

The build of new short-term social action DB using RGB-D cameras

Ho-Sub Yoon · Jaeyoon Jang · Jaehong Kim
 Electronics and Telecommunications Research Institute, Daejeon, Korea
 {yoonhs, jangjy, jhkim504}@etri.re.kr

Abstract There are so many action DBs including 3D depth information for the test and verification of action recognition technology. However, there are rare databases of the focused on social behavior recognition between people and people or people and robots. In this paper we aim to construct a new action databased that is specified to these social behaviors. To do this, we define various short-term social behaviors in groups according to their meaning, and describe the surrounding environment for acquiring them. Also, in order to develop a recognition system based on deep learning algorithm in the future, a DB generation method considering various ages and angles of cameras is described in order to acquire a large-capacity database. Finally, our short-term social action DB proposed in this paper will play an important role in recognizing and understanding the social interaction between human and robot.

Keywords Social action DB, RGB-D cameras, Short-term social action

1 Introduction

In *sociology*[1], **social action**, also known as "**Weberian social action**", refers to an act which takes into account the actions and reactions of **individuals** (or '**agents**'). According to **Max Weber**, "an Action is 'social' if the acting individual takes account of the behavior of others and is thereby oriented in its course". Our research aims to study social behaviors between human and human beings and to enable social behaviors between human and robots being based on them. To do this, we need to build a social behavior database of the first person viewpoint between people and people.

There are many existing databases of human behavior, not social behavior. Example, the kinetics human action video dataset from Will Kay[2] introduce a new, large, video dataset for human action classification. The dataset contains 400 human action classes, with at least 400 video clips for each action from a different YouTube video. Another recent human action datasets are shown table 1.

Table 1. Statistics for recent action recognition dataset

Dataset	Year	Actions	Clips	Total	Videos
HMDB-51	2011	51	min 102	6,766	3,312
UCF-101	2012	101	min 101	13,320	2,500
ActivityNet-200	2015	200	avg 141	28,108	19,994
Kinetics	2017	400	min 400	306,245	306,245

The "Head banging, shaking hand, robot dancing, riding bike, playing violin, braiding hair, dribbling basketball

and stretching leg" are example classes from the kinetics dataset. These type of action class is aimed for general kinetics human action recognition.

2 Definition and Grouping of social action DB

2.1 Definition

Short-term social action DB means the acting DB individual takes account of the behavior of others and is thereby oriented in its course. In this case, short-term means temporal behavior within 10 seconds that can easily identify intent. Contrary, long-term means that can be gained through long-term observations such as habits and attitudes.

2.2 Grouping

We defines 6 groups of social action as shown as below. Although this method of grouping is not widely defined, this study proposes a way to classify behaviors with similar social meanings into one group. That is, a group consists of a set of different behaviors with similar meanings.

- 1) Greeting(4):
{Bow one's head, Shake hands, Give out the handshake, Snuggle}
- 2) Agreement(3):
{Nod, Rolling head, Tilting head}
- 3) Encouragement(4):
{Clench fist, Stick out thumb, Make O.K finger sign, Hug}
- 4) Thinking(3):
{Fold arms, Scratch head, Chin in hand}
- 5) Eye contact(3):
{View the philtrum, View the pupil, View the air}
- 6) Facial emotion expression(6):
{Smiling, Laugh, Frown, Angry, Sad, Surprise}

3 Making of social action DB

In order to create a social action database, we wanted to be able to get video from many generations of people. From elementary school students to 60s or older, the images are collected for a total of 140 subjects, 10 men and 10 women at 10-year intervals. At this time, each person repeats action 10 times for 23 social actions of the above 6 groups. The agent is a person who act as a robot. If we had a humanoid robot that played a perfect human role, then the agent would have become humanoid robot, but it is impossible now. Each of the 140 consumers is tagged with their respective profile information. For example, information such as age, gender, hair style, lipstick color, clothes style, clothes color,

accessories, and the like are stored. Such human profile information can give important clues to understand the person's social behavior.

Figure 1 shows the environment of a social action DB generation. The set 1 datasets are used for training and verification based on robot viewpoint. The set 2 datasets are aimed for learning based on observer's view. These dataset are captured by RGB-D cameras as MS Kinect 2 and Intel Realsense D435 with high resolution 1280x720.

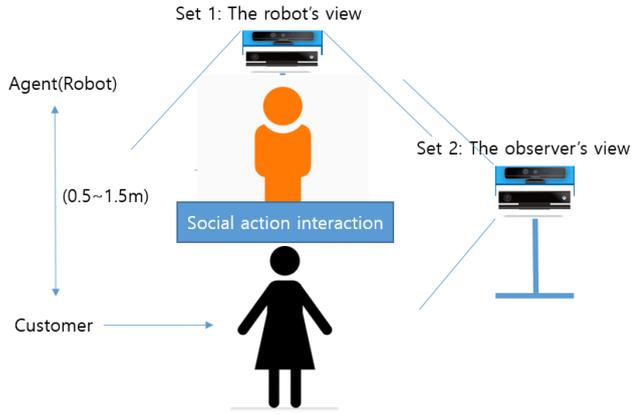


Fig. 1. The environment of a social action DB generation

4 Conclusions

In this paper we aim to construct a new action databased that is specified to social behaviors. To do this, we define various short-term social behaviors in groups according to their meaning, and describe the surrounding environment for acquiring them. Our short-term social action DB will play an important role in recognizing and understanding the social interaction between human and robot.

Acknowledgement

This work was supported by the IT R&D program of MOTIE/KEIT [10077553], Development of Social Robot Intelligence for Social Human-Robot Interaction of Service Robots.

References

1. Fadul, J. and Estoque, R. *A Textbook for an Introductory Course in Sociology*. Lulu Press, 2010.
2. Will Kay and etc., "The Kinetics Human Action Video Dataset", arXiv:17050v1, 19 May, 2017.