Segmentation & Classification of Surgical Gestures

Divya Bhaskara, Lingling Tao, Effrosyni Mavroudi, René Vidal
The Vision Lab
Directed by Dr. René Vidal
Activity Recognition Group
What is Activity Recognition?
Activity Recognition in the Real-world
Surgical Applications
Machine Learning Approach to Activity Recognition

Kinematic time-series data $X$ → Model → Sequence of Labels $Y$
Spatial vs. Temporal Models of Activities

Jumping Jacks

Walking

Spatial Model

Temporal Model
My Research Goals

Spatial Model: Discriminative Dictionary Learning
Sefati et al. [1]

Temporal Model: Skip-Chain CRF
Lea et al. [2]

Spatio-temporal Model

Simple Classifier

Kinematic time-series data $X$ → Frame-wise Classifier $\omega$ → Sequence of Labels $Y$

Optimizing $\omega$: 66.07% Prediction Accuracy (Suturing)
Spatial Model (Discriminative Dictionary Learning) Sefati et al.

Kinematic time-series data \( X \) → Mid-level motion primitives \( \Psi \) \( \rightarrow \) \( Z = f(\Psi, X) \) → Frame-wise Classifier \( \omega \) → Sequence of Labels \( Y \)

Optimizing \((\Psi, \omega)\) then \( \omega \): 73.91% Prediction Accuracy (Suturing)
Temporal Model (Skip-Chain CRF) \textsuperscript{Lea et al.}

Kinematic time-series data $X$ → Frame-wise Classifier $w$ → Temporal Model $d$ → Sequence of Labels $Y$

$p = \begin{bmatrix}
    p_{1,1} & p_{1,2} & p_{1,3} & . & . \\
    p_{2,1} & p_{2,2} & . & . & . \\
    p_{3,1} & . & . & . & . \\
    . & . & . & . & . \\
    . & . & . & . & . \\
\end{bmatrix}$

Optimizing $(w, p)$ with $d = 1$:

74.55\% Prediction Accuracy (Suturing)
Spatio-temporal Model

Comparison of Results (Suturing)

<table>
<thead>
<tr>
<th>Simple Classifier</th>
<th>Spatial Model (Sefati et al.)</th>
<th>Temporal Model (Lea et al.)</th>
<th>Spatio-temporal model (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimizing $\psi$</td>
<td>Optimizing $(\psi, \omega)$, then $\omega$</td>
<td>Optimizing $(\omega, p)$ with $d = 1$</td>
<td>Optimizing $(\psi, \omega)$ then $(\omega, p)$ with $d = 1$</td>
</tr>
<tr>
<td>66.07%</td>
<td>73.91%</td>
<td>73.41%</td>
<td><strong>78.73 %</strong></td>
</tr>
</tbody>
</table>

Note: The table shows the comparison of results for different models and optimization strategies in suturing tasks. The best performing model is highlighted in blue.
Varying the Skip-chain Length $d$

- Small $d$: frame-level transitions
- Large $d$: gesture-level transitions

Pairwise matrices $\mathbf{p}$
# Results: Suturing LOUO

<table>
<thead>
<tr>
<th>Simple Classifier</th>
<th>Spatial Model (Sefati et al.)</th>
<th>Temporal Model (Lea et al.)</th>
<th>Spatio-temporal model (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimizing $\psi$</td>
<td>Optimizing $(\psi, w)$, then $w$</td>
<td>Optimizing $(w, p)$ with $d = 1$</td>
<td>Optimizing $(\psi, w)$, then $(w, p)$ with $d = 1$</td>
</tr>
<tr>
<td>66.07%</td>
<td>73.91%</td>
<td>73.41%</td>
<td>78.73%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>77.64%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>80.21%</td>
</tr>
</tbody>
</table>
More experiments...

Different metrics

Other experimental setups

**Accuracy**

**Edit Score**
Summary & Conclusions

Spatial Model: Discriminative Dictionary Learning
Sefati et al.

Temporal Model: Skip-Chain CRF
Lea et al.

Kinematic time-series data \( \mathbf{X} \) → Mid-level motion primitives \( \mathbf{\Psi} \) → Frame-wise Classifier \( \mathbf{W} \) → Temporal Model \( \mathbf{P} \) → Sequence of Labels \( \mathbf{Y} \)
Acknowledgements

Dr. René Vidal

Lingling Tao and Efi Mavroudi

Colin Lea and Shahin Sefati

NSF REU Program

Anita Sampath
Ongoing work: Joint learning

How can we optimize $\psi$, $\omega$, and $\rho$ together?

Kinematic time-series data $X$ → Mid-level motion primitives $\psi$ → Frame-wise Classifier $\omega$ → Temporal Model $\rho$ → Sequence of Labels $Y$