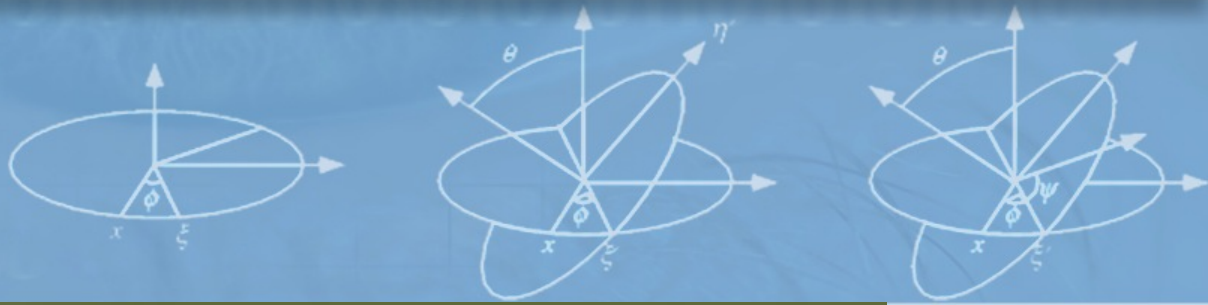
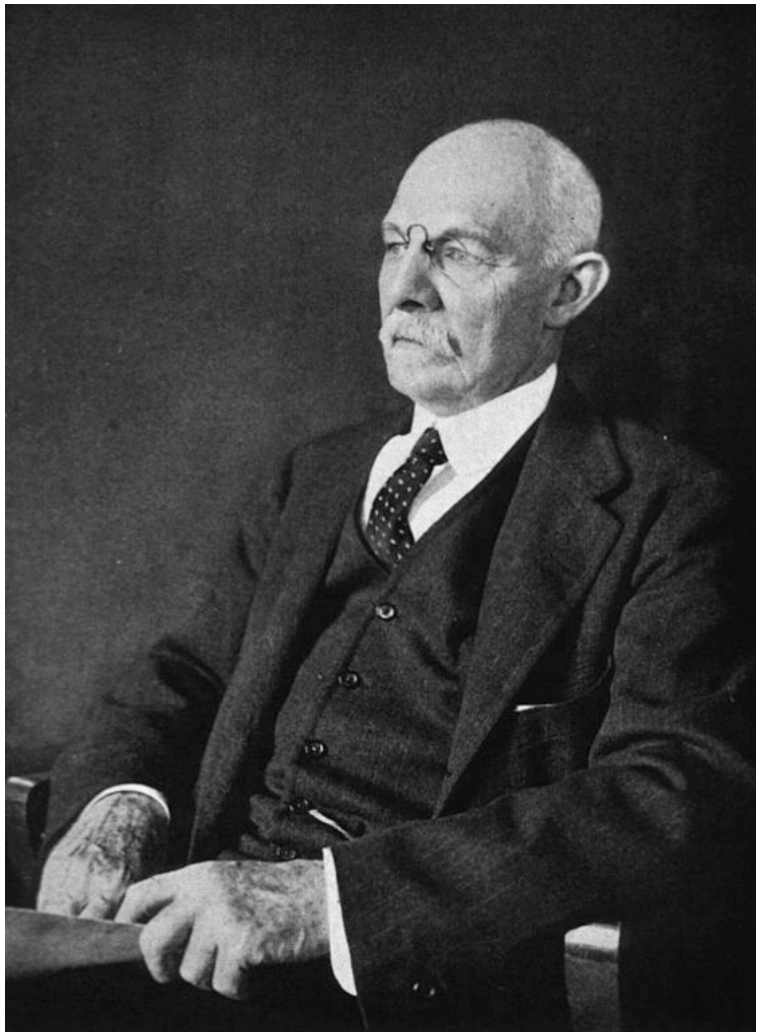
JHU vision lab

Discovering the Language of Surgery

René Vidal
Herschel L. Seder Professor of Biomedical Engineering
Director of the Mathematical Institute for Data Science
Johns Hopkins University



How Is Surgery Taught?



William S. Halsted, JHU 1889
"see one, do one, teach one"

C. Franzese and S. Stringer. The Evolution of Surgical Training: Perspectives on Educational Models from the Past to the Future. Otolaryngologic Clinics of North America, 40(6):1227–1235, 2007.

Medical Injuries

- US National Center for Health Statistics [1]
  - 40-45 million procedures/year were performed in 2000-2005.
- US Agency for Healthcare Research and Quality [2]
  - 20% of US hospitals in 2000.
  - 18 types of medical injuries.
  - 2.4 million extra days of hospitalization.
  - $9.3 billion in excess charges.
  - 32,591 attributable deaths.

Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries During Hospitalization

Chunliu Zhan, MD, PhD
Marlene R. Miller, MD, MSc

Context Although medical injuries are recognized as a major hazard in the health care system, little is known about their impact.

Objective To assess excess length of stay, charges, and deaths attributable to medical injuries during hospitalization.

Design, Setting, and Patients The Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators (PSIs) were used to identify medical injuries in 7.45 million hospital discharge abstracts from 994 acute-care hospitals across 28 states in 2000 in the AHRQ Healthcare Cost and Utilization Project Nationwide Inpatient Sample database.

Main Outcome Measures Length of stay, charges, and mortality that were recorded in hospital discharge abstracts and were attributable to medical injuries according to 18 PSIs.

Results Excess length of stay attributable to medical injuries ranged from 0 days for injury to a neonate to 10.89 days for postoperative sepsis, excess charges ranged from $0 for obstetric trauma (without vaginal instrumentation) to $57727 for postoperative sepsis, and excess mortality ranged from 0% for obstetric trauma to 21.96% for postoperative sepsis ($P<.001$). Following postoperative sepsis, the second most serious event was postoperative wound dehiscence, with 9.42 extra days in the hospital, $40323 in excess charges, and 9.63% attributable mortality. Infection due to medical care was associated with 9.58 extra days, $38656 in excess charges, and 4.31% attributable mortality.

Conclusion Some injuries incurred during hospitalization pose a significant threat to patients and costs to society, but the impact of such injury is highly variable.

JAMA. 2003;290:1868-1874 www.jama.com

DESPITE RECOGNITION OF medical injuries as a leading cause of death and patient safety as a critical area for improvement,[1-4] the overall approach to patient safety (eg, focusing on medical injuries)[5,6] and definitional issues (eg, what is considered preventable)[7,8] remains debated. Medical injuries can happen during all stages of the complicated process of care, vary widely in nature, and are relatively infrequent. The lack of standard taxonomy, in addition to definitional issues, in large part explains why so little is known about the prevalence, adverse outcomes, and effective prevention of medical injuries.[7,9-12]

The limited research on medical injuries has primarily relied on medical record abstraction conducted ad hoc and on a small scale.[2,13-17] Medical records contain rich clinical details that allow identification of various injuries and close calls and analysis of circumstances and causes. However, transforming medical records into useful research data on medical injuries is resource intensive and requires exceptional knowledge and skills in medical context and research methods. Alternative systems for research include mandatory and voluntary reports, drug safety surveillance, nosocomial infection surveillance, and medical malpractice data.[2,18] All of these data systems have limitations, and obtaining access for research purposes may be difficult. For example, approximately 20 US states mandate reporting of serious adverse events,[19] but no published study has ever used these data, most likely because they are strictly guarded from the public and researchers.

Administrative data are a potential source of information on medical injuries. Administrative data are regularly collected and maintained for reimbursement and management purposes; are computer readable, inexpensive to analyze, and longitudinal; and cover large populations. These data have been used to reveal startling small-

For editorial comment see p 1917.

Author Affiliations: Center for Quality Improvement and Patient Safety, Agency for Healthcare Research and Quality, Department of Health and Human Services, Rockville, Md (Dr Zhan); Department of Pediatrics, Johns Hopkins University, Baltimore, Md (Dr Miller).
Corresponding Author and Reprints: Chunliu Zhan, MD, PhD, Center for Quality Improvement and Patient Safety, Agency for Healthcare Research and Quality, 540 Gaither Rd, Rockville, MD 20850 (e-mail: czhan@ahrq.gov).

[1] C. DeFrances and M. Hall. 2005 national hospital discharge survey. Advance data from vital and health statistics, 385:1–19, 2007.
[2] C. Zhan and M. Miller. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA, 290(14):1868–1874, 2003.

Need for an Improvement

- Pressures from government and insurance companies to reduce the cost of deaths due to iatrogenic causes [1].
- Economic pressures on medical schools to reduce the costs of training surgeons [1].
- New labor laws that limit resident work hours [1, 2].
- According to [3], residents consider very helpful observing experts and practice (even with simulators).

Table 1: Types of laparoscopic instruction incorporated into surgical skills laboratories

Answer option	Response		
	Not helpful (%)	Very helpful (%)	N (%) of responses*
Observation of procedure by instructor	21	79	183 (91)
Viewing of instructional videos	33	67	166 (82)
Discussion of instrumentation and laparoscopic theory	21	79	164 (81)
Basic dissection techniques	13	87	174 (86)
Basic intracorporeal suturing techniques	3	97	195 (97)
Use of surgical simulators	11	89	176 (87)
Live animal wet labs	7	93	170 (84)

*Number who answered question = 202

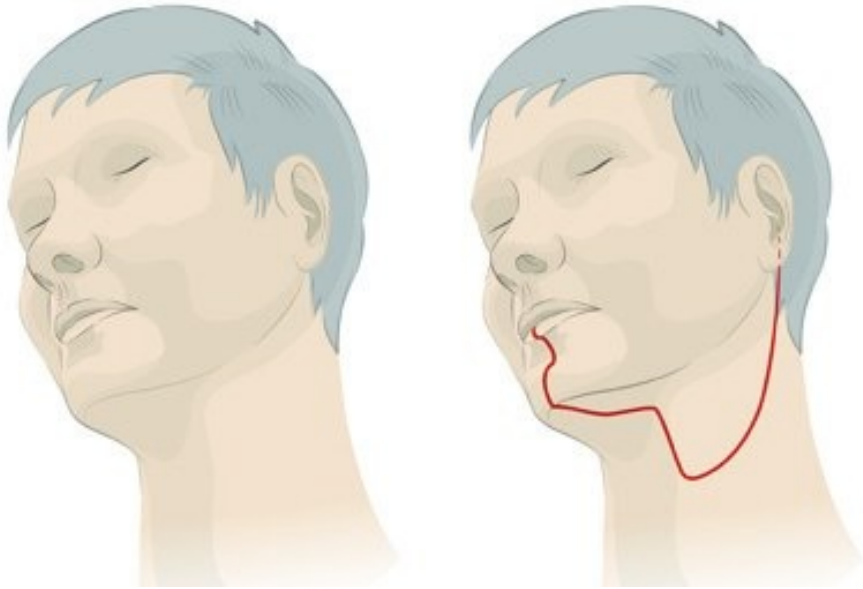
[1] Richard Reznick, "Teaching Surgical Skills – Changes in the Wind". NEJM 2006
[2] C. Barden, M. Specht, M. McCarter, J. Daly, and T. Fahey. Effects of limited work hours on surgical training. Obstetrical & Gynecological Survey, 58(4):244–245, 2003.
[3] Qureshi, Vergis, Jimenez, Green, Pryor, Schlachta, Okrainec. "A National Survey of General Surgery Residents". Surg. Endosc. 2011

Robotic Minimally Invasive Surgery (RMIS)

Figure taken from http://www.davincisurgery.com/

Pros and Cons of RMIS

- Advantages of RMIS [1]:
  - Better precision.
  - Smaller incisions.
  - Decreased need of blood transfusions.
  - Less postoperative surgical complications.
  - Quicker recovery time.



da Vinci Surgery Incision: no insicision, no scars

Open Surgery Incision

- Challenges of RMIS [2]:
  - Steep learning curve for surgeons.
  - Lack of fair, objective, and effective criteria for judging acquired skills with RMIS.
  - Novel rapid, cost-effective methods to quantify surgical skill are required to train 100,000 new surgeons between now and 2030 [3].

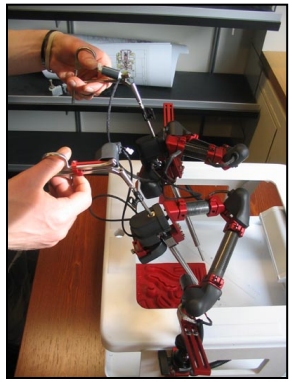
[1] W. Lowrance, E. Elkin, L. Jacks, D. Yee, T. Jang, V. Laudone, B. Guillonneau, P. Scardino, J. Eastham. Comparative effectiveness of prostate cancer surgical treatments: A population based analysis of postoperative outcomes. The Journal of Urology 183, 1366-1372, 2010.
[2] Lenihan, J., Kovanda, C., Seshadri-Kreaden, U., What is the learning curve for robotic assisted gynecologic surgery? Journal of Minimally Invasive Gynecology 15, 589–594, 2008.
[3] Thomas Lendvay, 2014.

Earlier Work: 2001-2006

- Imperial College Surgical Assessment Device
  - Electromagnetic markers to track a subject's hands during a standardized task [1] (simulation).

- Minimally Invasive Surgical Trainer - Virtual Reality
  - Movements of two standard laparoscopic instruments are tracked. Low level analysis of positions, forces and times [2] (simulation).

- Generalized approach
  - MIS are modeled as a stochastic process using a discrete Markov model [3] (real task).



[1] Datta et al. The use of electromagnetic motion tracking analysis to objectively measure open surgical. skill in laboratory-based model". Journal of the American College of Surgery, 2001.
[2] Darzi, Mackay. Skills assessment of surgeons. Surgery 2002;131(2):121–124.
[3] Rosen et al. "Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model." Trans.in Biomedical Engineering, 53(3):399–413, 2006.

The Language of Surgery Project @ JHU

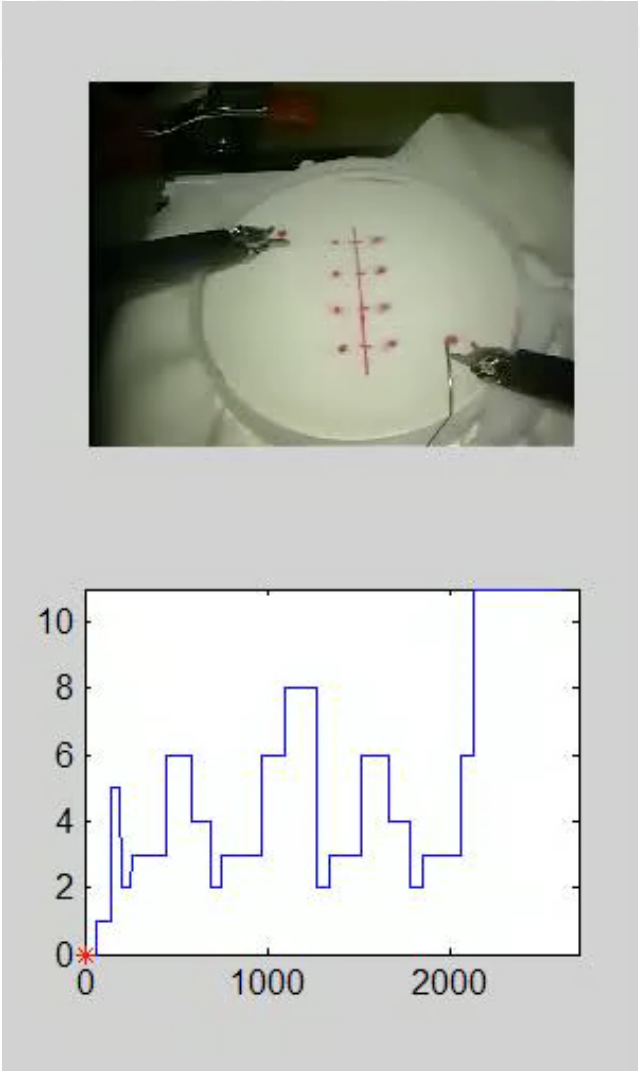
- RMIS has the potential to revolutionize our understanding of modeling, teaching and evaluating human manipulation skills.

- The goal of the project is to develop quantitative methods for modeling surgical tasks and evaluation of surgical skill.

JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

JHU-ISI Gesture Skill Assessment Working Set

- JIGSAWS Dataset:
  - Video data: 640x480 pixels,
  - Kinematic data: 78 dimensional position and velocity data,
  - Length: 1~5 min sampled at 30 Hz (1779~9012 samples per trial).
- 3 skill levels:
  - Expert (3 surgeons),
  - Intermediate (2 surgeons),
  - Novice (3 surgeons).
- 3 surgical tasks:
  - Suturing (39 trials),
  - Knot tying (26 trials),
  - Needle passing (36 trials).

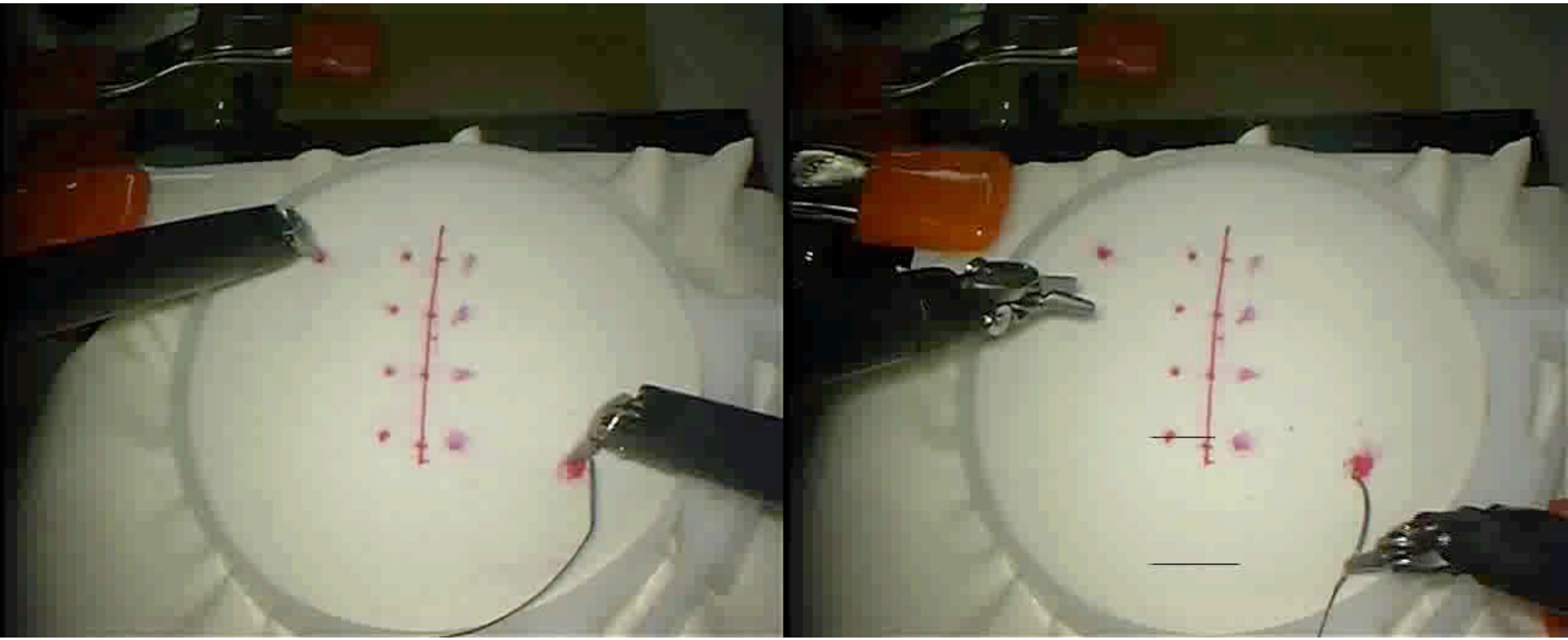


Video and surgeme ground-truth of an expert

Narges Ahmidi, Lingling Tao, Shahin Sefati, Yixin Gao, Colin Lea, Benjamın Bejar Haro, Luca Zappella, Sanjeev Khudanpur, Rene Vidal, Gregory D. Hager. A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery, Transaction of Biomedical Engineering, 2017

Example: Expert vs Expert

Example: Expert vs Novice



Modeling the Language of Surgery

- Similar to speech, surgical motion is not random:
  - A procedure is composed of tasks (incision, suturing, knot tying, etc.)
  - A task is composed of gestures (insert needle, pull needle, etc.)
  - Procedures, tasks and gestures follow a grammar.

Reiley, C.E., Lin, H.C., Varadarajan, B., Vagolgyi, B., Khudanpur, S., Yuh, D.D., Hager, G.D.: Automatic recognition of surgical motions using statistical modeling for capturing variability. In: Medicine Meets Virtual Reality. (2008) 396–401

Modeling the Language of Surgery

- Similar to speech, surgical motion is not random:
  - A procedure is composed of tasks (incision, suturing, knot tying, etc.)
  - A task is composed of gestures (insert needle, pull needle, etc.)
  - Procedures, tasks and gestures follow a grammar.

- **Goal**: develop methods to discover surgical phonemes and the surgical grammar underlying a surgical procedure.

Reiley, C.E., Lin, H.C., Varadarajan, B., Vagolgyi, B., Khudanpur, S., Yuh, D.D., Hager, G.D.: Automatic recognition of surgical motions using statistical modeling for capturing variability. In: Medicine Meets Virtual Reality. (2008) 396–401

Talk Outline

- Surgical Gesture Classification Before Deep Learning [1,2]
  - Linear Dynamical Systems (LDSs).
  - Bag of Spatio-Temporal Features (BoSTF).
  - Multiple Kernel Learning (MKL) to combine kinematic data and video.

- Surgical Gesture Segmentation Before Deep Learning [3-6]
  - Factor Analyzed Hidden Markov Models (FA-HMMs) [3]
  - Sparse Hidden Markov Models (SHMM) [4].
  - Conditional Random Field Models (CRF) [5].
  - CRFs + Deformable Part Models (DPMs) [6].

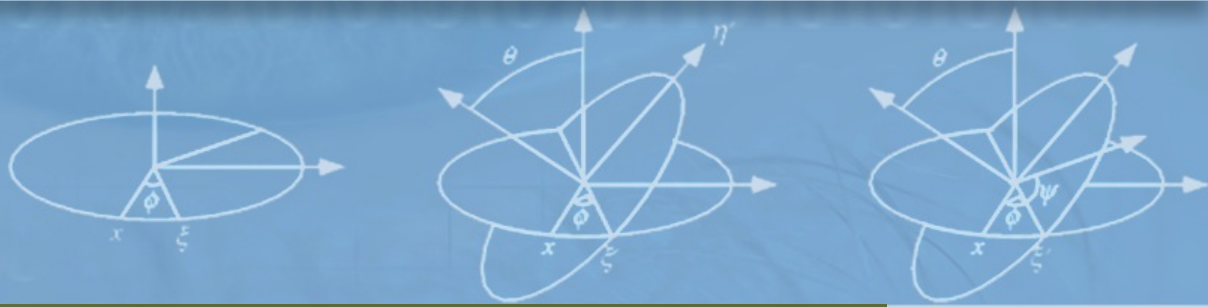
- Surgical Gesture Segmentation Using Deep Learning [7,8]
  - Segmental Spatio-Temporal CNNs [7].
  - Temporal Convolutional Networks [8].

[1] Bejar, Zapella, Vidal. Surgical Gesture Classification from Video Data, MICCAI 2012 **(Best Paper Award).**
[2] Zapella, Bejar, Hager, Vidal. Surgical Gesture Classification from Video Data, Medical Image Analysis, 2013.
[3] Varadarajan. Learning and Inference Algorithms for Dynamical System Models of Dextrous Motion. PhD thesis, JHU, 2011.
[4] Tao, Elhamifar, Khudanpur, Hager, Vidal. Sparse HMMs for Surgical Gesture Classification and Skill Evaluation. IPCAI, 2012.
[5] Tao, Zapella, Hager, Vidal. Surgical Gesture Segmentation and Recognition, MICCAI 2013.
[6] Lea, Hager, Vidal. Improved Model for Segmentation & Recognition of Fine-grained Activities with Application to Surgical Training Tasks. WACV 15.
[7] Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016
[8] Lea, Flynn, Vidal, Reiter, Hager. Temporal Convolutional Networks for Action Segmentation and Detection. CVPR 2017.

JHU vision lab

Surgical Gesture Classification Before Deep Learning

Benjamín Béjar[1], Luca Zappella[1], Gregory Hager[2] and René Vidal[1]
[1]Center for Imaging Science and [2]Laboratory for Computational Sensing and Robotics
Johns Hopkins University



JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins

Center for IMAGING SCIENCE

Linear Dynamical Systems (LDSs)

$$z(t+1) = Az(t) + Bv(t)$$
$$I(t) = C^0 + Cz(t) + w(t)$$

S03: push needle through tissue

S06: pull needle

$$C^0 \in \mathbb{R}^p$$
$$C \in \mathbb{R}^{p \times n}$$

APPEARANCE

$$A \in \mathbb{R}^{n \times n}$$
$$B \in \mathbb{R}^{n \times n_v}$$
$$z_0 \in \mathbb{R}^n$$

DYNAMICS

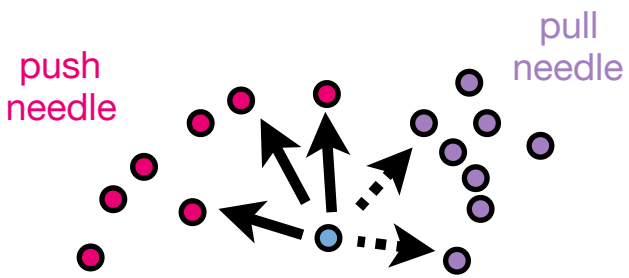
$$v(t) \sim \mathcal{N}(0, Q)$$
$$w(t) \sim \mathcal{N}(0, R)$$

NOISE

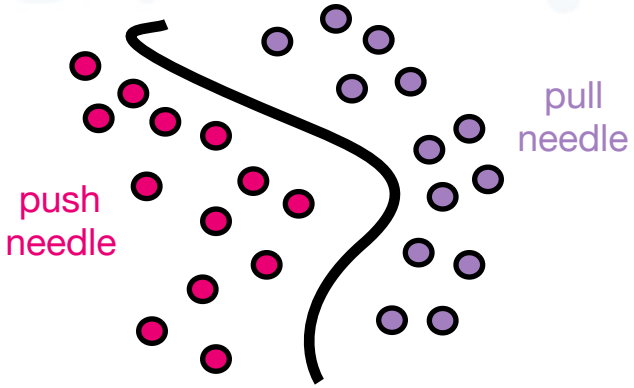
$$C \equiv \{C^1, \ldots, C^n\} \qquad \{C^0, C^1, \ldots, C^n\}$$

Dynamic Appearance Images (DAI)

Classification of Linear Dynamical Systems



k-NN



Kernel SVM

$$d_M\left(\blacksquare, \blacksquare\right)$$

push needle

pull needle

- Subspace Angles [1,2]
- KL-Divergence [3]
- Binet-Cauchy Kernels [4]
- Align Distances [5]

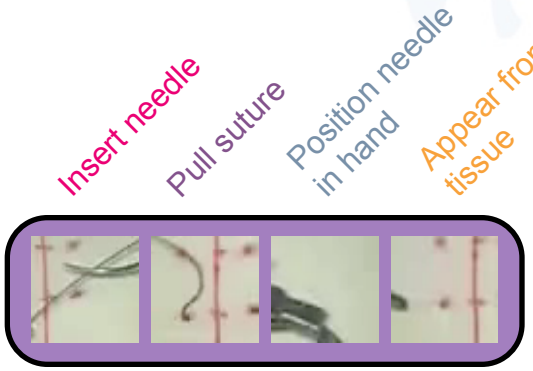
[1] Martin. A Metric for ARMA Processes, Transactions on Signal Processing, 2000.
[2] De Cook, De Moor Subspace angles Between ARMA Models. Systems and Control Letters, 2002.
[3] Chan, Vasconcelos. Probabilistic Kernels for the Classification of Auto-regressive Visual Processes CVPR'05, TPAMI'07.
[4] Vishwanathan, Smola, Vidal. Binet Cauchy Kernels on Dynamical Systems. International Journal of Computer Vision, 2007.
[5] Afsari, Chaudhry, Ravichandran, Vidal. Group Action Induced Distances for Averaging and Clustering Linear Dynamical Systems. CVPR'12.

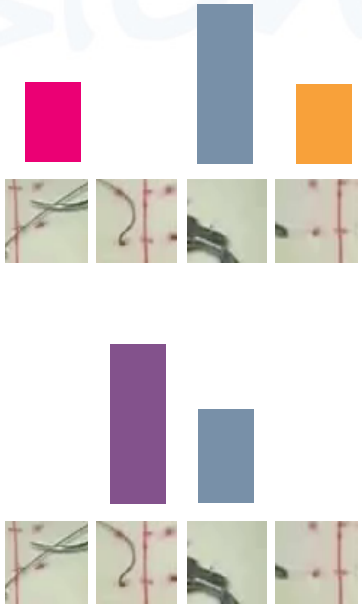
JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

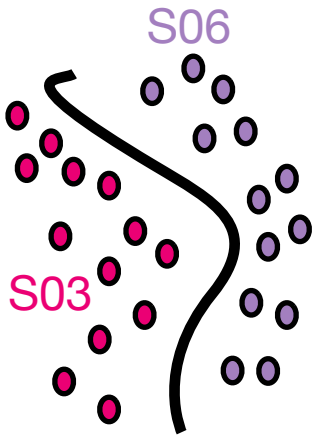
Bag of Spatio-Temporal Features (BoSTF)

S03: push needle through tissue

S06: pull needle







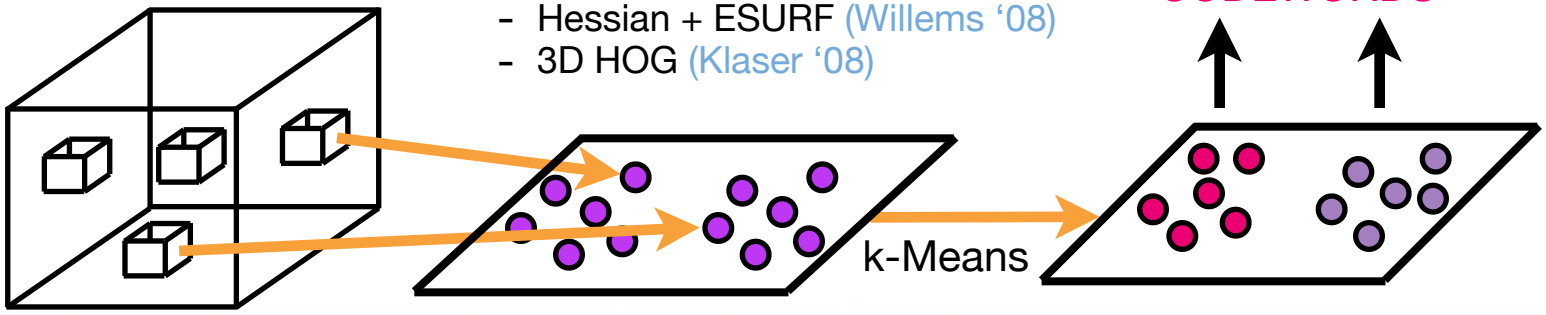
VIDEO SEQUENCE

$I(t)$

Codebook Construction

- 3D Harris + JET (Laptev '03)
- ST Filter + Gradient (Dollar '05)
- Hessian + ESURF (Willems '08)
- 3D HOG (Klaser '08)

Representation



Classifier

- K-NNs
- SVMs
- Naive Bayes

Laptev. On space-time interest points. International journal of computer vision 64 (2-3), 107-123, 2005.
H Wang, MM Ullah, A Klaser, I Laptev, C Schmid. Evaluation of local spatio-temporal features for action recognition. BMVC 2009

Multiple Kernel Learning (MKL)

- How can we combine both kinematic and video data?
- Construct a kernel from kinematic data: k_{kin}
- Construct a kernel from video data: k_{vid}
- Combine two kernels as

$$k(x,y) = \phi(x)^\top \phi(y) = \mu_{kin} k_{kin}(x,y) + \mu_{vid} k_{vid}(x,y)$$

- Jointly learn the kernel weights and the classifier parameters using Multiple Kernel Learning (MKL) [1]

$$\min_{\mathbf{w},\{\xi_i\},b,\boldsymbol{\mu}\geq\mathbf{0}} \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{w} + C\sum_{i=1}^{n} l(\mathbf{w},\mathbf{x}_i,y_i,b) + \sum_{l}\mu_l^2$$

where $$l(\mathbf{w},\mathbf{x}_i,y_i,b) = \max(0, 1 - y_i(\mathbf{w}^\mathsf{T}\phi(\mathbf{x}_i) + b))$$

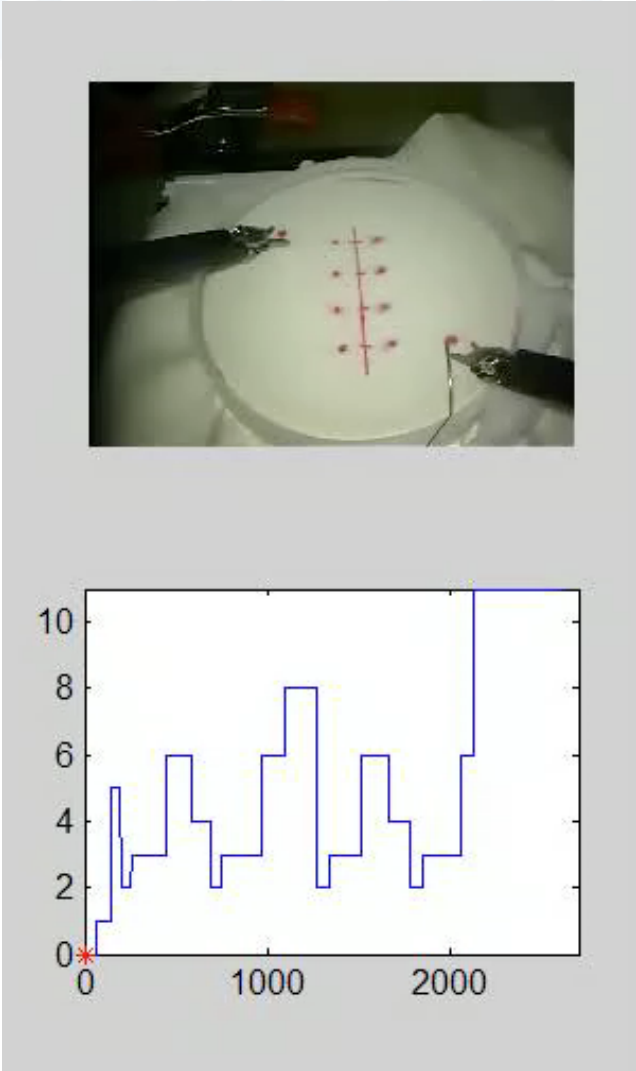
[1] Varma, M., Babu, R.: More generality in efficient multiple kernel learning. ICML (2009) 1065–1072

Surgical Gesture Classification Results

- JIGSAWS Dataset:
  - Video data: 640x480 pixels,
  - Kinematic data: 78 dimensional
  - 3 surgical skill levels, 8 surgeons
  - 3 surgical tasks, ~ 40 trials

- LOSO: leave one super trial out.
  - Training: 32 sequences.
  - Testing: 8 sequences.

- LOUO: leave one user out.
  - Training: 35 sequences (1 user is never seen).
  - Testing: 5 sequences (from unseen user).



Video and surgeme ground-truth of an expert

Narges Ahmidi, Lingling Tao, Shahin Sefati, Yixin Gao, Colin Lea, Benjamın Bejar Haro, Luca Zappella, Sanjeev Khudanpur, Rene Vidal, Gregory D. Hager, A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery, Transaction of Biomedical Engineering, 2017

Surgical Gesture Classification Results

Classification accuracy (%)		Kinematic	
		KSVD-HMM	LDS Fro.
Suturing	LOSO	79.37	87.25
	LOUO	60.85	74.95

[1] Bejar, Zapella, Vidal. Surgical Gesture Classification from Video Data, MICCAI 2012 **(Best Paper Award).**
[2] Zapella, Bejar, Hager, Vidal. Surgical Gesture Classification from Video Data, Medical Image Analysis, 2013.

Surgical Gesture Classification Results

Classification accuracy (%)		Kinematic		Video		
		KSVD-HMM	LDS Fro.	BoF	LDS	BoF+ LDS
Suturing	LOSO	79.37	87.25	90.68	87.15	91.79
	LOUO	60.85	74.95	79.95	74.22	81.17

[1] Bejar, Zapella, Vidal. Surgical Gesture Classification from Video Data, MICCAI 2012 **(Best Paper Award).**
[2] Zapella, Bejar, Hager, Vidal. Surgical Gesture Classification from Video Data, Medical Image Analysis, 2013.

Surgical Gesture Classification Results

Classification accuracy (%)		Kinematic		Video			Both	
		KSVD-HMM	LDS Fro.	BoF	LDS	BoF+LDS	BoF+LDS(kin)	BoF+LDS(all)
Suturing	LOSO	79.37	87.25	90.68	87.15	91.79	93.52	**93.95**
	LOUO	60.85	74.95	79.95	74.22	81.17	86.28	**86.56**

[1] Bejar, Zapella, Vidal. Surgical Gesture Classification from Video Data, MICCAI 2012 **(Best Paper Award).**
[2] Zapella, Bejar, Hager, Vidal. Surgical Gesture Classification from Video Data, Medical Image Analysis, 2013.

Surgical Gesture Classification Results

Classification accuracy (%)		Kinematic		Video			Both	
		KSVD-HMM	LDS Fro.	BoF	LDS	BoF+LDS	BoF+LDS(kin)	BoF+LDS(all)
Suturing	LOSO	79.37	87.25	90.68	87.15	91.79	93.52	**93.95**
	LOUO	60.85	74.95	79.95	74.22	81.17	86.28	**86.56**
Needle Passing	LOSO	76.43	78.77	74.14	68.91	77.84	85.32	**86.04**
	LOUO	45.26	67.28	65.53	58.77	66.88	80.08	**80.16**
Knot Tying	LOSO	86.78	85.07	88.39	87.25	90.76	**93.75**	92.76
	LOUO	71.94	78.89	84.92	77.36	86.7	90.08	**90.38**

[1] Bejar, Zapella, Vidal. Surgical Gesture Classification from Video Data, MICCAI 2012 **(Best Paper Award).**
[2] Zapella, Bejar, Hager, Vidal. Surgical Gesture Classification from Video Data, Medical Image Analysis, 2013.

Conclusions

- Three methods for classification of surgical gestures that combine BoF and LDS.
  - Video data can be as discriminative as kinematic data.
  - The integration of both kinematic and video data improves results.

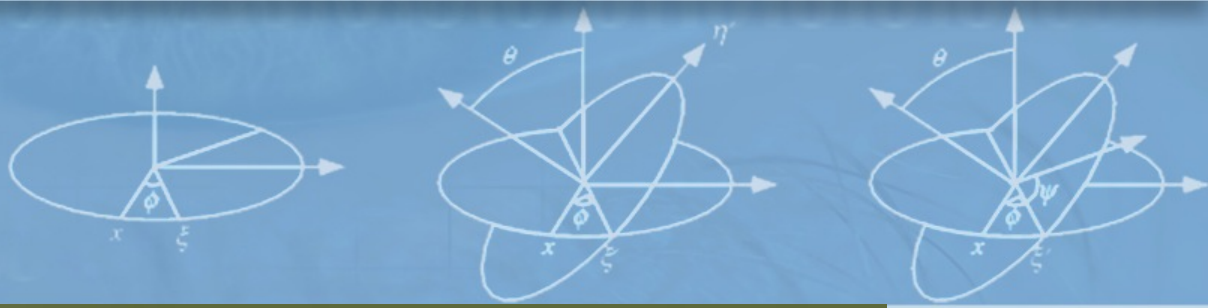
- So far we have assumed that gesture boundaries are known.
  - In practice, surgical tasks can be decomposed into sequences of gestures and we do not know the segmentation.
  - Need methods to segment kinematic and video data into segments and classify each segment.

JHU vision lab

Surgical Gesture Segmentation Before Deep Learning

Reiley[2], Varadarajan[3], Tao[1], Zappella[1], Lea[2], Khudanpur[3], Hager[2] and Vidal[1]

[1]Center for Imaging Science, [2]Laboratory for Computational Sensing and Robotics,
[3]Center for Speech and Language Processing, Johns Hopkins University

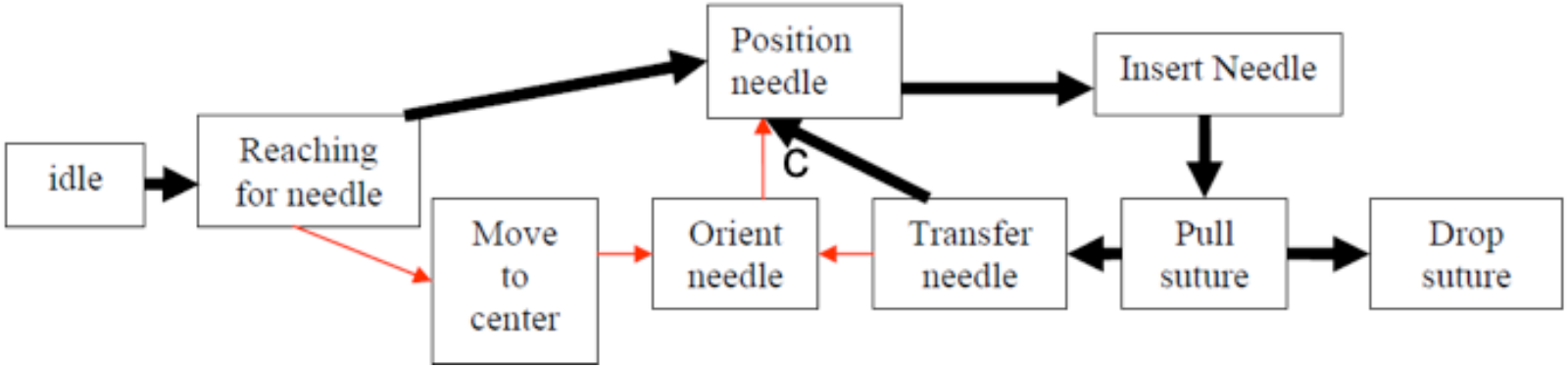


JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins

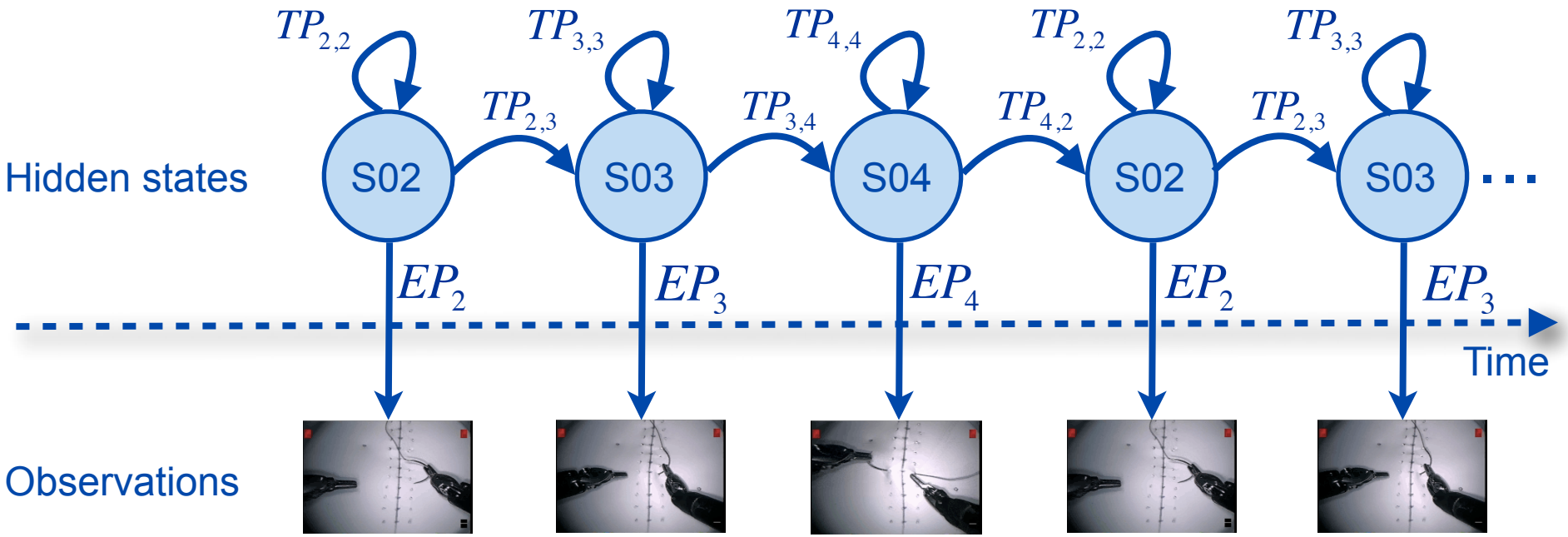
Center for IMAGING SCIENCE

Modeling the Language of Surgery



Hidden Markov Models (HHMs)

- Gesture at time t is defined by the state of a Markov chain

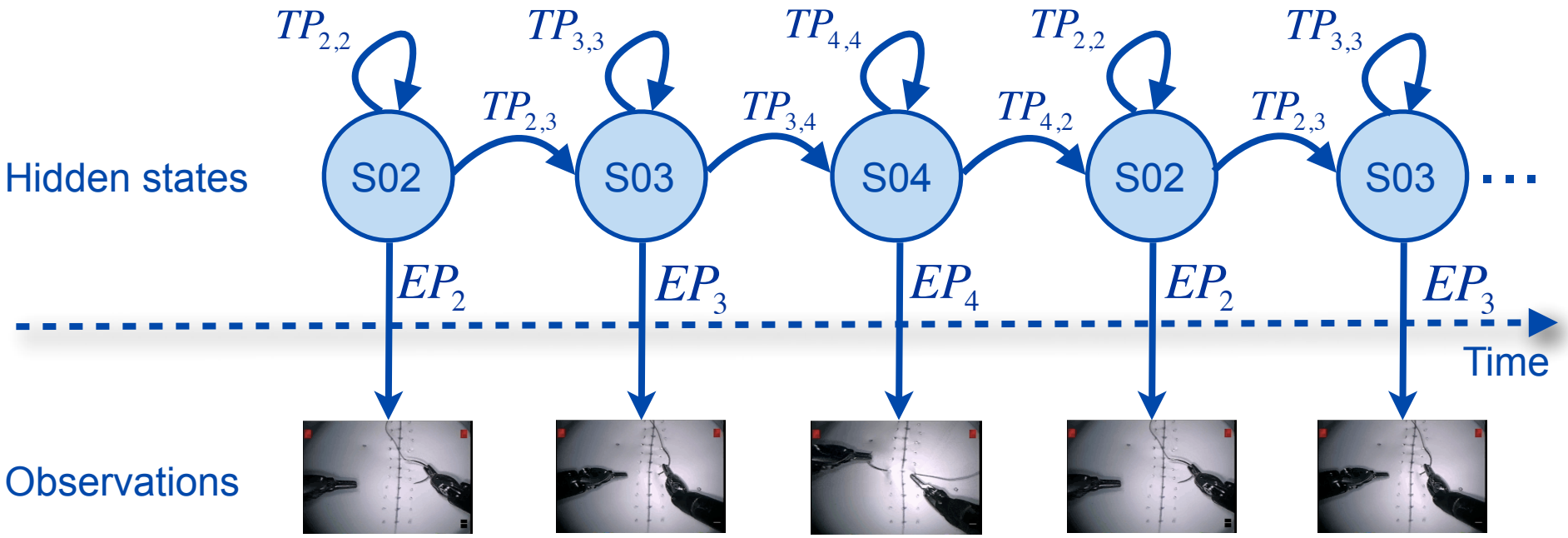


- **Learning**: parameters found using Baum Welch algorithm.

Rosen et al. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Trans. Biomedical Eng., 2001
Rosen et al. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Computer Aided Surgery 7(1):49–61, 2002
Reiley, C.E., Hager, G.D.: Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. MICCAI, 2009

Hidden Markov Models (HHMs)

- Gesture at time t is defined by the state of a Markov chain



- **Learning**: parameters found using Baum Welch algorithm.
- **Inference**: sequence of states found using Viterbi algorithm.
  - Gesture recognition: find the most likely sequence of states.
  - Skill recognition: find the most likely model (one HMM per skill level).

Rosen et al. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Trans. Biomedical Eng., 2001
Rosen et al. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Computer Aided Surgery 7(1):49–61, 2002
Reiley, C.E., Hager, G.D.: Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. MICCAI, 2009

Surgical Gesture Segmentation Results

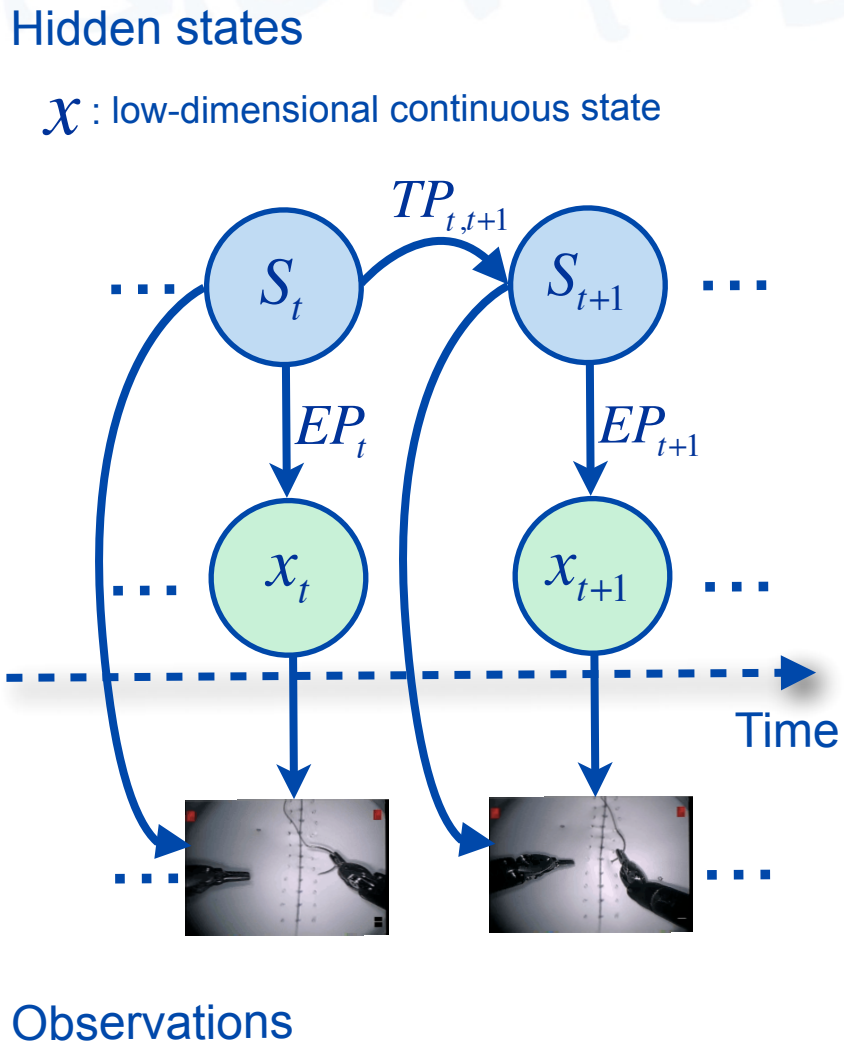
- Learning: given training trials, find the model parameters.
- Testing: given a test trial, find its sequence of surgemes.
- LOSO: leave one super trial out.
  - Training: 32 sequences.
  - Testing: 8 sequences.
- LOUO: leave one user out.
  - Training: 35 sequences (1 user is never seen).
  - Testing: 5 sequences (from unseen user).

JIGSAWS		HMM
Suturing	LOSO	68.5%
	LOUO	N/A
Needle Passing	LOSO	54.5%
	LOUO	N/A
Knot Tying	LOSO	69.8%
	LOUO	N/A

Varadarajan. Learning and Inference Algorithms for Dynamical System Models of Dextrous Motion. PhD thesis, JHU, 2011

Factor Analyzed HMM (FA-HMM)

- A kinematic observation lies in a high-dimensional space.
- Directly modeling the observations requires a large number of parameters.
- However, number of degrees of freedom might be lower.
- FA-HMM: model "the most relevant dimensions":
  - Learning: Baum-Welch.
  - Inference: Viterbi.



Varadarajan. Learning and Inference Algorithms for Dynamical System Models of Dextrous Motion. PhD thesis, JHU, 2011

Surgical Gesture Segmentation Results

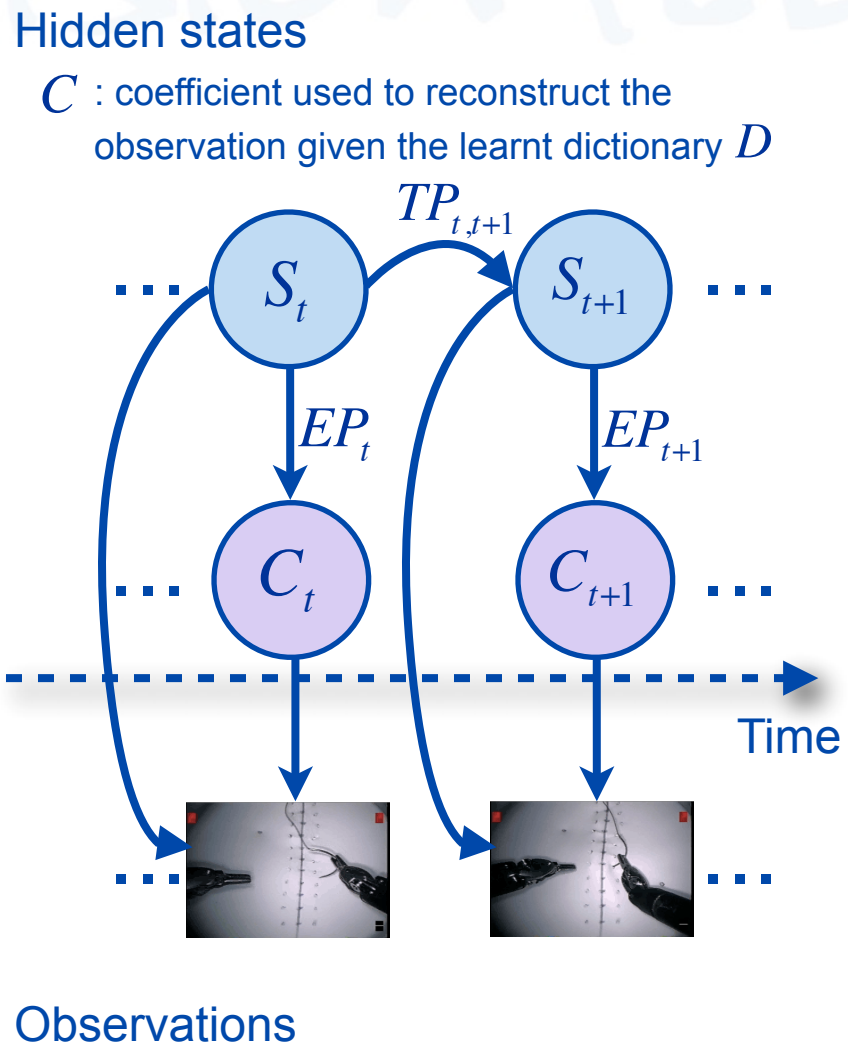
- Learning: given training trials, find the model parameters.
- Testing: given a test trial, find its sequence of surgemes.
- LOSO: leave one super trial out.
  - Training: 32 sequences.
  - Testing: 8 sequences.
- LOUO: leave one user out.
  - Training: 35 sequences (1 user is never seen).
  - Testing: 5 sequences (from unseen user).

JIGSAWS		HMM	FA-HMM
Suturing	LOSO	68.5%	78.2%
	LOUO	N/A	57.2%
Needle Passing	LOSO	54.5%	71%
	LOUO	N/A	42.7%
Knot Tying	LOSO	69.8%	82.8%
	LOUO	N/A	67%

Varadarajan. Learning and Inference Algorithms for Dynamical System Models of Dextrous Motion. PhD thesis, JHU, 2011

Sparse Hidden Markov Model (KSVD-HMM)

- So far, there is a Gaussian model for each gesture's EP.
- S-HMMs capture variability in gesture execution using a dictionary of surgical motions.
  - Each observation is a sparse combination of few motions.
  - EP depends on how well a new observation is reconstructed by the dictionary.
- KSVD-HMMs:
  - Learning: KSVD.
  - Inference: Viterbi++.



L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. "Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation", IPCAI, 2012

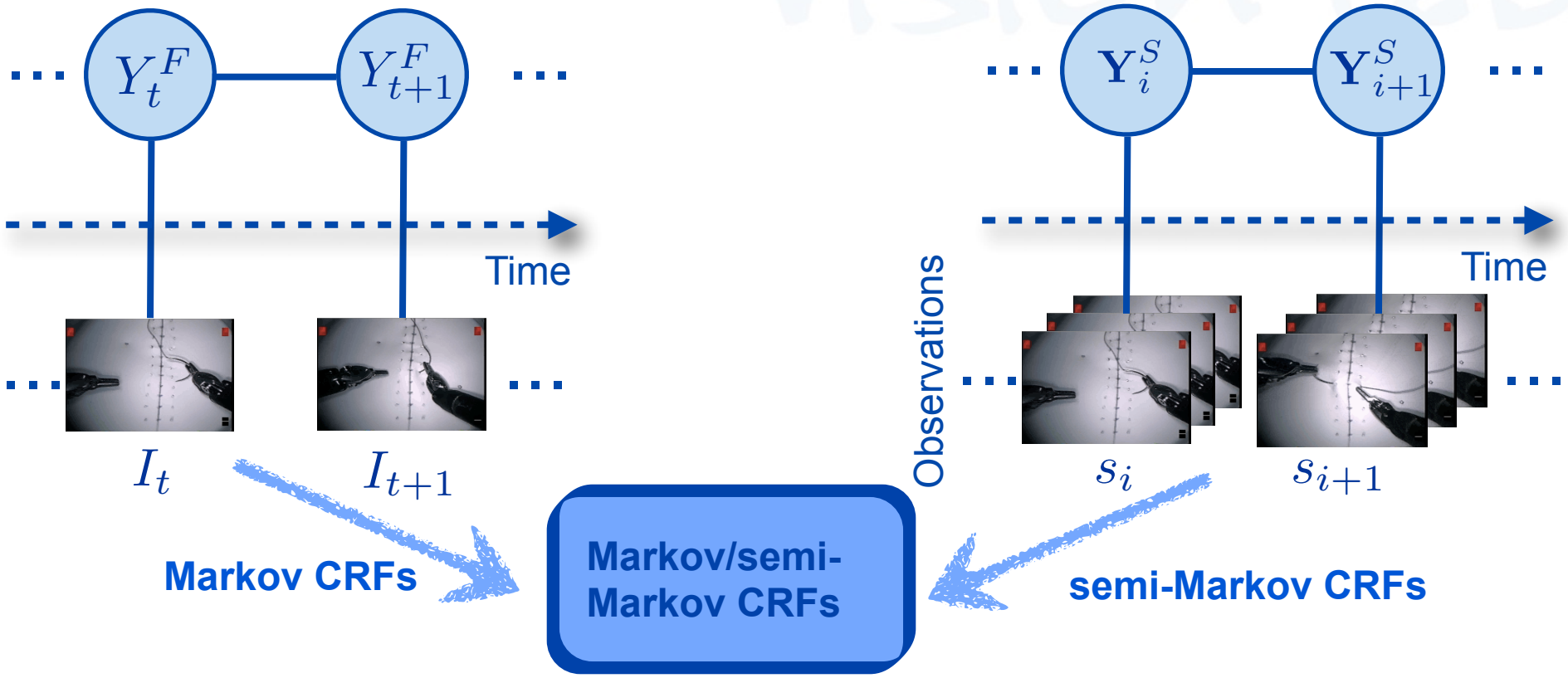
Surgical Gesture Segmentation Results

- Learning: given training trials, find the model parameters.
- Testing: given a test trial, find its sequence of surgemes.
- LOSO: leave one super trial out.
  - Training: 32 sequences.
  - Testing: 8 sequences.
- LOUO: leave one user out.
  - Training: 35 sequences (1 user is never seen).
  - Testing: 5 sequences (from unseen user).

JIGSAWS		HMM	FA-HMM	KSVD-HMM
Suturing	LOSO	68.5%	78.2%	82%
	LOUO	N/A	57.2%	67.8%
Needle Passing	LOSO	54.5%	71%	74.6%
	LOUO	N/A	42.7%	59.3%
Knot Tying	LOSO	69.8%	82.8%	81.1%
	LOUO	N/A	67%	65.7%

Varadarajan. Learning and Inference Algorithms for Dynamical System Models of Dextrous Motion. PhD thesis, JHU, 2011
L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. "Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation", IPCAI, 2012

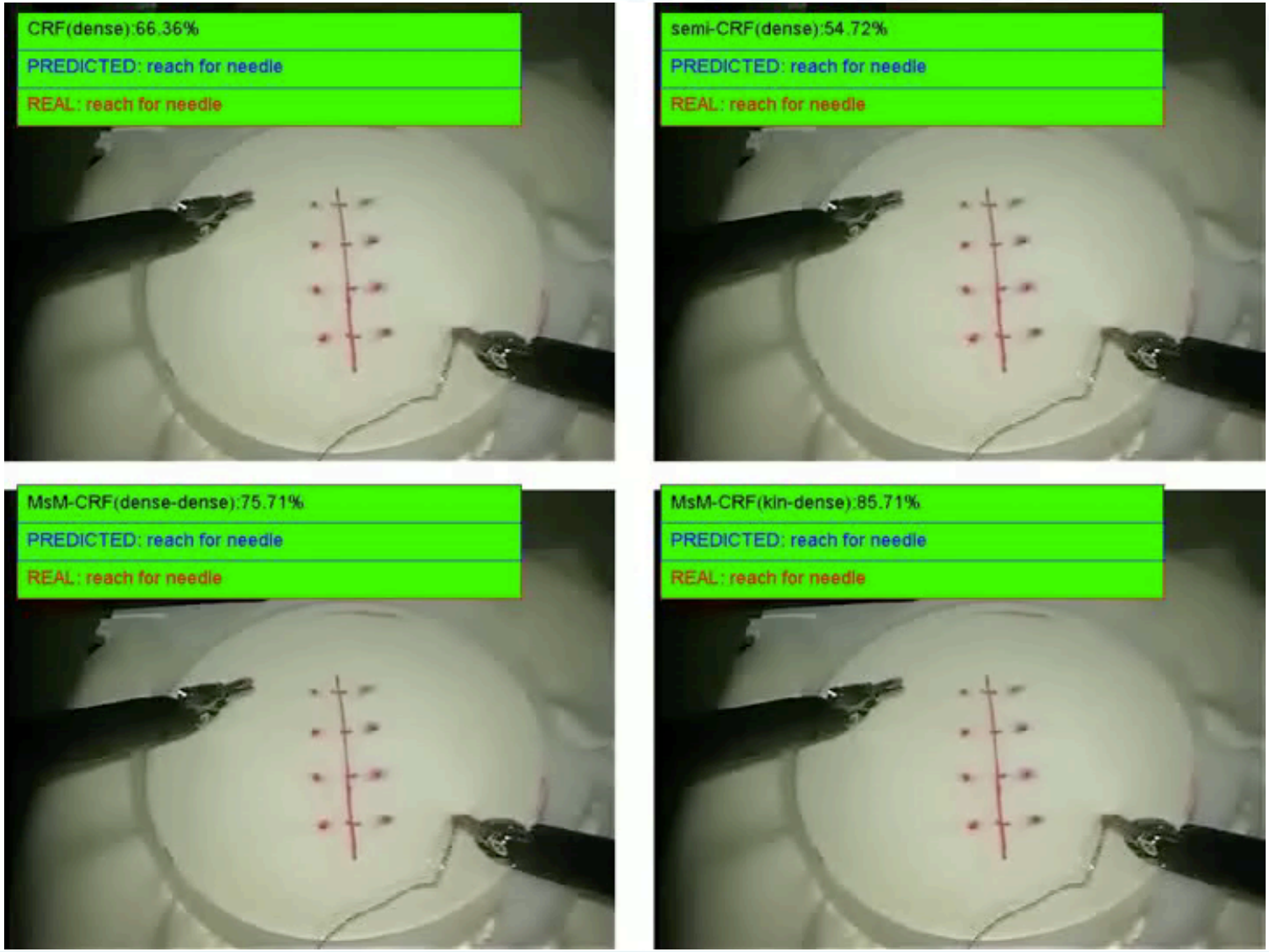
Markov Semi-Markov CRFs (MsM-CRFs)



- Inference: find sequence of gestures using a modified Viterbi
- Learning: find parameters using structural output SVMs

L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. "Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation", IPCAI, 2012

CRF Results for Suturing



L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. “Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation”, IPCAI, 2012

Surgical Gesture Segmentation Results

		Kinematic		
		CRF	semi-CRF	MsM-CRF
Suturing	LOSO	81.62	63.20	80.99
	LOUO	68.65	62.24	69.03

L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. “Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation”, IPCAI, 2012

Surgical Gesture Segmentation Results

		Kinematic			Video		
		CRF	semi-CRF	MsM-CRF	CRF	semi-CRF	MsM-CRF
Suturing	LOSO	81.62	63.20	80.99	76.51	65.83	79.04
	LOUO	68.65	62.24	69.03	68.80	59.41	71.76

L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. “Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation”, IPCAI, 2012

Surgical Gesture Segmentation Results

		Kinematic			Video			Both
		CRF	semi-CRF	MsM-CRF	CRF	semi-CRF	MsM-CRF	MsM-CRF
Suturing	LOSO	81.62	63.20	80.99	76.51	65.83	79.04	**82.81**
	LOUO	68.65	62.24	69.03	68.80	59.41	71.76	**72.60**

L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. “Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation”, IPCAI, 2012

Surgical Gesture Segmentation Results

		Kinematic			Video			Both
		CRF	semi-CRF	MsM-CRF	CRF	semi-CRF	MsM-CRF	MsM-CRF
Suturing	LOSO	81.62	63.20	80.99	76.51	65.83	79.04	**82.81**
	LOUO	68.65	62.24	69.03	68.80	59.41	71.76	**72.60**
Needle Passing	LOSO	74.56	54.15	74.85	62.23	56.22	68.81	**76.82**
	LOUO	46.44	38.36	52.39	54.52	46.89	**60.39**	57.08
Knot Tying	LOSO	81.06	39.20	79.39	69.16	44.82	72.04	**81.10**
	LOUO	67.38	44.28	64.28	60.17	41.46	66.94	**68.83**

L. Tao, E. Elhamifar, S. Khudanpur, G. Hager, and R. Vidal. “Sparse Hidden Markov Models for Surgical Gesture Classification and Skill Evaluation”, IPCAI, 2012

Conclusions

- Proposed segmentation methods based on HMMs and CRFs
- Classification accuracy reduces by ~ 10% when temporal segmentation is unknown
- Gap between LOSO and LOUO setups grows from 2-9% to 10-20%

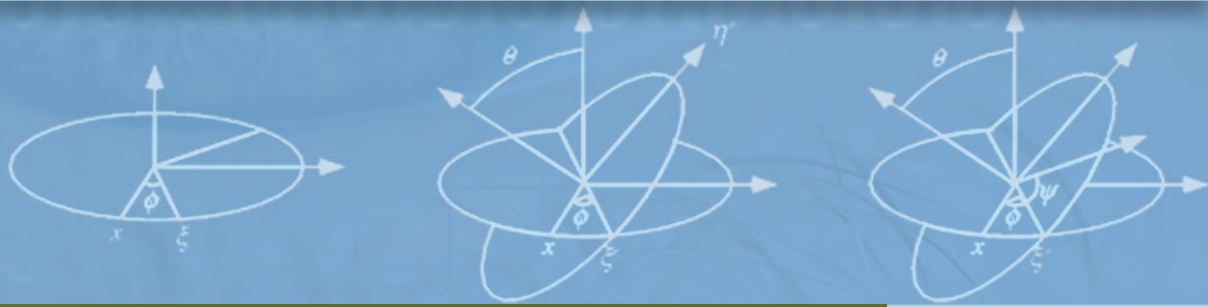
		Classification	Segmentation
JIGSAWS		BoF+LDS(all)	MsM-CRF
Suturing	LOSO	93.95%	82.81%
	LOUO	86.56%	72.60%
Needle Passing	LOSO	86.04%	76.82%
	LOUO	80.16%	57.08%
Knot Tying	LOSO	92.76%	81.10%
	LOUO	90.38%	68.83%

JHU vision lab

Surgical Gesture Segmentation Using Segmental Spatio-Temporal Networks

Colin Lea[2], Reiter, Gregory Hager[2] and René Vidal[1]

[1]Center for Imaging Science, [2]Laboratory for Computational Sensing and Robotics,
[3]Center for Speech and Language Processing, Johns Hopkins University



JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE

THE DEPARTMENT OF BIOMEDICAL ENGINEERING
The Whitaker Institute at Johns Hopkins

Center for IMAGING SCIENCE

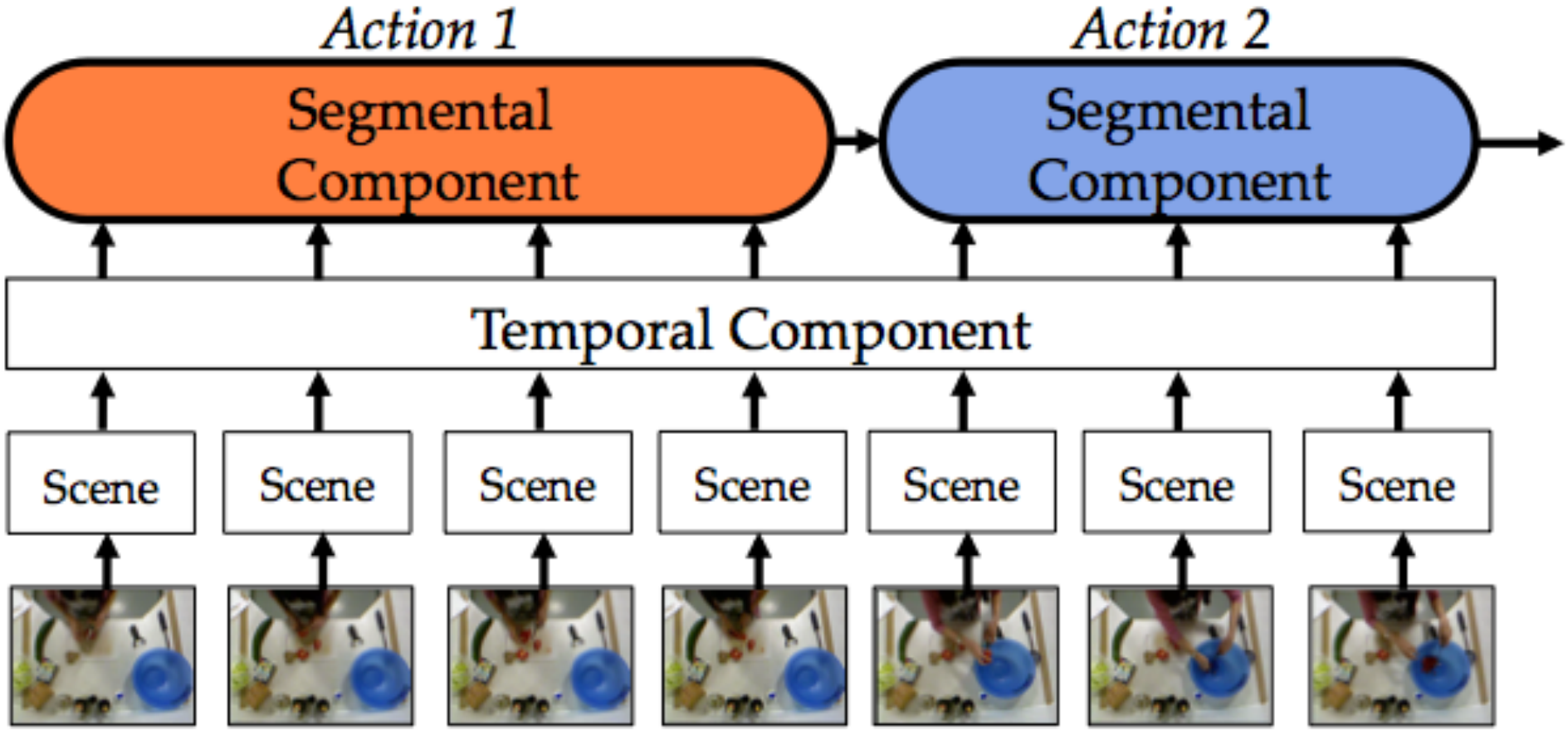
Motivations for a Deep Learning Approach

- There is a ~10% gap in segmentation performance between classification and segmentation, and ~10% gap between LOSO and LOUO setups.

- Video-based methods for surgical gesture classification and segmentation rely primarily on BoF model, thus fail to capture spatial relationships among objects in the scene.

- Temporal models rely primarily on pairwise potentials in a CRF, which fail to capture long-range temporal interactions.

Segmental Spatio-Temporal CNNs

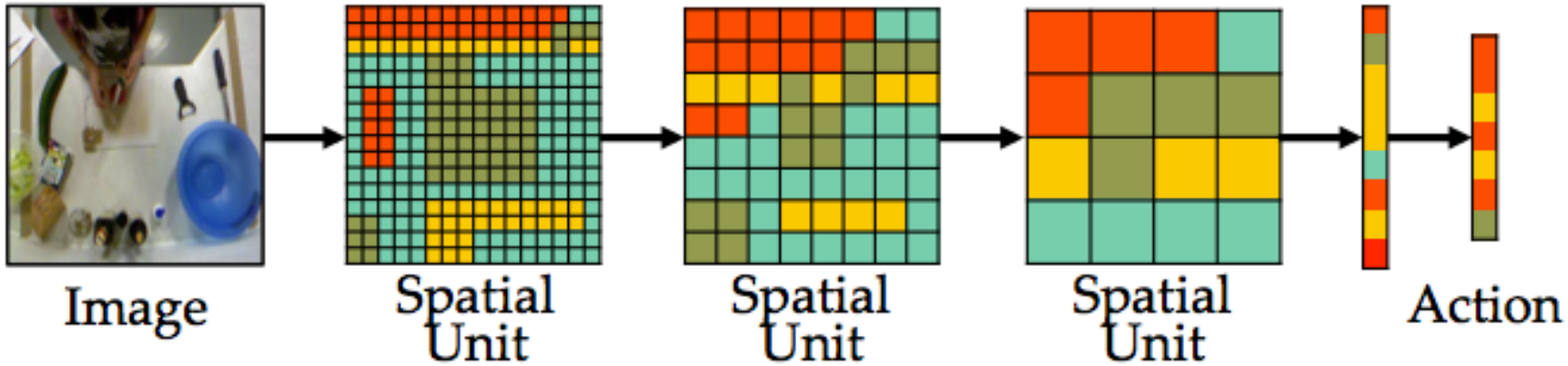


Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016.

Spatial Component of S-ST-CNN Model

- A 3-layer spatial convolutional neural network (VGG inspired)

- Spatial units capture objects in each region

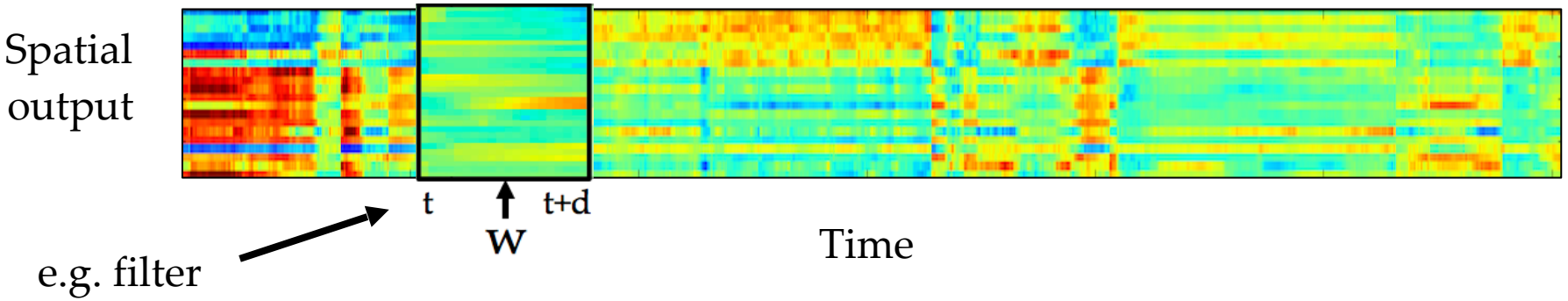


- Fully connected units capture relationships between regions

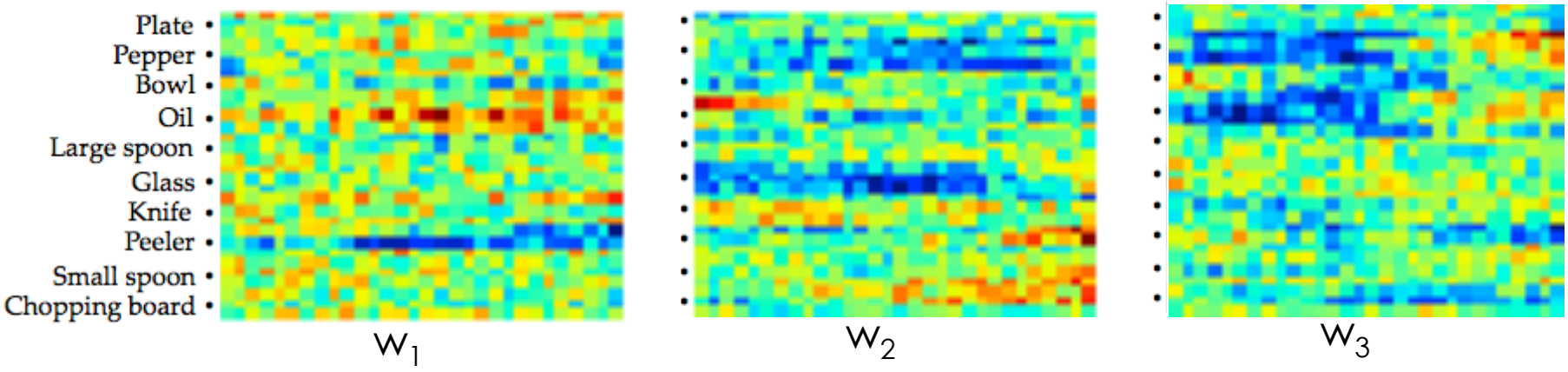
Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016.

Temporal Component of S-ST-CNN Model

- Concatenate spatial response at each frame.



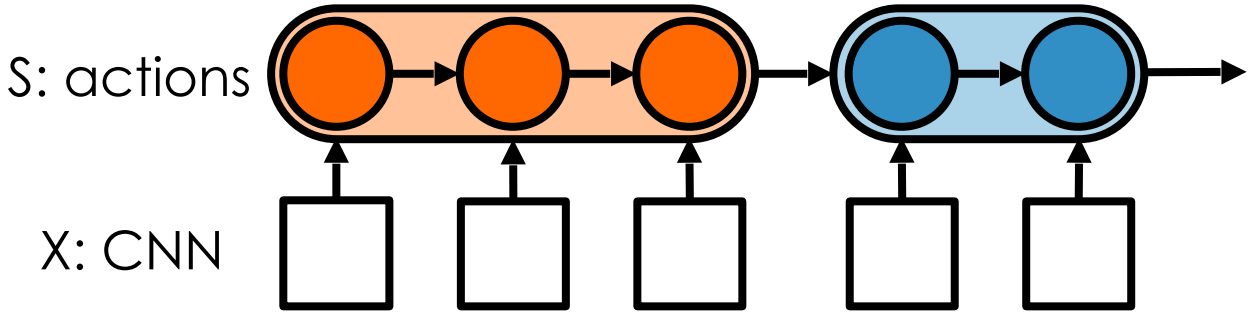
- Learn 1D temporal convolutional filters that capture sub-actions (start-middle-end).



Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016.

Segmental Component of S-ST-CNN Model

- Issue: HMMs, CRFs, CNNs typically predict actions per-frame
- Solution: Compute start/end/class of each action per-segment



- Energy

$$E(S, X) = \sum_{m=1}^{M} f(X, s_m, e_m, c_m) + p(c_{m-1}, c_m)$$

sum of CNN output

rules/pairwise

- Segment Function

$$f(X, s_m, e_m, c_m) = \sum_{t=s_m}^{e_m - 1} X_{t, c_m}$$

Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016.

Segmental Component of S-ST-CNN Model

- Issue: When actions are long, inference is costly
- Solution: Reframe problem; optimize over # segments
- Prior Approach: Segmental Viterbi
  - Optimize over durations (D)
  - Complexity: $O(DTC^2)$
- Our Approach:
  - Optimize over # segments (K)
  - Complexity: $O(KTC^2)$

Labels	Dur	#Segs	Speedup
Low	2289	65	35x
Mid	3100	25	124x
Eval	3100	24	129x
High	11423	6	1902x
JIGSAWS	1107	37	30x

true # segments

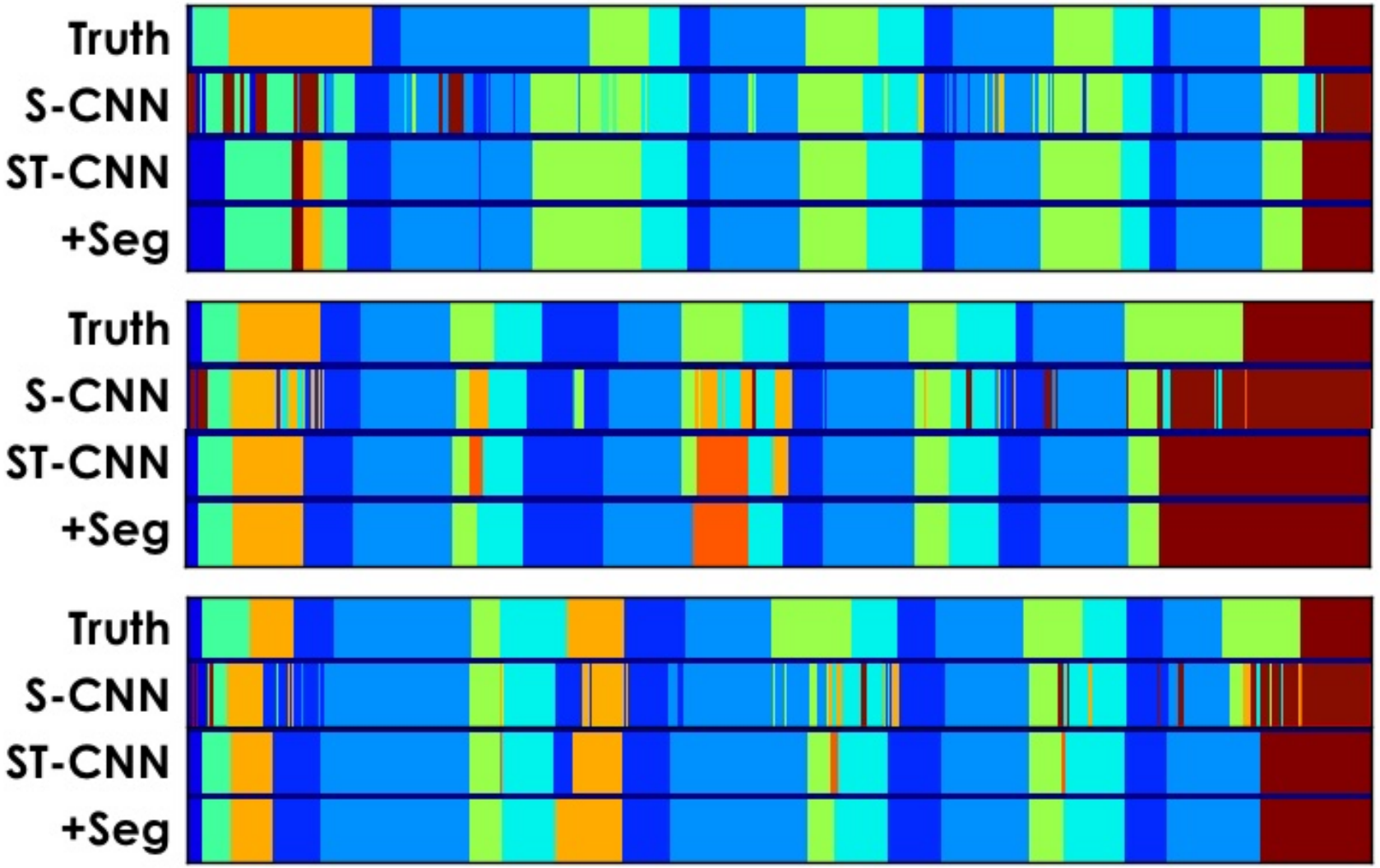
upper bound

$$\hat{S} = \arg\max_{S,M} E(S,X) \quad \text{s.t.} \quad 0 < M \leq K$$

Recursion: $$V_{t,c}^{m} = \max_{c' \in \mathcal{Y}} V_{t-1,c'}^{m'} + P_{c',c} + X_{t,c}$$

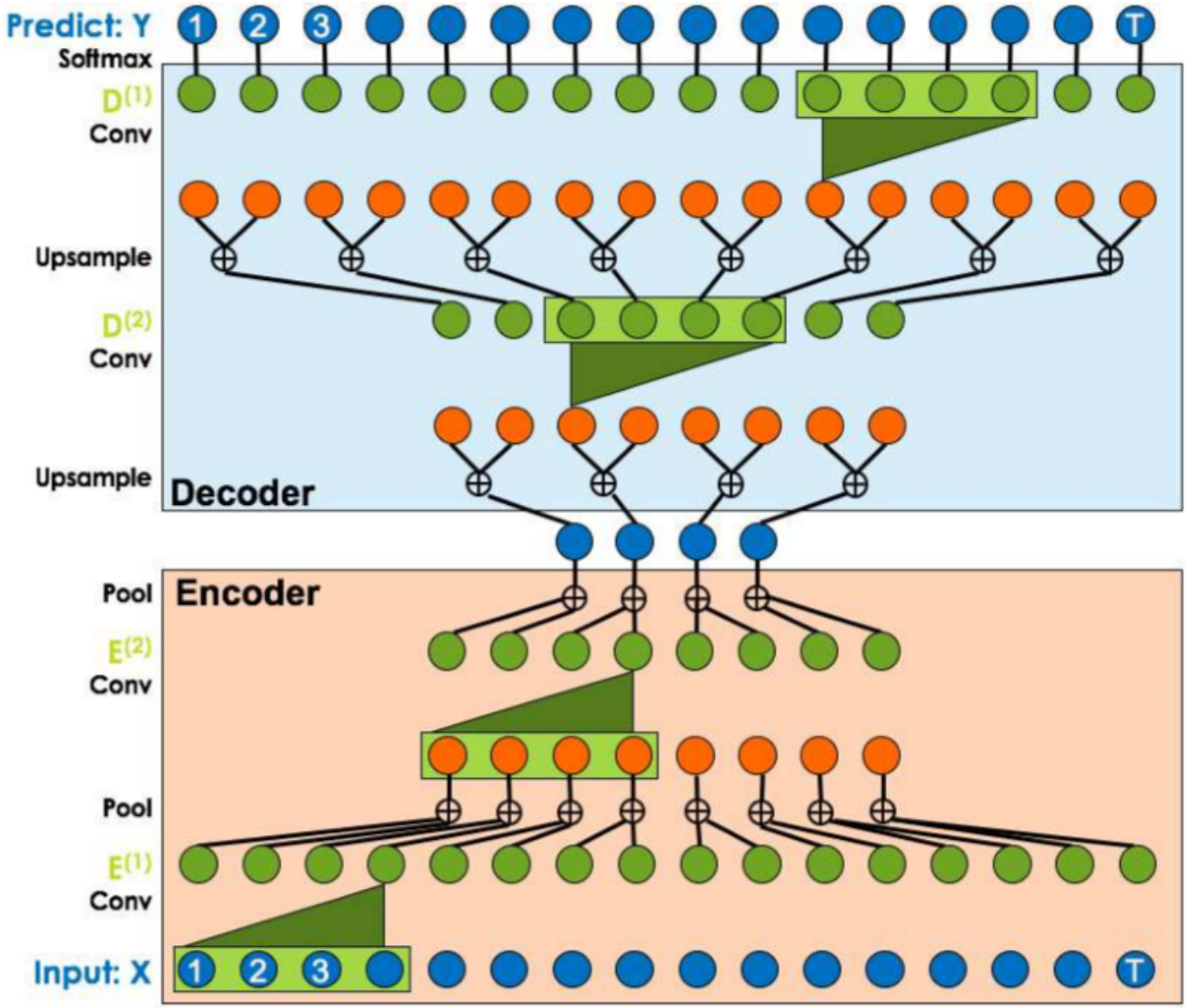
Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016.

Results of S-ST-CNN on JIGSAWS



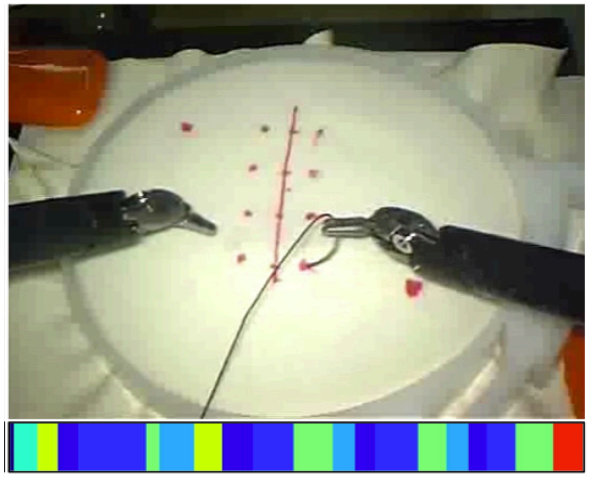
Lea, Reiter, Vidal, Hager. Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation. ECCV 2016.

Encoder-Decoder Temporal Conv Nets



Lea, Flynn, Vidal, Reiter, Hager. Temporal Convolutional Networks for Action Segmentation and Detection. CVPR 2017.

Results of ED-TCN



	Model	Accuracy	Edit
Sensors	LC-SC-CRF	81.9*	78.4*
	S-ST-CNN	79.2	82.6
	ED-TCN	82.4	**89.3**
Video	LC-SC-CRF	-	-
	S-ST-CNN	74.2	66.6
	ED-TCN	**78.3**	**85.6**

	Model	Accuracy	Edit
Sensors	LC-SC-CRF	81.8	58.5
	S-ST-CNN	82.1	55.5
	T-CNN	**84.8**	**76.9**
Video	LC-SC-CRF	-	-
	S-ST-CNN	**72.0**	**62.0**
	ED-TCN	71.0*	**62.0***

Lea, Flynn, Vidal, Reiter, Hager. Temporal Convolutional Networks for Action Segmentation and Detection. CVPR 2017.

Conclusions

- The future of robotic surgery
  - Quantitative skill assessment
  - Automated assistance
  - Improved surgical education

- Computer vision and machine learning approaches provided promising results.
  - CRFs and switched linear systems improve over HMMs
  - Deep temporal models improve over switched linear systems

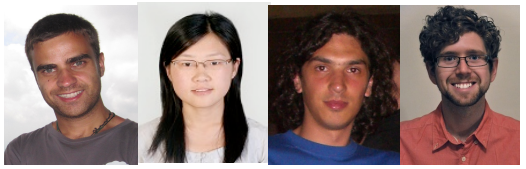
- Need more data, need more annotations, need unsupervised methods

Acknowledgements

- Collaborators
  - Greg Hager, JHU
  - Sanjeev Khudanpur, JHU


- Funding
  - NSF 0941362, NSF 0931805
  - Talentia Fellowship, Government of Andalusia

- Students and Postdocs
  - Benjamín Bejar, BME
  - Lingling Tao, ECE
  - Luca Zappella, Postdoc
  - Colin Lea, CS



  - Carol Reiley, CS
  - Rizwan Chaudhry, CS
  - Avinash Ravichandran, ECE
  - Balakrishnan Varadarajan, ECE

Vision Lab @ Johns Hopkins University

http://www.vision.jhu.edu

Thank You!

Center for IMAGING SCIENCE

Surgical Gesture Classification from Video Data

Benjamín Béjar, Luca Zappella and René Vidal

Center for Imaging Science, Johns Hopkins University