Ontologies of Action and Object in Home Environment towards Injury Prevention

Satoshi Nishimura
Shusaku Egami
satoshi.nishimura@aist.go.jp
s-egami@aist.go.jp
National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo, Japan

Takanori Ugai
ugai@fujitsu.com
National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo, Japan
Fujitsu Limited
Kawasaki, Japan

Mikiko Oono
Koji Kitamura
Ken Fukuda
mikiko-oono@aist.go.jp
k.kitamura@aist.go.jp
ken.fukuda@aist.go.jp
National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo, Japan

ABSTRACT

It is one of the critical applications for human-centric artificial intelligence that surveys the risky situation and inferring ways to prevent it by storing the situation information from surveillance cameras. Recognition of human activities in daily situations is an emerging topic in the computer vision domain. Significantly, the context information, such as objects involved in activities and the relationships between the objects and the activities, are attractive to improve the accuracy of the activity recognition task. However, the existing labels for actions and objects are not well considered for describing daily activities. This short research paper provides the ontologies of actions and objects in the home environment, so-called Primitive Action ontology, and Home Object ontology. The Primitive Action ontology contains a minimal set of primitive actions designed to abstract actions and discards objects and methods. The Home Object ontology has object types and properties such as affordance and attributes to describe daily situations. The properties represent both normal and abnormal effects, including intentional function and incidents in the home environment. We also discuss the prospect of using these ontologies as the conclusion of this paper.

CCS CONCEPTS

• **Computing methodologies** → **Ontology engineering**; **Semantic networks**.

KEYWORDS

ontology; knowledge graph; daily living activity; older adults

ACM Reference Format:

Satoshi Nishimura, Shusaku Egami, Takanori Ugai, Mikiko Oono, Koji Kitamura, and Ken Fukuda. 2021. Ontologies of Action and Object in Home Environment towards Injury Prevention. In *The 10th International Joint Conference on Knowledge Graphs (IJCKG'21), December 6–8, 2021, Virtual Event, Thailand*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3502223.3502239

IJCKG'21, December 6–8, 2021, Virtual Event, Thailand

ACM ISBN 978-1-4503-9565-6/21/12…$15.00
https://doi.org/10.1145/3502223.3502239

1 INTRODUCTION

One of the critical applications for human-centric artificial intelligence (AI)[3, 8, 10] is surveying the risky situation and inferring ways to prevent the danger by storing the situation information from surveillance cameras. Recognition of human activities in daily situations is an emerging topic in the computer vision domain that can be used as one component of human-centric AI. Context information helps the recognition by machine learning methods improve its accuracy. Sigurdsson et al. provide an experiment on what features are weighed by the latest deep neural network (DNN) based methods[18]. The result shows that the combination of co-occurrent objects and activities contributes to improving accuracy. Ji et al. added the relation information between an actor and the objects on the results, and they proved such relation contributes to improving the activity recognition[9].

This short research paper proposes ontology construction of actions and objects. First, we propose a Primitive action ontology that can abstract the activity label of the activity recognition dataset by separating actions and objects. Second, we propose a Home object ontology that contains objects in the home environment. It is intended to use with Primitive action ontology to represent the situation of daily life. Third, we designed properties of objects to describe attributes, functions, and affordances. This information describes the semantic perspective in activity recognition tasks; for instance, standing on a bed rarely occurs but lying on a bed occurs more often. In section 2, we propose the way to construct ontology for action and object. In section 3, we describe the outline of constructed ontology, which are Primitive action ontology and Home object ontology. Section 4 describes some preliminary results to apply Primitive action ontology for activity recognition. Section 5 describes related works regarding ontologies for action and object. Section 6 describes how the ontology helps to improve activity recognition as our potential application. Finally, in section 7, we give a summary and discuss future works. This paper shows a preliminary result and prospects related to the ontologies. For the first perspective, an evaluation of Primitive action ontology is provided. The Others are remaining issues.

2 ONTOLOGY CONSTRUCTION

2.1 Ontology construction of actions

The criteria of Primitive action are as follows. First, we collect the candidate vocabulary. We extracted the candidate vocabulary from HomeOntology[22], which was created based on the data used in the simulation environment VirtualHome[17] and International Classification of Functioning, Disability and Health (ICF)[16], which the World Health Organization created to describe health-related information. Primarily, we used the "activity and participation" part of the ICF to focus on action. Second, we classified the vocabulary into six classes, considering the criteria to limit the conceptual space in focus. The classification was performed based on the number of objects and persons that participated in the action. In this study, objects refer to abstract and concrete things, such as conversation content and a chair. Third, we took two hours per the focused vocabulary to discuss the criteria of Primitive action. The discussion was performed based on each term in the candidate vocabulary, and we decided which term is classified as Primitive action. After the discussion, we obtained the inclusion/exclusion criteria of Primitive action as the following list.

(1) Include actions that change or maintain the object state.
(2) Include actions that are performed consciously.
(3) Exclude actions that are specialized by their target objects.
(4) Exclude actions that include interpretations of the observer.
(5) Exclude actions that consist of other Primitive actions.

The first criterion means that the Primitive action does not contain how to realize the action, such as cut by scissors. The second criterion excludes actions performed unconsciously, such as "sleep" and "wake". "sleep" can be observed as "lay down" and "wake" can be observed as "sit" (from laying down to sitting state). The third criterion means that we do not specialize Primitive action by its target object. For instance, "cup" and "upper limb" and other objects can be targeted by the action "grasp." However, we do not specify the "grasp" action to "grasp cup" and "grasp upper limb" because the vocabulary increases according to the number of objects. The fourth criterion means that we exclude the observer's interpretation. For instance, whether a person performs an action called "think" is an interpretation of the observer since only the actor knows the truth. Intentions of the actor and interpretations of the observer are challenging targets in the activity recognition task. Thus, we added this criterion to separate controversial actions from Primitive actions. The fifth criterion means that Primitive actions should be atomic. For instance, action "buy" can be decomposed into a sequence of actions, including more fine-grained actions such as "pick up money," "carry money," and "carry item." Moreover, there are other options to implement "buy," such as using digital cash. We excluded such coarse-grained terms from Primitive action.

2.2 Ontology construction of objects

2.2.1 Criteria for object collection. We extracted the candidate terms to make clear what exists in a home environment. To decide the candidate terms, we used assets data in VirtualHome[17], objects defined in Charades[19], and objects that occurred in the videos in the video archive called Elderly Behavior Library[15].

We set criteria to exclude object terms from the ontology as follows.

(1) Exclude objects classified by their attributes such as colour, size, and shape.
(2) Exclude objects classified by their purpose for what to be used.
(3) Exclude objects characterized by relative attributes.
(4) Exclude objects conceptualized by their material.

The first criterion aims to limit the number of object classes. For instance, we interpreted that yellow, small, and rounded boxes are the same class "box." Such detailed attributes will be handled after instantiation.

An example of the second criterion is that we do not specify a TV remote controller and a video player controller. We use one class, "remote controller." Another example is a container for a meal and a container for an accessory. Its purpose classifies these objects, but it is hard to recognize for computer vision and even humans.

The third criterion means that we do not conceptualize the object from a context-dependent perspective. For instance, when we assume a bi-parting door, we can recognize right-side and left-side doors. However, we cannot call the right-side door without the left-side door. Therefore, we do not make such relative concepts but call them merely "door."

The fourth criterion means that we do not classify the objects by their material, such as pottery. The concept denotes its material is clay but does not tell any shape, function, or usage of the objects.

2.2.2 Properties of an object. Robotics research uses a property "affordance" to make a robotic agent's plan[24]. The property inheres in the object and makes clear what the robot can perform to the object. Therefore, we chose "affordance" as one of the properties of the object. Function takes an important role to describe objects in the ontology domain[2]. Borgo et al. provide several definitions of technical artifact, and the function is an essential property to define the artifact from the engineering viewpoint. They also provide other properties such as "design proposition" and "plan to use" to distinguish technical artefacts from integrated perspectives, including philosophy and ontology. Our target domain is daily life; therefore, we chose only the essential property "function" to describe the object. Tarumi et al. provide a distinction of "attribute" and "property" to describe the quality of nanomaterials[21]. The distinction provides the integrated description of quantitative attributes and qualitative property that are inhered to the object. We employed the concept to describe the object flexibly.

In the end, we defined the following 5 properties: affordance, function, state, attribute, and property to characterize the object from the perspective of representing accident cases.

(1) Affordance is an action induced by an object, such as climbing up towards a step.
(2) Function is an occurrence performed by an object, such as the heating of a stove.
(3) State is a time-indexed quality that inheres in an object, such as open/close towards a door.

(4) Attribute is a quality that inheres in an object that is not changed during time-lapse, such as the height of a step.
(5) Property is a complex concept of attribute and value interpreted from the perspective of a certain threshold, such as high for a step that is a combination of attribute height and its value 50 cm with the threshold that more than 30 cm step is high for older adults.

3 CONSTRUCTED ONTOLOGIES

3.1 Primitive action ontology

We extracted the candidate terms from ICF and HomeOntology and decide 79 terms are Primitive actions according to the criteria described in Section 3.1. After the extraction, we also added two terms to the extracted terms. The ontologies are formulated in RDF[1] and published on Github[2].

3.2 Home object ontology

As explained in section 2.2.1, we extracted candidate terms from the existing resources. First, we extracted 332 objects from the asset data in VirtualHome and 37 objects from Charades. The total number of objects is 348 by removing duplicate objects. In addition to that, we extracted objects that occur in 16 videos in the Elderly Behavior Library. The objects were counted manually, and 148 new object kinds were found.

We tried to find whether the increasing number of unknown terms is saturated by gradually increasing the extracted terms. First, we extracted 7 videos from the Elderly Behavior Library. Then, the number of occurred objects and new objects are 155 and 38, respectively. The rate of the new objects and the occurred objects is 0.25. Second, we examined another nine videos. Then the number of the occurred objects and the new objects are 190 and 27, respectively. The rate of the new objects and the occurred objects is 0.14. After that, we decided the number of objects was saturated enough. Then, we got 568 object classes. Home Object Ontology is written in OWL[3], and properties are described as restrictions of the domain and the range. It is also published in Github[4].

4 PRELIMINARY EXPERIMENTAL RESULTS

4.1 Application to activity recognition

We evaluated Primitive action ontology with performing the task of activity recognition. Context information such as objects in the scene, the relation between a person and the objects, and intention are known to improve the accuracy of the activity recognition task[18]. We assume that Primitive action ontology provides abstract context to the dataset and contributes to capturing features of activities on the video. We evaluated our assumption on the Charade dataset[19] that contains 9,848 videos of around 30 seconds in length. Each video records more than two activities performed by at least one person. The number of activity classes is 157 classes. We chose Long-term Feature Bank (LFB)[23] as a baseline method. LFB captures visual cues along one video and combines them with

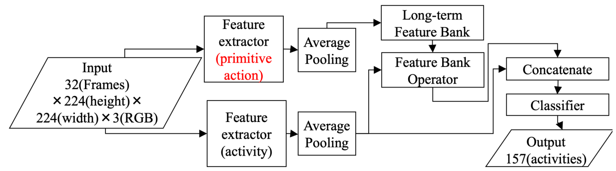

Figure 1: Network diagram

Table 1: Activity recognition on the validation set in mAP (%)

Method	mAP(%)
LFB with original training schedule	39.5
LFB with our training schedule	40.4
Ours	40.6

the focused video frame. The detector uses the long-term feature of the video to recognize activities in the specific frames. We use our ontology to annotate new labels to the Charades dataset to represent abstract action. Figure 1 shows the network diagram.

We used the Charades dataset as input data along with their division of train and validation sets. The original 157 Charades activity label trains the baseline, and the extended labels train ours with Primitive action ontology that contains 181 labels in total. We extended the training schedule from the original 24,000 iterations to 40,000 iterations because we used more labels than the original. We tried to train the baseline with the original training schedule and our training schedule to compare the results.

Table 3 reports the performance of baseline and our models using mean average precision (mAP) on Charades validation set. Extending the activity label with our Primitive action ontology outperforms the baseline model, but the difference is tiny with only 0.2 points gap. We observed that our model tends to improve the AP of the activity that the baseline could not recognize well but tends to get worse than the baseline AP where the baseline already achieves high scores. We expected our model to improve the activity recognition with a small sample by binding the features from the abstract perspective. However, the result does not show such a contribution.

5 RELATED WORKS

5.1 Smart Home

Ontologies are popularly used in the smart home domain. Sensor data is one of the significant resources to obtain daily human information. Several projects aim to develop ontologies for the Internet of Things (IoT) devices[6, 7] to make the data interoperable. The aim of these projects is mainly to provide Application Programming Interfaces (APIs). World Wide Web Consortium proposed Semantic Sensor Network Ontology[4] as sub-ontologies to modularize the interest domains. Their platforms and ontologies provide only fundamental functions and schema to handle low-level sensory data. They do not provide specific concepts to represent the context of daily life, such as actions and objects with properties.

[1] https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/
[2] https://github.com/aistairc/PrimitiveActionOntology
[3] https://www.w3.org/OWL/
[4] https://github.com/aistairc/HomeObjectOntology

For a more specific domain, such as assessing dementia people in smart homes, we can find the effort of the Aging Neuro-Behavior Ontology[13] and iKnow platform[14]. The former is an ontology to represent the activities performed by the older adults who have memory problems obtained by several kinds of sensors. The ontology does not aim to capture sensory information directly but to represent interpretations such as "get a glass from the cupboard" and "fill the glass" from motion sensors and water flow rate. iKnow platform provides recognition of activities of daily living from the sensor data in a similar way. It also models the activities by interpreting the sensory data. Compared to their ontologies, our proposed ontologies contain broader concepts in daily life.

5.2 Robotics

Several efforts were made to handle the semantics around the action and the object[1, 24] in the robotics domain. Yamanobe et al. pointed out that the concept "affordance" inhering to the concept "object" is used for robotics to recognize the object type and plan to manipulate the object. They reviewed more than 57 papers related to affordance events, concentrating on the robotic agent's "grasping" action. However, they aim to clarify how affordance is dealt with in the robotics domain and not humans' actions.

Beßler et al. provide a formal model of affordances for flexible robotic task execution[1]. The research is performed under the Everyday Activity Science and Engineering project to enable robots to perform everyday activities. They provided an OWL-based model to represent the affordance for the robots, and the competency was evaluated based on simulation. Although the model is valid enough to answer necessary queries for a robotic agent to handle unseen situations with its knowledge, they do not provide a way to extend their ontology to apply more general daily life situations.

5.3 Activity recognition

Datasets for activity recognition provide a controlled vocabulary to describe general activities. Early datasets such as HMDB[11] and UCF101[20] are simple vocabulary without dividing the action and its target object. These vocabularies depend on the videos, and these datasets cover only trivial situations that rarely occur in daily life. Charades[19] accumulate daily life situations by crowdsourcing. They tried to provide peripheral information. The activity labels still combine the action part, which represents state change, and the object part, which represents the target object of the action. The later study pointed out that the peripheral information helps the detector improve their accuracy[18], and richer information, including the relation between objects and a person, contribute to recognizing the activities[9]. Our Primitive action ontology was also evaluated based on the activity recognition task as mentioned in section 3. Our Home object ontology will contribute to activity recognition by using the similar setting of [9].

6 POTENTIAL APPLICATIONS

Collecting data on daily life situations costs a lot, and the data can be unbalanced. For instance, the Charades dataset has a significant variance among the activity labels. Such variance prevents improving the accuracy of the detector. Ontology-based data augmentation can be a solution for this problem. We have integrated our ontology as a controlled vocabulary for VirtualHome2KG[5] platform that can generate and augment knowledge graphs of daily activities based on the VirtualHome[17]. We believe that augmenting the low sample data by integrating our ontology with VirtualHome2KG will improve activity recognition accuracy.

As mentioned in section 4.3, the combination of action and its object will provide rich features for activity recognition. Ji et al. use the feature of relation between actors and objects in the same frame[9]. Li et al. use affordance information to detect the location of the human who is sitting down on a chair[12]. Such context information will help recognize rare but possible situations such as a person standing on a bed or sitting on a light. Home Object ontology will provide affordance information on each object type that relates to Primitive action ontology. The combination of both ontologies will be used for describing the dataset for activity recognition.

7 SUMMARY AND FUTURE WORKS

We proposed two ontologies to describe actions and objects in daily life situations. These ontologies have three characters as follows. First, the Primitive action ontology contains a minimal set of primitive actions. The primitive actions are designed to abstract actions discarding the object and the method how to perform. We evaluate the ontology in the activity recognition task based on the assumption that abstraction of activity improves the accuracy of activity recognition. Second, we proposed the Home object ontology that contains object types that occurred in the home environment. The candidate concepts were collected from the video archive[15] that captures the daily life of older adults in Japan. The number of concepts that we collected was 568 in the end. Third, we proposed a property set of the objects to describe their characters. We defined 5 fundamental properties based on the conventional research in multiple domains. In the current phase, we provide only a simple definition to distinct these properties. Further considerations of the detailed definition of the properties and extending the classes are remaining issues.

We discussed applying the ontologies to activity recognition. Traditionally, ontology study focused on the specification of how humans conceptualize things, the expressiveness and interoperability for computer systems. However, the trade-off between how rich the representation is for humans and how easy the representation is for Artificial Intelligence (in this context, deep neural network architectures) to "understand" the concept remains an uncovered research area to explore. We will use these to help logical inference detect dangerous situations in the home environment and prevent injury. Evaluation in this application domain is the most critical future work. Comparison to use ontology (structured-knowledge-based) with machine learning techniques (unstructured-large-amount-data-based) is also one of our future works.

ACKNOWLEDGEMENT

This paper is based on results obtained from projects, JPNP20006 and JPNP180013, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] Daniel Beßler, Robert Porzel, Mihai Pomarlan, Michael Beetz, Rainer Malaka, and John A. Bateman. 2020. A Formal Model of Affordances for Flexible Robotic

Task Execution. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020) (Frontiers in Artificial Intelligence and Applications, Vol. 325)*, Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang (Eds.). IOS Press, 2425–2432. https://doi.org/10.3233/FAIA200374

[2] Stefano Borgo, Maarten Franssen, Paweł Garbacz, Yoshinobu Kitamura, Riichiro Mizoguchi, and Pieter E. Vermaas. 2014. Technical Artifacts: An Integrated Perspective. *Appl. Ontol.* 9, 3–4 (jul 2014), 217–235.

[3] Ginevra Castellano. 2020. What Kind of Human-Centric Robotics Do We Need? Investigations from Human-Robot Interactions in Socially Assistive Scenarios. In *Proceedings of the 8th International Conference on Human-Agent Interaction* (Virtual Event, USA) *(HAI '20)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3406499.3422313

[4] Michael Compton, Payam Barnaghi, Luis Bermudez, Raúl García-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, Vincent Huang, Krzysztof Janowicz, W. David Kelsey, Danh Le Phuoc, Laurent Lefort, Myriam Leggieri, Holger Neuhaus, Andriy Nikolov, Kevin Page, Alexandre Passant, Amit Sheth, and Kerry Taylor. 2012. The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics* 17 (2012), 25–32. https://doi.org/10.1016/j.websem.2012.05.003

[5] Shusaku Egami, Satoshi Nishimura, and Ken Fukuda. 2021. VirtualHome2KG: Constructing and Augmenting Knowledge Graphs of Daily Activities Using Virtual Space. In *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24-28, 2021 (CEUR Workshop Proceedings, Vol. 2980)*, Oshani Seneviratne, Catia Pesquita, Juan Sequeda, and Lorena Etcheverry (Eds.). CEUR-WS.org. http://ceur-ws.org/Vol-2980/paper381.pdf

[6] Maria Fazio, Antonio Celesti, Fermín Galán Márquez, Alex Glikson, and Massimo Villari. 2015. Exploiting the FIWARE cloud platform to develop a remote patient monitoring system. In *2015 IEEE Symposium on Computers and Communication (ISCC)*. 264–270. https://doi.org/10.1109/ISCC.2015.7405526

[7] Sten Hanke, Christopher Mayer, Oliver Hoeftberger, Henriette Boos, Reiner Wichert, Mohammed-R. Tazari, Peter Wolf, and Francesco Furfari. 2011. *universAAL – An Open and Consolidated AAL Platform*. Springer Berlin Heidelberg, Berlin, Heidelberg, 127–140. https://doi.org/10.1007/978-3-642-18167-2_10

[8] Bin Huang, Siao Tang, Guangyao Shen, Guohao Li, Xin Wang, and Wenwu Zhu. 2020. Commonsense Learning: An Indispensable Path towards Human-Centric Multimedia. In *Proceedings of the 1st International Workshop on Human-Centric Multimedia Analysis* (Seattle, WA, USA) *(HuMA'20)*. Association for Computing Machinery, New York, NY, USA, 91–100. https://doi.org/10.1145/3422852.3423484

[9] Jingwei Ji, Ranjay Krishna, Li Fei-Fei, and Juan Carlos Niebles. 2020. Action Genome: Actions As Compositions of Spatio-Temporal Scene Graphs. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10233–10244. https://doi.org/10.1109/CVPR42600.2020.01025

[10] Kristiina JOKINEN. 2020. Situated Agents, Personality and Human-Centric AI. *Proceedings of the Annual Conference of JSAI2020* JSAI2020 (2020), 2G4ES402–2G4ES402. https://doi.org/10.11517/pjsai.JSAI2020.0_2G4ES402

[11] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. 2011. HMDB: A large video database for human motion recognition. In *2011 International Conference on Computer Vision*. 2556–2563. https://doi.org/10.1109/ICCV.2011.6126543

[12] Xueting Li, Sifei Liu, Kihwan Kim, Xiaolong Wang, Ming Hsuan Yang, and Jan Kautz. 2019. Putting humans in a scene: Learning affordance in 3D indoor environments. In *Proceedings - 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019 (Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition)*. IEEE Computer Society, United States, 12360–12368. https://doi.org/10.1109/CVPR.2019.01265

[13] Fernando Martínez-Santiago, M Rosario García-Viedma, John A Williams, Luke T Slater, and Georgios V Gkoutos. 2020. Aging neuro-behavior ontology. *Applied Ontology* 15, 2 (2020), 219–239.

[14] Georgios Meditskos and Ioannis Kompatsiaris. 2017. iKnow: Ontology-driven situational awareness for the recognition of activities of daily living. *Pervasive and Mobile Computing* 40 (2017), 17–41. https://doi.org/10.1016/j.pmcj.2017.05.003

[15] National Institute of Advanced Industrial Science and Technology. 2020. Elderly Behavior Library. http://www.behavior-library-meti.com/behaviorLib/homes/about

[16] World Health Organization. 2001. International Classification of Functioning, Disability and Health (ICF). https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health

[17] Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. 2018. VirtualHome: Simulating Household Activities Via Programs. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8494–8502. https://doi.org/10.1109/CVPR.2018.00886

[18] Gunnar A. Sigurdsson, Olga Russakovsky, and Abhinav Gupta. 2017. What Actions are Needed for Understanding Human Actions in Videos?. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 2156–2165. https://doi.org/10.1109/ICCV.2017.235

[19] Gunnar A. Sigurdsson, Gül Varol, Xiaolong Wang, Ali Farhadi, Ivan Laptev, and Abhinav Gupta. 2016. Hollywood in Homes: Crowdsourcing Data Collection for Activity Understanding. In *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 510–526.

[20] Khurram Soomro, Amir Roshan Zamir, Mubarak Shah, Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. [n.d.]. UCF101: A dataset of 101 human actions classes from videos in the wild. *CoRR* ([n. d.]), 2012.

[21] Shinya Tarumi, Kouji Kozaki, Yoshinobu Kitamura, and Riichiro Mizoguchi. 2010. A Consideration of Quality-Attribute-Property for Interoperability of QualityData (In Japanese). *Transactions of the Japanese Society for Artificial Intelligence* 25, 5 (2010), 579–592. https://doi.org/10.1527/tjsai.25.579

[22] Alexandros Vassiliades, Nick Bassiliades, Filippos Gouidis, and Theodore Patkos. 2020. A Knowledge Retrieval Framework for Household Objects and Actions with External Knowledge. In *Semantic Systems. In the Era of Knowledge Graphs*, Eva Blomqvist, Paul Groth, Victor de Boer, Tassilo Pellegrini, Mehwish Alam, Tobias Käfer, Peter Kieseberg, Sabrina Kirrane, Albert Meroño-Peñuela, and Harshvardhan J. Pandit (Eds.). Springer International Publishing, Cham, 36–52.

[23] Chao-Yuan Wu, Christoph Feichtenhofer, Haoqi Fan, Kaiming He, Philipp Krähenbühl, and Ross Girshick. 2019. Long-Term Feature Banks for Detailed Video Understanding. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 284–293. https://doi.org/10.1109/CVPR.2019.00037

[24] Natsuki Yamanobe, Weiwei Wan, Ixchel G. Ramirez-Alpizar, Damien Petit, Tokuo Tsuji, Shuichi Akizuki, Manabu Hashimoto, Kazuyuki Nagata, and Kensuke Harada. 2017. A brief review of affordance in robotic manipulation research. *Advanced Robotics* 31, 19-20 (2017), 1086–1101. https://doi.org/10.1080/01691864.2017.1394912 arXiv:https://doi.org/10.1080/01691864.2017.1394912