Surgical Action Retrieval for Assisting Video Review of Laparoscopic Skill

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
Behind the scene
Laparoscopic Surgery Training


Motivation
Assessment of Technical Skills

• Endoscopic video recordings are reviewed
• Global rating with Objective-Structured Assessment of Technical Skills (OSATS)
  • Overall performance of a surgeon, e.g. respect for tissue, time and motion, instrument handling
• Local rating with Generic Error Rating Tool (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 \( f_i = (x, y, L, a, b, c, e) \).
- \( S: \mathbb{F} \rightarrow \mathbb{R} \) subject to \( |R_S| < \infty \), where \( R_S = \{ f \in \mathbb{F} \mid S(f) \neq 0 \} \).

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

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


Experimental Setup
Optimal Frame Sampling
Retrieval Performance Analysis
Comparison of Signature-based Approaches
# Retrieval Performance Analysis

## Comparison of the retrieval methods

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>FS</th>
<th>CNNA</th>
<th>CNNE</th>
<th>DFS130</th>
<th>DFSF50</th>
<th>MoHISTF125</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical Action / Distance</td>
<td>SMD1</td>
<td>SQFD1</td>
<td>L2</td>
<td>SMD1</td>
<td>SQFD1</td>
<td>L1</td>
</tr>
<tr>
<td>Abdominal Access</td>
<td>0.319</td>
<td>0.243</td>
<td>0.430</td>
<td>0.154</td>
<td>0.431</td>
<td>0.235</td>
</tr>
<tr>
<td>Blunt dissection</td>
<td>0.105</td>
<td>0.118</td>
<td>0.113</td>
<td>0.098</td>
<td>0.090</td>
<td>0.104</td>
</tr>
<tr>
<td>Coagulation</td>
<td>0.230</td>
<td>0.228</td>
<td>0.237</td>
<td>0.267</td>
<td>0.225</td>
<td>0.248</td>
</tr>
<tr>
<td>Cutting (cold)</td>
<td>0.144</td>
<td>0.150</td>
<td>0.136</td>
<td>0.176</td>
<td>0.145</td>
<td>0.136</td>
</tr>
<tr>
<td>Cutting (thermal)</td>
<td>0.184</td>
<td>0.147</td>
<td>0.129</td>
<td>0.089</td>
<td>0.120</td>
<td>0.108</td>
</tr>
<tr>
<td>Endobag</td>
<td>0.115</td>
<td>0.090</td>
<td>0.101</td>
<td>0.080</td>
<td>0.103</td>
<td>0.105</td>
</tr>
<tr>
<td>Hysteroscopy</td>
<td>0.537</td>
<td>0.343</td>
<td>0.594</td>
<td>0.238</td>
<td>0.507</td>
<td>0.286</td>
</tr>
<tr>
<td>Injection</td>
<td>0.133</td>
<td>0.101</td>
<td>0.117</td>
<td>0.096</td>
<td>0.126</td>
<td>0.094</td>
</tr>
<tr>
<td>Morcellate</td>
<td>0.095</td>
<td>0.089</td>
<td>0.111</td>
<td>0.077</td>
<td>0.086</td>
<td>0.084</td>
</tr>
<tr>
<td>Rinse</td>
<td>0.286</td>
<td>0.257</td>
<td>0.279</td>
<td>0.286</td>
<td>0.247</td>
<td>0.225</td>
</tr>
<tr>
<td>Suture</td>
<td>0.200</td>
<td>0.162</td>
<td>0.187</td>
<td>0.153</td>
<td>0.166</td>
<td>0.155</td>
</tr>
<tr>
<td>Averaged MAP</td>
<td>0.213</td>
<td>0.175</td>
<td>0.221</td>
<td>0.156</td>
<td>0.204</td>
<td>0.162</td>
</tr>
</tbody>
</table>
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