35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Super-Image Mosaic of Infant Retinal Fundus: Selection and Registration of the Best-quality Frames from Videos

Enea Poletti, Giulio Benedetti, and Alfredo Ruggeri, *Senior Member, IEEE*

Abstract— Wide-field retinal fundus cameras are commercially available devices that allow acquiring videos of a wide area of infants' eye, considered of clinical interest in screening for ROP (Retinopathy of Prematurity). Many frames of the video are often altered by defects such as artifacts, interlacing and defocus, which make critical and time consuming the search and choice of the good frames to be analyzed. We developed a computerized system that automatically selects the best still frames from the video and builds a mosaic from these images. It will allow clinicians to examine a single large, best quality image. The best frames are identified using several image quality parameters that measure *sharpness* and *steadiness,* and then registered to obtain a single mosaic image. A custom blending procedure is then applied in order to provide a final image with homogeneous luminosity and contrast, devoid of the dark areas typically present in the outer regions of single frames. The best-frame selection module showed a PPV of 0.92, while the visual inspection of resulting mosaics confirmed the remarkable capability of the proposed system to provide higher quality images.

I. Introduction

Retinopathy of Prematurity (ROP) [1] is an eye disease that affects prematurely born infants. It can be mild and resolve spontaneously, but in more serious cases it becomes very aggressive: new blood vessel formation progresses to scarring, retinal detachment and possibly blindness. ROP is categorized by zone, stage, and presence of plus disease, and its severity is characterized by different signs: arterial tortuosity and venous dilation at the posterior pole, vitreous haze, and iris rigidity.

A. The Need

Wide-field retinal cameras (130° of field of view, e.g., RetCam by Clarity Medical Systems, Pleasanton, CA, USA) are recent commercially available devices that allow inspecting the most peripheral area of the eye, where vessels grow during the last weeks of gestation. Several studies have assessed their clinical value in screening for ROP [2]. The main differences of RetCam images (Fig. 1) with respect to images provided by standard adults fundus cameras are: 1) low contrast, 2) presence of interlacing artifacts, as images are actually single frames extracted from a video, 3) narrow blood vessels, due to the wide-field of view coupled with the 640x480 pixel resolution, 4) non uniform illumination in the captured wide field of view, 5) high visibility of choroidal vessels, related to the lack of pigmentation of the infant choroid [3, 4].

All these aspects require the clinicians to spend a significant amount of time in viewing and selecting from the video the "best" frame, which will then subject to their clinical evaluation, all of this being a critical and subjective step. If the selected frame does not exhibit a sufficient level of quality or does not contain sufficient information, the analysis of the image becomes quite challenging (e.g., as regards vessel tracing [3, 5]) or not fully reliable (e.g., the extraction of parameters of clinical interest [6]), especially if it is performed by a computerized system.

Because of the aforementioned problems, a method to identify high quality frames and discard the low quality ones is needed. For our aims, the *quality* of a frame is expressed by the capability of recognizing the structures that compose the retina, mainly vessels.

Image quality is hampered by artifacts, defocus, movement, etc. Artifacts can be introduced by poor lens-eye contact, presence of bubbles in the contact gel, light reflexes, poor dilation of the pupil, optical misalignment (Fig. 1) [4]. All these defects have the ultimate effect of decreasing the *sharpness* of the image and hence the recognizability of the vessels.

The speed of the relative motion between camera and eye directly affects the amount of interlacing in the frame, hampering its *steadiness.* There are plenty of de-interlacing algorithms in literature, each producing different problems or artifacts on its own [7]. In general, these methods combine the even and odd fields to provide better looking frames, but the resulting quality is anyhow worse than an ideal acquisition without motion.

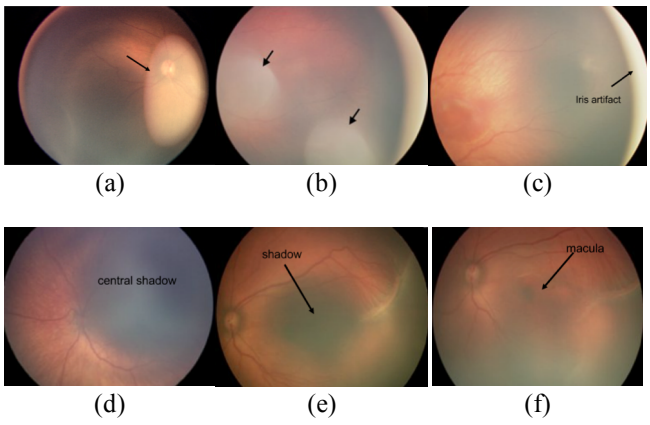


Fig 1. Different examples of artifacts that may appear in a ROP video. (a) Artifact due to poor contact with gel. (b) Artifacts due to bubbles within the coupling gel. (c) Iris artifact. (d, e) Central shadows due to insufficient pupil dilation, which (e) may be moved, e.g., inferiorly to reveal the macula, by tilting the probe.

All authors are with the Department of Information Engineering, University of Padova, Padova, 35131 Italy (phone: (+39)049-827-7758; fax: (+39)049-827-7699; e-mail: enea.poletti@dei.unipd.it).

(a) (b) (c)

Fig 2. Detail of a vessel in three different quality state. (a) Interlaced frame. (b) Unfocused frame. (c) Focused and not interlaced frame. With relation to Eq 1 and Eq 2, (a) will have the lowest st and (b) will have the lowest sh among the three examples.

B. Our Proposal

In this work we present a computerized system that automatically selects the best still frames from a RetCam video and builds with them a comprehensive mosaic image. This will both allow clinicians to examine a single large, best quality image and also make possible the processing with a computerized system for retinal analysis.

The selected frames are registered by applying rotation and translation movements to achieve the best overlapping in all the areas of intersection, until a single mosaic image is obtained. A custom blending procedure will provide a final image with homogeneous luminosity and contrast, devoid of the dark areas typically present in the outer regions of single frames.

II. Materials

Eighteen videos of retinal fundus were acquired in premature infants with the RetCam fundus camera (Clarity Medical Systems Inc., CA, USA) with a 130° field of view and 640×480 pixels frame size. Videos are composed of a number of frames that ranges from 850 to 2200 (1800 on average).

Three of the 18 videos were randomly selected for the composition of the ground truth: all the frames of these videos were manually labeled as either "*high quality"* or "*normal-to-poor quality"*, for a total of 5523 labeled frames. This ground truth dataset has been used to validate the best-frame selection stage.

III. Methods

The super-image mosaicking system is organized as follows.

At first, each frame of the video is analyzed in order to identify the frames with the best indexes of *sharpness* (recognizability of the vessels) (Sec. III-A-1) and *steadiness* (absence of motion artifacts) (Sec. III-A-2). An *adaptive combined threshold* approach assures that a fixed number of frames is always provided, regardless of the overall quality of the video (Sec. III-A-3).

Once the best quality frames have been identified, registration is accomplished sequentially between pairs of frames (Sec. III-B-1). In order to optimally merge the content from the regions of overlap between frames, a custom weighting function has been devised (Sec. III-B-2).

A. Selection of the Best Frames

We propose two parameters to describe the overall *quality* level of a frame: *sharpness* and *steadiness*. While *sharpness* is hampered by defocus and local artifacts present in the image, *steadiness* measures the quality loss due to the relative motion of the eye with respect to the camera.

We therefore developed a *sharpness* detector and a *steadiness* detector. Each provides its own quality index, which is computed for each frame of a video. As it will be shown in Sec. IV, *Results,* both detectors should be used, since *sharpness* and *steadiness*, as defined in this work, are independent concept.

1) *Sharpness Detection*

The *sharpness* index sh of a frame f is mathematically defined as the average energy of the maximum response over different scales of the convolutions between the subimage f_{roi} and the multi-scale filter LoG. $f_{roi} \in f$ is a circular ROI that excludes the black pixels, while LoG is a bank of 2nd order Laplacian of Gaussian filters with different scales, so that

$$sh(f) = mean(\, max\{f_{roi} \otimes LoG_s \mid s \in scales\}\,) \quad (1)$$

where $\otimes$ is the operation of convolution. Gaussian variance and filter scales have been properly sized to fit vessels' shape ($\sigma = \{1.5, 2, 2.5\}$ and $scales = \{3, 5, 7\}, in\ pixels$).

While filters commonly used to detect contrast are square windows, we need to use here a different approach since the interlacing artifact, which may affect the frame under analysis, can modify the index by increasing the vertical contrast. In order to avoid this possible dependency, LoG is composed only of horizontal filters.

2) *Steadiness Detection*

The *steadiness* index st of a frame f has been expressed as the ratio of the mean correlation between adjacent horizontal rows and the mean correlation between odd (or even) horizontal rows.

Since the interlacing artifact displaces the even rows with respect to the odd rows, correlation between odd-only rows (or even-only rows) is affected only by image-related features and hence not influenced by motion issues. This value is then assumed as the *control*. Correlation between two adjacent rows (one even and one odd) is influenced by both image-related feature and motion artifact, and it is considered to be the actual *measure*. The index of steadiness is defined as:

$$st(f) = \frac{measure}{control} = \frac{mean\ (corr\{\,pairs\ of\ adjacent\ rows \in f\})}{mean\ (corr\{\,pairs\ of\ odd\ rows \in f\})} \quad (2)$$

It is intuitive that a frame without interlacing artifact will have a *steadiness* very close to 1, while the larger the artifact, the smaller the value of $st(f)$.

3) *Adaptive Threshold*

RetCam videos exhibit a wide range of overall quality (i.e., a single video could be entirely composed only of high quality or only of low quality frames). In order to always identify a given number N of best-frames (with N chosen by the user) the algorithm has been provided with *adaptive* selection thresholds. Let th_{sh} and th_{st} respectively be the *sharpness* and *steadiness* thresholds, and B_{sh} and B_{st} the two sets of frames with best *sharpness* and best *steadiness,* i.e. $B_{sh} = \{f \mid sh(f) > th_{sh}\}$ and $B_{st} = \{f \mid st(f) > th_{st}\}$.

Given N, the user defined number of frames, the two thresholds th_{sh} and th_{st} are determined so that:

$$th_{sh}, th_{st} = argmin_{th_{sh},th_{st}} \ |\#(B_{sh} \cap B_{st}) - N| \qquad (3)$$

The two thresholds are computed by means of a naïve gradient search algorithm.

B. Mosaicking

1) *Frame Registration*

Registration in ROP images generally involves relatively large translations and small rotations (due to tilting of the head and ocular torsion), but, as opposed to adult fundus acquisition, the need for scaling is negligible.

A broad range of image registration methods have been proposed for retinal fundus [8]. The best results in our set of images were obtained with an extension of the phase correlation method, previously proposed in [9], which uses the Fourier Transforms of the two images to compute the translation and rotation to be applied. The method is characterized by an outstanding robustness against noise and disturbances, such as those related to non-uniform illumination.

2) *Overlap Region Weighting*

Let I_1 and I_2 be two correctly registered frames. The combination of the two frames will compose the so-called "mosaic image", I_{mos}, whose pixels' values are provided by the following expression:

$$I_{mos}(x,y) = \begin{cases} f(I_1, I_2, x, y) & \forall (x,y) \in I_{over} = \{I_1 \cap I_2\} \\ I_1(x,y) & \forall (x,y) \in I_1 \backslash I_{over} \\ I_2(x,y) & \forall (x,y) \in I_2 \backslash I_{\text{over}} \end{cases} \qquad (4)$$

Since pixels that belong to the overlap region I_{over} are likely to have different values in I_1 and I_2, a function f to determine these value of $I_{mos}(x,y)$ is needed.

A naïve method to assign pixels' values in the overlap region would be computing the average of the original pixels' values:

$$f(I_1, I_2, x, y) = \frac{I_1(x,y) + I_2(x,y)}{2} \qquad (5)$$

The result of such an approach is shown in Fig. 4-a. Unfortunately, the peculiar features of retinal fundus images make simple solution like Eq. 5 unsuitable. This is due to the high non-uniformity in luminosity and contrast between different frames (even if they are from the same region) and to the distinctive darkening that each frame exhibits in its outer regions due to limitations of optical design. In order to overcome this issue, we devised a suitable weighting function:

$$f(I_1, I_2, x, y) = \ w_1(x,y) I_1(x,y) + \ w_2(x,y) I_2(x,y) \qquad (6)$$

where

$$w_1(x,y) = \frac{d_1^n(x,y)}{d_1^n(x,y) + d_2^n(x,y)} \qquad (7)$$

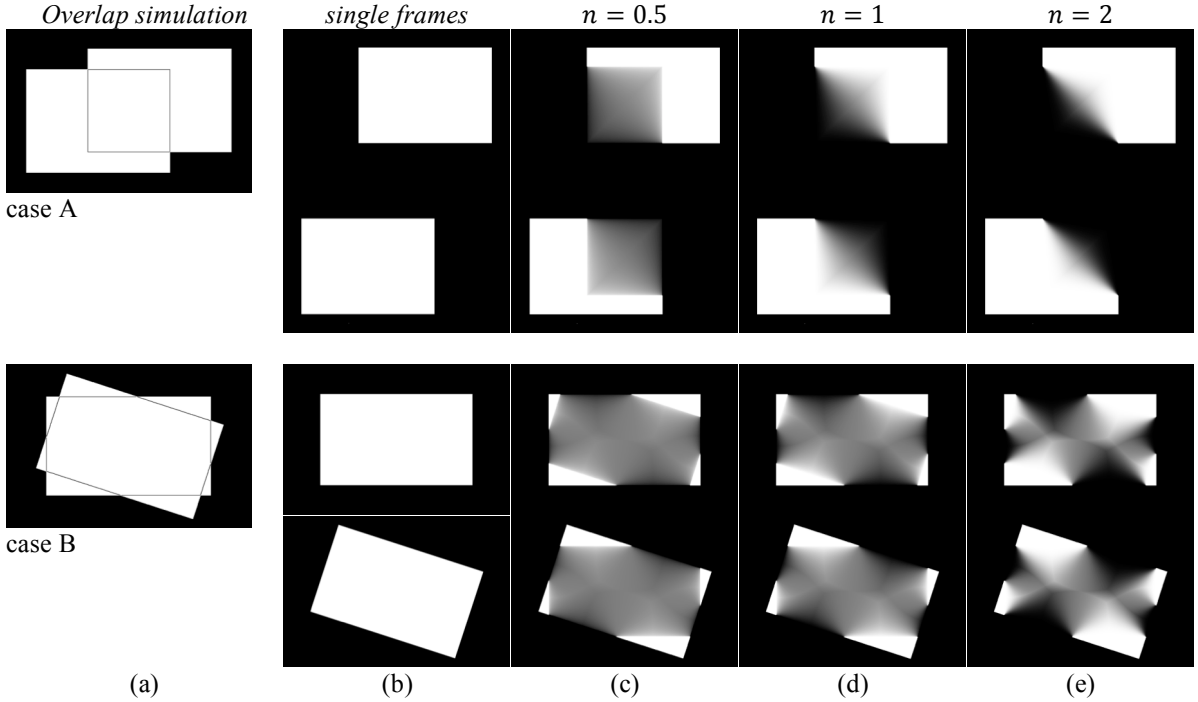
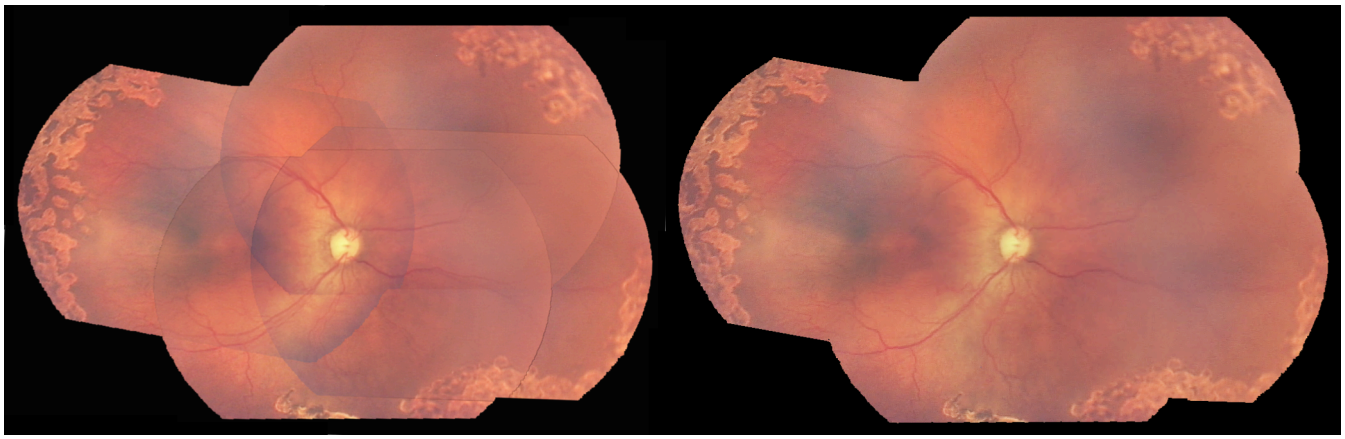


Fig 3. (a) Two simulated cases (A and B) of overlapping frames; (b-e) first and second rows are related to frame I_1 and I_2 of case A, while third and fourth rows are related to frame I_1 and I_2 of case B. (b) the single frames of the pair to be composed. Values of w_1 and w_2 (see Eq. 6) with (c) $n = 0.5$, (d) $n = 1$, and (e) $n = 2$.



(a) (b)

Fig 4. Example of mosaicking with 4 frames. (a) Result obtained using average as weighting function (Eq 5). (b) Result obtained by using our proposed weighting function (Eq 6-8) with $n = 2$.

$$w_2(x,y) = \frac{d_2^n(x,y)}{d_1^n(x,y) + d_2^n(x,y)} \quad (8)$$

so that $w_1(x,y) = 1 - w_2(x,y)$ is always satisfied. The function $d_k(x,y)$ is defined as the minimum Euclidian distance between the point (x,y) and the region $I_k \backslash I_{over}$:

$$d_k(x,y) = min\{\, distance(\,(x,y),\ I_k \backslash I_{over})\,\} \quad (9)$$

The value of the power n in Eq. 7 and 8 has been empirically chosen with the aim of obtaining the best visual appearance in the final mosaic (see Fig. 4). It is worth noting that if $n = 0$ then f becomes the naïve average function in Eq. 5.

IV. Results

5523 frames from the 3 videos of ground truth set were used to validate the best-frame selection stage by computing its PPV (Positive Predicted Value), which is the fraction of *actual* high quality frames over the number of *estimated* high quality frames.

Different PPV have been obtained with different value of N, the user-defined number of wanted best frame extracted. With $N = 10$, the *sharpness detector* alone shows a $PPV_{N=10}^{sh} = 0.81$, the *steadiness detector* alone shows a $PPV_{N=10}^{st} = 0.76$, while the PPV of their effective coupling (Eq. 3) is 0.92. With $N = 20$, $PPV_{N=20}^{sh} = 0.74$, $PPV_{N=20}^{st} = 0.69$, and $PPV_{N=20} = 0.88$.

As far as registration and frame blending are concerned, the results were assessed by visual inspection. The final mosaics of the 18 video analyzed confirmed the remarkable capability of the proposed system to provide high quality images (see example in Fig. 4 a-b).

V. Conclusion

We presented a computerized system that automatically selects a user-defined number of best still frames from a RetCam video by effectively coupling sharpness and steadiness detectors. High quality frames were then registered into a single mosaic image: merging of the overlapping areas was carried out with a custom weighting method, designed to remove the peripheral quality degradation typical of ROP images.

The resulting mosaic will allow clinicians to examine a single wide, best quality image that contains all the relevant information from the central and peripheral regions of the retina. For this reason, we also expect a much better performance when a computerized system for the analysis of retinal fundus in ROP is applied to these mosaic images.

Acknowledgements

We would like to thank Barry Linder and Leslie MacKeen (Clarity Medical Systems, Pleasanton, CA, USA) for having kindly provided the RetCam videos.

References

[1] The Committee for the Classification of Retinopathy of Prematurity, "An international classification of retinopathy of prematurity," Arch Ophthalmol, Vol. 102(8), pp. 1130–1134, 1984.

[2] Committee for the Classification of Retinopathy of Prematurity, "The International Classification of Retinopathy of Prematurity revisited," Arch Ophthalmol, Vol. 123(7), pp. 991–999, 2005.

[3] E. Poletti et al., "Automatic vessel segmentation in wide-field retina images of infants with Retinopathy of Prematurity," EMBC Annual International Conference of the IEEE, pp. 3954-3957, 2011.

[4] http://www.claritymsi.com/us/retcamtraining.html, Clarity Medical System website, RetCam training module.

[5] E. Poletti, A. Ruggeri, "Segmentation of vessels through supervised classification in wide-field retina images of infants with Retinopathy of Prematurity," CBMS, 25th International Symposium, pp. 1-6, 2012.

[6] E. Poletti, E. Grisan, A. Ruggeri, "Image-level tortuosity estimation in wide-field retinal images from infants with Retinopathy of Prematurity," EMBC Annual International Conference of the IEEE, pp. 4958-4961, 2012.

[7] Y.-R. Chen, S.-C. Tai, "True Motion-Compensated De-Interlacing Algorithm," Circuits and Systems for Video Technology, IEEE Transactions on, Vol.19(10), pp. 1489-1498, Oct. 2009.

[8] L. Xiaoqi et al., "A review of algorithm research progress for non-rigid medical image registration," Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on, pp.3863-3866, 16-18 April 2011

[9] M. De Luca, "New techniques for the processing and analysis of retinal images in diagnostic ophthalmology," PhD Thesis, http://paduaresearch.cab.unipd.it/443/, 2008.