Surgical Action Retrieval for Assisting Video Review of Laparoscopic Skill

ACM MM 2017
MultiEdTech Workshop
Mountain View, California, USA

Sabrina Kletz, Klaus Schoeffmann, Bernd Münzer, Manfred J. Primus
Alpen-Adria-University, Klagenfurt - Institute of Information Technology

Heinrich Husslein
Medical University Vienna – Department of Gynecology and Obstetrics

October 27, 2017

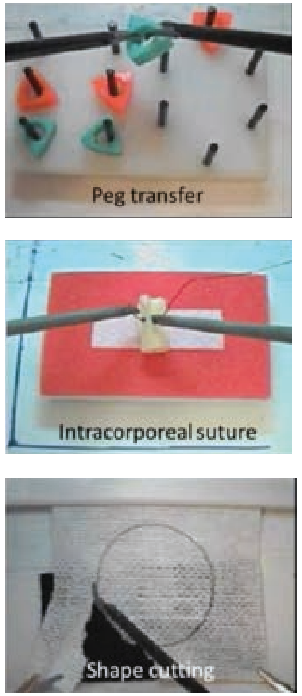
Bendet, S. (2005). Laparoscopic stomach surgery. Wikipedia Online. Available: https://upload.wikimedia.org/wikipedia/commons/c/c3/Laparoscopic_stomach_surgery.jpg

Behind the scene

Laparoscopic Surgery Training

[1]

[2]



[3]

[1] Ashraf, A., Collins, D., Whelan, M., O'Sullivan, R., & Balfe, P. (2015). Three-dimensional (3D) simulation versus two-dimensional (2D) enhances surgical skills acquisition in standardised laparoscopic tasks: A before and after study. *International Journal of Surgery*, *14*, 12–16.

[2] Zhou, M., Tse, S., Derevianko, A., Jones, D. B., Schwaitzberg, S. D., & Cao, C. G. L. (2012). Effect of haptic feedback in laparoscopic surgery skill acquisition. *Surgical Endoscopy and Other Interventional Techniques*, *26*(4), 1128–1134.

[3] Islam, G., Kahol, K., Li, B., Smith, M., & Patel, V. L. (2016). Affordable, web-based surgical skill training and evaluation tool. Journal of Biomedical Informatics, 59, 102–114.

Motivation

Assessment of Technical Skills

- Endoscopic video recordings are reviewed
- Global rating with **O**bjective-**S**tructured **A**ssessment of **T**echnical **S**kills (**OSATS**)
  - Overall performance of a surgeon, e.g. respect for tissue, time and motion, instrument handling
- Local rating with **G**eneric **E**rror **R**ating **T**ool (**GERT**)
  - Task-specific error rate, e.g. abdominal access, use of energy devices, grasping and dissection
- Surgical actions are the basis for assessments, e.g. grasping, suturing, cutting

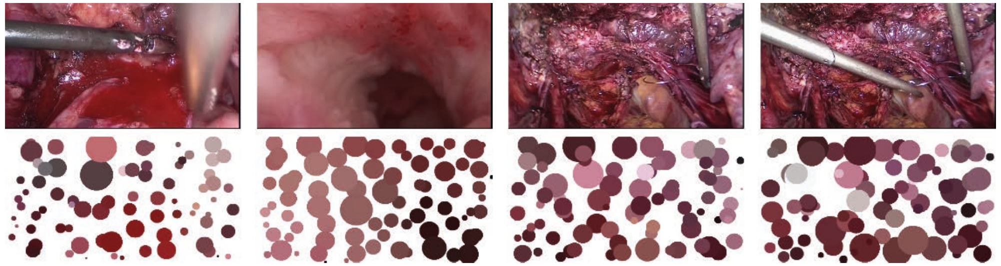
Related Work

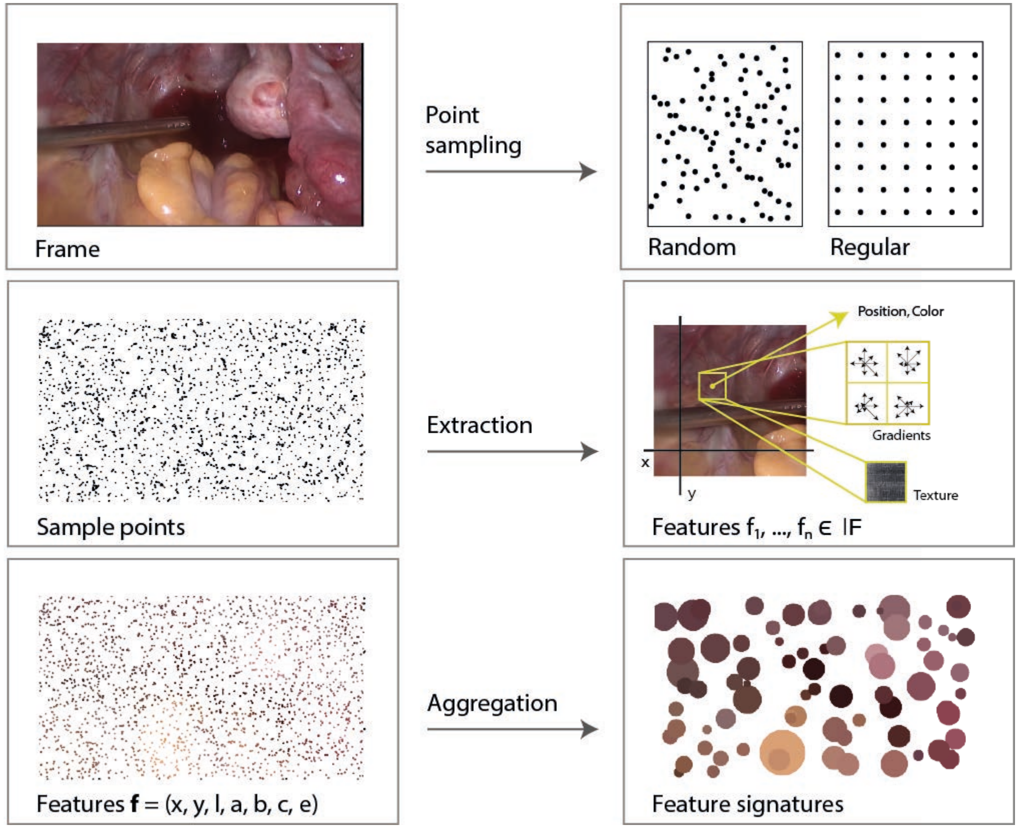
Feature Signatures

- Each point is described by a feature vector $f_i \in \mathbb{F}$ including color, texture, and position information
- Features $f_1, \ldots, f_n \in \mathbb{F}$ are aggregated using k-means
- Each cluster is described via 7-dimensional vector

  $\boldsymbol{f_i = (x, y, L, a, b, c, e)}$
- $S: \mathbb{F} \rightarrow \mathbb{R}$ subject to $|R_S| < \infty$, where $R_S = \{f \in \mathbb{F} \mid S(f) \neq 0\}$

Beecks, C., Schoeffmann, K., Lux, M., Uysal, M. S., & Seidl, T. (2015). Endoscopic Video Retrieval: A Signature-based Approach for Linking Endoscopic Images with Video Segments. In Proc. of the IEEE International Symposium on Multimedia, 1–6.





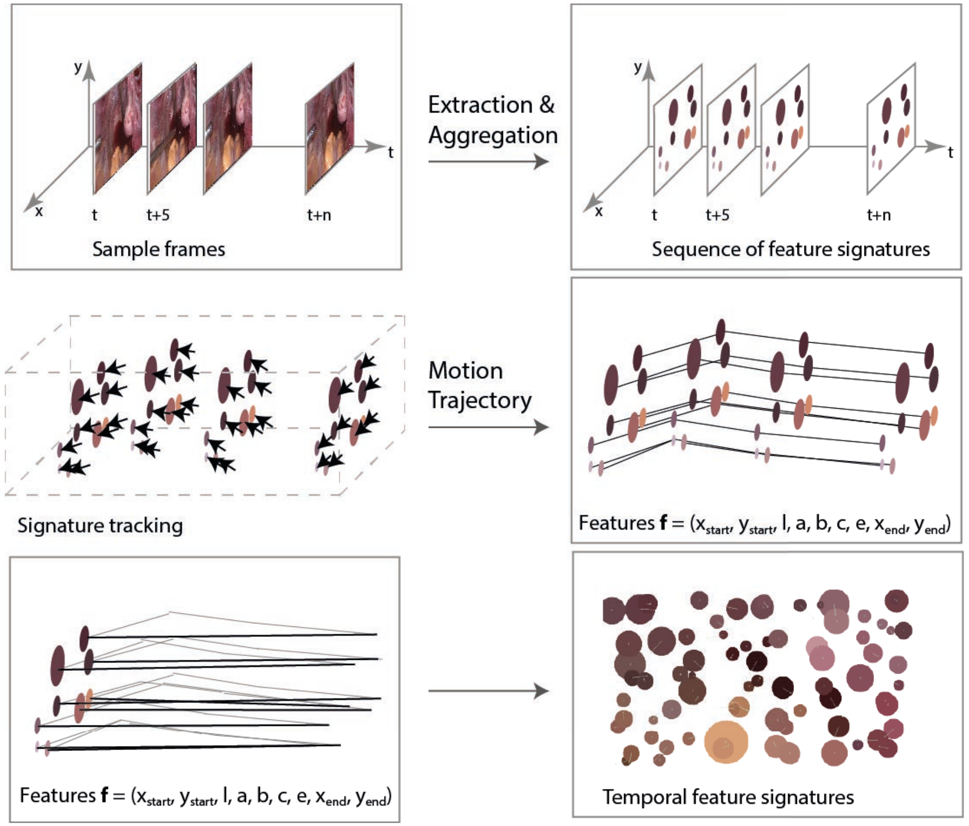
Proposed Approach

Dynamic-based Feature Signatures

Track-Based Feature Signatures

- Track each cluster and extend the vector by their end position
- For each cluster the most similar cluster is searched in the previous frame: Nearest Neighbor Matching
- Each displacement vector is summed up to its start position
- 9-dimensional vector

$$f_i = (x_{start}, y_{start}, L, a, b, c, e, x_{end}, y_{end})$$



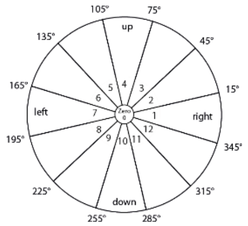
Proposed Approach

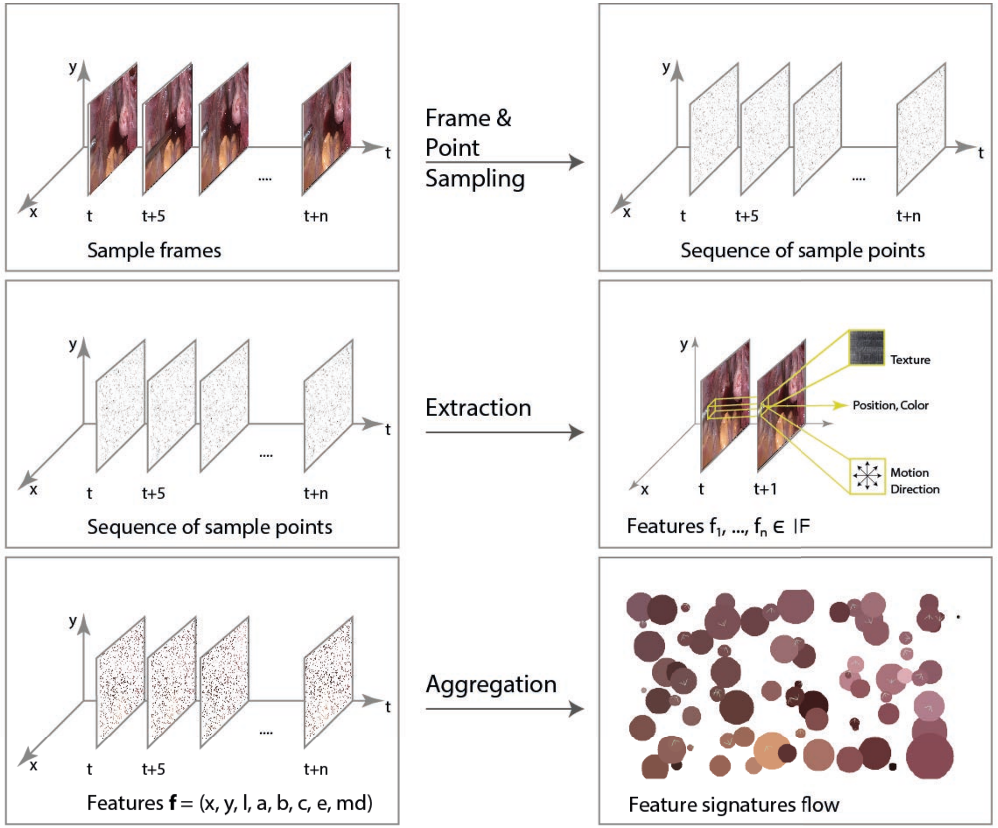
Dynamic-based Feature Signatures

Flow-based Feature Signatures

- Create clusters for adjacent frames and extend the initial vector by motion information
- Calculate the optical flow and determine the motion directions
- 8-dimensional vector

  f_i = (x, y, L, a, b, c, e, md)





Experiments

Dataset

- 18 hours of endoscopic video content
- 966 manually annotated video segments
- Classified into 11 different surgical actions for surgical skill assessment
- Retrieval Performance Measure: Mean Average Precision (MAP) for each surgical action and retrieval method

Surgical Action	Abdominal Access	Blunt dissection	Coagulation	Cutting (cold)	Cutting (thermal)	Endobag	Hysteroscopy	Injection	Morcellate	Rinse	Suture
Segments	108	53	177	103	74	16	50	57	28	189	111
Frames	5,400	69,689	85,656	208,210	139,007	45,264	173,917	61,726	159,493	95,423	552,649
Avg. Frames/Segment	50	1,315	484	2,021	1,878	2,829	3,478	1,083	5,696	505	4,979

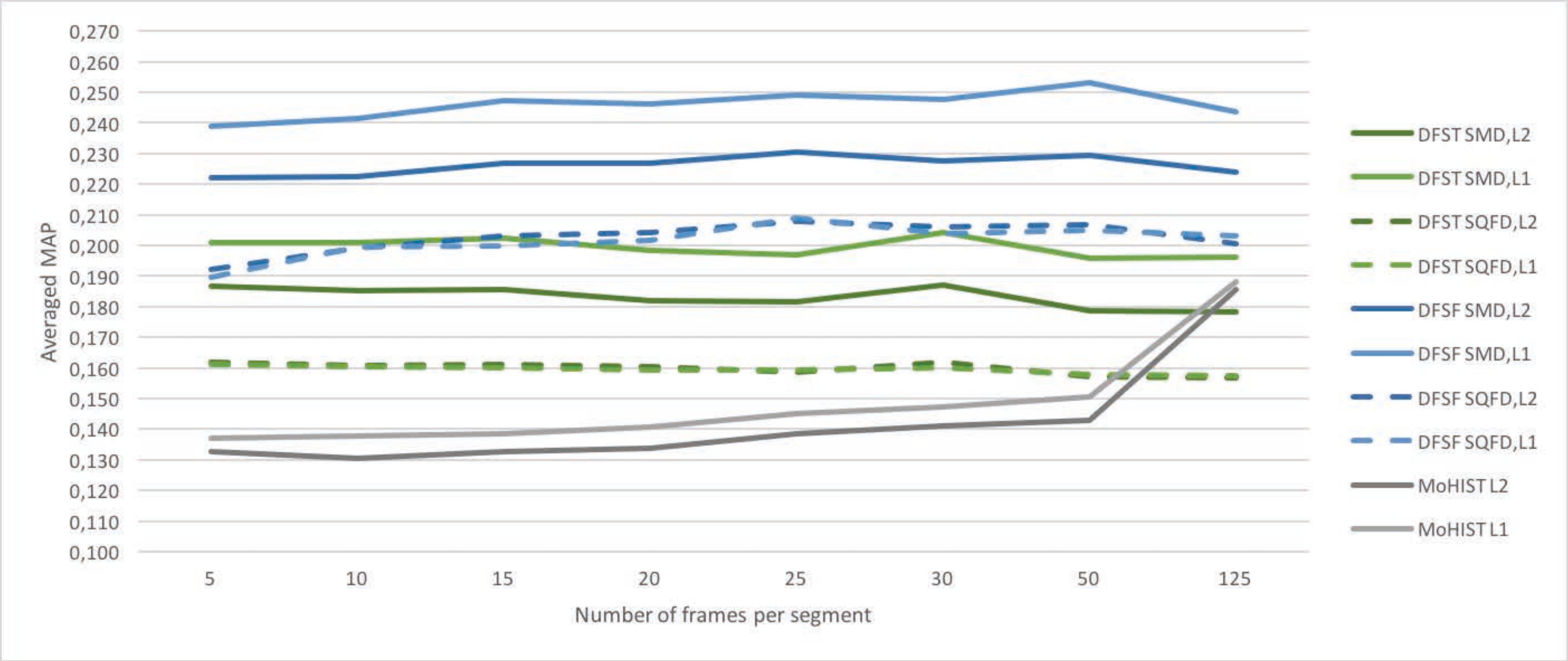
a) Abdominal Access b) Blunt dissection c) Coagulation d) Cutting (cold) e) Cutting (thermal) f) Endobag g) Hysteroscopy h) Injection i) Morcellate j) Rinse k) Suture

Experiments
Implementation Details

- Dynamic Feature Signatures
- Static Feature Signatures
- Motion Histogram[1]
- CNN Features (last fully-connected layer)
  - AlexNet (4096-dimensional feature vector)
  - EndoNet (fine-tuning of the AlexNet on a laparoscopic dataset)[2]
- Similarity Measures:
  - Signature Quadratic Form Distance
  - Signature Matching Distance
  - L1 and L2-Distance as Ground Distance

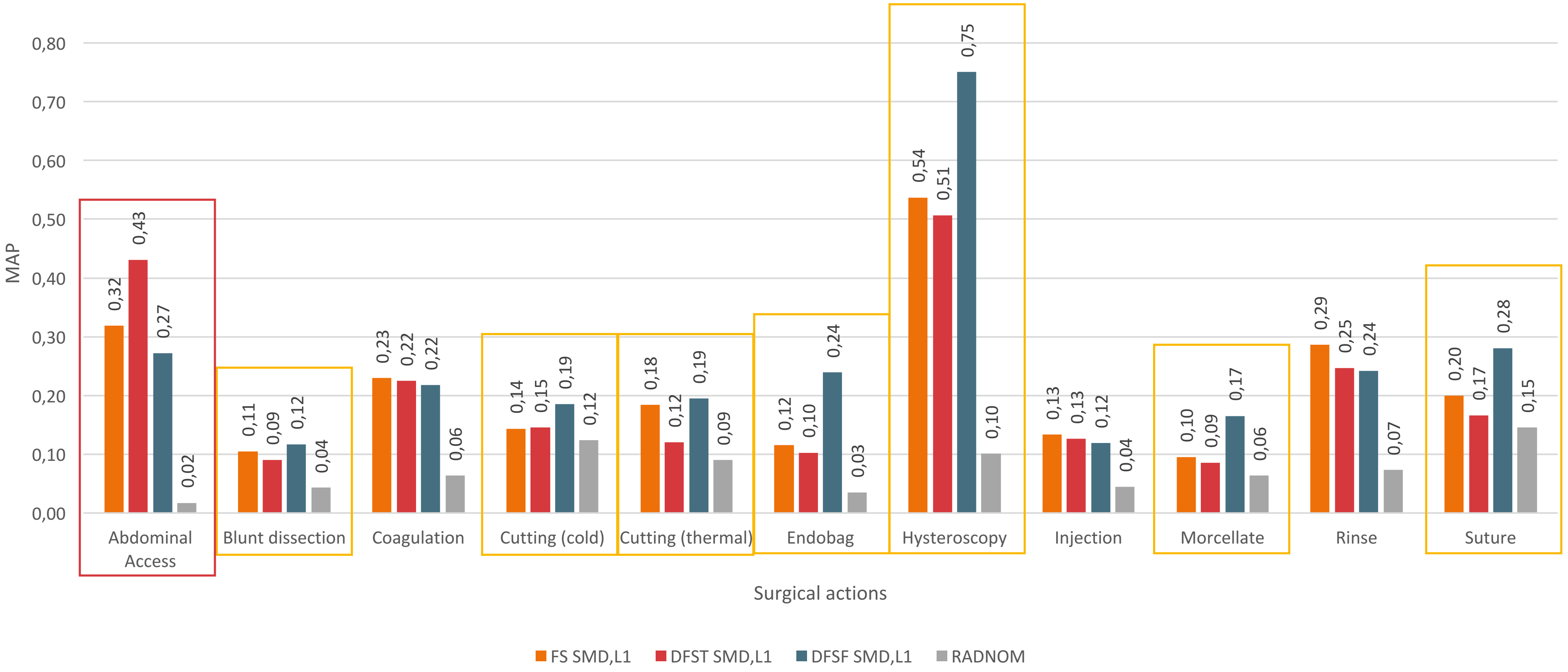
[1] Schoeffmann, K., Lux, M., Taschwer, M., & Boeszoermenyi, L. (2009). Visualization of Video Motion in Context of Video Browsing. In 2009 IEEE International Conference on Multimedia and Expo (pp. 658–661). IEEE.
[2] Petscharnig, S., & Schöffmann, K. (2017). Learning laparoscopic video shot classification for gynecological surgery. In MultiMedia Modeling: 23rd In. Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proc., Part I. Springer Int. Publishing, 702–713.

Experimental Setup
Optimal Frame Sampling



Retrieval Performance Analysis

Comparison of Signature-based Approaches



Retrieval Performance Analysis

Comparison of the retrieval methods

a) Abdominal Access b) Blunt dissection c) Coagulation d) Cutting (cold) e) Cutting (thermal) f) Endobag g) Hysteroscopy h) Injection i) Morcellate j) Rinse k) Suture

Descriptor	**FS**		**CNNA**	**CNNE**	**DFST$_{f=30}$**		**DFSF$_{f=50}$**		**MoHIST$_{f=125}$**
Surgical Action / Distance	SMD_{L1}	$SQFD_{L1}$	L2	L2	SMD_{L1}	$SQFD_{L2}$	SMD_{L1}	$SQFD_{L2}$	L1
Abdominal Access	0.319	0.243	0.430	0.154	0.431	0.235	0.272	0.252	0.390
Blunt dissection	0.105	0.118	0.113	0.098	0.090	0.104	0.117	0.098	0.105
Coagulation	0.230	0.228	0.237	0.267	0.225	0.248	0.218	0.216	0.203
Cutting (cold)	0.144	0.150	0.136	0.176	0.145	0.136	0.185	0.158	0.185
Cutting (thermal)	0.184	0.147	0.129	0.089	0.120	0.108	0.195	0.130	0.118
Endobag	0.115	0.090	0.101	0.080	0.103	0.105	0.239	0.139	0.116
Hysteroscopy	0.537	0.343	0.594	0.238	0.507	0.286	0.751	0.615	0.260
Injection	0.133	0.101	0.117	0.096	0.126	0.094	0.119	0.097	0.162
Morcellate	0.095	0.089	0.111	0.077	0.086	0.084	0.165	0.119	0.097
Rinse	0.286	0.257	0.279	0.286	0.247	0.225	0.242	0.231	0.206
Suture	0.200	0.162	0.187	0.153	0.166	0.155	0.281	0.219	0.228
Averaged MAP	**0.213**	**0.175**	**0.221**	**0.156**	**0.204**	**0.162**	**0.253**	**0.207**	**0.188**

Result Showcase

Conclusion

- Addressing the problem of **searching similar surgical actions using query-by-example segments**
- Investigation of **dynamic feature signatures**
- Manually **annotated dataset** containing typical **surgical actions**
- Comparison of **three state-of-the-art content descriptors** with **static and dynamic feature signatures**
- Dynamic feature signatures can be used for surgical action retrieval

Thank you! **Questions?**

sabrina@itec.aau.at