Retrieval Evaluation and Distance Learning from Perceived Similarity between Endomicroscopy Videos

Poster ID: P06-1-W

Barbara André [1,2], Tom Vercauteren [1], Anna M. Buchner [3], Michael B. Wallace [4], Nicholas Ayache [2]

[1] Mauna Kea Technologies (MKT), Paris, France
[2] Asclepios Team, INRIA Sophia-Antipolis, France
[3] Hospital of the University of Pennsylvania, Philadelphia, USA
[4] Mayo Clinic, Jacksonville, Florida, USA

CONTEXT & MOTIVATIONS

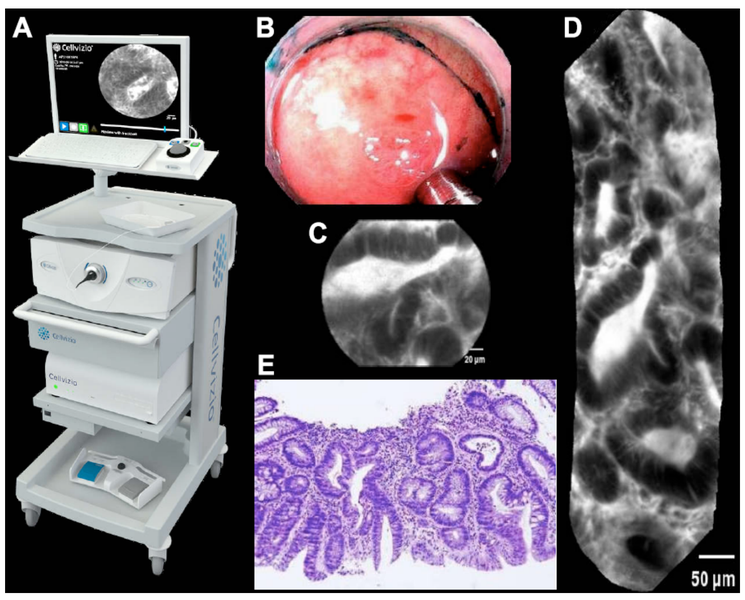
probe-based Confocal Laser Endomicroscopy pCLE
imaging the epithelium *in vivo et in situ* at microscopic level & real-time frame rate

Problem *In vivo* pCLE diagnosis is still a challenge for many endoscopists

Similarity-based Reasoning: physicians rely on visually similar cases they have seen in the past

⇩

Investigate Content-Based Retrieval to support the interpretation of pCLE videos



A. Setup of pCLE imaging system;
B. Endoscopic image of a colonic polyp; pCLE miniprobe;
C. Acquired pCLE image;
D. Associated pCLE mosaic image;
E. Corresponding histological image.

Preliminary work (ISBI'10)

- Dense bag-of-visual-words method "*Dense-Sift*" for the retrieval of pCLE videos
- Indirect retrieval evaluation using pathological classification

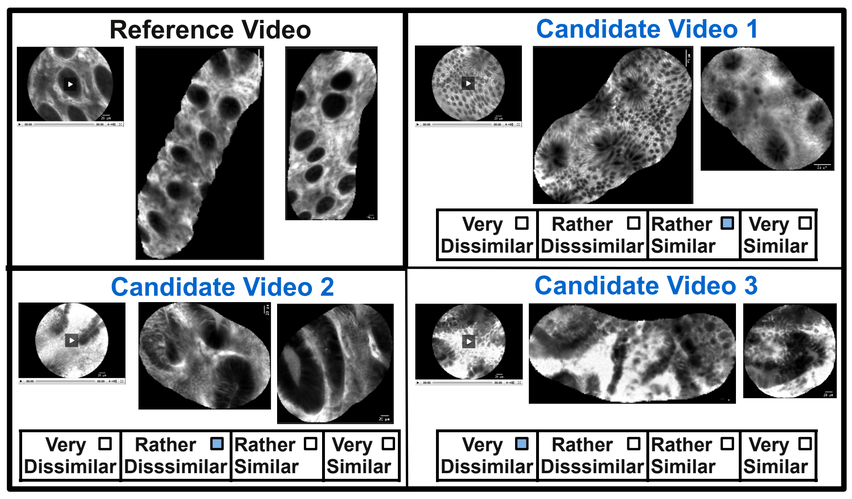
Objective of this study

Learning the visual similarity perceived between pCLE videos of colonic polyps to improve retrieval performance

⇩

Requires a perceived similarity ground truth (allowing for direct retrieval evaluation)

PERCIEVED SIMILARITY GROUND TRUTH



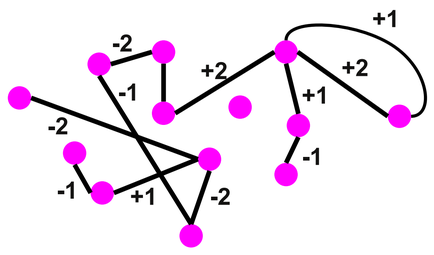
Scheme of the online survey tool (http://smartatlas.maunakeatech.com) allowing endoscopists to individually score the visual similarity that they perceive between pCLE videos of colonic polyps.

The probability of drawing a video couple (I,J) is proportional to the inverse of the density of the retrieval distance $d_{prior}(I,J)$ computed by the CBVR method "*Dense-Sift*".

Four-points Likert scale

"very dissimilar" (L = -2)
"rather dissimilar" (L = -1)
"rather similar" (L = +1)
"very similar" (L = +2)

Pairwise Similarity Graph corresponding to the sparse ground truth



pCLE video

L — L-score between two pCLE videos

SIMILARITY DISTANCE LEARNING

Margin-based cost function

$$f(W,\beta,\gamma) = \frac{1}{N+}\sum_{(I,J)\in D_+} g(\beta - d(W\cdot\mathscr{S}(I), W\cdot\mathscr{S}(J))) + \gamma\,\frac{1}{N_-}\sum_{(I,J)\in D_-} g(d(W\cdot\mathscr{S}(I), W\cdot\mathscr{S}(J)) - \beta)$$

margin; trade-off; transformation matrix; visual word signature of video I

D+ is the set of N+ training video couples scored with L = +2 ("very similar")
D- is the set of N- training video couples scored with L = +1, -1 or -2 (not "very similar")

$g(z) = log(1 + e^{-z})$ *is the logistic-loss function*

$$d(W\cdot\mathscr{S}_{\mathrm{Vis}}(I), W\cdot\mathscr{S}_{\mathrm{Vis}}(J)) = \chi^2\left(\frac{W\cdot\mathscr{S}_{\mathrm{Vis}}(I)}{||W\cdot\mathscr{S}_{\mathrm{Vis}}(I)||_{L^1}}, \frac{W\cdot\mathscr{S}_{\mathrm{Vis}}(J)}{||W\cdot\mathscr{S}_{\mathrm{Vis}}(J)||_{L^1}}\right)$$

Learned similarity distance between pCLE videos I and J

$$d^{learn}(I,J) = d(W^{opt}\cdot\mathscr{S}(I), W^{opt}\cdot\mathscr{S}(J))$$

RESULTS

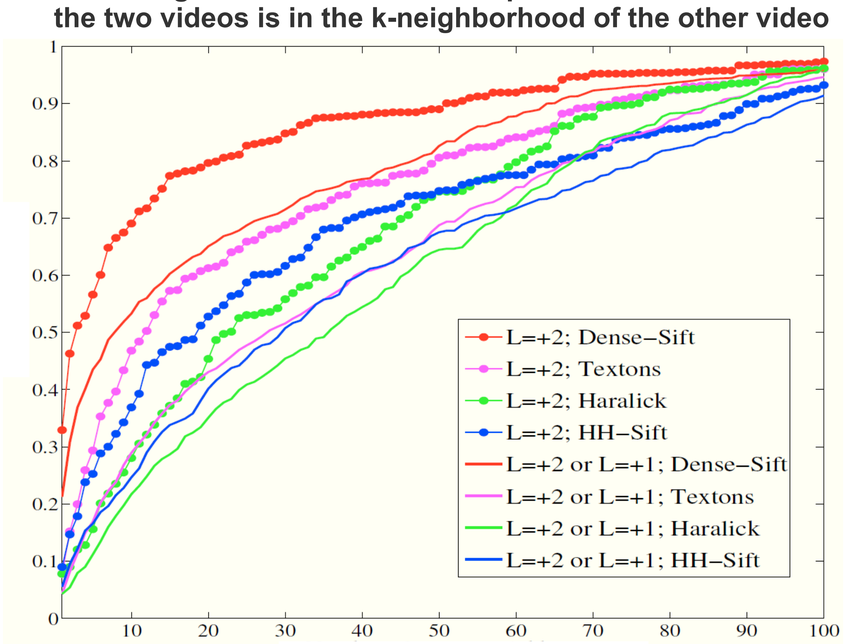
		Dense-Sift *(proposed method)*	Textons	Haralick	HH-Sift
Indirect evaluation using classification	*LOPO** Classification Accuracy	93 %	78 %	79 %	74 %
	Sensitivity	95 %	80 %	72 %	70 %
	Specificity	89 %	72 %	84 %	78 %
	McNemar's test p-value < 0.05	> Textons > Haralick > HH-Sift			
Direct evaluation w.r.t. perceived similarity	Pearson corr.	49 %	33 %	34 %	16 %
	Spearman corr.	52 %	35 %	34 %	22 %
	Kendall corr.	47 %	32 %	31 %	19 %
	Steiger's Z-test on Kendall corr. p-value < 0.05	> Textons > Haralick > HH-Sift	> HH-Sift	> HH-Sift	

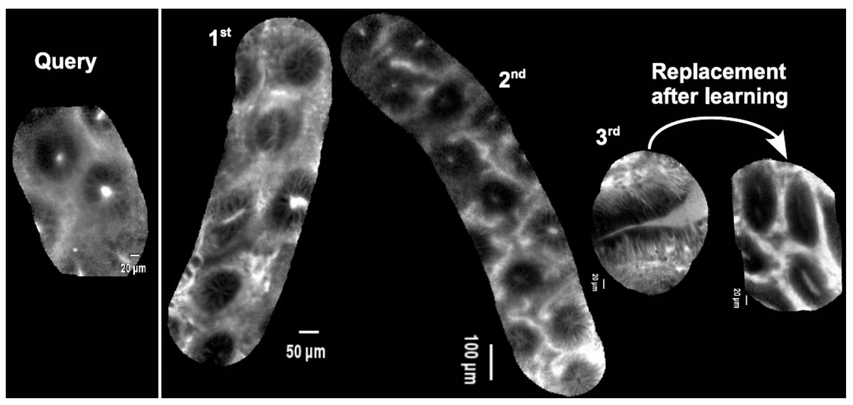
** LOPO: Leave-One-Patient-Out cross-validation*
" > " means "outperforms with statistical significance"

with 30x3-fold cross-validation		30x3 Dense-Sift + Distance Learning	30x3 Dense-Sift
Direct evaluation w.r.t. perceived similarity	Pearson corr.	53 %	46 %
	Spearman corr.	57 %	49 %
	Kendall corr.	53 %	45 %
	Steiger's Z-test on Kendall corr. p-value < 0.05	> 30x3 Dense-Sift	

Sparse Recall Curves

Percentage of L-scored video couples for which one of the two videos is in the k-neighborhood of the other video





Example of pCLE video query, represented by a mosaic, with its 3 nearest neighbors retrieved by "Dense-Sift" before and after similarity distance learning.

CONTRIBUTIONS

(1) Construction of an adequate sparse ground truth for perceived similarity between pCLE videos

(2) Direct evaluation of pCLE retrieval

⇨ In terms of visual similarity, our CBVR method significantly outperforms several state-of-the-art methods

(3) Generic visual-word-weighting-based method for perceived similarity distance learning

⇨ Significant improvement of pCLE retrieval performance

ONGOING WORK

- Enlarge database of perceived similarity ground truth
- Investigate more sophisticated distance learning techniques
- Clinically evaluate how pCLE similarity estimation could assist the endoscopists for *in vivo* pCLE diagnosis

References

- B. André et al., "A smart atlas for endomicroscopy using automated video retrieval", MedIA 2011
- B. André et al., "Endomicroscopic video retrieval using mosaicing and visual words", ISBI 2010
- Visual Similarity Scoring (VSS): http://smartatlas.maunakeatech.com, login:MICCAI-User, password: MICCAI2011
- J. Philbin et al., "Descriptor learning for efficient retrieval", ECCV 2010
- A.M. Buchner et al., "Comparison of probe based confocal laser endomicroscopy with virtual chromoendoscopy for classification of colon polyps", Gastroenterology 2009

Contact: barbara.andre@maunakeatech.com