ISSN 2186-7437

NII Shonan Meeting Report

No. 2015-12

Challenges for Real-time activity recognition

Sozo Inoue (Kyusyu Institute of Technology, Japan)
Stephan Sigg (Aalto University, Finland)

September 14–17, 2015

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Challenges for Real-time activity recognition

Sozo Inoue (Kyusyu Institute of Technology, Japan)
Stephan Sigg (Aalto University, Finland)

September 14–17, 2015

Over the last decade, the recognition of human activities or situations has seen remarkable progress. This was driven by several strong developments in related areas. First of all, sensing hardware has been greatly improved (e.g. size, accuracy and also new sensing modalities and sense-able quantities), enabling an enhanced perception of the world through sensors. Also, machine learning has celebrated great successes (algorithms, toolboxes) and has become a mainstream ability that attracts a huge user base towards activity recognition. Furthermore, rapid development in wireless protocols and near-global coverage of wireless technologies (e.g. UMTS, LTE, Bluetooth, Wlan) enabled the transmission at higher data rates and new usage areas through wireless communication. Last, but not least, novel applications have spread that promote the publishing and sharing of all kinds of data (e.g. Facebook, Line, WhatsApp), which led to novel valuable inputs for machine learning and activity recognition. With these opportunities, new challenges and application areas for activity recognition evolve. The amount of data to be analysed is growing rapidly, in particular, towards personal information, which results in an increasing demand for the accurate sensing of social and sentiment information of individuals and crowds.

Fuelled by the advancing Internet of Things and opportunistic as well as participatory sensing campaigns, activity recognition is going Big Data. New devices like fitness accessory, smart watches or instrumented glasses promise continuous collection of data for monitored subjects. Sensing systems are developing capabilities to monitor virtually everybody, everywhere and with device-free sensing paradigms, the necessity to equip individuals with sensing hardware or specifically installing system components at any particular physical location diminishes. In summary, a tremendous amount of data for human activity recognition is provided by the above mentioned appliances. A seminal part of this data is related to personal information so that there is a potential to extend the perception of sensing devices towards sentiment states. Sentiment sensing focuses on peoples mental state, intention or emotion, for instance, by interpreting eye-gaze information, body gesture or pose, for emotion, intention or attention, and thereby directs and extends the perception of a sensing system inwards. The recognition of sentiment is actively discussed in academia and industry recently. Mainly, text-mining has yet been employed but clearly also other sensing modalities are able to capture aspects of sentiment. In particular, there is a large body of work on the recognition of emotion from body gesture and pose – two characteristics that can be well estimated from traditional on-body or also environmental sensing hardware.

In the scope of this meeting, we focus on questions that arise for an actual instrumentation of such Big-Data-based human activity recognition and open issues regarding sentiment recognition, such as

- Suitable sensing modalities, features and algorithms for sentiment sensing
- Relation of gestures, action and activities to Sentiment classes
- Sentiment expression through action and activities
- Differences/similarities among various human subjects
- Necessity of a central store for this data in order to utilise it
  - How should such a central store be organised?
  - If distributed over various devices, how to access data it and keep it concise?
- Dealing with various noise levels of different device classes for the same modality

The tremendous amount of continuously acquired information requires new methods and algorithms. Ongoing projects allow a glimpse on the challenges stemming from the handling of such massive amounts of data, but the information available when subjects are continuously monitored by multiple devices will exceed this greatly. This challenge of large scale database construction is also focused by the Human Activity Sensing Consortium (HASC), where data collection, feature and algorithm development, as well as algorithm and tool standardization for human activity sensing are focused.

The purpose of this Shonan meeting is to bring together researchers from the fields of activity recognition, mobile sensing, pervasive computing and information processing to discuss the technical challenges, possible societal impact, as well as promising industrial applications for emerging applications in activity recognition. The seminar puts a clear focus on crowd, social and sentiment sensing in continuous sensing platforms and Big Data in activity recognition. By gathering experts from these respective fields, we envision to identify most pressing open research questions and to come up with novel, interdisciplinary approaches. We hope that through discussions of the individual researchers, this Shonan meeting can help spur interdisciplinary discussions and foster joint activities.

1 List of participants

- Prof. Yutaka Arakawa, NAIST, Japan
- Anja Bachmann, Karlsruhe Institute of Technology, Germany
- Dr. Nils Hammerla, Newcastle University, UK
- Samuli Hemminki, University of Helsinki, Finland
- Prof. Sozo Inoue, Kyusyu Institute of Technology, Japan
- Prof. Yusheng Ji, NII, Japan

- Prof. Shoji Kobashi, University of Hyogo, Japan
- Prof. Shinichi Konomi, University of Tokyo, Japan
- Prof. Kai Kunze, Keio University, Japan
- Prof. Takuya Maekawa, Osaka University, Japan
- Prof. Kazuya Murao, Ritsumeikan University, Japan
- Prof. Petteri Nurmi, University of Helsinki, Finland
- Prof. Ren Ohmura, Toyohashi University of Technology, Japan
- Dr. Till Riedel, Karlsruhe Institute of Technology, Germany
- Philipp Scholl, Albert-Ludwigs University Freiburg, Germany
- Shuyu Shi, NII, Japan
- Prof. Stephan Sigg, Aalto University, Finland
- Prof. Masaki Shuzo, University of Tokyo / ATR Japan
- Dr. Kalika Suksomboon, KDDI R&D Laboratories, Japan
- Prof. Tsutomu Terada, Kobe University, Japan
- Prof. Takuro Yonezawa, Keio University, Japan
- Prof. Moustafa Youssef, Egypt-Japan University of Science and Technology, Egypt

2 Program

Sunday, September 13th, 2015	
15:00 -	Check-in
19:00 – 20:30	Welcome banquet
Monday, September 14th, 2015	
07:30 – 09:00	Breakfast
09:00 – 09:10	Introduction of NII Shonan meetings
09:10 – 10:00	Introduction to the seminar (Sozo Inoue, Stephan Sigg)
10:00 – 10:15	Break
10:15 – 12:00	5 Minutes Self-introduction and position talks of each participant
12:00 – 13:30	Lunch
13:30 – 14:00	Group photo
14:00 – 15:15	5 Minutes Self-introduction and position talks of each participant
15:15 – 15:30	Break
15:30 – 16:30	Identifying topics for Break-out sessions
16:30 – 18:00	Break-out Sessions (important issues and new research directions)
18:00 – 19:30	Dinner

Tuesday, September 15th, 2015

07:30 – 09:00	Breakfast
09:00 – 10:45	Break-out sessions continued
10:45 – 11:00	Break
11:00 – 12:00	Presentations: Results from Break-out sessions
12:00 – 13:30	Lunch
13:30 – 14:30	Identifying interdisciplinary challenges and collaboration opportunities
14:30 – 15:15	Break-out Sessions (interdisciplinary challenges and collaboration opportunities)
15:15 – 15:30	Break
15:30 – 16:30	Discussion: Content of the report
16:30 – 18:00	Break-out sessions continued
18:00 – 19:30	Dinner

Wednesday, September 16th, 2015

07:30 – 09:00	Breakfast
09:00 – 10:45	Presentations: Results from Break-out sessions
10:45 – 11:00	Break
11:00 – 12:00	Brainstorming towards potential position papers
12:00 – 13:30	Lunch
13:30 – 21:00	Group Excursion to Kamakura

Thursday, September 17th, 2015

07:30 – 09:00	Checkout and breakfast
09:00 – 12:00	Seminar session with coffee break
	Idea marketplace and future collaborations
	Final organizer presentation and wrap up
12:00 – 13:30	Lunch

3 Results

3.1 Large Scale Emotion Sensing

Anja Bachmann, Samuli Hemminki, Shin'ichi Konomi, Stephan Sigg

Our group discussed several ideas and envisioned a system that is able to detect emotions based on sensor measurements not only on a personal and small scale but also on a larger scale such as a whole city.

We decided to focus on sensing devices that are related to one or more of the following categories: stationary, mobile, wearable or personal. This includes systems such as *stationary* wifi access points, *mobile* drones and transportation systems, *wearable* inertial sensors, *personal* devices such as smartphones or even *combinations* of different categories, e.g. smartwatches and smartglasses as personal wearable devices.

3.1.1 Objective

The overall objective is to develop a vision paper that reflects the ideas collected within the group. This paper is supposed to motivate the need for a **La**rge **S**cale **E**motion **S**ensing system (*LaSES*) and to present related work. As a next step, we intend to identify technical challenges based on gaps in the related work and to derive a taxonomy for emotion sensing. The LaSES system is based on these

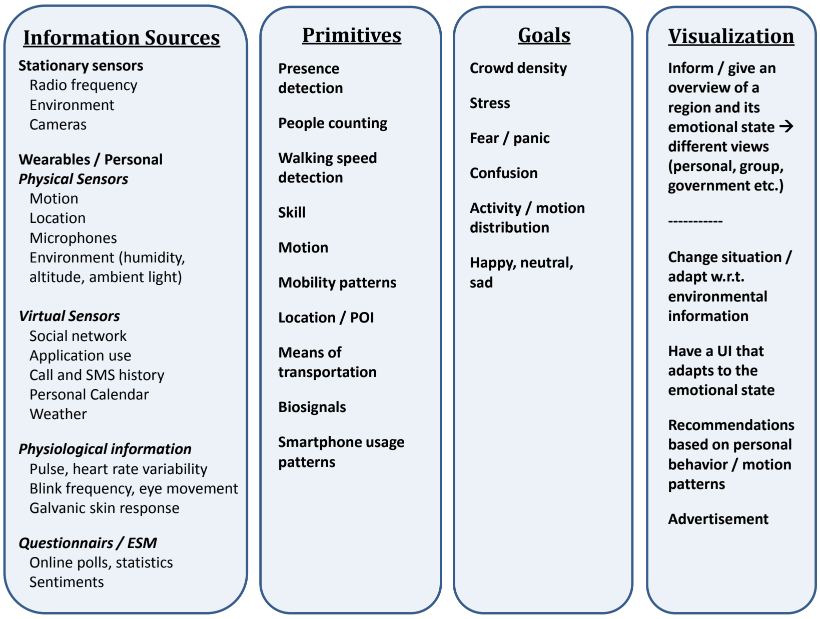

Figure 1: Overview of the LaSES system

findings. The vision paper is supposed to present a preliminary system architecture and introduce opportunities for environmental sensing, mobile sensing, and personal and wearable sensing. we will discuss these opportunities and derive future work and challenges.

3.1.2 Workflow

Emotions At first, we collected definitions on what "emotions" are and how they are distinguished from other phenomena such as "affect" or "mood". Afterwards, we collected examples on emotions [4, 1] and were faced with the six basic emotions of Ekman [3] from which we adapted happy, sad, fear and anger. In addition, we included a neutral state or a state of confusion. Apart from emotions, we also considered other mental states such as mood [5], mental illnesses [6], stress [8], cognitive load [7], a feeling of safety [1], and social behvior [2].

Information Sources and Primitives The sensing systems we identified contain different useful sensors and information sources. They can be categorized as *stationary*, *physical*, *virtual* sensors and additional information sources such as *physiological* or based on *quesionnaires and experience sampling*.

Next, we discussed which primitives we want to derive from the sensor sources. Primitives describe specific features that can be derived from measurements gathered from the before-mentioned information sources.

An overview of the information sources we identified and the derived primitives can be found in Figure 1.

[1]http://streetscore.media.mit.edu/

Technologies We reviewed related work to detect approaches that already apply mobile sensing for recognition of those states. Those papers serve as a basis to identify technological challenges. In a next step, we described opportunities for emotion sensing leveraging technologies from the area of environmental, mobile, personal and wearable sensing systems.

Next steps What is left to do is to 1) review the collected related work and identify the gap; 2) merge the related work and opportunity sections; 3) identify technical challences; 4) present a few selected use cases and demonstrate the feasibility of a corresponding sensing system; and 5) find a venue to present this vision.

3.2 Large-scale behaviour change

Kai Kunze, Sozo Inoue, Tsutomu Terada, Takuro Yonezawa

Background: Behavior change, which can refer to any transformation or modification of human behavior, is an important concept which could be applied for many domains such as smart community, healthcare, and education. Many of trials for behavior changes in each domain.

However, there is a large gap between personal-level behavior changes and city-level behavior changes. For example, personal behavior change targets smaller goals such as personal health keeping, having good abilities in sports, and good scores in exams, whereas city-level behavior changes include living in harmony with environments, energy saving, and civic engagements (c.f. OECD reports of the city-level satisfaction indicators).

Activity recognition has been researched so far in the UbiComp community, mainly applied to personal-level and group-level. For the personal level, it is also useful to achieve each users goal by recognizing, visualizing, or giving a feedback to the user, such as the activity tracker for fitness use.

Objective: Here, it would be also fruitful and challenging to answer the question of can we apply and incorporate activity recognition to city-level goals? In this group, we talked about a virtual research proposal which is named HyBAR: Hybrid Behavior Change with Activity Recognition.

The basic idea is to formalize the optimization function combining personal-level satisfaction (or achievement) and city-level one, each of which will be measured or sensed by activity recognition technology, urban sensing technology, or any level of statistic survey, and develop a system to solve the function and intermediate between personal-level and city-level goals.

Items: We formed the tasks to 4 components:

- Personal level sensing (and feedback)
- Household level sensing (and feedback)
- City level sensing sensing (and feedback)
- Feedback coordination and optimization (and feedback)

Scenarios: We listed up two scenarios as the following.

Power consumption – health :

- Social level purpose: decreasing peak demand of electric power
- Personal level purpose: good health (by walking, exercising)
  - detection of electric power shortage: collection of data from city open-data and ubiquitous sensors in home
  - change the health-care scenario on smart phone: detection of human behaviors using wearable sensors
  - behavior recommendation explicitly/implicitly
  - saving electricity by behavior change

Garbage – navigation :

- Social level purpose: no garbage on road
- personal level purpose: navigation to the destination
  - detection of having garbage for general people
  - change the route of navigation, where trash bins are along the route

The technical challenges will be: real time activity recognition, activity prediction, social sensing, city-level context recognition, integration of context from sensors and context from open data in the city, clarification of the relationship between information presentation and behavior change, and translation method from social task to personal task.

3.3 Novel Applications Group

Syoji Kobashi, Takuya Maekawa, Petteri Nurmi, Shuyu Shi, Moustafa Youssef

3.3.1 Position Statement

The focus of the subgroup was to investigate innovative applications in different domains and their associated challenges covering novel sensors, leveraging existing devices in our daily lives to detect new activities, sensors fusion, collaborative sensing, quantifying activities, encouraging/helping behavior change, detecting abnormalities, among others. The subgroup discussed different application domains including smart spaces (office, factory, home, hospital, school, sports, city, etc) as well as structure health monitoring, interdisciplinary work with other sciences, education, non-human activity recognition, and smart transportation systems. Different challenges were identified such as users incentives; handling noisy sensors to obtain accurate, robust, and reliable estimates; compensating for low density data (e.g. through compressive sensing); energy efficiency aspects; preserving user privacy through non-invasive techniques; unobtrusive sensing techniques; unaware sensing (especially for elderly); new ways of obtaining ground truth; social acceptance and implications; as well as data scale. Two motivating cases where discussed to spark innovative applications in activity recognition: Needs-based (where the applications are initiated by a society need) and Technology-driven (where new hardware or software technologies encourage the development of new applications that build on them). The group identified three main innovative applications of interests: Food activity recognition using standard Wi-Fi devices, apnea classification, and sensors topology optimization for optimal accuracy.

3.3.2 Potential applications and future collaboration

According to the above results, we propose several innovative activity recognition applications.

Food activity recognition using off-the-shelf-smartphone Our survey related to innovative activity recognition applications revealed that many existing applications rely on state-of-the-art sensing technologies. Also, we found that healthcare is a promising domain of research on innovative activity recognition applications. Based on the facts, we propose an innovative food activity recognition application using Wi-Fi signals. We assume that a smartphone is inserted into a users chest pocket and also a Wi-Fi receiver is placed at the back of the user. The Wi-Fi receiver receives Wi-Fi signals from the smartphone and analyzes CSI information of the signal to detect such eating activities as swallowing water and food. The analysis method will be based on Wi-Fi device free passive indoor localization/activity recognition techniques. Because this approach does not require such wearable sensors as microphone inserted to the ear or EEG sensors attached to the neck, this will be a practical solution for unobtrusive eating monitoring.

Application to SAS detection and type classification There are many potential sleep apnea syndrome (SAS) patients who are not diagnosed in hospitals and are not aware by themselves. Because current clinical products and in-home products require special device and/or special skills to diagnose SAS, it is difficult to use them by general persons. Many sensors including smartphones, Wi-Fi devices, activity watches, etc. have been already installed in our home. A future application of HAR should be to find a risk of such diseases using the already existing sensors. In case of SAS application, smartphones beside the bedside will measure breathing using Wi-Fi signal fluctuation, sound signals, etc.

Sensors topology optimization for optimal accuracy Recently, some research community has been attracted to discuss on RF-based device-free passive (DfP) localization techniques, because such kind of localization solutions do not require subjects to wear a radio device. Instead, some RF devices are required to be deployed in the interested area to track the coordinate of human. For reducing the localization errors, some issues must be addressed including interference, multipath effect, signal noise and so on. These challenges require us to find an optimal topology for RX-TX deployment, however, it will induce burdensome effort to conduct experiments to find it. We propose a novel simulation tool, which can output an optimal TX-RX deployment strategy based on the floorplan of a room and the positions of house furnishings.

3.4 Future Direction of Activity Recognition Algorithms (Towards Human Emotion and Activity Recognition Research Community)

Prof. Yusheng Ji, Prof. Masaki Shuzo, Prof. Kazuya Murao, Dr. Kalika Suksomboon, Prof. Ren Ohmura

In our group, we had another discussion of human emotion recognition. First two days our discussion started from algorithm of only human activity. Then, last two days, the discussion region was extended to emotional recognition because the pattern recognition might be usually same as activity recognitions. Finally, we got the report which titled in Future Direction of Activity Recognition Algorithms. And this document is as an optional result paper.

There will be various applications by using combined information of human activity and emotion such as stress check, care for elderly, and so on. Also a commercially available humanoid has a function about emotion recognition, of which algorithm and accuracy cannot be opened and usage will not be decided yet.

As we know, nowadays, if we have appropriate data set (training data with ground truth label), high precision recognition can be done with the machine learning technique. As for activity recognition, good research environment has been achieved in the world (ie. HASC). However, as for emotion recognition, there still exit several small groups.

As for the methods for emotional recognition, there are several ways by using face data, biological signals, texts in the web, and so on. Also, human voice data will be easy to get in a smart phone and useful information for emotional recognition. Although voice data is suitable for a mobile environment, there are few papers.

In fact, many researchers will hesitate for some difficulties in emotional recognition; f.g. sample number, individual differences, difference in sex, and so on. So, although we have a nice application using emotional information, it cannot be easily achieved. If our human activity recognition group can develop and provide a prototype tool of emotional recognition such as the HASC tools, exciting big opportunities will be waiting for us.

So, how about develop and provide a prototype classifier for all researchers? And, how about organize the emotional research community like HASC, and lets gather large scale data (mainly voice data, but welcomed additional data)? It should not be limited only in Japanese. Multiple language data will be appreciated. Maybe we can easily provide a data collection application for users like HASC logger.

Of course, we should have a sample application, and critical exit strategy. But at this time, its not necessary to share with us. Even if hobby use and entertainment use will be welcomed for us because our community will be encouraged. Anyway, in near future, we want to have an ideathon (idea contest) with university students (maybe in this year). Also a mobile game company as a sponsor will be interested in the possibility of emotional application.

Summarising, we have discussed about the future direction of activity recognition algorithms. A goal of activity recognition algorithm is to be 'self-adaptive' and 'self-optimizing'.

The main problems of current algorithms are 1) dealing with only static label set, 2) dealing with no ambiguity of labels, and 3) performing with fixed sensor modality.

To solve these problems, activity recognition researchers should tackle with the following issues;

- Dynamic changes of activity set
- Location, time environment and etc.
- Add new/ remove old activities
- Ambiguous interpretation of an activity
- Hierarchical model
- Ontology model
- Mixed activity label
- Location
- Situation
- With or without object
- With or without someone else
- Very short/long activity
- Sequence of activities (temporal relation)
  - E.g. <0.5 second
- Large number of recognizing activities
  - E.g.1000+
- Group activity
- Accurate Recognition (No FP/No FN)
- Tolerable delay (Quick response)
- From 0.1 to 0.5 sec (<=Human perception)
- Prediction
- Generalization
- Reuse existing data/model(Transfer learning)
- Extend existing data/model
- Personalization
- Extract the characteristics of a person

- Personalize the model
- Model maintenance
- Follow the change of sensors, activity set, environment, and etc.
- Real-time model construction
- Large dataset
  - E.g. Peta-byte
  - From a lot of people, from a lot of sensors on a person, or a lot of data for a person
- Feature selection
  - Deep learning
- Adaptive sensing
  - Sampling frequency
  - Resolution
  - Range
- Meta-classifier(Classifier creator)
- Emotion recognition

3.5 Systems Group Report

Nils Hammerla, Till Riedel, Philipp M Scholl, Yutaka Arakawa

Our discussion started off with an exchange of experiences on the practicalities of data collection and analysis for activity recognition. We found that isolated datasets suffer from task and domain-specific bias, and knowledge transfer across them is extremely limited. Similar fields, like computer vision, put more effort into curating datasets and their interoperability to increase the comparability of different analytical approaches. Due to complex data collection protocols, heterogeneous sensor modalities and different goals, curating datasets for Activity Recognition is challenging. We intend to reinvigorate the datasets conversation in the Activity Recognition community in an open-forum of discussion and collaboration.

Our primary goal is to, as a community, set a direction toward which we can collectively work in creating a loosely federated dataset infrastructure (e.g. by the means of the proceedings of a workshop), as well as collecting best practices and reproducible experiments. Having better supplementary publications for datasets will allow for transparency across individual datasets and annotations, experimental benchmarks with community-set corpora and metrics, and a web-based infrastructure to cultivate continued developed of Activity Recognition datasets. We will discuss dataset availability, dataset creation and annotation issues, descriptive statistics on datasets, cross-dataset challenge problems and usage potential, as well as practical dataset infrastructure and reuse issues.

Recently, large population-scale data collection projects in the public health domain have produced immensely large, but un-annotated data-sets (such as the UK biobank and similar endeavours). It is an open question how these rich sources of e.g. movement data can be used for system development and evaluation in ubiquitous computing. Crucial for incorporating these sources is that data collection, sensor selection and placement, preprocessing techniques and data storage formats are, to an extent, standardized. One major goal is to establish a set of guidelines or best practices for data collection in ubiquitous computing, and to establish requirements for tools and frameworks used in the field that would allow for this interoperability.

The Shonan meeting provided us with the opportunity to make future plans on curating best practises and exchange experiences on data collection.

References

[1] Xuan Bao, Songchun Fan, Alexander Varshavsky, Kevin Li, and Romit Roy Choudhury. Your reactions suggest you liked the movie: Automatic content rating via reaction sensing. In *UbiComp'13*. ACM, 2013.

[2] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: Sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4), 2006.

[3] Paul Ekman. Are there basic emotions? 1992.

[4] Tom Geller. How do you feel?: Your computer knows. *Commun. ACM*, 57(1):24–26, January 2014.

[5] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. Moodscope: Building a mood sensor from smartphone usage patterns. In *MobiSys 2013*, MobiSys '13. ACM, 2013.

[6] Oscar Mayora, Bert Arnrich, Jakob Bardram, Carsten Dräger, Andrea Finke, Mads Frost, Silvia Giordano, Franz Gravenhorst, Agnes Grunerbl, Christian Haring, et al. Personal health systems for bipolar disorder anecdotes, challenges and lessons learnt from monarca project. Pervasive-Health'13, 2013.

[7] Matei Negulescu, Jaime Ruiz, Yang Li, and Edward Lank. Tap, swipe, or move: Attentional demands for distracted smartphone input. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 173–180, New York, NY, USA, 2012. ACM.

[8] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–14. ACM, 2014.