# HAPRec: Hybrid Activity and Plan Recognizer

**Roger Granada** and **Ramon Fraga Pereira** and **Juarez Monteiro** and **Leonardo Amado**
**Rodrigo Barros** and **Duncan Ruiz** and **Felipe Meneguzzi**
School of Informatics - Pontifical Catholic University of Rio Grande do Sul
Porto Alegre - RS, Brazil
{roger.granada, ramon.pereira, juarez.santos, leonardo.amado}@acad.pucrs.br
{rodrigo.barros, duncan.ruiz, felipe.meneguzzi}@pucrs.br

## Abstract

Computer-based assistants have recently attracted much interest due to its applicability to ambient assisted living. Such assistants have to detect and recognize the high-level activities and goals performed by the assisted human beings. In this work, we demonstrate activity recognition in an indoor environment in order to identify the goal towards which the subject of the video is pursuing. Our hybrid approach combines an action recognition module and a goal recognition algorithm to identify the ultimate goal of the subject in the video.

Figure 1: Pipeline of the hybrid architecture for activity and plan recognition.

## 1  Introduction

Activity recognition can be understood as the task of recognizing the independent set of actions that generates an interpretation of a movement that is being performed. On the other hand, plan recognition can be understood as the task of recognizing agent goals and plans based on observed interactions in an environment. These observed interactions can be either events provided by sensors or actions/activities performed by an agent.

Although much research effort focuses on activity and plan recognition as separate challenges, comparatively less effort focused on attempting to identify higher-level plans from activities in video sequences, *i.e.*, try to understand the overarching goal of subjects within a video and make the correct inference from the observed activities. Rafferty *et al.* (2017) use sensor based approach to implement assistive smart homes. Their approach is based upon an intention recognition mechanism that uses sensors affixed to objects and an ontological rule-based goal recognition system. Massardi *et al.* (2019) performs plan recognition using plan libraries from learned activities. Their top-down approach uses a particle filter with a population of plan trees to deal with noisy observations, producing a quick reliable solution.

In this work, we develop a hybrid approach that comprises both activity and plan recognition that identifies, from a set of candidate plans, which plan a human subject is pursuing based exclusively on still-camera video sequences. To recognize such plan, we employ an activity recognition algorithm
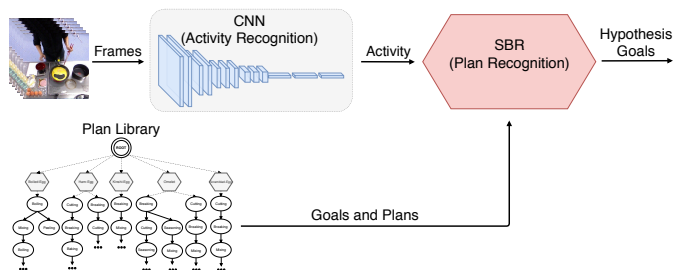
based on convolutional neural networks (CNN), which generates a sequence of activities that are checked for temporal consistency against a plan library using a symbolic plan recognition approach modified to work with a CNN. As supplemental material we provide a video demonstration of our architecture[1].

## 2  A Hybrid Architecture for Activity and Plan Recognition

Our hybrid architecture is divided in two main parts: i) CNN-based activity recognition, and ii) CNN-backed symbolic plan recognition. The first part consists of training a Convolutional Neural Network (CNN) using video frames as input, with the activity being done in the video as the expected output. Our CNN architecture is based on GoogLeNet architecture (Szegedy et al. 2015) and computes a probability score for all possible classes (*softmax* output) for each frame. If two classes contain a high probability and the difference between them is lower than a threshold ($\theta$), we use an heuristic approach to disambiguate classes. This heuristic consists of assigning the class of the last frame to the current frame in case one of the two classes is equal to the class of the last frame. Otherwise, the current frame receives the class that contains the highest probability, disregarding the threshold.

After using the CNN to identify the activity being pursued, we use a plan recognizer that returns a set of possible plans that are temporally consistent with what is recognized

---

[1]Link to our video: http://shorturl.at/agtBN

from the input frames. To perform the task of plan recognition, we use a symbolic plan recognition approach called Symbolic Behavior Recognition (SBR). SBR (Avrahami-Zilberbrand and Kaminka 2005) is a plan recognition approach that takes as input a plan library and a sequence of observations, in this case, a sequence of observed feature values. Feature values are used as a set of conditions to execute a plan-step in a plan library. To match observed features with plan-steps in a plan library, we use an efficient matching step that maps observed features with matching plan-step nodes in a plan library. To do so, they use a feature decision tree (FDT) that maps observable features to plan-steps in a plan library. As output, SBR returns set of hypotheses plans such that each hypothesis represents a plan that achieves a top-level goal in a plan library. Instead of using the FDT to match observations with consistent plan-steps in the plan library, we modify the SBR and replace the FDT with the CNN-backed Activity Recognition. For instance, given a video frame, the CNN-based Activity Recognition returns which activity such video frame corresponds, and subsequently, we take this activity as input to the SBR, as shown in Figure 1. Note that to recognize goals and plans using the SBR, we must model a plan library containing a set of possible sequence of activities (*i.e.*, plan) that achieves goals. In this paper, a plan library corresponds to a model that contains a set of plans to achieve cooking menus.

## 3  Application

We create *HAPRec* (Granada et al. 2017) for demonstrating that it is possible to perform goal recognition using CNN-based activity recognition and plan libraries with real-world data (images). To demonstrate our work, we use the activities from ICPR 2012 Kitchen Scene Context based Gesture Recognition dataset (KSCGR) (Shimada et al. 2013), which contains video sequences of five menus for cooking eggs in Japan: *Ham and Eggs*, *Omelet*, *Scrambled Egg*, *Boiled Egg*, and *Kinshi-Tamago*. Each menu is performed by 7 subjects: 5 actors in training datasets and 2 actors in evaluation datasets, *i.e.*, 5 cooking scenes are available for each training menu. Eight cooking gestures composes the dataset: *breaking*, *mixing*, *baking*, *turning*, *cutting*, *boiling*, *seasoning*, *peeling*, and *none*, where *none* means that there is not an activity being performed in the current frame. We chose the KSCGR dataset since it contains the activity being performed in each frame (*e.g.*, *breaking*, *baking* and *turning*) as well as the goal achieved in the whole video sequence (*e.g.*, preparing *Ham and Eggs*, *Omelet*, *Scrambled Egg*, *etc.*). Thus, we can carry out activity recognition using activities performed in each frame and plan recognition using the steps to achieve the recipe in each video.

For recognizing goals and plans, we model a plan library containing knowledge of the agent's possible goals and plans based on the dataset, where each recipe is a top-level goal in the plan library. Based on videos from the training set, we model all possible plans for achieving each possible menu (*i.e.*, top-level goal). We consider that a sequence of cooking gestures is analogous to a sequence of plan-steps, *i.e.*, a plan in the plan library. Figure 2 illustrates the demo screen, showing the current image of the dataset, its frame id, the
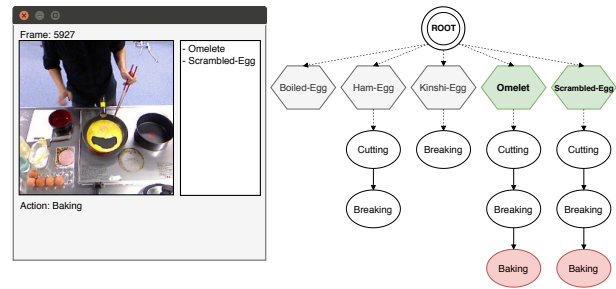


Figure 2: Demo screen showing the activity identified in the current frame, the plan-steps achieved up to that frame, and the set of candidate goals.

action predicted in that frame (*Baking*), and the list of candidate goals (*Omelet* and *Scrambled-Egg*). On the right side, the sequence of plan-steps identified and the top-level goals, where the candidate goals are highlighted in green.

## 4  Conclusion

We presented *HAPRec*, a tool that performs both activity and plan recognition using real-world data. Our architecture is based on CNNs and a modified symbolic approach to plan recognition. We demonstrated how the algorithm works by testing using a kitchen scene environment containing actions performed by subjects and plans (recipes).

## Acknowledgement

## References

Avrahami-Zilberbrand, D., and Kaminka, G. A. 2005. Fast and Complete Symbolic Plan Recognition. In *IJCAI'05*.

Granada, R.; Pereira, R. F.; Monteiro, J.; Barros, R.; Ruiz, D.; and Meneguzzi, F. 2017. Hybrid Activity and Plan Recognition for Video Streams. In *The AAAI 2017 Workshop on Plan, Activity, and Intent Recognition*.

Massardi, J.; Gravel, M.; and Beaudry, E. 2019. Error-tolerant anytime approach to plan recognition using a particle filter. In *ICAPS 2019*, 284–291.

Rafferty, J.; Nugent, C. D.; Liu, J.; and Chen, L. 2017. From activity recognition to intention recognition for assisted living within smart homes. *IEEE THMS* 47(3):368–379.

Shimada, A.; Kondo, K.; Deguchi, D.; Morin, G.; and Stern, H. 2013. Kitchen scene context based gesture recognition: A contest in icpr2012. In *Advances in depth image analysis and applications*. Springer. 168–185.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2015. Going deeper with convolutions. In *CVPR 2015*.