Endoscopic Video Retrieval: A Signature-based Approach for Linking Endoscopic Images with Video Segments

Christian Beecks[1], Klaus Schoeffmann[2], Mathias Lux[2], Merih Seran Uysal[1] and Thomas Seidl[1]

***Abstract*— In the field of medical endoscopy more and more surgeons are changing over to record and store videos of their endoscopic procedures, such as surgeries and examinations, in long-term video archives. In order to support surgeons in accessing these endoscopic video archives in a content-based way, we propose a simple yet effective signature-based approach: the *Signature Matching Distance* based on *adaptive-binning feature signatures*. The proposed distance-based similarity model facilitates an adaptive representation of the visual properties of endoscopic images and allows for matching these properties efficiently. We conduct an extensive performance analysis with respect to the task of linking specific endoscopic images with video segments and show the high efficacy of our approach. We are able to link more than 88% of the endoscopic images to their corresponding correct video segments, which improves the current state of the art by one order of magnitude.**

***Keywords*— Endoscopic Video Retrieval; Endoscopic Images; Feature Signatures; Signature Matching Distance;**

I. INTRODUCTION

In the field of medical endoscopy more and more surgeons are changing over to record and store videos of their endoscopic procedures, such as surgeries and examinations, in long-term video archives. The recorded *endoscopic videos* are used later (i) as valuable source of information for follow-up procedures, (ii) to give information about the procedure to the patients, and (iii) to train young surgeons and teaching of new operation techniques. Sometimes these videos are also used for manual inspection and assessment of the technical skills of surgeons, with the ultimate goal of improving surgery quality over time [1], [2].

Although some surgeons record the entire procedure as video, for example in the Netherlands where it is enforced by law, many surgeons frequently record only the most important video segments. During the course of the procedure surgeons additionally capture static images that show specific situations or operation phases, which may be important for later inspection (or for showing them to the patient). These static images are frequently insufficient for further explanations and conclusions, thus surgeons may need to access video files based on specific static images. Unfortunately, this is a very awkward and time-consuming process when performed manually by the surgeons, since in practice there is currently no link between a static image and the corresponding exact position in a video. Though this is merely an organizational issue the time to market in this field is high due to the extensive quality assurance tests and the prohibitive costs of medical computing equipment. We don't expect an implementation of a meta data scheme and its application to solve this problem to be widely employed in at least the next five years. In that sense all videos taken up to now and until then need to be processed and indexed in the proposed way.

One way to support surgeons in accessing endoscopic video archives in a content-based way, i.e. in searching for a specific frame in an endoscopic video, is to automatically segment the video [3], remove irrelevant content [4], extract diverse keyframes [5], and provide an interactive browsing tool, e.g. with hierarchical refinement [6].

The problem of linking static images with the corresponding segments in the endoscopic videos is addressed by Roldan Carlos et al. [7] by means of standard techniques from content-based image retrieval. In their work, the authors have investigated the performance with respect to local and global image descriptors and have shown that they are able to correctly link more than 79% of the static images by using the local SIMPLE [8] descriptor.

In this paper, we propose to address the issue of linking static images with endoscopic video segments by means of *adaptive-binning feature signatures* [9], [10] and the *Signature Matching Distance* [11]. Our approach individually aggregates and matches the visual characteristics of static images and endoscopic video frames and is allows for different matching strategies. By making use of a local, efficiently computable 7-dimensional feature descriptor reflecting the spatial color and texture distribution of the images, we achieve an improvement in performance of up to 88.2% of correctly linked images.

Our main contributions are summarized as follows:

- We propose an adaptive-binning feature signature model for the purpose of content-based access into endoscopic video archives.
- We present different variants of the Signature Matching Distance and investigate their strengths and weaknesses.
- We conduct a performance analysis on the recently introduced endoscopic video database [7] indicating the high efficacy of our approach.

The remainder of this paper is structured as follows: Section II outlines related work with respect to medical image retrieval. Our approach is proposed in Section III. The results of the performance analysis are given in Section IV,

[1]Christian Beecks, Merih Seran Uysal and Thomas Seidl are with the Data Management and Exploration Group, RWTH Aachen University, Germany {beecks,uysal,seidl}@cs.rwth-aachen.de

[2]Klaus Schoeffmann and Mathias Lux are with the Institute of Information Technology, Alpen-Adria-Universität Klagenfurt, Austria {ks,mlux}@itec.aau.at

Preprint of the paper to appear in the IEEE Proceedings of the International Symposium on Multimedia 2015 (ISM 2015), Miami, Florida, USA, Dec. 2015

while we conclude this paper with an outlook on future work in Section V.

II. Related Work

Visual retrieval in medical data can be applied to medical images as well as videos. One focus in medical imaging lies on gray scale images such as X-rays or magnetic resonance imaging (MRI). For instance in [12], potential applications of medical image retrieval are described and existing medical CBIR systems are reviewed. The authors of [13] introduce different types of medical images used in CBIR systems as well as a large variety of techniques, potential applications and future research directions. In [14] a more recent review emphasizing the multi-dimensional (2D and 3D) and multi-modality nature of the medical retrieval scenario is provided. In [7], the problem of linking static images with endoscopic video segments is solved with feature fusion of the three global descriptors CEDD [15], [16], color correlograms, and PHOG [17]. In addition, the authors propose to utilize the local SIMPLE [8] descriptor, which turns out to outperform the aforementioned global descriptors.

There is a wide variety of approaches in medical image retrieval ranging from low-level wavelet-based visual signatures [18] to high level concept detectors [19]. Moreover, computer vision algorithms approaches like [20] create textual descriptions and labels and use these labels to support text-based information retrieval. Besides automatic content-based methods, also interactive approaches have been published. In [5] the authors select frames as summary of endoscopic videos. While this does not offer a solution to the retrieval problem, it tries to drastically reduce the visual information in order to allow for fast assessment by surgeons and experts without watching the video as a whole.

Visual retrieval in medical image and video databases still remains a challenging research area. The ImageCLEF benchmarking initiative [21], for instance, is a platform for comparing and discussing different approaches for retrieval of medical images. A task for image-based retrieval in medical scenarios was organized from 2004 to 2013. For this task each medical case was represented by a text and 1-7 images accompanying the text. All in all in 2013 the best textual run achieved the same performance as the best technique using both textual and visual features [22], [23]. As in previous years and other scenarios, visual-only approaches achieved significantly lower results than textual and multi-modal methods. However, it is worth noting that the best visual-only approach as presented in [24] was based on global image features. The authors employed the Color and Edge Directivity Descriptor CEDD [16], a fuzzy color and texture histogram and a color layout descriptor.

Currently, medical retrieval systems try to become much more accessible on the web, typically being multi-modal by supporting both textual and visual queries. Examples for web-based search engines are NovaMedSearch [25] and GoldMiner [26].

Going beyond still images and frames the authors of [27] propose a framework that uses principal video shots for video content representation and feature extraction. The classification is based on elementary semantic medical concepts, such as “Traumatic surgery” or “Diagonosis”. Another example for medical video retrieval is presented in [28]. There, a framework to retrieve short videos in real-time by modeling the motion content with a polynomial model is outlined.

One main issue that has not yet been adressed in visual retrieval of endoscopic images to the best of our knowledge is that there are primary colors in the videos taken inside human bodies, ie. mostly tones of red and yellow for tissue and fat and to a smaller extent tones of gray, blue and green for medical equipment. Therefore, we deem our adaptive-binning feature signature model fit for this tasks, especially as it adapts to this color and texture distribution peculiarties of this use case and hypothesize that it leads to better results than approaches not adaptive to the color characteristics, ie. with fixed quatization tables.

III. Signature-Based Approach to Endoscopic Image Retrieval

In this section, we first describe the feature signature representation of endoscopic images and we then present our signature-based approach in detail.

A. Representing Endoscopic Images via Feature Signatures

The visual properties of endoscopic images are frequently described by means of features $f_1, \ldots, f_k \in \mathbb{F}$ in a feature space $(\mathbb{F}, \delta)$ that is additionally endowed with a function $\delta : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}$ for comparing two individual features. By indicating the relevance of each feature, an endoscopic image is mathematically modeled by a *feature representation* $X : \mathbb{F} \rightarrow \mathbb{R}$. Thus, a feature representation X is a function which relates each feature $f \in \mathbb{F}$ with a *weight* $X(f) \in \mathbb{R}$, where the weight of zero is designated for features which are not relevant for the corresponding endoscopic image. Features with a weight unequal to zero are denoted as *representatives* and are defined for a feature representation X by the set $\mathrm{R}_X = \{f \in \mathbb{F} | X(f) \neq 0\} \subseteq \mathbb{F}$. Restricting a feature representation X to a finite set of representatives $\mathrm{R}_X \subseteq \mathbb{F}$ yields a *feature signature*, whose formal definition is given below.

Definition 1: (Feature signature)
Let $(\mathbb{F}, \delta)$ be a feature space. A *feature signature* $X \in \mathbb{R}^{\mathbb{F}}$ is defined as:

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

A feature signature X defines a finite set of contributing features, i.e. features with a weight unequal to zero, individually for each endoscopic image. Frequently, the representatives are obtained by clustering the extracted features of an endoscopic image and taking the cluster centroids as representatives. In this way, the feature representations of endoscopic images comply with those of generic multimedia objects [9] and can be further adapted to individual perceptual similarity by changing the underlying features or the aggregation process.

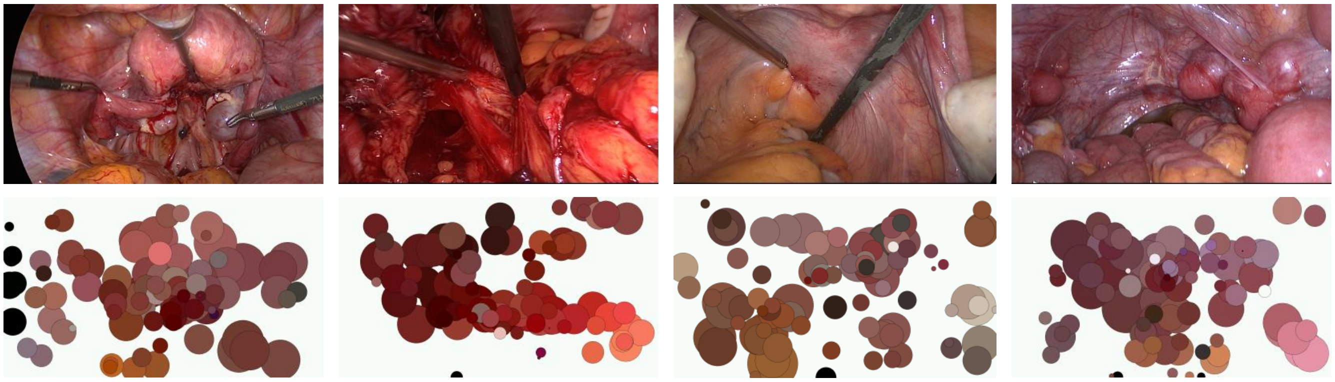

Fig. 1. Four example endoscopic images and the corresponding feature signatures.

In order to illustrate the idea of a feature signature, we depict four examples of endoscopic images and their corresponding feature signatures in Figure 1, where the representatives of the feature signatures are visualized by colored circles with diameters indicating their weights. The underlying features are based on position, color, and texture information. As can be seen in this example, feature signatures are able to visually approximate the content of endoscopic images by utilizing individual representatives. More theoretical details regarding this feature representation model can be found for instance in [9].

B. Signature Matching Distance

Defining the similarity between two endoscopic images based on their feature signatures is a challenging task. In fact, there exists a rich amount of distance-based similarity measures applicable to feature signatures [9], [29], [30], [10], [31], [32] with their individual strengths and weaknesses. Well-known measures are the transformation-based *Earth Mover's Distance* [10], the correlation-based *Signature Quadratic Form Distance* [33], the matching-based *Hausdorff Distance* [34] and its variants [35], [36] as well as the *Signature Matching Distance* [11].

In this work, we focus on the Signature Matching Distance for the following reasons: (i) its matching-based character facilitates partial matching which is particular useful in the domain of endoscopic images due to their high degree of homogeneity, (ii) it enables an asymmetric matching-based definition of similarity which can be used in combination with query feature signatures differing in size and structure, and (iii) it has been shown to outperform other signature-based distance functions in terms of accuracy and efficiency [11].

The idea of the Signature Matching Distance is to match the representatives of two feature signatures with respect to their visual characteristics which are evaluated by means of a ground distance function δ within the feature space $(\mathbb{F}, \delta)$. Possible ground distance functions include for instance the family of Minkowski Distances that is defined with respect to parameter $p \in \mathbb{R}^+$ between two features $f, g \in \mathbb{F} = \mathbb{R}^d$ from a d-dimensional Euclidean feature space $\mathbb{R}^d$ as $\mathrm{L}_p(f,g) = \left(\sum_{i=1}^{d} |f[i] - g[i]|^p\right)^{\frac{1}{p}}$. Based on a ground distance function $\delta : \mathbb{F} \times \mathbb{F} \to \mathbb{R}$, similar representatives between two feature signatures $X, Y \in \mathbb{S}$ are matched according to the principle of the δ*-Nearest-Neighbor Matching* $\mathrm{m}_{X \to Y}^{\delta\text{-NN}} \subseteq \mathbb{F} \times \mathbb{F}$ which is defined as

$$\mathrm{m}_{X \to Y}^{\delta\text{-NN}} = \{(f,g) | X(f) > 0 \wedge Y(g) > 0 \wedge g = \mathrm{argmin}_{h \in \mathbb{F}} \delta(f,h)\}.$$

The δ-Nearest-Neighbor Matching $\mathrm{m}_{X \to Y}^{\delta\text{-NN}}$ between two feature signatures X and Y matches each representative $f \in \mathrm{R}_X$ to one or more representatives $g \in \mathrm{R}_Y$. The size of the δ-Nearest-Neighbor Matching is thus restricted by the number of representatives of both feature signatures, i.e. it holds that $|\mathrm{m}_{X \to Y}^{\delta\text{-NN}}| \leq |\{X(f) > 0\}_{f \in \mathbb{F}}| \cdot |\{Y(g) > 0\}_{g \in \mathbb{F}}|$. The δ-Nearest-Neighbor Matching is the most intuitive and at the same time naive way of matching two feature signatures and provides the possibility of distance-based indexing for instance through *spatial access methods*, *metric access methods* or even *ptolemaic access methods*.

While the matching strategy determines the way of how the representatives between two feature signatures are assigned to each other, it does not define the influence of the matching representatives to the distance value. This is modeled by a cost function $\mathrm{c} : 2^{\mathbb{F} \times \mathbb{F}} \to \mathbb{R}$. In general, a cost function takes into account the differences of matching representatives and assigns each matching a single figure value. In this work, we utilize the following distance-based cost function with respect to a δ-Nearest-Neighbor Matching $\mathrm{m}_{X \to Y}^{\delta\text{-NN}}$ between two feature signatures X and Y:

$$\mathrm{c}_\delta(\mathrm{m}_{X \to Y}^{\delta\text{-NN}}) = \sum_{(f,g) \in \mathrm{m}_{X \to Y}^{\delta\text{-NN}}} X(f) \cdot \delta(f,g).$$

Based the δ-Nearest-Neighbor Matching between two feature signatures and the cost function defined above, we propose to utilize the following three simple yet effective variants of the Signature Matching Distance between a query feature signature $Q \in \mathbb{R}^{\mathbb{F}}$ and a database feature signature $O \in \mathbb{R}^{\mathbb{F}}$:

- *bidirectional variant*:

$$\mathrm{SMD}_\delta^{\leftrightarrow}(Q,O) = \mathrm{c}_\delta(\mathrm{m}_{Q \to O}^{\delta\text{-NN}}) + \mathrm{c}_\delta(\mathrm{m}_{O \to Q}^{\delta\text{-NN}})$$

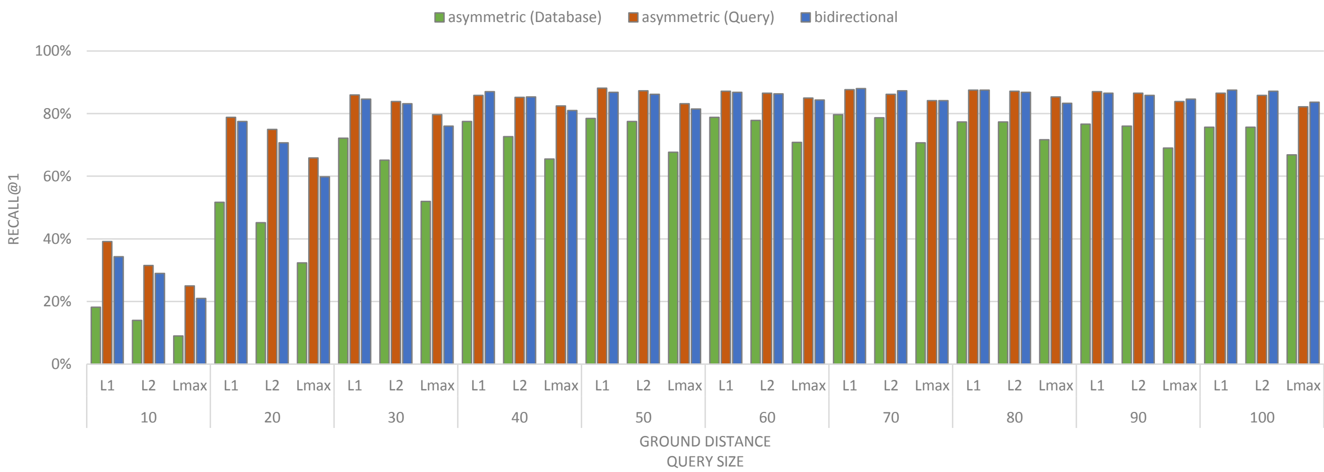

Fig. 2. Recall@1 values in percentage for the bidirectional variant of the Signature Matching Distance $\text{SMD}_\delta^{\leftrightarrow}$ and its asymmetric variants $\text{SMD}_\delta^{\rightarrow}$ and $\text{SMD}_\delta^{\leftarrow}$ as a function of the ground distance $\delta \in \{\text{L}_p\}_{p=1,2,\infty}$ and query signature size between 10 and 100.

- *asymmetric query variant*:

$$\text{SMD}_\delta^{\rightarrow}(Q, O) = \text{c}_\delta(\text{m}_{Q\rightarrow O}^{\delta\text{-NN}})$$

- *asymmetric database variant*:

$$\text{SMD}_\delta^{\leftarrow}(Q, O) = \text{c}_\delta(\text{m}_{O\rightarrow Q}^{\delta\text{-NN}})$$

While the bidirectional variant $\text{SMD}_\delta^{\leftrightarrow}(Q, O)$ takes into consideration the costs of both matching directions, i.e. $\text{c}_\delta(\text{m}_{Q\rightarrow O}^{\delta\text{-NN}})$ and $\text{c}_\delta(\text{m}_{O\rightarrow Q}^{\delta\text{-NN}})$, the asymmetric variants $\text{SMD}_\delta^{\rightarrow}(Q, O)$ and $\text{SMD}_\delta^{\leftarrow}(Q, O)$ are restricted to only one of the directions. This corresponds to the idea of partially matching representatives from feature signatures with different cardinalities and is especially useful when varying the feature signature size on the query-side.

How the proposed approaches compete with the state of the art provided by Roldan Carlos et al. [7] in the context of endoscopic video retrieval is empirically evaluated in the following section.

IV. PERFORMANCE ANALYSIS

In order to evaluate the retrieval performance of the proposed signature-based approach in the context of endoscopic video retrieval, we have selected video content from about 33 hours of anonymized laparoscopic procedures. More precisely, these videos were recorded from 48 procedures (i.e., different patients) in full HD quality (1920x1080@25p). The procedures were not entirely stored as videos, instead each surgeon stored about 25 segments per procedure in average, resulting in a total number of 1,276 video files and about three millions of video frames. The video files are of short duration; 70% of files contain video segments of less than 120 seconds, no segment is longer than 250 seconds.

Based on these endoscopic videos, we extracted feature signatures with a rate of five frames per second, i.e. we gathered every fifth frame, since the videos use a frame rate of 25 fps, resulting in an endoscopic image database of 593,446 frames. This database is the same as the one utilized in [7]. The feature signatures were computed based on a random sampling of 40,000 pixels by first extracting local feature descriptors describing the relative spatial information of a pixel, its CIELAB color value, and its coarseness and contrast values as described in [11] and then clustering the extracted feature descriptors by the k-means algorithm. In this way, we obtained database feature signatures with up to 40 representatives over a 7-dimensional feature space $\mathbb{F} = \mathbb{R}^7$ for each single frame. We implemented all signature-based approaches in Java 1.8 and conducted the experiments on a single-core 3.4 GHz machine equipped with 16 GB of main memory.

As already mentioned in Section I, surgeons regularly capture additional static images of important positions/situations in the procedures for later explanations and investigations. We selected 600 of those captured static images (12.5 per procedure in average) in order to form our query images and extracted feature signatures with 10 up to 100 representatives as mentioned above. These query images are the same as those used in [7]. Based on the query images, the main evaluation objective consists in linking each query image with its corresponding video segment within the 1,276 endoscopic video files. It is worth noting that the correct frame number is not available in the recorded data and can thus not be used for evaluation purposes. In addition, it is actually not important in the scope of the targeted practical scenario of endoscopic video retrieval to find exactly the same frame but rather a video segment that can be interactively inspected in the surrounding area by a surgeon. Thus, we utilized the information of the video segments as ground truth and average Recall@1, Recall@2, and Recall@3 over all query images. In doing so, a returned frame is considered to be relevant with respect to a query image if both are from the same video segment.

TABLE I

COMPARISON TO THE STATE OF THE ART. THE RECALL VALUES ARE AVERAGED OVER 600 QUERIES AND REPORTED IN PERCENTAGE. THE QUERY SIGNATURE SIZE IS GIVEN WHERE APPROPRIATE.

Approach	Recall@1	Recall@2	Recall@3
(CEDD, col.c., PHOG) [7]	78.3%	81.8%	84.2%
(SIMPLE) [7]	79.8%	83.3%	84.7%
$\mathrm{SMD}^{\leftrightarrow}_{\mathrm{L}_1}$	88.0% (70)	**89.5%** (70)	**90.5%** (50)
$\mathrm{SMD}^{\leftrightarrow}_{\mathrm{L}_2}$	87.3% (70)	88.7% (70)	89.7% (70)
$\mathrm{SMD}^{\leftrightarrow}_{\mathrm{L}_\infty}$	84.7% (90)	87.0% (70)	88.3% (70)
$\mathrm{SMD}^{\rightarrow}_{\mathrm{L}_1}$	**88.2%** (50)	**89.5%** (70)	90.0% (70)
$\mathrm{SMD}^{\rightarrow}_{\mathrm{L}_2}$	87.3% (50)	89.3% (80)	89.7% (80)
$\mathrm{SMD}^{\rightarrow}_{\mathrm{L}_\infty}$	85.3% (80)	87.3% (60)	88.0% (70)
$\mathrm{SMD}^{\leftarrow}_{\mathrm{L}_1}$	79.7% (70)	82.5% (70)	84.2% (80)
$\mathrm{SMD}^{\leftarrow}_{\mathrm{L}_2}$	78.7% (70)	81.8% (70)	83.7% (70)
$\mathrm{SMD}^{\leftarrow}_{\mathrm{L}_\infty}$	71.7% (80)	75.8% (60)	78.8% (70)

A. Results

The resulting Recall@1 values in percentage for the proposed bidirectional variant of the Signature Matching Distance $\mathrm{SMD}^{\leftrightarrow}_{\delta}$ and its asymmetric variants $\mathrm{SMD}^{\rightarrow}_{\delta}$ and $\mathrm{SMD}^{\leftarrow}_{\delta}$ are summarized in Figure 2 as a function of the ground distance $\delta \in \{\mathrm{L}_p\}_{p=1,2,\infty}$ and the query signature size between 10 and 100 over the 600 aforementioned queries. As can be seen in the figure, the asymmetric database variant $\mathrm{SMD}^{\leftarrow}_{\delta}$ on average shows the lowest performance regardless of the query signature size. For a small query signature size below 40, the asymmetric query variant $\mathrm{SMD}^{\rightarrow}_{\delta}$ shows better performance than the bidirectional variant $\mathrm{SMD}^{\leftrightarrow}_{\delta}$. The highest Recall@1 value of 88.2% is achieved by means of the asymmetric query variant $\mathrm{SMD}^{\rightarrow}_{\delta}$ based on the Manhattan Distance L_1 and a query signature size of 50. With these parameters, our approach is able to obtain a Recall@2 value of 89% and a Recall@3 value of 89.5%. The highest Recall@2 value of 89.5% and Recall@3 value of 90.5% are obtained when utilizing the bidirectional variant $\mathrm{SMD}^{\leftrightarrow}_{\delta}$ with the Manhattan Distance L_1 and a query signature size of 70 and 50, respectively.

We finally compare the results of our approach with the state of the art as proposed in [7] in Table I. As can be seen in the table, the Signature Matching Distance achieves the highest performance values across all three recall levels. We thus conclude, that our proposal improves the state of the art in endoscopic video retrieval.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a signature-based approach for the purpose of content-based access to endoscopic video archives. To this end, we have shown how to model and compare endoscopic images by means of adaptive-binning feature signatures and the Signature Matching Distance. We investigated a bidirectional and two asymmetric variants of the latter and evaluated their performance with respect to the task of linking endoscopic images to video segments. The performance analysis reveals that our approach is able to outperform the state of the art and achieves more than 88% of correctly linked images.

As future work, we intend to develop and investigate indexing and query processing approaches in order to increase the efficiency of our proposal.

ACKNOWLEDGMENTS

This work is partially funded by DFG grant SE 1039/7-1.

REFERENCES

[1] 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*, pp. –, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1931720415001300

[2] 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*, vol. 100, no. 8, pp. 1080–1088, 2013. [Online]. Available: http://dx.doi.org/10.1002/bjs.9168

[3] 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*, June 2013, pp. 223–228.

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

[5] K. Schoeffmann, M. Del Fabro, T. Szkaliczki, L. Böszörmenyi, and J. Keckstein, "Keyframe extraction in endoscopic video," *Multimedia Tools and Applications*, pp. 1–20, 2014. [Online]. Available: http://dx.doi.org/10.1007/s11042-014-2224-7

[6] J. Lokoc, K. Schoeffmann, and M. del Fabro, "Dynamic hierarchical visualization of keyframes in endoscopic video," in *MultiMedia Modeling*, ser. Lecture Notes in Computer Science, X. He, S. Luo, D. Tao, C. Xu, J. Yang, and M. Hasan, Eds. Springer International Publishing, 2015, vol. 8936, pp. 291–294. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-14442-9_31

[7] 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*, June 2015, pp. 1–6.

[8] C. Iakovidou, N. Anagnostopoulos, Y. Boutalis, S. Chatzichristofis *et al.*, "Searching images with mpeg-7 (& mpeg-7-like) powered localized descriptors: the simple answer to effective content based image retrieval," in *Content-Based Multimedia Indexing (CBMI), 2014 12th International Workshop on*. IEEE, 2014, pp. 1–6.

[9] C. Beecks, "Distance-based similarity models for content-based multimedia retrieval," Ph.D. dissertation, RWTH Aachen University, 2013, available online: http://darwin.bth.rwth-aachen.de/opus3/volltexte/2013/4807/.

[10] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[11] C. Beecks, S. Kirchhoff, and T. Seidl, "Signature matching distance for content-based image retrieval," in *International Conference on Multimedia Retrieval, ICMR'13, Dallas, TX, USA, April 16-19, 2013*, R. Jain, B. Prabhakaran, M. Worring, J. R. Smith, and T. Chua, Eds. ACM, 2013, pp. 41–48. [Online]. Available: http://doi.acm.org/10.1145/2461466.2461474

[12] 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*, pp. 258–291, 2006.

[13] 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*, vol. 73, no. 1, pp. 1–23, 2004.

[14] 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*, vol. 26, no. 6, pp. 1025–1039, 2013.

[15] 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.
[16] 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*. Springer, 2008, pp. 312–322.
[17] M. Lux, "Lire: open source image retrieval in java," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 843–846.
[18] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Medical image analysis*, vol. 14, no. 2, pp. 227–241, 2010.
[19] 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*, vol. 15, no. 4, pp. 640–646, 2011.
[20] 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*. ACM, 2010, pp. 165–174.
[21] H. Müller, P. Clough, T. Deselaers, and B. Caputo, *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*, 1st ed. Springer Publishing Company, Incorporated, 2010.
[22] 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*, vol. 39, pp. 46–54, 2015.
[23] 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*, vol. 2013, pp. 1–15, 2013.
[24] 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.
[25] A. Mourão, F. Martins, and J. Magalhães, "Multimodal medical information retrieval with unsupervised rank fusion," *Computerized Medical Imaging and Graphics*, vol. 39, pp. 35–45, 2015.
[26] C. E. Kahn Jr and C. Thao, "Goldminer: a radiology image search engine," *American Journal of Roentgenology*, vol. 188, no. 6, pp. 1475–1478, 2007.
[27] J. Fan, H. Luo, and A. K. Elmagarmid, "Concept-oriented indexing of video databases: toward semantic sensitive retrieval and browsing," *Image Processing, IEEE Transactions on*, vol. 13, no. 7, pp. 974–992, 2004.
[28] G. Quellec, M. Lamard, G. Cazuguel, Z. Droueche, C. Roux, and B. Cochener, "Real-time retrieval of similar videos with application to computer-aided retinal surgery," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 4465–4468.
[29] C. Beecks, S. Kirchhoff, and T. Seidl, "On stability of signature-based similarity measures for content-based image retrieval," *Multimedia Tools and Applications*, pp. 1–14, 2013.
[30] C. Beecks, M. S. Uysal, and T. Seidl, "A comparative study of similarity measures for content-based multimedia retrieval," 2010, pp. 1552–1557.
[31] M. S. Uysal, C. Beecks, J. Schmücking, and T. Seidl, "Efficient Similarity Search in Scientific Databases with Feature Signatures," in *SSDBM*, 2015, pp. 30:1–30:12.
[32] 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 *CIKM*, 2014, pp. 979–988.
[33] C. Beecks, M. S. Uysal, and T. Seidl, "Signature quadratic form distance," 2010, pp. 438–445.
[34] F. Hausdorff, *Grundzüge der Mengenlehre*. Von Veit, 1914.
[35] D. P. Huttenlocher, G. A. Klanderman, and W. Rucklidge, "Comparing images using the hausdorff distance," vol. 15, no. 9, pp. 850–863, 1993.
[36] B. G. Park, K. M. Lee, and S. U. Lee, "Color-based image retrieval using perceptually modified hausdorff distance," *EURASIP Journal of Image and Video Processing*, vol. 2008, pp. 4:1–4:10, 2008.