Normalizing Videos of Anterior Eye Segment Surgeries

Gwénolé Quellec, Katia Charrière, Mathieu Lamard, Béatrice Cochener and Guy Cazuguel

***Abstract*— Anterior eye segment surgeries are usually video-recorded. If we are able to efficiently analyze surgical videos in real-time, new decision support tools will emerge. The main anatomical landmarks in these videos are the pupil boundaries and the limbus, but segmenting them is challenging due to the variety of colors and textures in the pupil, the iris, the sclera and the lids. In this paper, we present a solution to reliably normalize the center and the scale in videos, without explicitly segmenting these landmarks. First, a robust solution to track the pupil center is presented: it uses the fact that the pupil boundaries, the limbus and the sclera / lid interface are concentric. Second, a solution to estimate the zoom level is presented: it relies on the illumination pattern reflected on the cornea. The proposed solution was assessed in a dataset of 186 real-live cataract surgery videos. The distance between the true and estimated pupil centers was equal to $8.0 \pm 6.9\%$ of the limbus radius. The correlation between the estimated zoom level and the true limbus size in images was high: $R = 0.834$.**

I. Introduction

In anterior eye segment surgeries, the surgeon wears a binocular microscope and the output of the microscope is video-recorded. Real-time analysis of these videos may be useful to automatically communicate information to the surgeon in due time (e.g. recommendations). However, several low-level issues make the development of such high-level decision support algorithms challenging. First, the eye is continually moving and, quite often, the pupil and the iris are only partially visible. Second, surgeons may change the zoom level multiple times during the surgery. In order to facilitate the development of the envisioned decision support algorithms, surgical videos need to be normalized, just like fundus photographs had to be normalized to allow efficient computer-aided diagnosis of the retina in the last decade [1].

In this paper, we focus on the most common anterior eye segment surgery: cataract surgery [2]. An algorithm was proposed by Lalys et al. to segment cataract surgery steps in videos [3]. In that algorithm, the pupil is segmented and visual features are extracted inside the pupil only [3]. However, a lot of relevant motion information appears outside the pupil. In particular, surgical tools usually enter the eye around the limbus, i.e. outside the pupil. Besides, even the pupil is hard to segment in many videos, due to occlusions, specular reflections, etc. Recently, we proposed an algorithm to categorize surgical steps in real-time: it was applied to epiretinal membrane surgery and cataract surgery [4], [5]. In that algorithm, motion information was processed regardless of the enumerated problems (eye motion, zoom level variations), which certainly limited performance. An adaptation to normalized cataract surgery videos is presented in a companion paper submitted to this conference [6].

All authors are with Inserm, UMR 1101, SFR ScInBioS, Brest, F-29200 France gwenole.quellec@inserm.fr

K. Charrière and G. Cazuguel are with INSTITUT TELECOM; TELECOM Bretagne; UEB; Dpt ITI, Brest, F-29200 France

M. Lamard and B. Cochener are with Univ Bretagne Occidentale, Brest, F-29200 France

B. Cochener is with CHRU Brest, Service d'Ophtalmologie, Brest, F-29200 France

II. Pupil Center Tracking

Detecting the pupil boundaries or the limbus in surgical videos is challenging, due to the variety of colors and textures in the pupil, the iris, the sclera and the lids (see Fig. 1). The variety of zoom factors makes it even more challenging. In particular, it is very difficult to differentiate the pupil boundaries, the limbus and the sclera / lid interface. However, all these structures are concentric. So, the pupil center, which is also the center of the limbus and the center of the lid / sclera interface, can be detected quite robustly.

A. Preprocessing

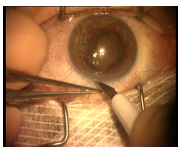
First, images are downsampled by a factor of two, to speed up computations and to get rid of interlacing artifacts. Then, the downsampled image is converted to a gray scale image I_t, where t denotes the time index. The proposed detection algorithm assumes that the pupil is darker than the sclera (see section II-C): this is true in the green channel (G_t) and the blue channel (B_t), but not always in the red channel (R_t). Therefore, the color channels were combined as follows: $I_t = 0.299B_t + 0.587G_t + 0.114R_t$, where the weights of B_t and R_t are switched compared to the commonly used luminance channel.

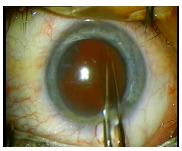
Specular reflections on the cornea, blood vessels, eye lashes and iris features produce edge information that may mislead the pupil center detection. A binary specular reflection mask $I_t^{(S)}$ is computed: $I_t^{(S)}(x,y)$ is true iff $I_t(x,y) \geq I_{min}$ ($I_{min} = 240$). Regions identified by $I_t^{(S)}$ were removed by inpainting [7]. As for blood vessels, eye lashes and iris features, they were removed using a median filter (radius: 5 pixels). Let $I_t^{(P)}$ denote the preprocessed image.

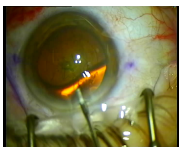
B. Accumulating Edge Information

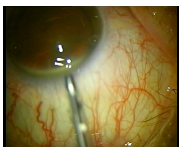
The proposed solution to detect the pupil center in $I_t^{(P)}$ relies on the Hough transform [8]. The Hough transform is not used to detect circles (a 3-D problem), but rather to detect circle centers (a 2-D problem), so a 2-D accumulator A_t is used. Besides complexity, the advantage of detecting circle centers directly is that the pupil boundaries, the limbus and the lids have approximately the same center, so their edge information will accumulate in the same region of A_t.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

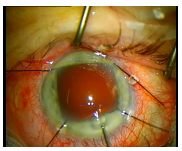

(a) incision

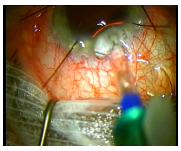

(b) rhexis

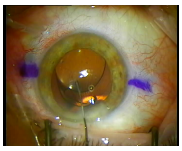

(c) hydrodissection

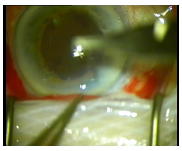

(d) phacoemulsification


(e) epinucleus removal


(f) viscous agent injection


(g) implant setting-up


(h) viscous agent removal


(i) stitching up


(j) miscellaneous

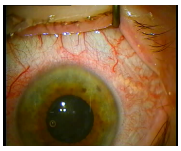

(k) idle

Fig. 1: Cataract surgery steps.

The usual Hough transform algorithm [8] is summarized hereafter and the proposed modifications are described in sections II-C and II-D. Edges are detected in $I_t^{(P)}$ using a Canny edge detector; a pixel with a Canny response greater than a given threshold σ (σ = 37.5) is referred to as an edge pixel. Edge information in each edge pixel $\mathbf{p}$ is then accumulated in the direction of the intensity gradient $\mathbf{d_p}$ at pixel $\mathbf{p}$. $\mathbf{d_p}$ is obtained by convolving $I_t^{(P)}$ with two directional Sobel filters, S_x and S_y, in the neighborhood of $\mathbf{p}$: $\mathbf{d_p} \propto ((I_t^{(P)} * S_x)[\mathbf{p}], (I_t^{(P)} * S_y)[\mathbf{p}]))$, $\|\mathbf{d_p}\| = 1$. Let R_{min} denote the minimum pupil radius (R_{min} = 30). Every pixel $\mathbf{u}$ of A_t such that $\mathbf{u} = \mathbf{p} + k\mathbf{d_p}$ and $|k| > R_{min}$ is incremented by 1.

C. Oriented Accumulation

In the absence of tools, the global maximum of A_t should be the center of the pupil (and of the limbus and of the lid / sclera interface). But in the presence of tools, which add strong edge information to the image, it is not always the case, in particular if the pupil and the iris are largely occluded. To overcome this problem, edge information is accumulated on one side of the edge pixel $\mathbf{p}$ only: on the darkest side (see section II-A). Let $\lambda^+(\mathbf{p})$ and $\lambda^-(\mathbf{p})$ denote the average intensity on both sides of $\mathbf{p}$, and let $\delta(\mathbf{p})$ denote the regional contrast:

$$\lambda^+(\mathbf{p}) = \frac{1}{R_{min}} \sum_{k=1}^{R_{min}} I_t^{(P)}[\mathbf{p} + k\mathbf{d_p}] \quad (1)$$

$$\lambda^-(\mathbf{p}) = \frac{1}{R_{min}} \sum_{k=-R_{min}}^{-1} I_t^{(P)}[\mathbf{p} + k\mathbf{d_p}] \quad (2)$$

$$\delta(\mathbf{p}) = |\lambda^+(\mathbf{p}) - \lambda^-(\mathbf{p})| \quad (3)$$

If $\lambda^+(\mathbf{p}) < \lambda^-(\mathbf{p})$ (resp. $\lambda^+(\mathbf{p}) \geq \lambda^-(\mathbf{p})$), then every pixel $\mathbf{u}$ of A_t such that $\mathbf{u} = \mathbf{p} + k\mathbf{d_p}$ and $k > R_{min}$ (resp. $k < -R_{min}$) is incremented by $\delta(\mathbf{p})$. The regional contrast is low on tool edges, but high on the limbus and even higher on the pupil boundaries. Information from the lid boundaries does not strenghten the pupil detection in this solution. But since the lids are concave objects, edge information from the lids accumulate in every direction, and not in one spot. So it does not create artifacts.

D. Spatiotemporal Accumulation Matrix Processing

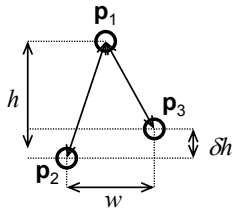
First, because the pupil boundaries and the limbus are not exactly circular and not exactly concentric, matrix A_t is smoothed using a median filter (radius: 8 pixels). Let $\widehat{A}_t$ denote the smoothed accumulation matrix. To increase the robustness of the pupil center detection further, we use the fact that the pupil center moves continuously over time. So the location of the pupil center at time t is not defined as the pixel location maximizing $\widehat{A}_t$, but rather as the pixel location maximizing $\tilde{A}_t$:

$$\tilde{A}_t = \begin{cases} \widehat{A}_t & \text{if } t = 0 \\ \alpha_c \tilde{A}_{t-1} + (1 - \alpha_c)\widehat{A}_t & \text{if } t > 0 \end{cases} \quad (4)$$

given a discount factor $0 \leq \alpha_c \leq 1$ ($\alpha_c = 0.8$).

III. Zoom Level Tracking

The limbus diameter is the best indicator of the zoom level in images: 1) it does not change over time, unlike the pupil diameter, and 2) it is little variant across the population [9]. So a straighforward solution to estimate the zoom level is to segment the limbus in images, as summarized in section III-A. However, it is hard to differentiate the pupil boundaries, the limbus and the sclera / lid interface, due to the variety of colors and textures in the pupil, the iris, the sclera and the lids (see Fig. 1). So a second solution is presented in section III-B. The second solution takes advantage of the fact that the radius of corneal curvature is also little variant across the population [9]. We propose to measure the illumination pattern reflected on the cornea: the three glints that are well visible in Fig. 1 (c), (f) or (k). This reflected illumination pattern only depends on the cornea shape and the distance between the lights and the cornea. Both parameters are approximately constant across surgeries: the size of the illumination pattern is mostly controlled (linearly) by the zoom factor.



$$\frac{\delta h}{h} \leq \frac{1}{4}$$
$$\frac{1}{5} \leq \frac{w}{h} \leq \frac{3}{2}$$
$$5 \leq h \leq 16$$

Fig. 2: Detecting corneal reflections. $\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$ denote the center of mass of three connected components in the specular reflection mask. h denotes the y-component of $\frac{\mathbf{p}_2+\mathbf{p}_3}{2} - \mathbf{p}_1$.

A. Baseline Solution — Limbus Detection

A graph-based region growing solution is proposed to jointly detect the pupil boundaries and the limbus. Each pixel in image $I_t^{(P)}$ is associated with one node in graph $\mathcal{G}$. Nodes associated with adjacent pixels in $I_t^{(P)}$, in the sense of 8-connexity, are connected by one weighted edge in $\mathcal{G}$: the edge is weighted by the intensity difference in $I_t^{(P)}$. The shortest path between the node associated with the pupil center and every other node in $\mathcal{G}$ is computed using the Dijkstra algorithm [10]. Then, the shortest distances are sorted in increasing order: the first jump d_p in this distance function should correspond to the pupil boundaries, the second jump d_l should correspond to the limbus. Finally, an ellipse is fitted to the boundaries of the region associated with distances less than d_l (supposedly the pupil + the iris), excluding the image boundaries if need be. The instant zoom factor z_t is given by $z_t = \frac{2\sqrt{ab}}{\overline{z}}$, where a and b are the major and minor axis of the fitted ellipse and $\overline{z}$ is a typical value for the numerator.

B. Proposed Solution — Corneal Reflection Detection

All groups of three connected components in the specular reflection mask $I_t^{(S)}$, whose configuration match the rules enumerated in Fig. 2, are selected. If at least one group is selected, the one closest to the pupil center is retained: the others are usually secondary reflections on the tear film, at the cornea / lid interface. If none is detected, it can be that the three glints are merged, because the image is out of focus. Therefore, while no suitable group of glints is detected, $I_t^{(S)}$ is eroded with a circular kernel of radius r: $r = 1$ initially and increases by 1 at each iteration. If $r > r_{max}$ ($r_{max} = 4$), then the algorithm stops: the detection failed. The instant zoom factor z_t is given by $z_t = \frac{\frac{1}{2}[\|\mathbf{p}_2-\mathbf{p}_1\|+\|\mathbf{p}_3-\mathbf{p}_1\|]}{\overline{z}}$ (see Fig. 2), where $\overline{z}$ is a typical value for the numerator.

C. Zoom Level Tracking

Whatever solution is used, instant zoom estimations z_t are noisy and not always available. The following tracking solution is used:

$$\tilde{z}_t = \begin{cases} 1 & \text{if } t = 0,\ z_t \text{ unavailable} \\ z_t & \text{if } t = 0,\ z_t \text{ available} \\ \tilde{z}_{t-1} & \text{if } t > 0,\ z_t \text{ unavailable} \\ \alpha_z \tilde{z}_{t-1} + (1-\alpha_z) z_t & \text{if } t > 0,\ z_t \text{ available} \end{cases} \quad (5)$$

given a discount factor $0 \leq \alpha_z \leq 1$ ($\alpha_z = 0.9$).

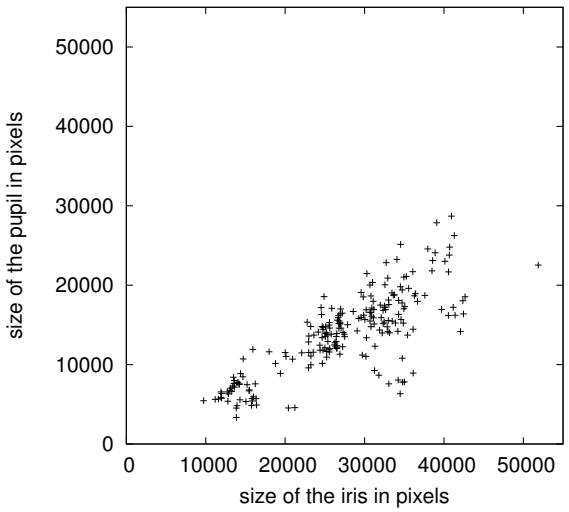


Fig. 3: Distribution of pupil sizes and "pupil + iris" sizes in the database.

IV. Cataract Surgery Dataset

A dataset of 186 videos from 186 consecutive cataract surgeries was collected at Brest University Hospital (Brest, France) between February and July 2011 [5]. Image definition is 720 x 576 pixels and the frame frequency is 25 FPS. A cataract expert manually indicated the beginning and the end of each surgical step in each video. The following steps were manually delimited in all videos: incision, rhexis, hydrodissection, phacoemulsification, epinucleus removal, viscous agent injection, implant setting-up, viscous agent removal and stitching up. Miscellaneous steps (iris retractor setting-up, iris retractor removal, angle measurement, landmark tracing, etc.) were also segmented in some videos. The remaining portions of the videos, where no surgical tools are visible, were assigned to the 'idle' category. 36 randomly selected surgeries were used for training. For evaluation purposes, 220 manually delimited videos were selected at random among the remaining 150 surgeries: 20 videos for each categories (the 9 usual surgical steps, 'miscellaneous steps', 'idle'). The 100$^{\text{th}}$ frame of each video was then segmented: the limbus and the pupil boundaries were manually delineated and an ellipse was fit to each segmented boundary. Examples of images that were manually segmented are given in Fig. 1. The distribution of pupil sizes and "pupil + iris" sizes, among the 220 manually segmented images, is reported in Fig. 3.

V. Results

The algorithm parameters (filter sizes, thresholds and discount factors) were adapted on the training set. The

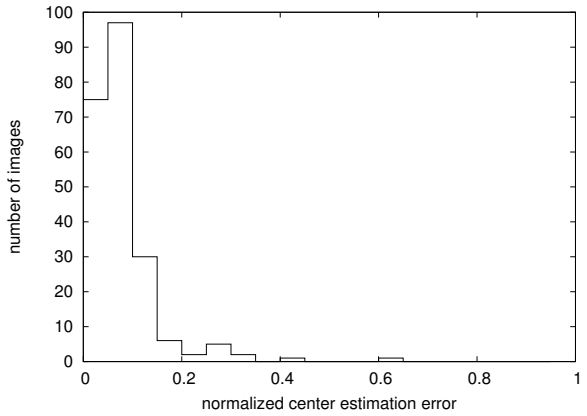


Fig. 4: Normalized pupil estimation error with tracking.

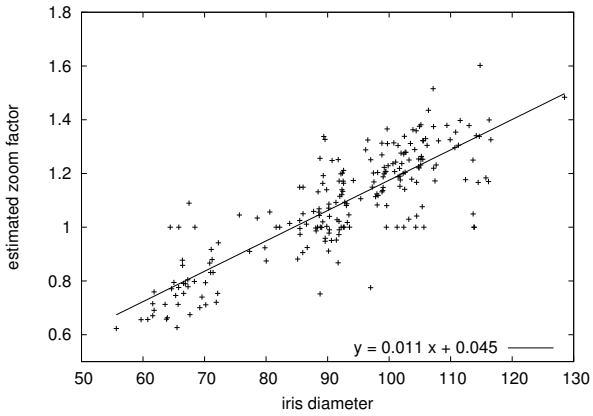


Fig. 5: Zoom level estimation using corneal reflections

results obtained on the 220 test videos are reported below. When tracking was used, the first 100 frames of each video were processed and the 100$^{\text{th}}$ was evaluated against the manual segmentation. Otherwise, only the 100$^{\text{th}}$ frame was processed.

Each manual limbus segmentation was fitted by an ellipse of center $\mathbf{c}$, of major axis a and minor axis b, and rotated by an angle θ (see section IV). Let $\mathbf{q} = (x, y)$ denote the projection $\mathbf{q} = R_\theta(\mathbf{p} - \mathbf{c})$ of the estimated pupil center $\mathbf{e}$ in the ellipse reference system, where R_θ is the rotation matrix of angle θ. We define the normalized pupil estimation error as $\sqrt{(\frac{x}{a})^2 + (\frac{y}{b})^2}$. A normalized pupil estimation error of zero indicates that $\mathbf{e}$ is at the center of the ellipse. A normalized pupil estimation error of 1 (resp. greater than 1) indicates that $\mathbf{e}$ is on the boundaries of (resp. outside) the ellipse. A normalized pupil estimation error of 0.080 ± 0.069 and 0.086 ± 0.077 was achieved with or without tracking, respectively; a histogram is reported in Fig. 4. The error was always less than 1. The normalized pupil estimation error was little correlated (R = 0.291) with the proportion of the "pupil + iris" region lying outside the camera's field of view, which was estimated using the fitted limbus segmentation. Note that the pupil and the iris were completely occluded in one image and the algorithm successfully found a very low maximum value in the accumulator matrix $\tilde{A}_{100}$ (significantly lower than in the other 219 videos).

The limbus diameter D_l was estimated by $D_l = 2\sqrt{ab}$, where a and b are the major and minor axis of the fitted ellipse. Using corneal reflections, the correlation between the estimated zoom level and D_l was 0.834 with tracking (see Fig. 5) and 0.739 without tracking. As a comparison, the correlation between D_l and the pupil diameter is 0.812. Using the automated limbus segmentation for zoom level estimation is much poorer: the correlation between the estimated limbus diameter and D_l was only 0.314. In fact, using the pupil segmentation was slightly better: the correlation between the estimated pupil size and D_l was 0.402 (the correlation between the estimated and the true pupil diameter was 0.578).

The average processing time was 31 ms per image (32.3 images per second): 20 ms were dedicated to preprocessing (see section II-A), 13 ms were dedicated to pupil center estimation in the preprocessed image and less than 1 ms was dedicated to zoom level estimation using corneal reflections. The algorithms were implemented in C++, using the OpenCV library[1].

VI. Conclusions

A robust solution to normalize anterior eye segment surgery videos in real-time, by tracking the pupil center and estimating the zoom level, was presented. Its performance in a real-live cataract surgery video dataset was quite good. This work will facilitate the development of real-time decision support tools for eye surgery [6].

References

[1] M. Foracchia, E. Grisan, and A. Ruggeri., "Luminosity and contrast normalization in retinal images." *Med Image Anal*, vol. 9, no. 3, pp. 179–190, 2005.

[2] X. Castells, M. Comas, M. Castilla, F. Cots, and S. Alarcón, "Clinical outcomes and costs of cataract surgery performed by planned ECCE and phacoemulsification," *Int Ophthalmol*, vol. 22, no. 6, pp. 363–367, 1998.

[3] F. Lalys, L. Riffaud, D. Bouget, and P. Jannin, "A framework for the recognition of high-level surgical tasks from video images for cataract surgeries," *IEEE Trans Biomed Eng*, vol. 59, no. 4, pp. 966–76, 2012.

[4] G. Quellec, M. Lamard, G. Cazuguel, Z. Droueche, C. Roux, and B. Cochener, "Real-time retrieval of similar videos with application to computer-aided retinal surgery," in *Proc IEEE EMBS*, 2011, pp. 4465–4468.

[5] G. Quellec, K. Charrière, M. Lamard, Z. Droueche, C. Roux, B. Cochener, and G. Cazuguel, "Real-time recognition of surgical tasks in eye surgery videos," *Med Image Anal*, 2014, in press.

[6] K. Charrière, G. Quellec, M. Lamard, G. Coatrieux, B. Cochener, and G. Cazuguel, "Automated surgical step recognition in normalized cataract surgery videos," in *Proc IEEE EMBS*, 2014.

[7] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proc CVPR'01*, vol. 1, 2001, pp. 355–62.

[8] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Patt Recog*, vol. 13, no. 2, pp. 111–122, 1981.

[9] R. C. Augusteyn, D. Nankivil, A. Mohamed, B. Maceo, F. Pierre, and J. M. Parel, "Human ocular biometry," *Exp Eye Res*, vol. 102, pp. 70–5, 2012.

[10] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.

[1] http://opencv.org/