Content-Based Retrieval in Videos from Laparoscopic Surgery

Klaus Schoeffmann[a], **Christian Beecks**[b], **Mathias Lux**[a], **Merih Seran Uysal**[b], **Thomas Seidl**[b]
[a]Institute of Information Technology, Alpen-Adria-Universität Klagenfurt, Austria
[b]Data Management and Exploration Group, RWTH Aachen University, Germany

Abstract. In the field of medical endoscopy more and more surgeons are changing over to record and store videos of their endoscopic procedures for long-term archival. These endoscopic videos are a good source of information for explanations to patients and follow-up operations. As the endoscope is the "eye of the surgeon", the video shows the same information the surgeon has seen during the operation, and can describe the situation inside the patient much more precisely than an operation report would do. Recorded endoscopic videos can also be used for training young surgeons and in some countries the long-term archival of video recordings from endoscopic procedures is even enforced by law. A major challenge, however, is to efficiently access these very large video archives for later purposes. One problem, for example, is to locate specific images in the videos that show important situations, which are additionally captured as static images during the procedure. This work addresses this problem and focuses on content-based video retrieval in data from laparoscopic surgery. We propose to use feature signatures, which can appropriately and concisely describe the content of laparoscopic images, and show that by using this content descriptor with an appropriate metric, we are able to efficiently perform content-based retrieval in laparoscopic videos. In a dataset with 600 captured static images from 33 hours recordings, we are able to find the correct video segment for more than 88% of these images.

Keywords: Content-based Video Retrieval, Feature Signatures, Earth Mover's Distance, Signature Quadratic Form Distance, Signature Matching Distance, Laparoscopy.

1 Summary

Feature signatures, as for instance investigated in the work of Beecks,[1] are a compact yet efficient way of describing the content of an image. They are obtained by clustering characteristic image features into a compact feature representation that adapts to individual image contents. In this paper, we propose to utilize feature signatures over a 7-dimensional feature space including position, color, and texture information for content-based retrieval in laparoscopic videos. To this end, we investigate different distance-based similarity measures including the Earth Mover's Distance,[2] the Signature Quadratic Form Distance,[3,4] and the Signature Matching Distance.[5] Through a large evaluation with video data from 33 hours of laparoscopic surgery we show that the proposed signature-based approaches are able to efficiently perform content-based retrieval in laparoscopic videos.

2 Description of Purpose

In endoscopic surgery, operating surgeons usually record video segments and capture static images during the operation, which they often need after the procedure.[6] However, while the static images can be used for later discussions with patients and colleagues, and can be easily integrated into operation reports, they lack of a complete and precise information, such as movements of operation instruments and the dynamics of performed actions. The recorded videos, on the other hand, contain all these important details, since they show exactly the same images that were visible on the monitors in the operation room, which were also used by the operation team for their decisions and movements. For this reason, more and more surgeons store recorded video segments, or even complete videos of full operations, for later inspection. This is particularly true for the field of laparoscopy, where the videos are used for technical quality assessment of the surgery.[7,8]

The later use of recorded videos from endscopic procedures is, however, currently not well supported by software tools, and has not been addressed intensively in research. While there are several works that focus on on-the-fly processing of the video stream from the endoscope (e.g., for detection of abnormalities[9,10]), post-processing of the recorded videos has only attracted a few research works so far. For example, there are works that focus on automatic extraction of relevant images from these recorded videos[11,12] and on the automatic removal of undesired content, such as blurry frames and out-of-patient recordings.[13] Also papers can be found in the literature that focus on the automatic segmentation of endoscopic videos into manageable units,[14] as well as on detection and tracking of specific operation instruments.[15] Although there are several works on content-based image retrieval in data from radiology (see[16] for a survey), content-based video retrieval in data from endoscopic procedures is sparsely addressed[17,18] and considered as an unsolved problem.

In this work we focus on a very relevant problem in practice: After the procedure surgeons often need to find the same images in their video recordings, which they have also captured as static images during the procedure. The reason for this is, that they often want to inspect the dynamics of the scene, i.e. see what exactly happened before and after this static image was shot. The problem, however, is that the search for one image in a set of many video files is a challenging task, in particular for large data sets, which is currently not supported well by software systems (for example, in our data set, which was collected through collaborating surgeons for several month and contains 2.25 million frames, the average number of segments per operation is 25). In practice, many surgeons use a common video player for that purpose, and the search process is very time-consuming and exhaustive.

We consider this problem as a video retrieval task, where the captured static images are used as query input to a *query-by-example* approach. For that purpose, we propose to utilize adaptive-binning feature signatures in order to model the image contents and distance-based similarity measures for quantifying the degree of image dissimilarity and evaluate their performance with respect to a large data set recorded from many different procedures of laparoscopic surgery.

3 Related Work

The potential of content-based retrieval in images and videos in the medical domain has been investigated by many groups over the last years. Focus of many publications are MRI (magnetic resonance imaging) or X-Ray images, being high resolution gray scale images. Early examples are described by Wei et al.[19] investigating potential applications of medical image retrieval as well as describing and reviewing existing medical content-based retrieval systems. Müller et al.[20] try to approach the problem from a different angle and investigate existing content-based retrieval systems for their capability of dealing with images from the medical domain. They also discuss a large variety of techniques, potential applications and future lines of research. More recently, Kumar et al.[21] provide a review emphasizing the multi-dimensional (2D and 3D) and multi-modality nature of the medical retrieval scenario.

As numerous approaches for content-based retrieval have been developed over the years,[22] a wide variety of approaches has been applied to medical image retrieval including for instance low-level wavelet-based visual signatures[23] as well as high level concept detectors.[24] From a computer vision perspective researchers tried to identify procedures, objects and contexts in images and applied labels for later text retrieval. One example for an early approach is presented by Kalpathy-Cramer and Hersh.[25] We assume that concept detection and labeling will become more important, especially due to the availability and success of deep learning frameworks.[26] One example for an approach based on deep learning and convolutional networks has already been published by Shin et al. for X-Ray images.[27]

Currently, medical retrieval systems try to become much more accessible on the web, typically being multi-modal in a way by supporting both textual and visual queries. Examples for web based search engines are NovaMedSearch[28] and GoldMiner.[29]

Still, visual retrieval in medical image and video databases remains an area of active research. Up to 2013 the ImageCLEF benchmarking initiative[30] for instance hosted a task for image-based retrieval in medical scenarios. For retrieval of medical cases each case was represented by a text and one to seven images. In 2013 neither multi-modal approaches incorporating both, text and visual information, nor purely visual approaches could outperform the best textual retrieval approach.[31,32] At this point we'd like to point out that the best achieving visual only approach in ImageCLEF 2013[33] was solely based on global features, especially the CEDD[34] descriptor. This supports our assumption that color is a necessary characteristic for retrieval, and that local features if considered, cannot be considered mutually independent, which makes feature signatures a great candidate for medical image retrieval. The strengths and weaknesses of the feature signature model have been investigated in various works[35–37] and its applicability reaches for instance from similarity search in multimedia databases[38–40] to scientific databases.[41,42]

4 Methods

4.1 Feature Signatures

For the purpose of content-based video retrieval in data from laparoscopy, we utilize feature signatures as adaptive content descriptors. These feature signatures can describe both very abstract information and local details of the images. More particularly, feature signatures describe the information contained in an image by means of features $f_1, \ldots, f_k \in \mathbb{F}$ in a feature space $\mathbb{F}$ comprising position, color, and texture information. By assigning each of these features a specific weight indicating the contribution to the corresponding image, we mathematically define a *feature signature* as follows:

$$X : \mathbb{F} \to \mathbb{R} \text{ subject to } |\{f \in \mathbb{F} | X(f) \neq 0\}| < \infty.$$

A feature signature X defines a finite set of contributing features, i.e. features with a weight unequal to zero, individually for each image. In this way, feature signatures adapt to individual image contents and can be further adapted by changing the underlying feature space $\mathbb{F}$. More details regarding feature representation models and in particular the feature signature model can be found for instance in the work of Beecks.[1]

An example of feature signatures is depicted in Figure 1. It shows three example images extracted from our data set and the visualizations of the corresponding feature signatures. The features are visualized by colored circles with diameters indicating their weights and are based on position, color, and texture information. As can be seen in this example, feature signatures are able to visually approximate image content.

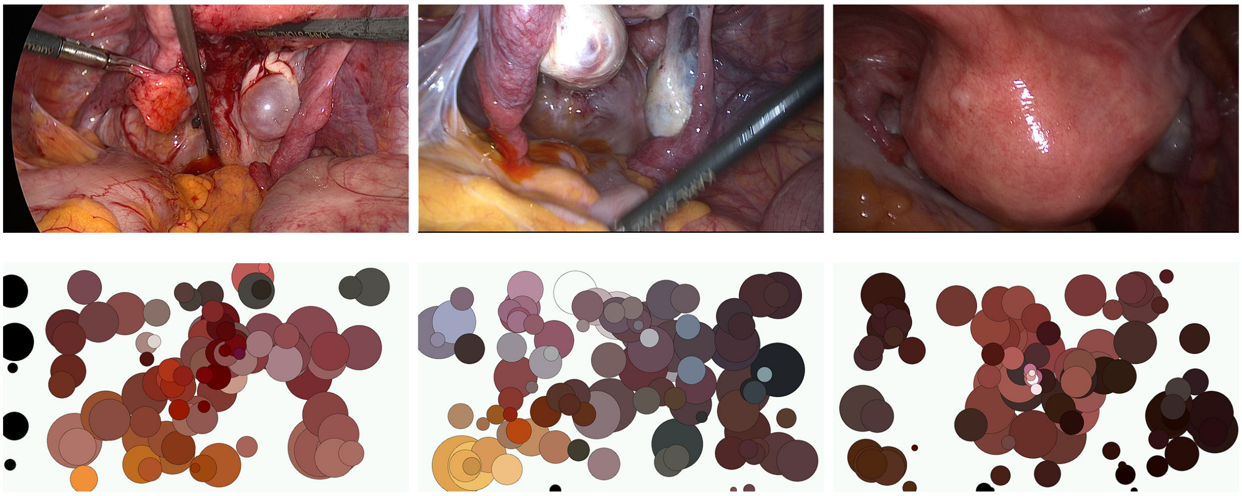

Fig 1 Three example images from laparoscopic videos and the corresponding feature signatures.

4.2 Distance-based Similarity Measures for Feature Signatures

Based on the feature signature model, we utilized different distance-based similarity measures in order to quantify the degree of dissimilarity between two images. These measures are based on a ground distance $\delta : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}$ which compares features from the underlying feature space $\mathbb{F}$.

The first approach for the comparison of feature signatures is the Earth Mover's Distance,[2] which is a transformation-based approach that measures the cost of transforming one feature signature into another one. The Earth Mover's Distance between two feature signatures $X, Y \in \mathbb{R}^{\mathbb{F}}$ is defined as a minimum-cost flow of all possible flows $F = \{f | f : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}\}$ as follows:

$$\mathrm{EMD}_{\delta}(X,Y) = \min_{f \in F} \left\{ \frac{\sum\limits_{x \in \mathbb{F}} \sum\limits_{y \in \mathbb{F}} f(x,y) \cdot \delta(x,y)}{\min\{\sum\limits_{x \in \mathbb{F}} X(x), \sum\limits_{y \in \mathbb{F}} Y(y)\}} \right\},$$

subject to the following constraints:

- non-negativity: $\forall x, y \in \mathbb{F} : \; f(x,y) \geq 0$
- source: $\forall x \in \mathbb{F} : \; \sum_{y \in \mathbb{F}} f(x,y) \leq X(x)$
- target: $\forall y \in \mathbb{F} : \; \sum\limits_{x \in \mathbb{F}} f(x,y) \leq Y(y)$
- total flow: $\sum\limits_{x \in \mathbb{F}} \sum\limits_{y \in \mathbb{F}} f(x,y) = \min\{\sum\limits_{x \in \mathbb{F}} X(x), \sum\limits_{y \in \mathbb{F}} Y(y)\}$

The second approach is the correlation-based Signature Quadratic Form Distance[3,4] which adapts the generic concept of correlation to feature signatures. The Signature Quadratic Form Distance is defined based on the similarity correlation $\langle X, Y\rangle_s = \sum_{x,y\in\mathbb{F}} X(x)\cdot Y(y)\cdot s(x,y)$ between two feature signatures with respect to a similarity function $s : \mathbb{F}\times\mathbb{F}\to\mathbb{R}$ as follows:

$$\mathrm{SQFD}_s(X,Y) = \sqrt{\langle X,X\rangle_s - 2\cdot\langle X,Y\rangle_s + \langle Y,Y\rangle_s}.$$

Throughout our experimental evaluation we utilized the Gaussian kernel with parameter $\sigma = 0.2$ as similarity function.

The third approach is the Signature Matching Distance.[5] It defines a distance value between two feature signatures $X, Y \in \mathbb{R}^{\mathbb{F}}$ by making use of a matching $\mathrm{m} \subseteq \mathbb{F}\times\mathbb{F}$ and a cost function $\mathrm{c} : 2^{\mathbb{F}\times\mathbb{F}} \to \mathbb{R}$. The Signature Matching Distance is defined for arbitrary matchings and cost functions as follows:

$$\mathrm{SMD}_\delta(X,Y) = \mathrm{c}(\mathrm{m}_{X\to Y}) + \mathrm{c}(\mathrm{m}_{Y\to X}) - 2\lambda\cdot\mathrm{c}(\mathrm{m}_{X\leftrightarrow Y}),$$

where the parameter $0 \le \lambda \le 1$ models the exclusion of bidirectional matches. Following our previous work,[43] we set $\lambda = 0$ and utilized the δ-Nearest-Neighbor Matching $\mathrm{m}^{\delta\text{-NN}}_{X\to Y} = \{(f,g) | X(f) > 0 \wedge Y(g) > 0 \wedge g = \mathrm{argmin}_{h\in\mathbb{F}}\delta(f,h)\}$ in combination with the linear cost function $\mathrm{c}_\delta(\mathrm{m}^{\delta\text{-NN}}_{X\to Y}) = \sum_{(f,g)\in\mathrm{m}^{\delta\text{-NN}}_{X\to Y}} X(f)\cdot\delta(f,g)$.

The question of which of the aforementioned approaches provides the highest retrieval performance is of highly empirical nature. To this end, we conduct an extensive performance analysis which is described in the following section.

5 Results

We perform an extensive evaluation with video data from 33 hours of laparoscopic surgery (2.25 million frames) and the corresponding static images captured by the surgeons. The data was collected from 48 procedures (i.e., different patients), with an average of 25 stored video files per procedure, recorded in full HD quality (1920x1080@25p). The additionally captured static images (in total 600 images for the whole data set, i.e. 12.5 per procedure in average, also stored in full HD quality) were used as example images for a query-by-example approach with our video retrieval engine. We evaluate the retrieval performance in terms of different recall levels, i.e., *Recall@1*, *Recall@2*, and *Recall@3*, against the provided ground truth, i.e., we check if the first results in the generated ranked lists of retrieval results comprise images from the same video file/sequence. It is worth noting that we cannot check whether it is also exactly the same video frame, since this data is not available in our ground truth.

The results for the Earth Mover's Distance are shown in Figure 2. As can be seen in the figure, an increase in the query signature size increases the retrieval performance. The maximum Recall@1 value of 65% is reached when utilizing a query signature size of 80 and the Manhattan ground distance L_1. The results for the Signature Quadratic Form Distance, which are shown in Figure 3, reveal a similar tendency. In general, increasing the query signature size improves the retrieval performance. The maximum Recall@1 value of 78% is reached when utilizing a query signature size of 80 and the Euclidean ground distance L_2 within the Gaussian kernel. Figure 4 depicts the results for the Signature Matching Distance. Among the aforementioned approaches, the Signature Matching Distance is able to achieve the highest retrieval performance. In fact, the

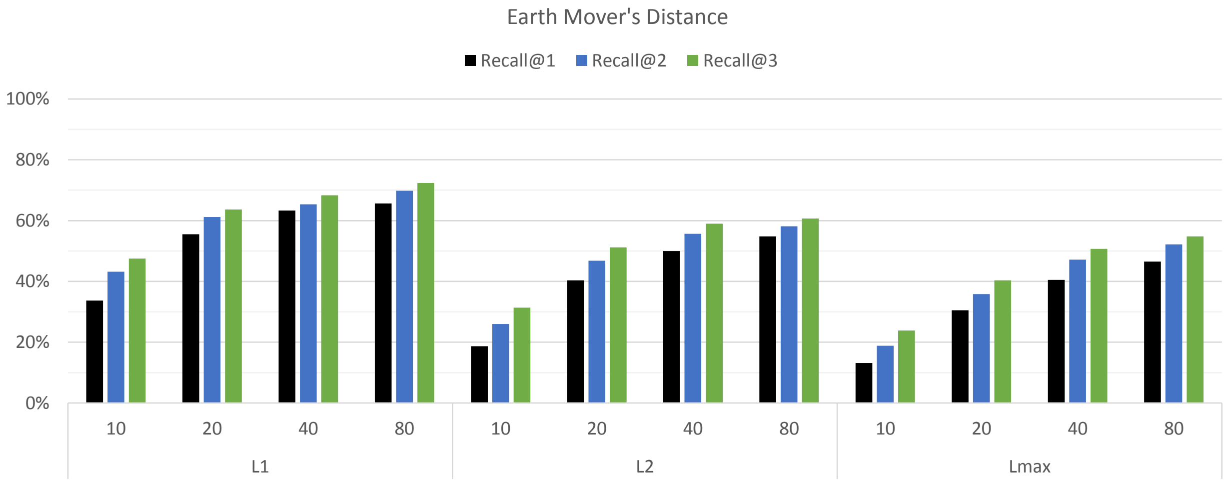

Fig 2 Recall@1, Recall@2, and Recall@3 values in percentage for the Earth Mover's Distance as a function of different query signature sizes between 10, 20, 40, and 80 and ground distances $\delta \in \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_{max}\}$.

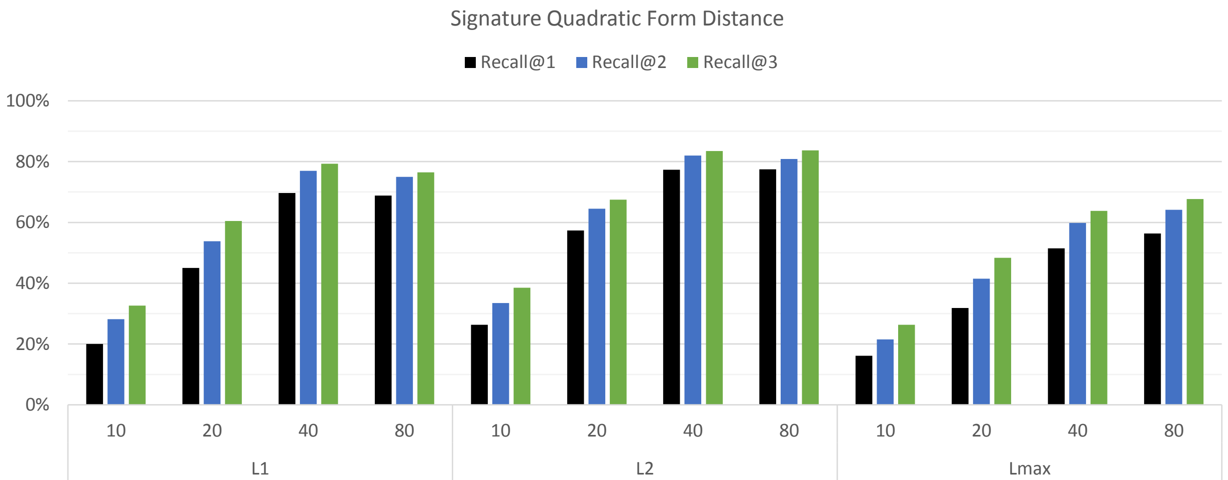

Fig 3 Recall@1, Recall@2, and Recall@3 values in percentage for the Signature Quadratic Form Distance as a function of different query signature sizes between 10, 20, 40, and 80 and ground distances $\delta \in \{\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_{max}\}$.

maximum Recall@1 value of 88% is reached when utilizing a query signature size of 80 and the Manhattan ground distance L_1.

In summary, our results show that the proposed approach of video retrieval using feature signatures is able to outperform existing methods of Roldan et al.,[44] which use a fusion approach of global image features as well as local image features. While the latter approach achieves a Recall@1 value of approximately 80%, the approach proposed in this work is able to improve the state of the art by more than 10% of retrieval performance. This complies with the findings presented in the work of Beecks et al.[43]

Additionally, our evaluation results show that feature signatures work best for the domain of laparoscopic surgery when used with the Signature Matching Distance (SMD) and L1 as ground dis-

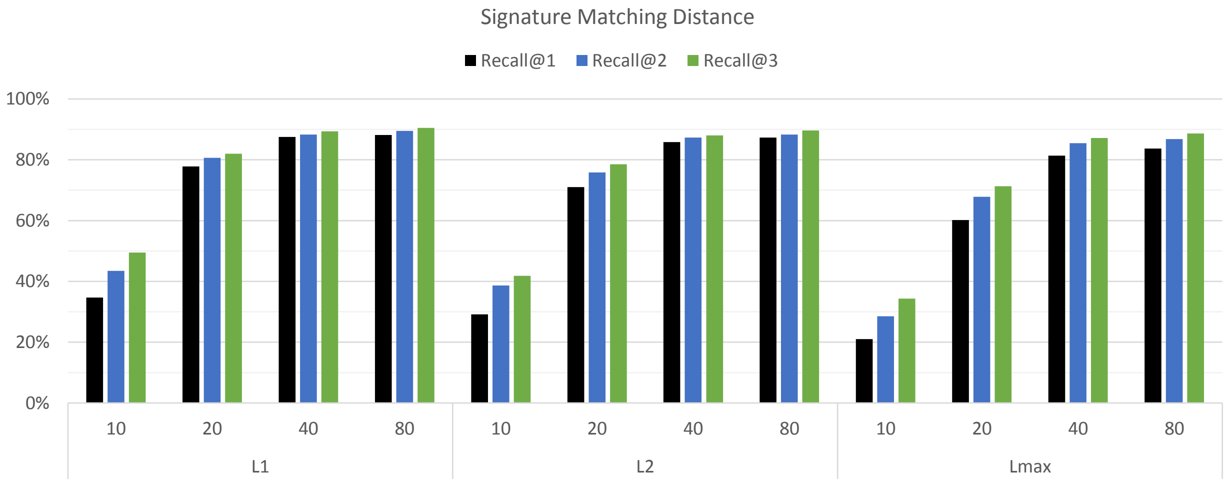

Fig 4 Recall@1, Recall@2, and Recall@3 values in percentage for the Signature Matching Distance as a function of different query signature sizes between 10, 20, 40, and 80 and ground distances $\delta \in \{L_1, L_2, L_{max}\}$.

tance. The achieved retrieval performance with both Signature Quadratic Form Distance (SQFD) and Earth Mover's Distance (EMD) is significantly lower. It is also worth noting that SMD already achieves a remarkably high retrieval performance with a Recall@1 value close to 80% at a query signature size of 20, which is almost twice as good as with SQFD. This means that the SMD will also allow for much faster retrieval in practice, which should be the subject of further investigations in this field.

6 Conclusions

In this work we have presented different signature-based approaches for content-based video retrieval in recordings from laparoscopic surgery. Our approaches utilize feature signatures based on low dimensional feature spaces in order to efficiently describe the endoscopic images. Further, we investigate the Earth Mover's Distance, the Signature Quadratic Form Distance, and the Signature Matching Distance to compare feature signatures, and to generate the ranked list of results for retrieval tasks. Through an evaluation with a large data set we have shown that these approaches achieve high retrieval performance and are able to outperform the current state of the art in the literature.

Our evaluation results have also shown that the Signature Matching Distance allows for video retrieval with high performance already with small-sized feature signatures, which are much faster to compare than larger ones. In future work we want to perform detailed run-time evaluations of the proposed approach, and investigate how to further reduce the required run-time. This will allow us to integrate the proposed approach into real-time interactive video retrieval tools, such as the one recently proposed by Hudelist et al.[45]

References

1 C. Beecks, *Distance-based similarity models for content-based multimedia retrieval*. PhD thesis, RWTH Aachen University (2013).

2 Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision* **40**(2), 99–121 (2000).

3 C. Beecks, M. S. Uysal, and T. Seidl, "Signature quadratic form distances for content-based similarity," in *Proceedings of the 17th International Conference on Multimedia*, 697–700 (2009).

4 C. Beecks, M. S. Uysal, and T. Seidl, "Signature quadratic form distance," in *Proceedings of the 9th ACM International Conference on Image and Video Retrieval*, 438–445 (2010).

5 C. Beecks, S. Kirchhoff, and T. Seidl, "Signature matching distance for content-based image retrieval," in *International Conference on Multimedia Retrieval*, 41–48 (2013).

6 B. Munzer, K. Schoeffmann, L. Boszormenyi, J. Smulders, and J. J. Jakimowicz, "Investigation of the impact of compression on the perceptional quality of laparoscopic videos," in *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, 153–158, IEEE (2014).

7 E. M. Bonrath, B. Zevin, N. J. Dedy, and T. P. Grantcharov, "Error rating tool to identify and analyse technical errors and events in laparoscopic surgery," *British Journal of Surgery* **100**(8), 1080–1088 (2013).

8 H. Husslein, L. Shirreff, E. M. Shore, G. G. Lefebvre, and T. P. Grantcharov, "The generic error rating tool: A novel approach to assessment of performance and surgical education in gynecologic laparoscopy," *Journal of Surgical Education* , – (2015).

9 A. Dahal, J. Oh, W. Tavanapong, J. Wong, and P. de Groen, "Detection of ulcerative colitis severity in colonoscopy video frames," in *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on*, 1–6 (2015).

10 R. Nawarathna, J. Oh, J. Muthukudage, W. Tavanapong, J. Wong, P. C. de Groen, and S. J. Tang, "Abnormal image detection in endoscopy videos using a filter bank and local binary patterns," *Neurocomputing* **144**, 70–91 (2014).

11 K. Schoeffmann, M. Del Fabro, T. Szkaliczki, L. Böszörmenyi, and J. Keckstein, "Keyframe extraction in endoscopic video," *Multimedia Tools and Applications* , 1–20 (2014).

12 E. Mendi, C. Bayrak, S. Cecen, and E. Ermisoglu, "Content-based management service for medical videos," *Telemedicine and e-Health* (2012).

13 B. Munzer, K. Schoeffmann, and L. Boszormenyi, "Relevance segmentation of laparoscopic videos," in *Multimedia (ISM), 2013 IEEE International Symposium on*, 84–91, IEEE (2013).

14 M. Primus, K. Schoeffmann, and L. Boszormenyi, "Segmentation of recorded endoscopic videos by detecting significant motion changes," in *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on*, 223–228 (2013).

15 M. J. Primus, K. Schoeffmann, and L. Böszörmenyi, "Instrument classification in laparoscopic videos," 1–6 (2015).

16 C. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *Journal of Digital Imaging* **24**(2), 208–222 (2011).

17 B. Andre, T. Vercauteren, A. Buchner, M. Wallace, and N. Ayache, "Learning semantic and visual similarity for endomicroscopy video retrieval," *Medical Imaging, IEEE Transactions on* **31**, 1276–1288 (2012).

18 B. Andre, T. Vercauteren, A. Buchner, M. Wallace, and N. Ayache, "Endomicroscopic video retrieval using mosaicing and visualwords," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, 1419–1422 (2010).

19 C.-H. Wei, C.-T. Li, and R. Wilson, "A content-based approach to medical image database retrieval," *Database Modeling for Industrial Data Management: Emerging Technologies and Applications. Idea Group, Hershey* , 258–291 (2006).

20 H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications - clinical benefits and future directions," *International journal of medical informatics* **73**(1), 1–23 (2004).

21 A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data," *Journal of digital imaging* **26**(6), 1025–1039 (2013).

22 R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (CSUR)* **40**(2), 5 (2008).

23 G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Medical image analysis* **14**(2), 227–241 (2010).

24 M. M. Rahman, S. K. Antani, and G. R. Thoma, "A learning-based similarity fusion and filtering approach for biomedical image retrieval using svm classification and relevance feedback," *Information Technology in Biomedicine, IEEE Transactions on* **15**(4), 640–646 (2011).

25 J. Kalpathy-Cramer and W. Hersh, "Multimodal medical image retrieval: image categorization to improve search precision," in *Proceedings of the international conference on Multimedia information retrieval*, 165–174, ACM (2010).

26 Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22Nd ACM International Conference on Multimedia*, *MM '14*, 675–678, ACM, (New York, NY, USA) (2014).

27 H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a very large-scale radiology database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1090–1099 (2015).

28 A. Mourão, F. Martins, and J. Magalhães, "Multimodal medical information retrieval with unsupervised rank fusion," *Computerized Medical Imaging and Graphics* **39**, 35–45 (2015).

29 C. E. Kahn Jr and C. Thao, "Goldminer: a radiology image search engine," *American Journal of Roentgenology* **188**(6), 1475–1478 (2007).

30 H. Müller, P. Clough, T. Deselaers, and B. Caputo, *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*, Springer Publishing Company, Incorporated, 1st ed. (2010).

31 A. G. S. de Herrera, R. Schaer, D. Markonis, and H. Müller, "Comparing fusion techniques for the imageclef 2013 medical case retrieval task," *Computerized Medical Imaging and Graphics* **39**, 46–54 (2015).

32 A. G. S. de Herrera, J. Kalpathy-Cramer, D. D. Fushman, S. Antani, and H. Müller, "Overview of the imageclef 2013 medical tasks," *Working notes of CLEF* **2013**, 1–15 (2013).

33 O. Ozturkmenoglu, N. M. Ceylan, and A. Alpkocak, “Demir at imageclefmed 2013: The effects of modality classification to information retrieval,” *Working Notes of CLEF* (2013).

34 S. A. Chatzichristofis and Y. S. Boutalis, “Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval,” in *Computer vision systems*, 312–322, Springer (2008).

35 C. Beecks, M. S. Uysal, and T. Seidl, “A comparative study of similarity measures for content-based multimedia retrieval,” in *Proceedings of theIEEE International Conference on Multimedia and Expo*, 1552–1557 (2010).

36 C. Beecks and T. Seidl, “On stability of adaptive similarity measures for content-based image retrieval,” in *Proceedings of the 18th International Conference on Multimedia Modeling*, **7131**, 346–357 (2012).

37 C. Beecks, S. Kirchhoff, and T. Seidl, “On stability of signature-based similarity measures for content-based image retrieval,” *Multimedia Tools Appl.* **71**(1), 349–362 (2014).

38 C. Beecks, M. S. Uysal, J. Hermanns, and T. Seidl, “Gradient-based signatures for efficient similarity search in large-scale multimedia databases,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 1241–1250 (2015).

39 C. Beecks, M. S. Uysal, and T. Seidl, “Gradient-based signatures for big multimedia data,” in *IEEE International Conference on Big Data, Big Data*, 2834–2835 (2015).

40 M. S. Uysal, C. Beecks, J. Schmücking, and T. Seidl, “Efficient filter approximation using the earth mover’s distance in very large multimedia databases with feature signatures,” in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 979–988 (2014).

41 M. S. Uysal, C. Beecks, J. Schmücking, and T. Seidl, “Efficient similarity search in scientific databases with feature signatures,” in *Proceedings of the 27th International Conference on Scientific and Statistical Database Management*, 30:1–30:12 (2015).

42 C. Beecks, M. Hassani, J. Hinnell, D. Schüller, B. Brenger, I. Mittelberg, and T. Seidl, “Spatiotemporal similarity search in 3d motion capture gesture streams,” in *Proceedings of 14th International Symposium on Spatial and Temporal Databases*, *Lecture Notes in Computer Science* **9239**, 355–372, Springer (2015).

43 C. Beecks, K. Schoeffmann, M. Lux, M. S. Uysal, and T. Seidl, “Endoscopic video retrieval: A signature-based approach for linking endoscopic images with video segments,” in *IEEE International Symposium on Multimedia*, (2015).

44 J. Roldan Carlos, M. Lux, X. Giro-i Nieto, P. Munoz, and N. Anagnostopoulos, “Visual information retrieval in endoscopic video archives,” in *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on*, 1–6 (2015).

45 M. A. Hudelist, C. Cobârzan, C. Beecks, R. Werken, S. Kletz, W. Hürst, and K. Schoeffmann, *MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part II*, ch. Collaborative Video Search Combining Video Retrieval with Human-Based Visual Inspection, 400–405. Springer International Publishing, Cham (2016).