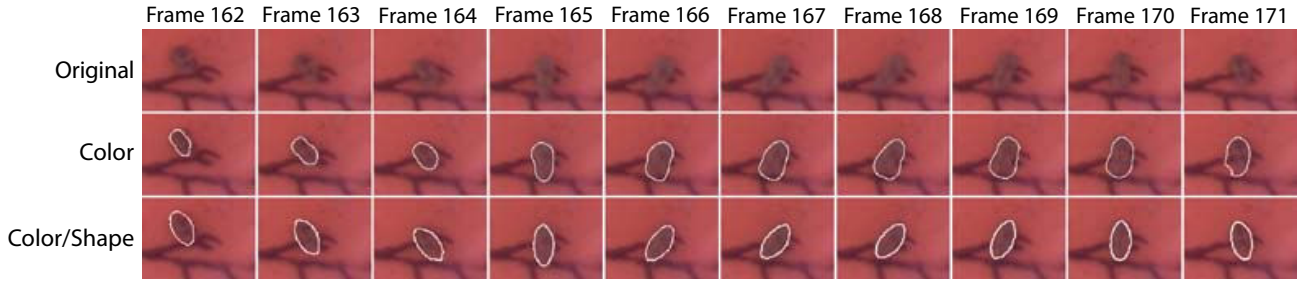


Fig. 7. Tracking using color information and color/shape information, for different frame sequences.

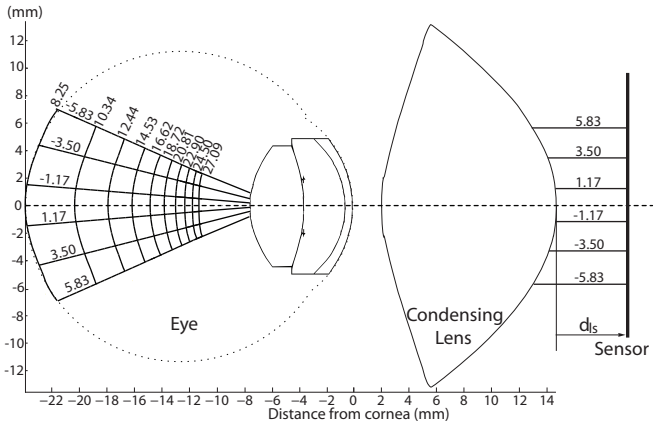


Fig. 8. Simulation for the localization algorithm of [6]. The different surfaces correspond to common sensor position d_{ls} in mm, for uniform sensor steps of ~ 2.1 mm. The curves correspond to common pixel distances from the optical axis for uniform steps of ~ 2.3 mm.

V. Conclusions

In this paper we presented a complete approach to tracking intraocular devices. Starting from colorspace evaluation, we selected the colorspace and channels that carry the most information. To decrease the chances of erroneous tracking, we introduced thresholds to maximize the object-from-background separation. We extended an available real-time level-set tracking algorithm to handle the thresholds and shape information. The colorspace and threshold selection, and the shape-prior extraction can be completed off-line. We showed the effectiveness of our approach by tracking a biomedical microrobot in silicone oil and in a model eye. For realism, we used a custom-made ophthalmoscopic system, available ophthalmic lenses, and an illumination apparatus based on the transscleral illumination principles.

This tracking system will be incorporated into a system to localize biomedical microdevices in the posterior of the human eye. Since it tracks robustly under defocus, it can provide focus information of the object of interest for input to the algorithm of [6]. The position of the in-focus sensor and the on-image position of the device will give the 6-DOF pose of the tracked microdevice (see Fig. 8).

References

[1] K. B. Yesin, K. Vollmers, and B. J. Nelson, "Modeling and control of untethered biomicrorobots in a fluidic environment using electromagnetic fields," *Int. J. Robotics Research*, vol. 25, no. 5-6, pp. 527–536, 2006.

[2] B. Mitchell, J. Koo, I. Iordachita, P. Kazanzides, A. Kapoor, J. Handa, G. Hager, and R. Taylor, "Development and application of a new steady-hand manipulator for retinal surgery," *IEEE Int. Conf. Robotics and Automation*, pp. 623–629, 2007.

[3] W. Wei, R. Goldman, N. Simaan, H. Fine, and S. Chang, "Design and theoretical evaluation of micro-surgical manipulators for orbital manipulation and intraocular dexterity," *IEEE Int. Conf. Robotics and Automation*, pp. 3389–3395, 2007.

[4] J. J. Abbott, O. Ergeneman, M. P. Kummer, A. M. Hirt, and B. J. Nelson, "Modeling magnetic torque and force for controlled manipulation of soft-magnetic bodies," *IEEE Trans. Robotics*, vol. 23, no. 6, pp. 1247–1252, 2007.

[5] Z. Nagy, O. Ergeneman, J. J. Abbott, M. Hutter, A. M. Hirt, and B. J. Nelson, "Modeling assembled-MEMS microrobots for wireless magnetic control," *IEEE Int. Conf. Robotics and Automation*, pp. 874–879, 2008.

[6] C. Bergeles, K. Shamaei, J. J. Abbott, and B. J. Nelson, "On imaging and localizing untethered intraocular devices with a stationary camera," *IEEE Int. Conf. Biomedical Robotics and Biomechatronics*, pp. 489–494, 2008.

[7] (2008) Model eye (8mm pupil) from Gwb International, Ltd. [Online]. Available: http://www.gwbinternational.com/model_eye.htm

[8] H. Stern and B. Efros, "Adaptive color space switching for tracking under varying illumination," *Image and Vision Computing*, vol. 23, no. 3, pp. 353–364, 2005.

[9] F. Moreno-Noguer, A. Sanfeliu, and D. Samaras, "A target dependent colorspace for robust tracking," *Int. Conf. Pattern Recognition*, vol. 3, pp. 43–46, 2006.

[10] C. Doignon, P. Graebling, and M. de Mathelin, "Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature," *Real-Time Imaging*, vol. 11, no. 5-6, pp. 429–442, 2005.

[11] L. Ascari, U. Bertocchi, C. Laschi, C. Stefanini, A. Starita, and P. Dario, "A segmentation algorithm for a robotic micro-endoscope for exploration of the spinal cord," *IEEE Int. Conf. Robotics and Automation*, pp. 491–496, 2004.

[12] M. Tjoa, S. Krishnan, C. Kugean, P. Wang, and R. Doraiswami, "Segmentation of clinical endoscopic image based on homogeneity and hue," *IEEE Int. Conf. Engineering in Medicine and Biology*, pp. 2665–2668, 2001.

[13] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. J. Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[14] S. Osher and R. P. Fedkiw, "Level set methods," *J. Computational Physics*, 1987.

[15] Y. Shi and W. C. Karl, "Real-time tracking using level sets," *IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 34–41, 2005.

[16] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, 2000.

[17] M. Rousson and N. Paragios, "Shape priors for level set representations," *Europ. Conf. Computer Vision*, vol. 2351, pp. 78–92, 2002.

[18] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, no. 2, pp. 156–158, 2008.

[19] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," *IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000.

[20] (2008) Digital High Mag® lens from Volk Optical Inc. [Online]. Available: http://www.volk.com/main/product_info.php?pName=digital-high-mag&osCsid=cfb7fffe1995c879f63871608e2f6bd4