Experts vs. Novices: Applying Eye-tracking Methodologies in Colonoscopy Video Screening for Polyp Search

Jorge Bernal*, F. Javier Sánchez, Fernando Vilariño
Centre de Visio per Computador & Universitat Autonoma de Barcelona
Mirko Arnold, Anarta Ghosh, Gerard Lacey
Graphics Vision and Visualisation group, School of Computer Science and Statistics, Trinity College, Dublin

Abstract

We present in this paper a novel study aiming at identifying the differences in visual search patterns between physicians of diverse levels of expertise during the screening of colonoscopy videos. Physicians were clustered into two groups -experts and novices- according to the number of procedures performed, and fixations were captured by an eye-tracker device during the task of polyp search in different video sequences. These fixations were integrated into heat maps, one for each cluster. The obtained maps were validated over a ground truth consisting of a mask of the polyp, and the comparison between experts and novices was performed by using metrics such as reaction time, dwelling time and energy concentration ratio. Experimental results show a statistically significant difference between experts and novices, and the obtained maps show to be a useful tool for the characterisation of the behaviour of each group.

CR Categories: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Tracking;

Keywords: Saliency, Eye-tracking, Experts, Novices, Medical Imaging, Polyp, Colonoscopy

1 Introduction

Colon cancer is nowadays the fourth most common cause of cancer death worldwide, showing a survival rate which strongly depends on the stage it is detected on [American Cancer Society 2013]. Colonoscopy is still considered as the gold standard for colon screening, although it presents some drawbacks being the most relevant the miss-rate, which has been reported to be at least of 6% [Bressler, B. et al. 2007]. Since this miss-rate results in deaths associated to the loss of polyps, the appropriate assessment of colon screening is a strong need.

We propose for this assessment the analysis of physicians visual search patterns. We show that the analysis of search patterns can be used to distinguish between physicians with different degrees of expertise, and this approach is proven as particulary successful when differentiating between experts and novices. In order to approximate physicians' visual attention models we propose the use of heat maps. These maps are generated by integrating physicians' fixations captured by an eye-tracker device during the screening of colonoscopy videos with the task of searching for a polyp. Our hypothesis is that differences that exist in the way physicians search for polyps are related to the degree of expertise -number of interventions- and that these differences can be objectively measured and tested.

Our approach allows a particular practical application related to the validation of automatic computer vision methods such as those proposed in recent works [Bernal et al. 2012; Dempere-Marco, Laura et al. 2011]. These methods present as output a computational saliency map related with the likelihood of presence of pathologies. Particularly, the experiments proposed in this paper can be straightforwardly extended to assess the performance of automatic polyp localization methods whose description and implementation is out of the scope of this contribution.

Finally, we introduce a novel ground truth consisting of an annotated database of video sequences. We use reaction time, dwelling time and concentration ratio to make the comparison between experts and novices. To assess the statistical significance of the results we validate our experiments using well-known statistical tests.

We present in Section 2 works related to the analysis of differences in search patterns related with eye-tracker devices. We detail the integration of fixations into heat maps in Section 3. We introduce the ground truth and the metrics of the experiment in Section 4. Experimental results are exposed in Section 5. We close this paper with the Conclusions and Future Work in Section 6.

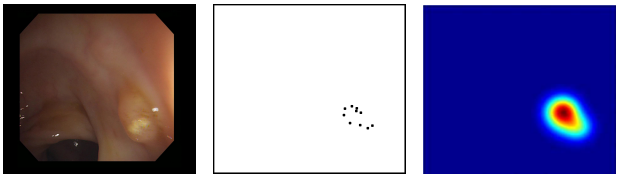
2 Related Work

The data captured by the eye tracker device is used in this paper as a seed to investigate the visual attention of the physicians -whose gaze is attracted by regions of interest in the image (RoIs) during the colonoscopy video screening-. The work of [Borji, A. et al. 2012] provides a formal definition for visual attention and its correspondence with visual saliency. Visual attention is defined as the process which makes either biological or computerized systems fixate on the most attractive region of an image. The attractiveness of a given region can be determined either by top-down factors related to a certain task to be performed or by bottom-up factors which highlight image regions which are different from their surroundings.

The experiment that we propose deals with top-down saliency as the attractive regions of the area are defined by the concrete task given to physicians -searching for a polyp-. There are some works related to the topic of integration of fixations into visual saliency maps, mainly with static images [Hu, X-P et al. 2003], but also in our particular fieldwork of video sequences [Chung, A.J. et al. 2005; Privitera, C.M. et al. 2000]. The contributions presented in [Harding, P. et al. 2009; Chen, H. et al. 2011] developed the concept of task driven saliency maps. The authors compared the performance of difference saliency maps by using thresholding on saliency levels and then check whether higher values correspond to positions of the object of interest.

The comparison of experts and novices using eye tracker data has been studied in fields such as threat assessment [Mann, C.M. et al. 2013] or the identification of potential burglars [Hillstrom, A.P. et al. 2013]. Closer to our domain, the work of

*e-mail: jbernal@cvc.uab.es



(a) (b) (c)

Figure 1: *(a) Original image; (b) Physicians' fixations; (c) Heat map. Hot colors represent attractive areas.*

[Khan, Rana et al. 2012] focuses on whether novice surgeons look at the same location as experts do in a laparoscopic operation. In the latter work physicians' gaze position is captured both at intervention time and at a posteriori screening of their own intervention.

3 Integration of Fixations

We aim at creating heat maps that approximate visual attention. These heat maps are created by combining by addition, for each frame of a video sequence, the fixations of the different subjects. For this aim, we propose to use the task driven saliency maps approximation depicted in [Chen, H. et al. 2011], in which the salience of an image is represented by a fixation density map: The heat map is created from a set of discrete fixation points (x_n^f,y_n^f), $n \in [1, N]$ where N is the total number of fixation points found in a frame and (x_n^f,y_n^f) is the location of the n-th fixation point. Those fixation points are interpolated by a Gaussian function to generate a fixation density map $s(x, y)$:

$$s\left(x,y\right) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{2\pi\sigma_s^2}\cdot exp\left(-\frac{\left(x-x_n^f\right)^2+\left(y-y_n^f\right)^2}{2\sigma_s^2}\right), \quad (1)$$

where x and y denote, respectively, the horizontal and vertical positions of an observation pixel and σ_s is the standard deviation of the Gaussian function, determined according to the visual angle accuracy of the eye tracking system. More precisely,

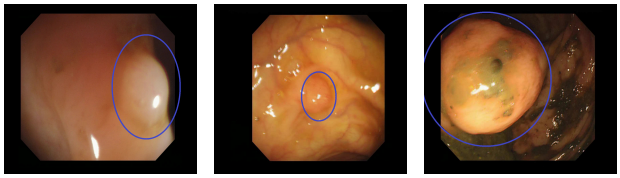
$$\sigma_s = L \times \tan\left(0.5\pi/180\right), \quad (2)$$

where L is the viewing distance between the subject and the display. L was set to 60 cm in this experiment. In this way, each fixation contributes to the heat map in a local neighborhood centered in the fixation position and with an area of influence defined by σ_s. Therefore, a pixel in a region densely populated by fixations has a brighter value than a pixel in a more diffuse area.

4 Experimental Setup

Experiments were run in order to observe the differences in search behaviour between experts and novices, comprising 22 physicians from Beaumont and St. Vincent's Hospitals, in Dublin, Ireland. The subjects were selected to show variability in the number of interventions performed: from more than 1000 intervention for senior physicians to no real intervention yet performed by novice trainees. We clustered physicians into experts and novices according to the number of procedures, using 100 as the threshold value to separate the clusters under the guidelines of the Joint Advisory Group on Gastrointestinal Endoscopy [Barton 2008].

We run an experiment consisting of the screening of 15 different video sequences by physicians -11 actual colonoscopy videos and 4 interlaced sequences for calibration-. Each video had an average of 1.500 frames with about 300 of them containing a polyp, and they were displayed at 25 fps. Physicians were asked to search for a polyp in the sequences and gaze position was captured using an EyeLink II eye-tracker device at 250 Hz. Physicians had



(a) (b) (c)

Figure 2: *Examples of the ground truth superimposed in three different colonoscopy frames.*

no interaction with the system and they were asked to view each whole sequence without any interruption. We created an *average expert* and *average novice* by following the methodology explained in Section 3. The result are two virtual subjects that do not have a single fixation by frame, as happens for the individual physicians, but a single heat map for each frame. In the context of the paper an *average expert/novice* is referred to the new subject created by the integration of physicians' fixations with no relation to calculation of average statistics. We show a complete example of heat map creation for a given video frame in Figure 1. Finally an elliptical mask for the polyp in each frame was provided by an expert annotation as a ground truth. Some examples of the ground truth are shown in Figure 2.

In order to compute performance metricss, we define a *Polyp Fixation Frame (PFF)* as a frame in which the maximum of the heat map falls under the polyp mask. Consequently the *First Polyp Fixation Frame (FPFF)* is the first PFF in the video sequence. We use the following metrics to compare *average expert* and *average novice*:

- *Reaction time (RT):* Number of frames between the first frame of the sequence where the polyp appears and the FPFF.
- *Dwelling time (DT):* Total number of PFF in a whole sequence.
- *Concentration Ratio (CR):* Percentage of the energy that falls under the polyp mask. In this context the energy is calculated as the sum of heat map values. The formal definition of CR is: $CR = 100 \times (E_p/E_f)$, where E_p corresponds to the total energy under the polyp mask, and E_f corresponds to the total energy of the map for the whole frame. A high CR value will correspond to a heat map focused on the polyp, whereas a low CR value will denote a more diffuse energy map.

We validate our experiments by using the well-known sign test and Wilcoxon signed rank test for paired data [Martinez, W.L. et al. 2001] to assess the statistical significance of the results of the comparison between the *average expert* and the *average novice*. We provide a p-value for each method (p_{st} for sign test, and p_{ws} for Wilcoxon signed rank test) when a significant result is obtained with a significance level of 0.05.

Finally, we put these results in correspondence with the individual analysis of fixations for each video. We analyze RT and DT using a linear mixed model with repeated measures with three co-variables: type of observer, video number and interaction between type of observer and video. For variables not following a normal distribution a $\log\left(x+1\right)$ transform was applied.

5 Results

5.1 Reaction time

Reaction time results are presented in Table 1. The analysis of the results shown in Table 1 provides a statistically significant difference between the *average expert* and the *average novice* -$p_{st} = 0.0039$ and $p_{ws} = 0.0039$-.

Video	Reaction Time			Dwelling Time		
	Avg. Expert	Avg. Novice	Diff.	Avg. Expert	Avg. Novice	Diff
1	3	9	-6	102	89	13
2	0	6	-6	88	83	5
3	0	16	-16	151	124	27
4	0	7	-7	101	86	15
5	0	14	-14	127	101	26
6	4	25	-21	83	93	-10
7	0	0	0	119	119	0
8	0	2	-2	79	69	10
9	0	1	-1	109	62	47
10	2	17	-15	108	88	20
11	0	0	0	92	52	40

Table 1: *RT and DT results for the average expert and the average novice. RT and DT are measured in number of frames.*

We can observe that for the majority of the videos *average expert* reaction time is 0. This is indeed a very interesting result which can be interpreted as the experts knowing where not to look at in the image, being prone in this way to be closer to the potential regions where a polyp can appear. This means that this *average expert* was already looking to the area where the polyp is, showing also that the degree of expertise has also a strong relationship with respect to the area of the image where the physicians place their attention.

5.2 Dwelling time

As can be seen from Table 1 for 10 of the 11 videos the dwelling time for experts is higher than for novices. The results show that again there is a statistically significant difference between experts and novices -$p_{st} = 0.0215$ and $p_{ws} = 0.0078$-. This means that, once the *average expert* finds the polyp in the image still places its fixation under the polyp mask for a high number of frames whereas the *average novice*, considering that for most of the cases it finds the polyp later than the *average expert*, places its fixation in a smaller number of frames. Differences in dwelling time may potentially be also related to sparser fixations for novices affecting the concentration ratio under the polyp mask, which is studied in the next subsection.

5.3 Concentration ratio

We use concentration ratio (CR) to assess whether heat maps are focused inside the polyp mask or scattered throughout the image. We make two different analysis regarding CR: the first explores differences in CR in the corresponding FPFF whereas the second extends the analysis for all the frames with a polyp.

Experimental results are presented in Table 2. The comparison of the CR values on the FPFF shows that there is an statistically significant difference between *average expert* and *average novice* -$p_{st} = 0.0386$ and $p_{ws} = 0.0269$-. As can be seen for 7 out of 11 videos CR is higher for the *average expert* than for the *average novice*. This can be interpreted as experts agreeing more on when the polyp appears in the image which is a result of having more fixations inside the polyp mask.

We present in Table 2 results on the comparison of mean CR for all the frames with a polyp. The results show a statistically significant difference between experts and novices for the signed rank test -$p_{ws} = 0.0244$-. We can observe from the Table that for 9 out of 11 videos the mean CR is higher for experts than for novices, which can be interpreted that experts, apart from finding the polyp earlier, have more confidence on where is the polyp in the image as their corresponding CR value is higher than for novices.

Video	CR in FPFF (%)			Mean CR (%)		
	Avg. Expert	Avg. Novice	Diff.	Avg. Expert	Avg. Novice	Diff
1	23.31%	49.61%	-26,3%	27.71%	25.90%	1,81%
2	40.36%	24.83%	15,53%	23.26%	21.63%	1,63%
3	21.43%	45.07%	-23,64%	29.69%	22.33%	7,36%
4	48.95%	29.04%	19,91%	32.99%	29.94%	3,05%
5	17.86%	8.44%	9,42%	24.37%	22.20%	2,17%
6	17.80%	9.69%	8,11%	28.98%	29.84%	-0,86%
7	98.71%	100.00%	-1,29%	84.49%	89.64%	-5,15%
8	50.38%	19.39%	30,99%	22.76%	16.91%	5,85%
9	70.21%	48.71%	21,5%	16.13%	11.77%	4,36%
10	13.71%	91.55%	-77,84%	40.44%	25.28%	15,16%
11	13.03%	10.89%	2,14%	11.70%	6.00%	5,7%

Table 2: *CR results.*

To illustrate better these differences we show in Figure 3 a comparison of the CR for experts and novices along a sequence of consecutive frames starting with the first polyp apparition. We have marked in the image some of the most interesting results: (1) We can observe how experts find the polyp earlier than novices and corresponding CR value is higher for experts than for novices; (2) CR of the average expert is higher than average novice for the majority of the sequence and whenever this difference is the opposite, it is not as high as the positive difference between experts and novices; (3) There are some frames of the video where the *average novice* map is completely focused outside the polyp whereas experts still concentrate energy inside; (4) CR is 0 since the polyp disappeared for several frames in the sequence.

5.4 Integration approach vs. individual analysis of fixations

The analysis of individual fixations using the linear mixed model confirms the differences between experts and novices regarding RT: the analysis of the co-variable type-of-observer provides $p = 0.0244$, assessing that individual experts localize polyps earlier than individual novices. Moreover, results confirm that differences in RT between experts and novices depend also on the particularities of the specific video, showing statistically significant results ($p = 0.0165$) for the co-variable interaction between type-of-observer and video. Regarding DT, the individual analysis of the fixations do not provide statistically significant differences between experts and novices, showing difference in mean DT for experts ($140.3\ [119.8, 160.8]$) and novices ($127.8\ [104.3, 151.3]$) at 95% confidence interval. The assessment of the co-variable video number is expectedly significant, illustrating that the total number of fixations, with independence of the type of observer, varies for each specific video independently of the type of the observer.

Both the analysis of individual fixations and the integration in heat maps approach share results in terms of confirming differences between experts and novices related to RT. The analysis of DT does not lead to the same conclusion, this being linked to heat maps approach also considering the influence of fixations that are close to the polyp but not strictly within the polyp mask. The heat maps approach integrates fixations not only as single-pixel coordinates but as a region of influence, which can be regarded as a rough initial approach to the region of foveal attention. This provides a more robust representation of the search patterns, since there is no practical difference between a fixation a few pixels outside or inside the polyp mask. In addition, this model also permits the direct comparison with computational saliency maps by naturally presenting them as the outcome obtained by a virtual expert.

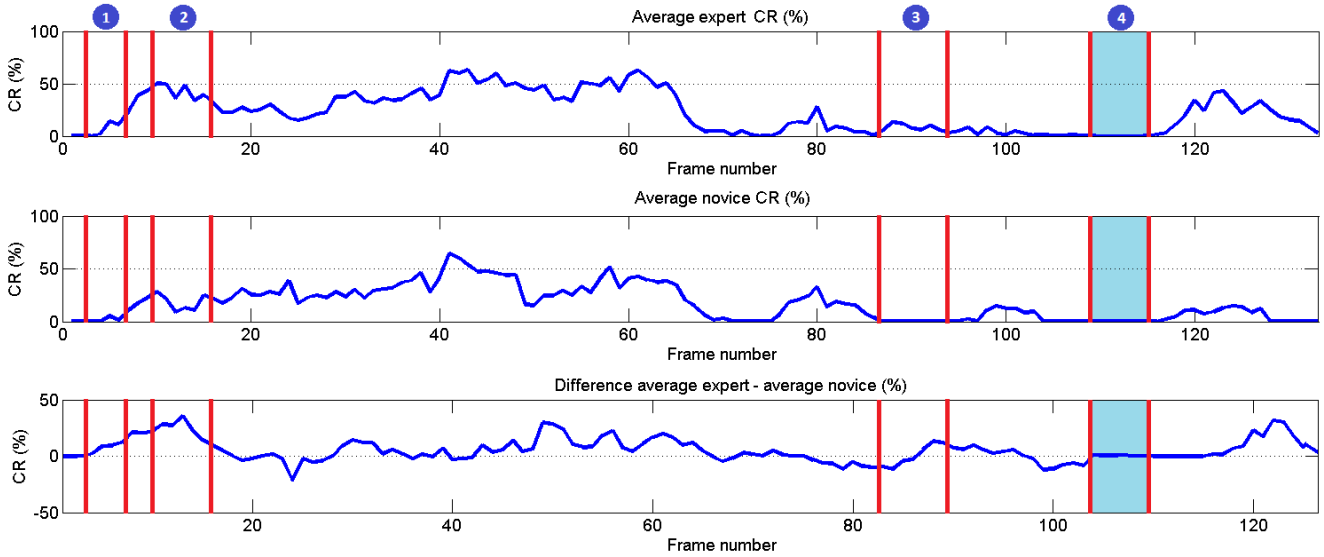


Figure 3: *Top to down: CR for average expert; CR for average novice; difference of CR.*

6 Conclusions and Future Work

We presented in this paper a novel study which aims at identifying potential differences between physicians of diverse degrees of expertise in visual search patterns when they were asked to localize polyps in colonoscopy videos. Physicians were clustered in two groups -experts and novices- according to the number of procedures performed. We modeled physicians' visual attention as heat maps created by the integration of their gaze position. These heat maps were validated on our proposed ground truth and results show a statistically significant difference between experts and novices. Experts react earlier to polyp presence, they provide more concentrated fixation patterns and, when localizing the polyp, the amount of energy inside polyp mask is higher than the one for novices.

The results of this study can potentially be used to assess the degree of expertise for a particular physician based on visual search patterns. This study can also be used to validate the performance of a computer-based polyp localization method by putting into correspondence the ROIs provided by the system [Bernal et al. 2012] with the regions provided by physicians visual attention.

Acknowledgements

This work was supported in part the Spanish Government through the founded projects "COLON-QA" ($TIN2009-10435$) and "FISIOLOGICA" ($TIN2012-33116$).

References

AMERICAN CANCER SOCIETY, 2013. How is colorectal cancer staged? American Cancer Society website, July.

BARTON, R. 2008. Accrediting competence in colonoscopy: validity and reliability of the uk joint advisory group/nhs bowel cancer screening programme accreditation assessment. *Gastrointestinal Endoscopy 67*, 5, AB77–AB77.

BERNAL, J., SÁNCHEZ, J., AND VILARIÑO, F. 2012. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition 45*, 9, 3166–3182.

BORJI, A. ET AL. 2012. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing 22*, 1, 55.

BRESSLER, B. ET AL. 2007. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. *Gastroenterology 132*, 1, 96–102.

CHEN, H. ET AL. 2011. Learning-based prediction of visual attention for video signals. *IEEE Transactions on Image Processing 20*, 99, 1–1.

CHUNG, A.J. ET AL. 2005. Extraction of visual features with eye tracking for saliency driven 2d/3d registration. *Image and Vision Computing 23*, 11, 999–1008.

DEMPERE-MARCO, LAURA ET AL. 2011. A novel framework for the analysis of eye movements during visual search for knowledge gathering. *Cognitive Computation 3*, 1, 206–222.

HARDING, P. ET AL. 2009. A comparison of feature detectors with passive and task-based visual saliency. *Image Analysis*, 716–725.

HILLSTROM, A.P. ET AL. 2013. Searching a house for valuables to steal: The influence of experience with burglary and other offences. *Book of Abstracts of the 17th ECEM 6*, 3, 244.

HU, X-P ET AL. 2003. Hot spot detection based on feature space representation of visual search. *IEEE Transactions on Medical Imaging, 22*, 9 (sept.), 1152 –1162.

KHAN, RANA ET AL. 2012. Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical endoscopy 26*, 12, 3536–3540.

MANN, C.M. ET AL. 2013. Rapidly imparting the skills of experts to novice participants in threat assessment tasks. *Book of Abstracts of the 17th ECEM 6*, 3, 245.

MARTINEZ, W.L. ET AL. 2001. *Computational statistics handbook with MATLAB*, vol. 2. Chapman & Hall/CRC.

PRIVITERA, C.M. ET AL. 2000. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 9, 970–982.