Hybrid Activity and Plan Recognition for Video Streams

Roger Granada, Ramon Fraga Pereira, Juarez Monteiro,
Rodrigo Barros, Duncan Ruiz, Felipe Meneguzzi

Pontifical Catholic University of Rio Grande do Sul, Brazil
{roger.granada, ramon.pereira, juarez.santos}@acad.pucrs.br
{rodrigo.barros, duncan.ruiz, felipe.meneguzzi}@pucrs.br

February, 2017

Introduction

- **Plan recognition**

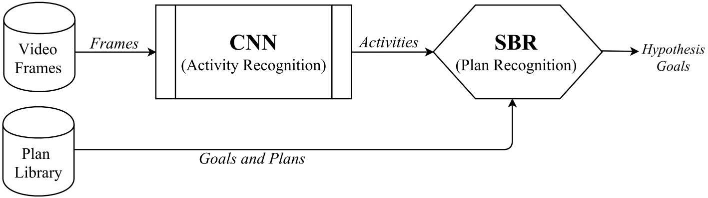
  Task of recognizing the plan (i.e., the sequence of actions) the observed agent is following in order to achieve his intention (Sadri, 2012)
- **Activity recognition**

  The task of recognizing the independent set of actions that generates an interpretation to the movement that is being performed (Poppe, 2010)
- Much research effort focuses on activity and plan recognition as separate challenges;
- We develop a hybrid approach that comprises both activity and plan recognition;
- The approach infers, from a set of candidate plans, which plan a human subject is pursuing based exclusively on fixed-camera video.

Poppe, R. A survey on vision-based human action recognition. Image and Vision Computing 28(6), pp. 976–990, 2010.
Sadri, Fariba. Intention Recognition in Agents for Ambient Intelligence: Logic-Based Approaches. Ambient Intelligence and Smart Environments, pp. 197-236, 2012.

A Hybrid Architecture for Activity and Plan Recognition

- **Conceptually divided in two main parts**
  - CNN-based activity recognition (CNN)
  - CNN-backed symbolic plan recognition (SBR)

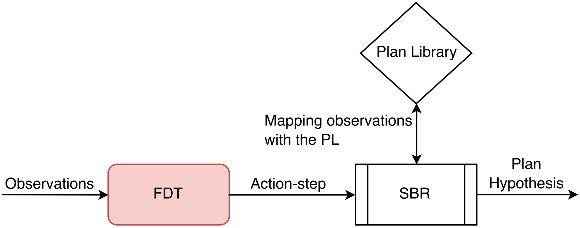


CNN-based Activity Recognition

- **Convolutional Neural Network**
  - Architecture: GoogLeNet
  - 22-layer deep network based on the Inception module
  - Input images: 224x224 (3 channels: RGB)
  - Output classes: 9 (activities)

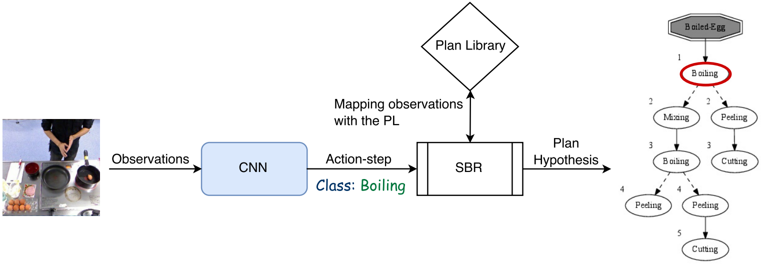
CNN-backed Symbolic Plan Recognition

- **Symbolic Behavior Recognition (SBR)**
  - A plan recognition approach that takes as input a plan library and a sequence of observations
  - Feature decision tree (FDT) maps observable features to plan-steps in a plan library
  - SBR returns set of hypotheses plans such that each hypothesis represents a plan that achieves a top-level goal in a plan library



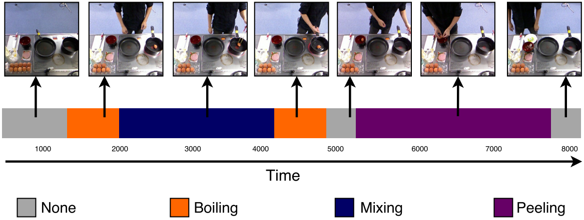
CNN-backed Symbolic Plan Recognition

- **Our Symbolic Behavior Recognition**
  - We modify the SBR and replace the FDT with the CNN-backed Activity Recognition
  - The CNN-backed Activity Recognition maps frames directly into nodes (activities) in the plan library used by SBR to compute sequential consistency of plan steps



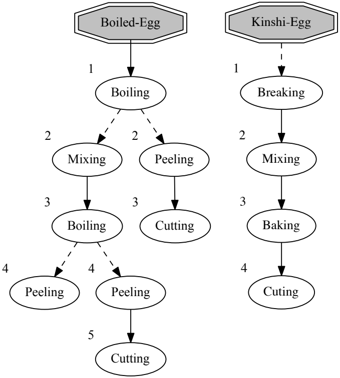
Experiments

- **Dataset**
  - ICPR 2012 Kitchen Scene Context based Gesture Recognition dataset (KSCGR)
- **5 recipes for cooking eggs in Japan**
  - Ham and Eggs, Omelet, Scrambled-Egg, Boiled-Egg and Kinshi-Tamago
  - Each recipe is performed by 7 subjects (5 in training set, 2 in testing set)
- **9 cooking activities composes the dataset**
  - Breaking, mixing, baking, turning, cutting, boiling, seasoning, peeling, and none



Experiments

- **Plan Library Modeling**
  - We model a plan library containing knowledge of the agent's possible goals and plans based on the KSCGR dataset
  - We consider that a sequence of cooking gestures is analogous to a sequence of a plan in the plan library



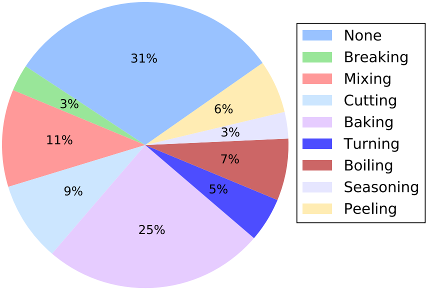
- **Activity Recognition results**
  - Precision, Recall, F-measure and Accuracy scores for each activity

Activity	Precision	Recall	F-measure	Accuracy
None	0.65	**0.97**	0.78	0.64
Breaking	0.44	0.41	0.42	0.27
Mixing	0.67	0.34	0.45	0.29
Baking	0.74	0.88	**0.80**	**0.67**
Turning	0.77	0.38	0.51	0.34
Cutting	0.87	0.63	0.73	0.58
Boiling	0.61	0.34	0.43	0.28
Seasoning	**0.89**	0.37	0.52	0.35
Peeling	0.72	0.10	0.18	0.09

Results

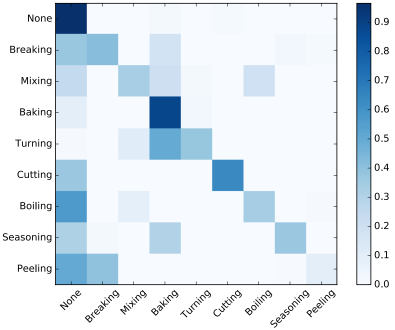
- **Activity Recognition results**
  - Accuracy scores for each activity and the distribution of frames in KSCGR dataset

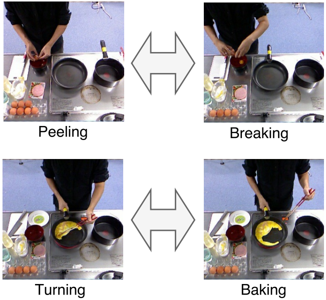
Activity	Frames	Accuracy
None	**31%**	**0.64**
Breaking	3%	0.27
Mixing	11%	0.29
Baking	**25%**	**0.67**
Turning	5%	0.34
Cutting	9%	0.58
Boiling	7%	0.28
Seasoning	3%	0.35
Peeling	6%	0.09



Results

- **Activity Recognition results**
  - Confusion matrix





  - Close position in the scene
  - Similar movements

- **Plan Recognition results**
  - We evaluate the whole pipeline using the number of hypotheses inferred by the plan recognizer
  - **Score** weights correct predictions by the number of hypotheses

$$Score = \frac{c}{\#recipesFromSBR}$$

  - c: 1 if the correct recipe was inferred, 0 otherwise
  - $\#$recipesFromSBR: Number of recipes yielded by the recognizer

Results

- **Plan Recognition results**

#	True Recipe	Predicted Recipes	Score
10	Boiled-Egg	Scramble-Egg, Omelette, Ham-Egg	0.00
	Ham-Egg	Scramble-Egg, Omelette	0.00
	Kinshi-Egg	Kinshi-Egg	1.00
	Omelette	Scramble-Egg, Omelette	0.50
	Scramble-Egg	Ham-Egg	0.00
11	Boiled-Egg	Kinshi-Egg, Omelette, Ham-Egg	0.00
	Ham-Egg	Scramble-Egg	0.00
	Kinshi-Egg	Scramble-Egg, Omelette, Ham-Egg	0.00
	Omelette	Kinshi-Egg, Scramble-Egg, Omelette, Ham-Egg	0.25
	Scramble-Egg	Kinshi-Egg	0.00
		Average:	0.18

Conclusion and Future Work

- We developed a hybrid architecture for activity and plan recognition
- Our pipeline includes:
  - a convolutional Neural Network (CNN) for activity recognition that feeds directly into
  - a modified Symbolic Behavior Recognition (SBR) approach that works with the CNN to identify the goal that describes the sequence of activities
- There are limitations of using a plan library in the plan recognizer

- Employ other deep learning architectures such as Long-Short Term Memory networks (LSTM) and 3D CNNs
- Use a more flexible approach for plan recognition, such as planning-based plan recognition
- Explore object recognition to provide additional clues of the activity that is being performed

- Demo video: `https://youtu.be/BoiLjU1vg3E`

Thank you!
Questions?