

論文 / 著書情報
Article / Book Information

題目(和文)	Plan-Intention-Event Framework for Scene Analysis Based on Robot Audition and Plan Recognition (summary)
Title(English)	ロボット聴覚とプラン認識に基づく環境理解のためのPlan-Intention-Eventフレームワーク (要旨)
著者(和文)	小島 諒介
Author(English)	Ryosuke Kojima
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10554号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:井村 順一,天谷 賢治,中島 求,早川 朋久,篠田 浩一,中臺 一博
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10554号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

専攻:	情報環境学	専攻
Department of		
学生氏名:	小島 諒介	
Student's Name		

申請学位 (専攻分野):	博士 (工学)
Academic Degree Requested	Doctor of
指導教員 (主):	井村 順一
Academic Advisor(main)	
指導教員 (副):	中臺 一博
Academic Advisor(sub)	

要旨 (英文 800 語程度)

Thesis Summary (approx.800 English Words)

This dissertation addresses scene analysis, which is essential for environmental monitoring and understanding. The main issues for scene analysis are (Issue 1) to extract attributes of an event and (Issue 2) to reconstruct scene structure from signal data obtained from sensors. The first issue has often been assessed for each attribute; for example, attributes of a sound event can be extracted by different techniques including sound source detection, localization, and identification. The second issue is challenging and only limited research has been reported. More sophisticated analysis requires appropriate integration of these techniques and reconstruction of scene structure in many applications. To fulfill these requirements, we proposed a ‘Plan-Intention-Event framework’ (PIE framework) by combining event extraction with plan/intention recognition. This framework provides design guidelines for scene analysis systems by taking Issues 1 and 2 into consideration. This dissertation also shows the effectiveness of applying the PIE framework to three attractive applications: cooking recognition, bird song analysis, and web session log analysis.

Regarding (Issue 1), this dissertation focuses on how to extract attributes of an event in the presence of sensory noises and obstacles. To solve this problem, the PIE framework provides two approaches: (1-a) multi-sensory and (1-b) multi-attribute approaches. The first approach (1-a) exploits multimodal information, e.g., utilizing multiple audio and visual sensors such as microphones and cameras. To evaluate this approach, an audio-visual multimodal cooking recognition system is introduced in this dissertation. This system extracts audio-visual events utilizing a Convolutional Neural Network (CNN). Experimental results showed the robustness of this system in noisy and/or occluded situations. The second approach (1-b) involves the integration of different methods of extracting attributes. This approach can be realized by extending and integrating existing technologies. This dissertation focuses on robot audition technologies, which provide extraction of sound attributes using sound source detection, localization, separation and identification. In robot audition, these technologies are performed separately, and connected in a cascade manner, which often results in poor performance. To tackle this problem, we propose a new Spatial-Cue-Based Probabilistic Model (SCBPM), which makes it possible to extract these sound attributes more precisely than the conventional methods dealing with attributes separately and their cascade-type integration. The effectiveness of the method is shown in a bird song analysis application.

(Issue 2) can be solved by the reconstruction of a scene by connecting the extracted events to be consistent with background knowledge. Plan/intention recognition is applicable to solve such an issue. However, two issues remain: (2-a) acquisition of background knowledge and (2-b) incompleteness of the events. Issue (2-a) can be solved by obtaining background knowledge from documents on the web. This task is challenging because these documents are not well-structured. The first step is the selection of a cooking domain; that is, a set of recipes on websites, followed by the construction of background knowledge based on a Hierarchical Hidden Markov Model (HHMM), which is often used in probabilistic plan/intention recognition. To show its effectiveness, this construction of background knowledge was applied to multimodal cooking recognition. The latter problem (2-b) occurs in realistic situations such as online plan/intention recognition. We propose a new method for plan/intention recognition that can deal with incomplete data using probabilistic context-free grammar (PCFG), a more flexible model than HHMMs. The proposed method introduces two types of computation for a prefix probability for PCFGs, and a plan and intention with the highest likelihood. This method was evaluated by web session log analysis, an application that can improve websites and monitor visitors of websites. The most likely plans were computed from logs of web sessions, and the intentions of website visitors were estimated. We successfully show the superiority of these methods compared with other methods without plan recognition.

Chapter 1 introduces the background, motivation, contributions, and organization of the dissertation. Chapter 2 provides a review of literature related to scene analysis. First, research related to visual and auditory processing is described briefly, followed by the description of fundamental techniques in scene analysis, including probabilistic models with SCBPM and plan/intention recognition. Chapter 3 describes the PIE framework and an example of this framework. Chapter 4 explains cooking recognition as an application of the PIE framework. Tasks in this application include (1-a) audio-visual event recognition using CNN and (2-a) construction of HHMM to represent information of recipes from websites. Chapter 5 describes an analysis of bird song that can aid ecologists. Robot audition techniques are integrated with the SCBPM (1-b). Chapter 6 addresses plan/intention recognition in relation to web session log analysis. A new method is proposed to deal with incomplete data (2-b) in realistic situations. Chapter 7 discusses the applicability of the PIE framework and remaining issues and concludes this dissertation.