2014 IEEE International Conference on Control System, Computing and Engineering, 28 - 30 November 2014, Penang, Malaysia

Human Activity Recognition: A Review

Ong Chin Ann
Faculty of Engineering, Computing & Science
Swinburne University of Technology
Kuching, Malaysia
cong@swinburne.edu.my

Lau Bee Theng
Faculty of Engineering, Computing & Science
Swinburne University of Technology
Kuching, Malaysia
blau@swinburne.edu.my

***Abstract*— Human Activity Recognition is one of the active research areas in computer vision for various contexts like security surveillance, healthcare and human computer interaction. In this paper, a total of thirty-two recent research papers on sensing technologies used in HAR are reviewed. The review covers three area of sensing technologies namely RGB cameras, depth sensors and wearable devices. It also discusses on the pros and cons of the mentioned sensing technologies. The findings showed that RGB cameras have lower popularity when compared to depth sensors and wearable devices in HAR research.**

***Index Terms*—Human activity recognition, sensing technology, depth sensor, wearable devices, RGB camera, Kinect.**

I. Introduction

Human Activity Recognition (HAR) is one of the active research areas in computer vision as well as human computer interaction [1]–[3]. However, it remains a very complex task, due to unresolvable challenges such as sensor motion, sensor placement, cluttered background, and inherent variability in the way activities are conducted by different human [4], [5]. In this paper, a total of thirty-two recent research papers on sensing technologies used in HAR are reviewed. The most commonly employed sensing technologies in HAR system regardless of the computational models or classification algorithms are analyzed. The pros and cons of each sensing technology have been discussed. This paper is concluded with some challenges for the most sophisticated sensing technologies.

The paper is organized as follows: Section II briefly introduces human activity recognition and its application in various contexts; Section III describes the types of sensing technologies used in HAR systems; Section IV discusses the findings; finally Section V concludes and posts a few open questions for further discussions.

II. Human Activity Recognition

Human activity recognition is an ability to interpret human body gesture or motion via sensors and determine human activity or action [6]. Most of the human daily tasks can be simplified or automated if they can be recognized via HAR system [7], [8]. Typically, HAR system can be either supervised or unsupervised [9]. A supervised HAR system required some prior training with dedicated datasets while unsupervised HAR system is being configured with a set of rules during development. HAR is considered as an important component in various scientific research contexts i.e. surveillance, healthcare and human computer interaction (HCI) [5], [10]–[12].

A. Surveillance System

In surveillance context, HAR was adopted in surveillance systems installed at public places i.e. banks or airports [7], [13], [14]. Ryoo [12] introduced a new paradigm of human activity prediction to prevent crimes and dangerous activities from occurring at public places. The findings confirmed the proposed approaches are able to recognize ongoing human-human interactions at the earlier stage. Lasecki et al. [15] proposed Legion:AR, a system that provides robust, deployable activity recognition by supplementing existing recognition systems with on-demand, real-time activity identification using inputs from the crowds at public places.

B. Healthcare

From most of the literature reviewed, HAR is employed in healthcare systems installed in residential environment, hospitals and rehabilitation centers. HAR is used widely for monitoring the activities of elderly people staying in rehabilitation centers for chronic disease management and disease prevention [16]. HAR is also integrated into smart homes for tracking the elderly people's daily activities [17], [18]. Besides, HAR is used to encourage physical exercises in rehabilitation centers for children with motor disabilities [19], post-stroke motor patients [20], patients with dysfunction and psychomotor slowing [21], and exergaming [22]. Other than that, the HAR is adopted in monitoring patients at home such as estimation of energy expenditure to aid in obesity prevention and treatment [23] and lifelogging [24]. HAR is also applied in monitoring other behaviours such as stereotypical motion conditions in children with Autism Spectrum Disorders (ASD) at home [25], abnormal conditions for cardiac patients [26] and detection for early signs of illness [27] and it provided the clinicians with opportunities for intervention. Other healthcare related HAR such as fall detection and intervention for elderly people using HAR are found in [28]–[30].

978-1-4799-5686-9/14/$31.00 ©2014 IEEE

C. Human Computer Interaction

In the field of human computer interaction, HAR has been applied quite commonly in gaming and exergaming such as Kinect [31]–[33], Nitendo Wii [34], [35], full-body motion-based games for older adults [36] and adults with neurological injury [37]. Through HAR, human body gestures are recognized to instruct the machine to complete dedicated tasks. Elderly people and adults with neurological injury can perform a simple gesture to interact with games and exergames easily. HAR also enables surgeons to have intangible control of the intraoperative image monitor by using standardized free-hand movements [38].

TABLE I. REVIEWED PAPERS ON HAR

Types	Papers	Total
RGB Camera	Roshtkhari and Levine[5], Noorit and Suvonvorn [11], Ryoo [12], Tamas [14], Wang et al. [30]	5
Depth Sensor	Ong et al. [8], Chaaraoui et al. [10], Lasecki et al. [15], Jalal et al., Chang et al., Hayes et al., González-Ortega et al., [18]–[21], Stone and Skubic [27], Auvinet and Meunier [28], Lange et al. [37], Xia et al., Amiri et al., Frontoni et al., Yang and Tian [39]–[42]	14
Wearable	Yang et al. [6], Banos et al. [16], Alshurafa et al., Sazonov et al., Khan, Paragliola and Coronato, Kantoch and Augustyniak [22]–[26], Vo et al. [29], Ustev et al., Zhang and Sawchuk, Reiss et al., Kreil et al., He and Bai [43]–[47]	13

III. SENSING TECHNOLOGIES

Generally, the sensor(s) in a conventional HAR plays an important role in recognizing human activity. Figure 1 illustrates the process of how a human activity is recognized when a body gesture is given as input. The sensor(s) capture the information acquired from human body gesture and the recognition engine analyzes the information and determines the type of activity has been performed.

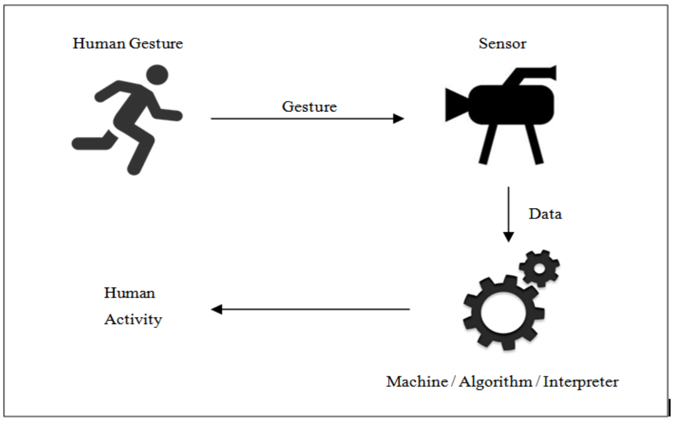

Fig. 1. General structure of HAR system

We reviewed 32 papers published recently (from 2011 to 2014) on different sensing technologies used in HAR. These technologies are classified as RGB camera-based, depth sensor-based and wearable-based as shown in Table I.

Recognizing human activity using RGB camera is simple but with low efficiency. A RGB camera is usually attached to the environment and the HAR system will process image sequences captured with the camera. Most of the conventional HAR systems using this sensing technology are built with two major components which is the feature extraction and classification [13], [48]. Besides, most of the RGB-HAR systems are considered as supervised system where trainings are usually needed prior to actual use. Image sequences and names of human activities are fed into the system during training stage. Real time captured image sequence are passed to the system for analysis and classification by dedicated computational/classification algorithms such as Support Vector Machine (SVM) [2].

The depth sensor also known as infrared sensor or infrared camera [49] is adopted into HAR systems for recognizing human activities. In a nutshell, the depth sensor projects infrared beams into the scene and recapture them using its infrared sensor to calculate and measure the depth or distance for each beam from the sensor. The reviews found that Microsoft Kinect sensor is commonly adopted as depth sensor in HAR [33]. Since the Kinect sensor has the capability to detect 20 human body joints with its real-world coordinate [40], many researchers utilized the coordinates for human activity classification.

HAR using wearable-based requires single or multiple sensors to be attached to the human body. Most commonly used sensor includes 3D-axial accelerometer, magnetometer, gyroscope and RFID tag [44], [45]. With the advancement of current smart phone technologies, many works uses mobile phone as sensing devices because most smart phones are equipped with accelerometer, magnetometer and gyroscope [29], [50]. A physical human activity can be identify easily through analysing the data generated from various wearable sensing after being process and determine by classification algorithm. Further to this, Kantoch and Augustyniak claims that GPS and temperature signal acquired from smart phone can be further feed into machine for healthcare monitoring purpose [26].

IV. FINDINGS AND DISCUSSIONS

Figure 2 shows the sensing technologies adopted in HAR from the reviewed papers. The review outcome indicates both depth sensor and wearable sensor technologies are gaining more popularity in HAR research recently. On the other hand, RGB camera has obtained less emphasis in HAR research, most probably due its imitation in capturing the scene and human motions in 3D space [39]. Besides, detecting and extracting human from image sequences is another constraint which requires high machine processing [28]. Thus, the performance of real time HAR system might be effected when lots of data are processes at a time [5]. Another concern rose while employing RGB camera into HAR system is the privacy issue. A human i.e. elderly may feel discomfort or intruded being watches all the time.

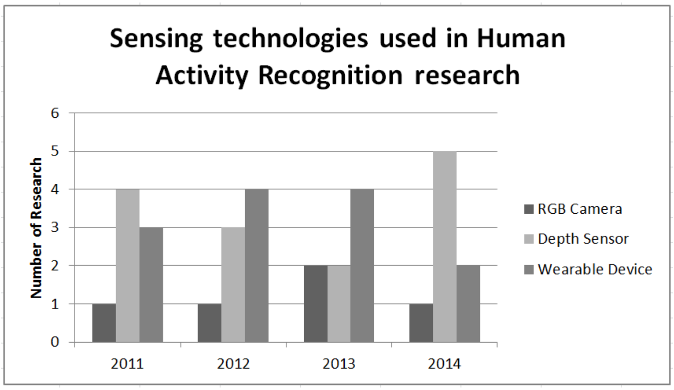

Fig. 2. Reviewed HAR research and the sensing technologies used

Both depth sensor and wearable sensor had its own pros and cons when employed in HAR system. The depth sensors have become popular due to low cost [19], [20], [27], [37], [39] , high sample rate and capability of combining visual and depth information [10]. The recognition processes are consider lighter, robust and inexpensive as compare with RGB camera [41]. However some common visual based issue and challenges still persist for depth sensor such as occlusion [39] as well as limitation of sensor viewpoint [30].

The emergence of wearable sensor systems could address occlusion and viewpoint limitation challenges which occurs in HAR system which employs either RGB camera and depth sensor or both [44]. The wearable sensors are well known being flexible in providing location independent and seamless human monitoring without affecting their daily lifestyle i.e. privacy issue [26]. According to Kreil et al. [46], another value added point to wearable sensor is it cost which is quite cheap, compact and low power consumption. The main drawback against wearable sensor is the recognition accuracy. Usually a wearable-based HAR system requires the subject to wear or attached with multiple sensors on various body parts [23]. This is so troublesome, intrusive and inconvenient for the subjects [8]. While Vo et al., and Reiss et al. [29], [45] indicated the wearable-based HAR could not work effectively as there is a tendency where the human subject can forget to put on or displace the dedicated sensor.

From the reviews, it seems that RGB camera is substituted with other sensors in HAR research due to its limitations [39]. As for depth sensor and wearable sensor, it is difficult to justify which sensor would be the best or most suitable to be employed in HAR because there is a deadlock between them. There is no specific indicator or measurement on whether depth sensor is better than wearable sensor or vice-versa as far as universal context is concern. It is suggested that both sensors have their own strengths and weaknesses depending on the human subject and context of use. Thus, researchers, practitioners and developers need to study the human subjects and their contexts of use before adopting the sensing technologies for the HAR.

The findings also indicated that many researchers who employ depth sensing technologies into their HAR use Microsoft Kinect sensor as the experiment tools. This is mainly motivated the cost and efficiency of Kinect sensor and HAR system can be developed easily with the support of its standard Software Development Kit (SDK) as well as public support via open forums. There are some questions for consideration when building a HAR system: 1) Is there any possibility to resolve occlusion problem in vision based system? 2) Is there any possibility to increase the range of viewpoint for without adding more sensors? 3) Is there a possibility to create a universal or standard file format to store sensing data in view of multiple sensing technologies could be employed in a HAR system?

V. CONCLUSIONS

A review has been completed on thirty two papers published in 2011-2014 for various sensing technologies used in HAR. We classify these technologies into three main categories namely RGB camera, depth sensor and wearable device. Our review found that the popularity of RGB camera in HAR research has dropped while both depth and wearable sensors are the substitutes. On the other hand, the use of Kinect sensor (depth sensor) into HAR system is promising. This could be a sign of the rise of Kinect as a popular sensing tool in HAR system.

REFERENCES

[1] A. Iosifidis, A. Tefas, and I. Pitas, “Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis,” *Signal Processing*, vol. 93, no. 6, pp. 1445–1457, Jun. 2013.

[2] D. Weinland, R. Ronfard, and E. Boyer, “A survey of vision-based methods for action representation, segmentation and recognition,” *Comput. Vis. Image Underst.*, vol. 115, no. 2, pp. 224–241, Feb. 2011.

[3] S. Ali and M. Shah, “Human action recognition in videos using kinematic features and multiple instance learning.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 288–303, Feb. 2010.

[4] A. Oikonomopoulos and M. Pantic, “Human Activity Recognition Using Hierarchically-Mined Feature Constellations,” pp. 150–159, 2013.

[5] M. Javan Roshtkhari and M. D. Levine, “Human activity recognition in videos using a single example,” *Image Vis. Comput.*, vol. 31, no. 11, pp. 864–876, Nov. 2013.

[6] J. Yang, J. Lee, and J. Choi, “Activity Recognition Based on RFID Object Usage for Smart Mobile Devices,” *J. Comput. Sci. Technol.*, vol. 26, no. 2, pp. 239–246, Mar. 2011.

[7] L. Chen, H. Wei, and J. Ferryman, “A survey of human motion analysis using depth imagery,” *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 1995–2006, Nov. 2013.

[8] W. Ong, L. Palafox, and T. Koseki, “Investigation of Feature Extraction for Unsupervised Learning in Human Activity Detection,” *Bull. Networking, Comput. Syst. Softw.*, vol. 2, no. 1, pp. 30–35, 2013.

[9] O. D. Lara and M. a. Labrador, “A Survey on Human Activity Recognition using Wearable Sensors,” *IEEE Commun. Surv. Tutorials*, vol. 15, no. 3, pp. 1192–1209, Jan. 2013.

[10] A. A. Chaaraoui, J. R. Padilla-López, P. Climent-Pérez, and F. Flórez-Revuelta, “Evolutionary joint selection to improve human action recognition with RGB-D devices,” *Expert Syst. Appl.*, vol. 41, no. 3, pp. 786–794, Feb. 2014.

[11] N. Noorit and N. Suvonvorn, “Human Activity Recognition from Basic Actions Using Finite State Machine,” in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, 2014, vol. 285, pp. 379–386.

[12] M. S. Ryoo, “Human activity prediction: Early recognition of ongoing activities from streaming videos,” in *2011 International Conference on Computer Vision*, 2011, no. Iccv, pp. 1036–1043.

[13] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, “Gait Recognition with Kinect,” in *Workshop on Kinect in Pervasive Computing at Pervasive 2012*, 2012.

[14] V. Tamas, “Human Behavior Recognition In Video Sequences,” Technical University of CLUJ-NAPOCA, 2013.

[15] W. S. Lasecki, Y. C. Song, H. Kautz, and J. P. Bigham, “Real-time crowd labeling for deployable activity recognition,” in *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW ’13*, 2013, p. 1203.

[16] O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas, “Daily living activity recognition based on statistical feature quality group selection,” *Expert Syst. Appl.*, vol. 39, no. 9, pp. 8013–8021, Jul. 2012.

[17] L. Chen, C. D. Nugent, and H. Wang, “A Knowledge-Driven Approach to Activity Recognition in Smart Homes,” *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, Jun. 2012.

[18] A. Jalal, Z. Uddin, J. T. Kim, and T. Kim, “Recognition of Human Home Activities via Depth Silhouettes and Â Transformation for Smart Homes,” no. February, pp. 467–475, 2011.

[19] Y.-J. Chang, S.-F. Chen, and J.-D. Huang, “A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities.,” *Res. Dev. Disabil.*, vol. 32, no. 6, pp. 2566–2570, 2011.

[20] A. Hayes, P. Dukes, and L. F. Hodges, “A Virtual Environment for Post-Stroke Motor Rehabilitation,” South Carolina, 2011.

[21] D. González-Ortega, F. J. Díaz-Pernas, M. Martínez-Zarzuela, and M. Antón-Rodríguez, “A Kinect-based system for cognitive rehabilitation exercises monitoring.,” *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 620–31, Feb. 2014.

[22] N. Alshurafa, W. Xu, J. Liu, M.-C. Huang, B. Mortazavi, C. Roberts, and M. Sarrafzadeh, “Designing a Robust Activity Recognition Framework for Health and Exergaming using Wearable Sensors.,” *IEEE J. Biomed. Heal. informatics*, no. c, pp. 1–11, Oct. 2013.

[23] E. S. Sazonov, G. Fulk, J. Hill, Y. Schutz, and R. Browning, “Monitoring of posture allocations and activities by a shoe-based wearable sensor.,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 4, pp. 983–90, Apr. 2011.

[24] A. M. Khan, “Human Activity Recognition Using A Single Tri-axial Accelerometer,” Kyung Hee University, Seoul, Korea, 2011.

[25] G. Paragliola and A. Coronato, “Intelligent Monitoring of Stereotyped Motion Disorders in Case of Children with Autism,” *2013 9th Int. Conf. Intell. Environ.*, pp. 258–261, Jul. 2013.

[26] E. Kantoch and P. Augustyniak, “Human activity surveillance based on wearable body sensor network,” in *Computing in Cardiology (CinC)*, 2012, pp. 325 – 328.

[27] E. Stone and M. Skubic, “Passive, In-Home Gait Measurement Using an Inexpensive Depth Camera: Initial Results,” in *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*, 2012, pp. 183–186.

[28] A. T. Nghiem, E. Auvinet, and J. Meunier, “Head detection using Kinect camera and its application to fall detection,” in *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, 2012, pp. 164–169.

[29] Q. V. Vo, G. Lee, and D. Choi, “Fall Detection Based on Movement and Smart Phone Technology,” in *2012 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future*, 2012, pp. 1–4.

[30] S. Wang, S. Zabir, and B. Leibe, “Lying Pose Recognition for Elderly Fall Detection,” in *Robotics:Science and Systems VII*, no. 1, H. Durrant-Whyte, N. Roy, and P. Abbeel, Eds. MIT Press, 2012, pp. 345 –353.

[31] J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with Microsoft Kinect sensor: a review.,” *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–34, Oct. 2013.

[32] J. Smisek, M. Jancosek, and T. Pajdla, “3D with Kinect,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, pp. 1154–1160.

[33] Z. Zhang, “Microsoft Kinect Sensor and Its Effect,” *IEEE Multimed.*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[34] D. T. G. Huynh, “Human Activity Recognition with Wearable Sensors,” Technische Universität Darmstadt, 2008.

[35] E. Lawrence, C. Sax, K. F. Navarro, and M. Qiao, “Interactive Games to Improve Quality of Life for the Elderly: Towards Integration into a WSN Monitoring System,” in *2010 Second International Conference on eHealth, Telemedicine, and Social Medicine*, 2010, pp. 106–112.

[36] K. Gerling, I. Livingston, L. Nacke, and R. Mandryk, “Full-body motion-based game interaction for older adults,” in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*, 2012, p. 1873.

[37] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, “Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor.,” *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2011, pp. 1831–4, Jan. 2011.

[38] K. Yoshimitsu, Y. Muragaki, T. Maruyama, M. Yamato, and H. Iseki, “Development and Initial Clinical Testing of ‘OPECT’: An Innovative Device for Fully Intangible Control of the Intraoperative Image-Displaying Monitor by the Surgeon,” *Neurosurgery*, vol. 10, 2014.

[39] L. Xia, C. Chen, and J. K. Aggarwal, “View invariant human action recognition using histograms of 3D joints,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 20–27.

[40] S. M. Amiri, M. T. Pourazad, P. Nasiopoulos, and V. C. M. Leung, “Human action recognition using meta learning for RGB and depth information,” *2014 Int. Conf. Comput. Netw. Commun.*, pp. 363–367, Feb. 2014.

[41] E. Frontoni, A. Mancini, and P. Zingaretti, “RGBD Sensors for Human Activity Detection in AAL Environments,” in *Ambient Assisted Living*, S. Longhi, P. Siciliano, M. Germani, and A. Monteriù, Eds. Cham: Springer International Publishing, 2014, pp. 127–135.

[42] X. Yang and Y. Tian, “Effective 3D action recognition using EigenJoints,” *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 2–11, Jan. 2014.

[43] Y. E. Ustev, O. Durmaz Incel, and C. Ersoy, “User, device and orientation independent human activity recognition on mobile phones,” in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication - UbiComp ’13 Adjunct*, 2013, pp. 1427–1436.

[44] M. Zhang and A. A. Sawchuk, “Motion primitive-based human activity recognition using a bag-of-features approach,” in *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI ’12*, 2012, no. 1, p. 631.

[45] A. Reiss, G. Hendeby, and D. Stricker, “A Competitive Approach for Human Activity Recognition on Smartphones,” in *21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013, no. Esann, pp. 455–460.

[46] M. Kreil, B. Sick, and P. Lukowicz, “Dealing with human variability in motion based, wearable activity recognition,” in *2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, 2014, pp. 36–40.

[47] Z. He and X. Bai, “A wearable wireless body area network for human activity recognition,” in *2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2014, pp. 115–119.

[48] B. Mirmahboub, S. Samavi, N. Karimi, and S. Shirani, “Automatic Monocular System for Human Fall Detection based on Variations in Silhouette Area.,” *IEEE Trans. Biomed. Eng.*, no. c, pp. 1–10, Nov. 2012.

[49] K. Khoshelham, “Accuracy Analysis of Kinect Depth Data,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2011, vol. XXXVIII, no. August, pp. 133–138.

[50] T. Mashita, K. Shimatani, M. Iwata, H. Miyamoto, D. Komaki, T. Hara, K. Kiyokawa, H. Takemura, and S. Nishio, “Human activity recognition for a content search system considering situations of smartphone users,” *2012 IEEE Virtual Real.*, pp. 1–2, Mar. 2012.