

Toward a structured training system for probe-based Confocal Laser Endomicroscopy (pCLE) on Barrett's Esophagus: a video retrieval approach to estimate diagnosis difficulty

Barbara André, Tom Vercauteren, Anna M. Buchner, Muhammad Waseem Shahid,
Michael B. Wallace, Nicholas Ayache

Background: pCLE (Cellvizio, Mauna Kea Technologies) allows the endoscopist to image the epithelial surface *in vivo*, at microscopic level and in real-time (12 frames per second) during an ongoing endoscopy. Early diagnosis of epithelial cancers with pCLE may be perceived as a challenging task for many new endoscopists. There is a crucial need to provide objective methods to diagnose neoplasia, estimate confidence levels, and to shorten the learning curve.

Aims: Our long-term objective is to develop a modular training system for pCLE diagnosis, by adapting the difficulty level according to the endoscopist's expertise. This study aims at providing an automated estimation of the diagnosis difficulty. As the understanding of pCLE video diagnosis is driven by perceived visual similarity, we propose a content-based video retrieval approach toward this goal.

Methods: Our database contains annotated pCLE videos of BE that were provided by the multicentric study NCT00795184. It includes 76 patients and 123 videos (62 benign, 61 neoplastic) split into 862 stable video subsequences. 20 of these videos (9 benign, 11 neoplastic) were graded offline by 21 endoscopists, including 9 pCLE experts and 12 non-experts, who individually established a blinded pCLE diagnosis for each lesion. A single expert GI pathologist reviewed all the biopsies acquired on the imaging spots and provided a reference diagnosis. The percentage of false pCLE diagnosis established on a video among the endoscopists is our "ground truth" for the diagnosis difficulty of the video. We first applied to the video database a video retrieval method that we developed especially for this task. We then used the retrieval results to extract a relevant difficulty criterion that measures contextual discrepancies between the video query and its most visually similar videos.

Results: Our video retrieval method, objectively evaluated using k-nearest neighbor classification, outperforms several state-of-the-art methods on the BE database (acc. 85.4%, sens. 90.2%, spec. 80.7%). Our estimated diagnosis difficulty has a correlation of 0.78 (p-value < 0.0002) with the ground truth difficulty measured for all the endoscopists, 0.63 (p-value < 0.003) with those measured for the experts only, and 0.80 (p-value < 0.0001) with those measured for the non-experts only.

Conclusion: Our experiments demonstrate that there is a noticeable relationship between our retrieval-based difficulty estimation and the difficulty experienced by the endoscopists. The complete video database with estimated difficulty could thus be used to identify lesions for which an optical diagnosis will be difficult, and to develop a training simulator that features difficulty level selection. Finally, a clinical validation will be required to assess whether such a structured training system will eventually help shorten the pCLE learning curve.