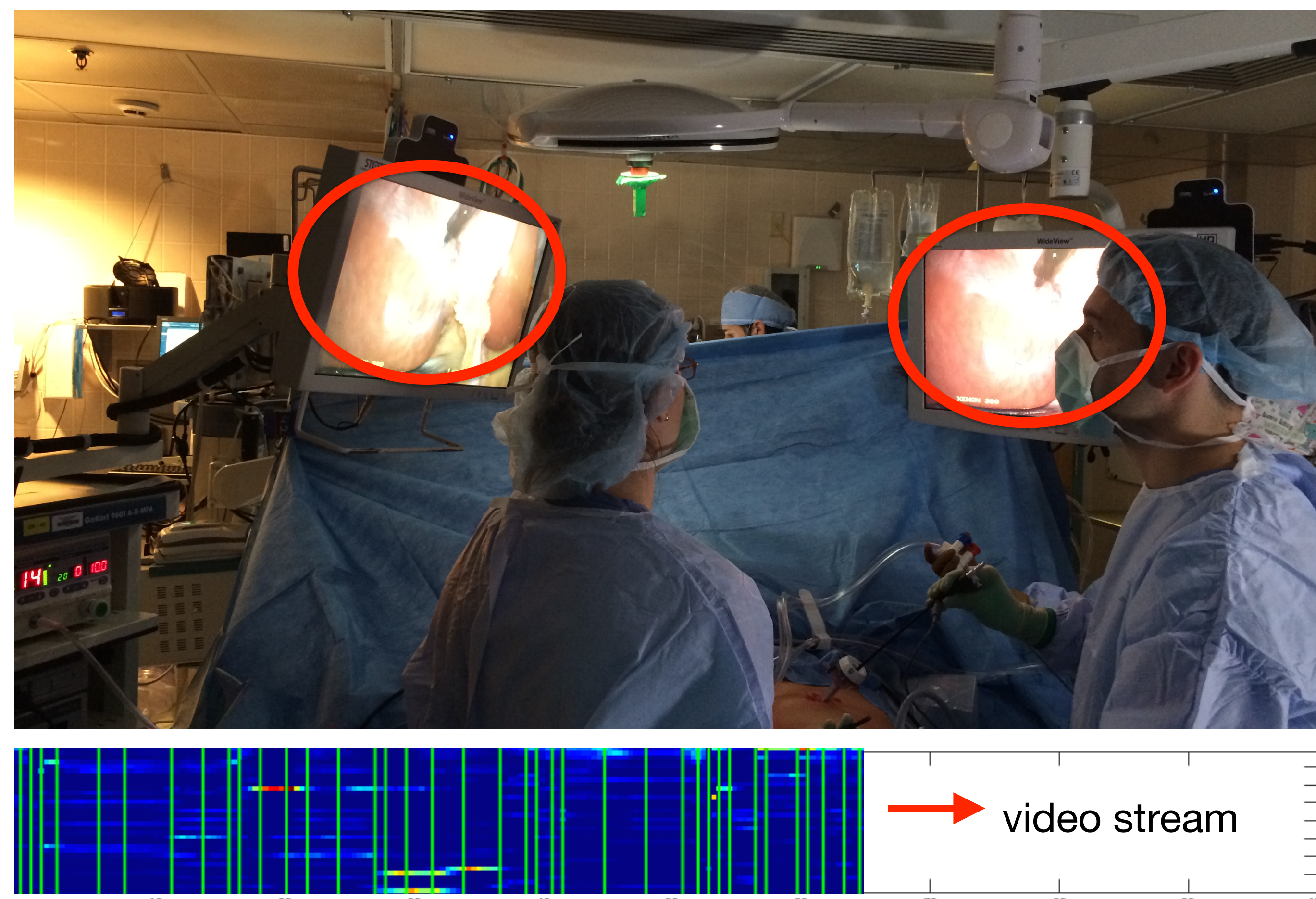


INTRODUCTION

- Video-based coaching and debriefing of laparoscopic surgical procedures has been demonstrated to contribute to enhanced surgical performance. While recording laparoscopic procedures is easy, reviewing video remains a time-consuming process.
- Context-aware video segmentation must be done manually, by viewing the video in its entirety or by browsing through the video to identify different phases of an operation.
- Manual video viewing and editing is time-consuming and error-prone. Moreover, hospital infrastructure is often prohibitive of recording and storing large amounts of video as a matter of routine.



OBJECTIVES

- In this study we present a system that can automatically segment laparoscopic surgical video according to its underlying surgical steps (phases).
- We leverage the technical expertise of surgeons to obtain a ground truth for a number of laparoscopic vertical sleeve gastrectomy (LSG) procedures.
- Using this ground truth we learn a visual feature space that describes the variability in the procedures. Using machine learning methods we show that we can learn the phase transitions and segment the videos according to their phases. We present a system that computes a segmentation automatically, online, and in real-time.
- We demonstrate that using data reduction constructs called coresets we obtain results of almost identical quality, while using only a fraction of the computational memory resources.
- We evaluate our system experimentally using video from LSG procedures, and present a blueprint for instantiating such a system in a real operating room environment.

DATA

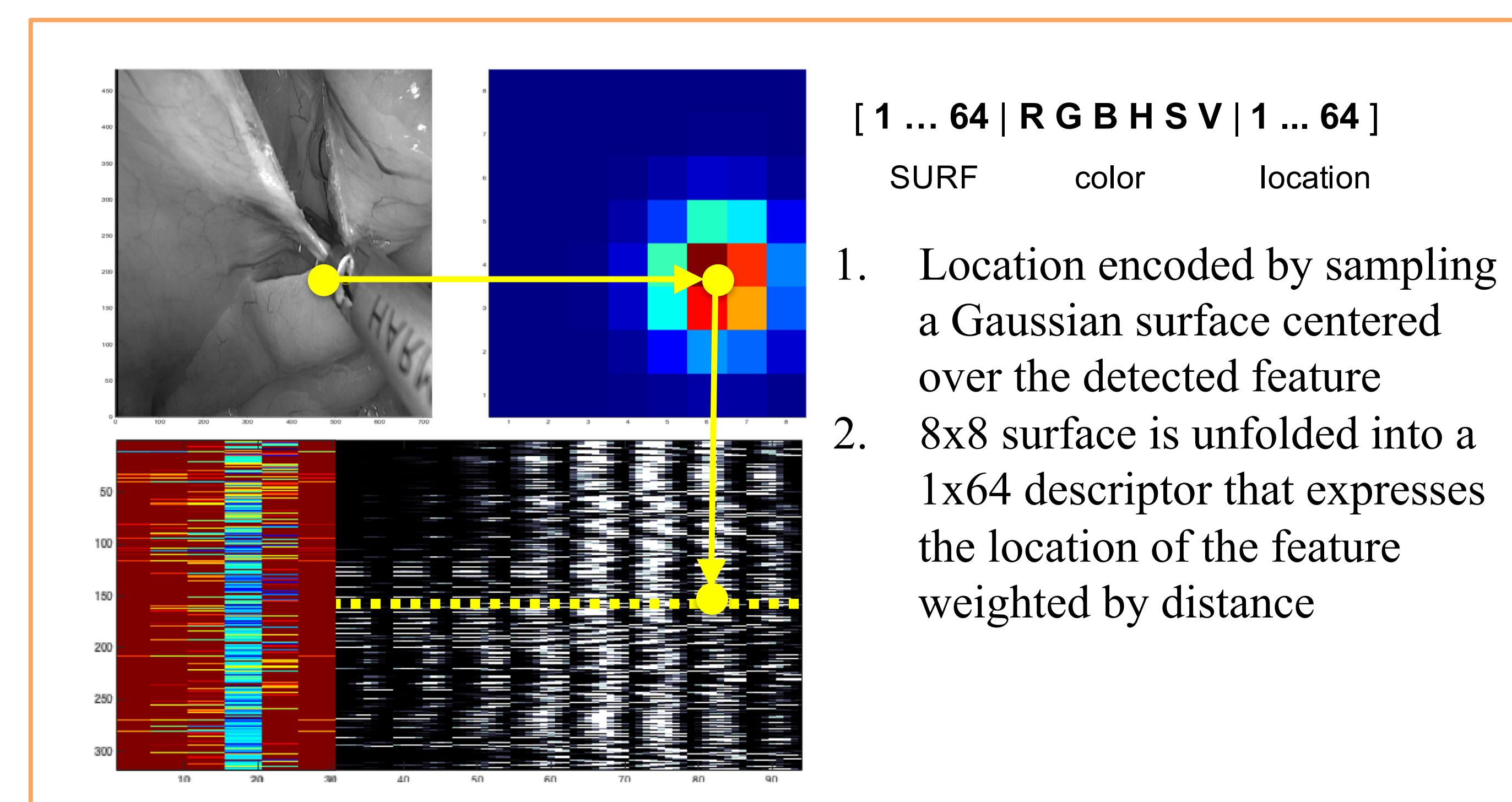
- IRB approved
- 10 videos, over 9 hours of laparoscopic surgery video recorded at MGH
- Learning 7 operational phases from LSG procedures:

Expert annotations (ground truth)

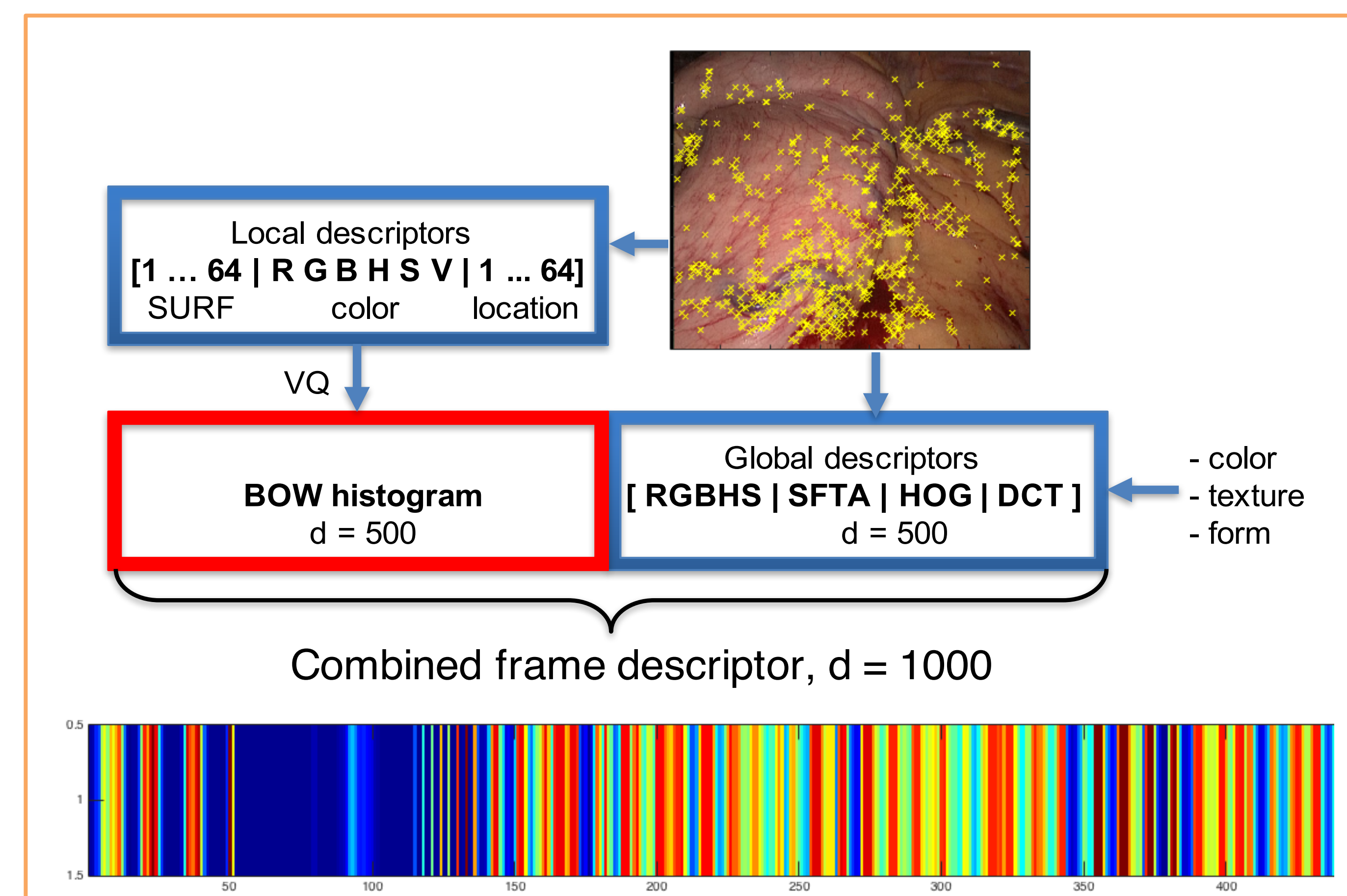
Phases	Video 1	Video 2	Video 3	...
1. port -----	Phase 1	$T(1,0)$	0	0
2. biopsy & liver retraction -----	Phase 2	$T(1,1)$	$T(1,1)$	$T(3,1)$
3. omentum removal, dissection along greater curve, hiatus inspection -----	Phase 3	$T(1,2)$	$T(2,1)$	$T(3,2)$
4. stapling -----	Phase 4	∞	$T(3,1)$	$T(3,3)$
5. bagging -----	Phase 5	∞	$T(4,1)$	$T(3,4)$
6. irrigation -----	Phase 6	∞	$T(5,1)$	$T(3,5)$
7. finish -----	Phase 7	∞	$T(6,1)$	$T(3,6)$

METHODS

- k -segment coreset using adapted visual feature space under the bags-of-words model (BOW)
- Support vector machine (SVM) classifier
- Augmented local descriptors:



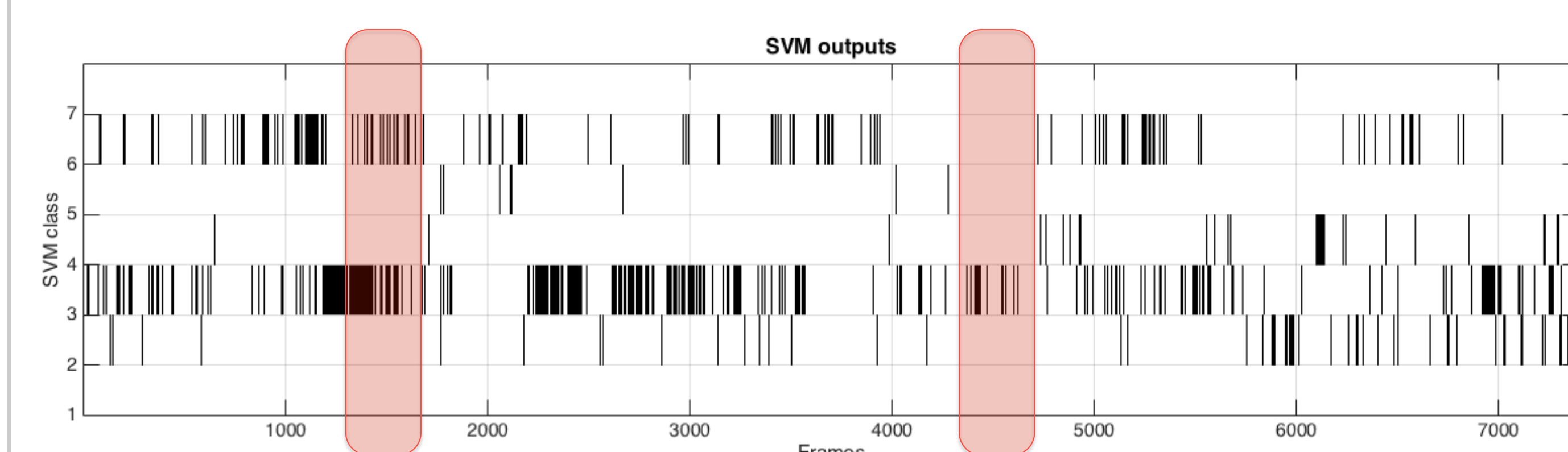
- Augmenting global descriptors:
 - Color: RGB/HSV histogram
 - Texture: SFTA texture descriptor
 - Form: HOG



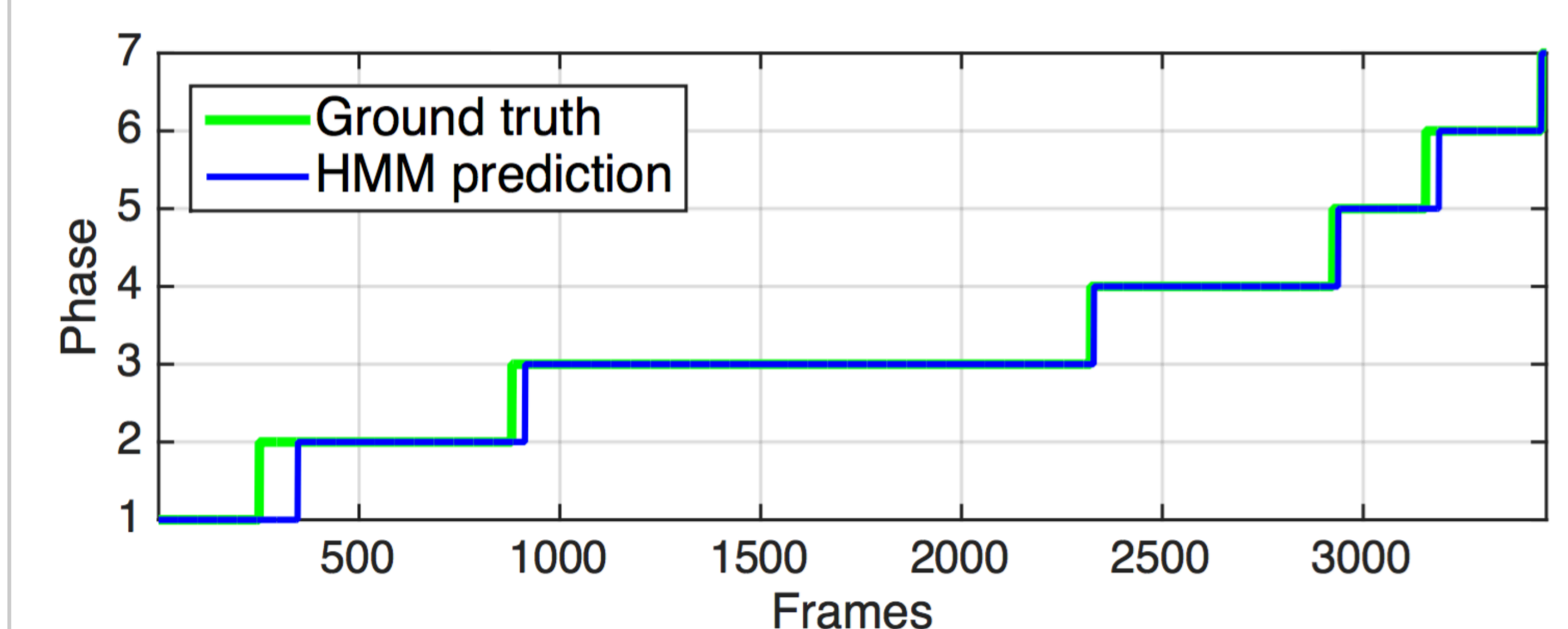
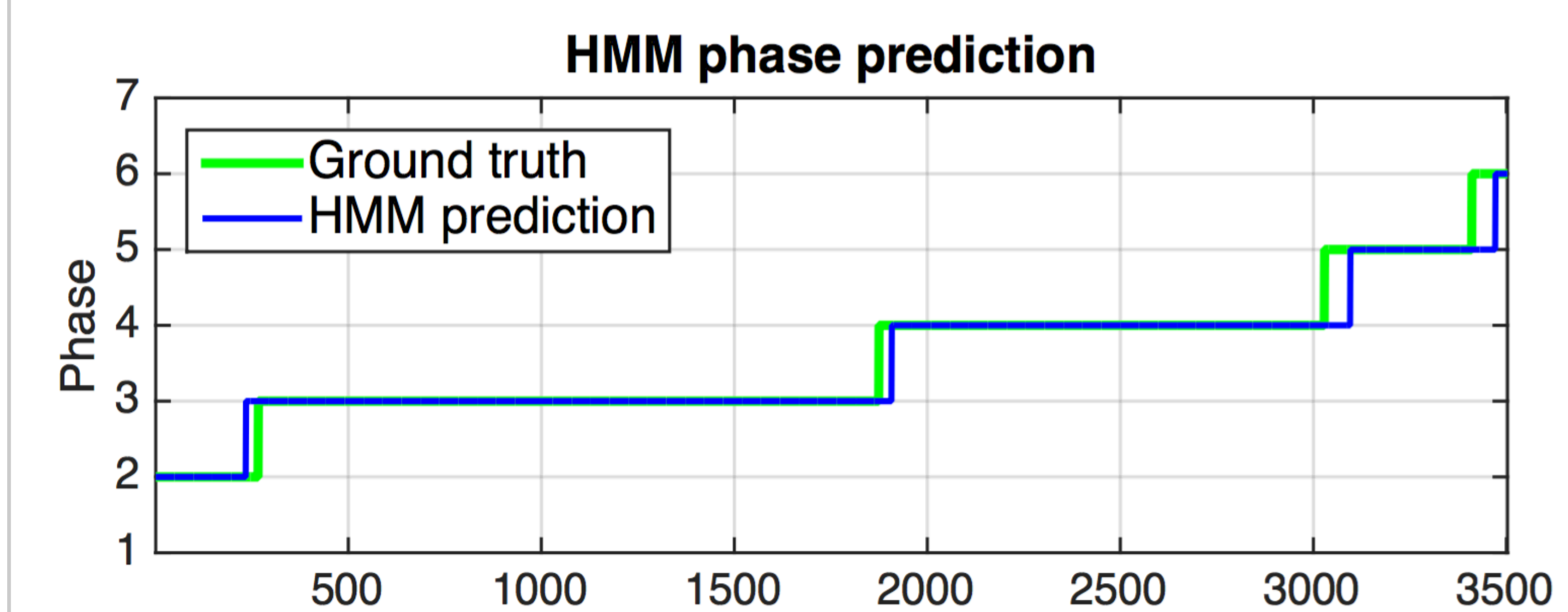
RESULTS

- Trained on each 9/10 videos (80/20 training/validation) and tested on 10th video
- Classifier: SVM cascade
- Each trained on phase $p = 1, P \setminus \{p\} = 0$
- Additionally, there are 2 ambiguous cases:

1. Multiple classifiers identify phase $y_i(t) = 1$
2. No outputs produced by any SVM



- Correcting predictions with HMM
- We use Viterbi algorithm on the emission sequence to find the most likely sequence of hidden states, given the emission sequence
- **90+% coreset compression**
- **93.6% prediction accuracy**



CONCLUSION

- Automated segmentation of surgical videos would allow important and interesting parts of large video sequences to be efficiently indexed for use in training and during time-critical expert consultations for targeted feedback and coaching.
- Using state-of-the-art machine learning methods trained on surgical video annotations, we demonstrate the use of coresets to automatically segment and classify a surgical video according to its constituent phases.
- We demonstrate 90%+ compression, 93.6% accuracy.