Media of Things: Supporting the Production of Metadata Rich Media Through IoT Sensing

Gerard Wilkinson[1,2], Tom Bartindale[1], Tom Nappey[1], Michael Evans[2], Peter Wright[1], Patrick Olivier[1]

[1]Open Lab, Newcastle University
Newcastle upon Tyne, UK
g.wilkinson@newcastle.ac.uk

[2]British Broadcasting Corporation
Salford, UK
mike.evans@bbc.co.uk

ABSTRACT
Rich metadata is becoming a key part of the broadcast production pipeline. This information can be used to deliver compelling new consumption experiences which are personalized, location-aware, interactive and multi-screen. However, media producers are struggling to generate the metadata required for such experiences, using inefficient post-production solutions which are limited in how much of the original context they can capture. In response, we present Media of Things (MoT), a tool for on-location media productions. MoT enables practical and flexible generation of sensor based point-of-capture metadata. We demonstrate how embedded ubiquitous sensing technologies such as the Internet of Things can be leveraged to produce context rich, time sequenced metadata in a production studio. We reflect on how this workflow can be integrated within the constraints of broadcast production and the possibilities that emerge from access to rich data at the beginning of the production lifecycle to produce well described media for reconfigurable consumption.

Author Keywords
Film; sensor-based productions; production roles; ambient sensing; metadata; IoT.

ACM Classification Keywords
H.5.2 [User Interfaces]: Input devices and strategies, Interaction styles.

INTRODUCTION
Broadcast TV technology has evolved at an incredible pace over the past century. Broadcast media is no longer limited to the TV placed in the living room, where the family gathers around a singular broadcast. The availability of content to myriad devices and platforms has enabled access to content never envisioned by earlier broadcast production pipelines, and as such, the industry is struggling to keep up with the array of available consumption scenarios. Current production workflows are predicated on a singular linear output, usually a TV broadcast, and so the media they produce is often without the necessary descriptive information at the correct granularity to be reconfigured. To support emergent consumption scenarios such as interactive narrative, personalized and multi-device content, new processes and workflows must be developed. Core to these workflows is a rich description of the media being reconfigured to enable reasonable and appropriate reconfiguration decisions to be made.

This move to treating media as reconfigurable descriptive objects is an approach adopted by many in both industry and academia, and is understood as Object Based Media (OBM) and the associated production processes as Object Based Broadcasting (OBB). Currently, what little OBM that is produced by broadcasters is generated using resource intensive post-production workflows for specific outputs (such as companion apps). By following non-destructive practices to media production and delivery, broadcasters can enable future reconfiguration and manipulation of the media allowing for emergent consumption scenarios. However, without detailed metadata associated with each media object, reconfiguration opportunities are significantly limited. Although the industry has responded by streamlining end-to-end production through; digitization and standardization of timecode; licensing and the standard inclusion of some semantic metadata [54], much descriptive information, and in particular - the context in which the source video is shot, is lost. This information is either never captured, captured in a non-machine-readable manner (i.e. clapper boards or hand-

Figure 1: MoT deployed on-location at a film shoot

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04.
https://doi.org/10.1145/3173574.3173780

written notes), or is lost during the traditionally destructive (i.e. original source media is not kept) production workflow. Combining this data and filling the gaps that exist in context and meaning of production assets is vital in the adoption and proliferation of OBM in the production pipeline.

Drawing from the field of Internet of Things (IoT), we present our open source [51], embedded, sensor-based, metadata capture solution which integrates within a film set. Through discussion of its design, we demonstrate how embedded ubiquitous sensing technologies can be leveraged to produce context rich, time sequenced metadata at point-of-capture in a studio scenario. Through reporting on a real-world production where we deployed this system in conjunction with traditional production roles, we discuss how such a system can semi-automate the process of context and more general metadata capture, and how such data-capture roles should be introduced as core parts of the production workflow.

THE ROLE OF METADATA

Reconfigurable media affords new opportunities for creativity in film making and broadcast media. Media is no longer created for one consumption scenario (such as TV broadcast), in one linear pre-packaged form; instead it is adapted dynamically in response to the context of consumption. For example, when viewing a cooking show on both a tablet and TV, close-ups of food preparation could be displayed on the tablet and the presenter could be displayed on the TV. Although some approaches to OBM generation have explored how to generate context based descriptions during capture, they are usually applied in response to pre-defined consumption scenarios such as branching-narrative [48,49]. As such, the type and scope of metadata required is carefully controlled, but lacks the flexibility to produce other types of reconfigurable media at a later date. The task of 'getting back' this lost context is thus a more complex and resource intensive process. The existing production pipeline produces myriad metadata throughout the workflow, indeed various roles are entirely based around the creation and maintenance of this information. As an example, the *script supervisor*, *continuity editor*, *production assistants* and *camera teams* all use intricate systems of metadata to maintain quality and consistency during a shoot, however this data is: destroyed as part of the post production workflow; and based around tasks rather than context of content of the media.

RELATED WORK

Object-based Broadcasting (OBB) [12,43] is a new approach to the production of broadcast media. Traditional broadcast pipelines in both radio and video broadcast a packaged, edited, single stream of information to all users, regardless of playback device or environmental factors.

OBB looks to rethink this workflow, offering media as a set of assets with associated metadata describing each asset. This metadata is defined as supplementary information that provides description of the scene, camera information, location information or contextual information about what is happening in each video shot.

In the most generic case, OBB assumes no pre-determined use case for any media. This enables elements of the consumption experience (such as timing, narrative and display size) to be decided not by the broadcaster during post-production, but by the consumer, their device or environment (or a combination of these) based on narrative rules or external parameters.

As an example; consider responsive web design, developers build their sites to resize, re-order and reconfigure based upon the available screen real-estate. A website will look different on a desktop compared to a mobile device due to a set of rules put in place by the developer. The developer did not need to redesign the website for every single display size when creating the content. OBB can be thought of similarly, where a set of rules or protocols defines how the media objects should be combined, allowing us to achieve a similar level of responsiveness by re-framing the media to fit the device. This could be as simple as presenting a traditional wide-screen shot to desktops but an appropriately reframed portrait feed could be shown on a mobile device. Some more complex examples may include: redistributing the content across an array of displays in an environment or other environmental devices such as lighting; adapting to individual user needs with character re-introductions in TV shows; or adapting a cookery show to accommodate multiple consumers cooking along together.

During a TV show, custom catch-up experiences could be presented to users combining salient sections of previous shows into the latest episode to allow users to more easily catch-up or understand the context behind scenes they missed and this could be personalised to each user based upon their viewing history. To achieve such a result, careful labelling of each asset (video feed, audio channel, pictures, text) is required, this information needs to be as detailed as possible to enable the device or user to make informed decisions about what to display. Each camera shot needs a careful description of the view, actions in the scene and potentially device (e.g. appliance) usage.

Previous work on OBB has approached this issue through manual labelling of media. Churnside *et al.* [11] experimented with reconfigurable radio shows, sections of elaboration could be varied depending on the available time for the user. This content was produced from an existing radio broadcast. Logical dependencies from parts of the story were produced and sections labelled as "introduces", "expands" or "resolves". A narrative graph was then produced to encapsulate the production. Cox *et al.* [13] looked to produce some bespoke OBB content. Their system, "CAKE", was an interactive cooking show which allowed viewers to select individual parts of a recipe to cook and swap some ingredients for others. The viewer experience reconfigures in response to these selections. Again, the metadata to support this experience was constructed

manually as a post-production process. This not only restricted the possibility for configuration as each scenario has to be known beforehand but also requires a significant amount of manual labelling of video in post-production. They successfully demonstrated an object-based production; however, they lacked a sustainable means to construct the rich content required for the consumption scenario.

Tools to Produce Metadata

The specificity of each consumption scenario is a recognised constraint of current OBB workflows, and recent work as part of the 2Immerse [28,55] project has approached how production tooling and workflows must adapt to new forms of media production to allow for a sustainable workflow. The requirement to generate rich context based metadata for use in these reconfigurable contexts is an ongoing research field, roughly split into two areas: post-production generation; and point-of-capture generation.

Post Production

Previous work in automated labelling and metadata gathering in video has focussed on Computer Vision. Object detection within regions of an image was originally explored by Duygulu *et al.* [16] and Barnard *et al.* [3], where the authors could identify distinct objects and demonstrated reasonable accuracy in generic objects (planes, animals, scenery). This was expanded to full scene descriptions by Kulkarni *et al.* [26] demonstrated a means to construct full scene descriptions combining previous work to describe objects within the scene and using a Constructed Conditional Random Field to produce full scene descriptions. This work was refined by Mitchell *et al.* [30] to reduce the often 'noisy' descriptions produced by Kulkarni's method and similar processes have been successfully commercialised. Google's Vision API [18]; and Microsoft's Cognitive Services both offer highly accurate image descriptions trained on their extensive image search metadata. Open Source Machine Intelligence project TensorFlow [45] demonstrates continuing improvements in image descriptions and performance [42]. Hu *et al.* [22] then expanded this work to video, using generated descriptions to automate the detection of unusual activity on CCTV cameras and real-time video scene analysis is also available through Microsoft's Cognitive Services. Although these techniques are now becoming highly accurate they lack detailed information on non-visible context such as performative action and intent, which can only be obtained during the capture process.

Point-of-Capture

To overcome this, Red-tag [4] takes a different approach to object identification. The authors look to detect which objects are currently within a scene using IR emitters flashing in an encoded pattern mounted on objects and people in the scene. This enables timestamp identification of labelled objects recorded on any camera pointed into the scene. In an alternative approach, Bartindale *et al.* [5] demonstrate that pre-defining some context based metadata before capturing video by using on-screen templates of shots on a mobile device can produce reliable context based metadata, but this approach is limited in the type of context that can be recorded (i.e. shot angle, label of who is in shot).

Automatic metadata generation systems are widely used in sports broadcasting. Soccer statistics [56] and goal-line technology [57] supply a wealth of real-time data that provides a compelling consumption experience for viewers and real-time assistance to referees in matches. In film production, applications to support script supervisors also demonstrate the opportunities for digitised metadata capture, albeit human powered. Tablet applications such as LockitScript [58] and Script Evolution [59] allow the digital recording of script notes and time cards at the point of capture. Such tools however are role specific and do not form part of a coordinated output of production metadata used outside of the shoot.

Current literature thus demonstrates advances in object detection and recognition, but identifying state and interactions being performed is out of reach. For example, in a cookery show, the speed of a blender, the amount of water used, the current utensil in use are all valuable pieces of information that are required to create rich consumption experiences, which are difficult to obtain with Computer Vision.

Production Usages of Context Based Metadata

Examples of using metadata to drive a production are also prominent in sports broadcasting. These solutions mostly relate to identifying the most salient parts of a stream such as goals, in order to generate automatic highlights [1,15]. Using specific features (goal posts, crowd movement etc) they can determine the most interesting parts of a game and automatically summarise them in a variety of formats.

Indeed much of the current literature focusses on attempting to understand where the most interesting segments of the video are and work such as Virtual Director [25] has attempted to fully automate the editing process by framing close-up shots from wide-angle high resolution camera images based on contextual metadata.

Similarly, Schofield *et al.*'s [40] 'Bootlegger' smartphone application coordinates a 'crew' of users covering a live event, assisting them in producing a quality film with shot overlay templates and other shot guidance. Recruited members choose from a set of roles defined by the shoot organiser and an 'auto-director' assigns shots to be taken to appropriate roles, in a live filming scenario the 'auto-director' ensures complete coverage of the event and provides a complete set of shots (from start to finish of the production) to be composed into a film later based on the context based metadata captured by the devices about what is being shot in real-time.

Rather than human powered, Kaiser *et al.* have investigated using Computer Vision to make shot decisions by framing shots out of a wide-angle camera placed within a soccer stadium [23]. They attempted to track the position of players, the ball and rapid movement, creating shot decisions based

on these parameters. However, they found after much experimentation that Computer Vision alone struggles to provide sufficient information to make informed shot decisions. In reality, the resulting media was mostly of the same thing, and did not relate directly to the action. Kaiser *et al.* [24] also explored how the narrative of a production is affected by the placement of 360° high-resolution cameras throughout a theatre production's stage. A performance was conducted to explore this new perspective on theatre productions, essentially moving the audience to the middle of the stage. The authors deployed similar Computer Vision techniques to attempt to automate shot selection and guide the viewers' attention. Furthermore, there has been some recent consumer products like Mevo [27], a 4K wide-angle camera with live shot framing for livestreaming services like Facebook Live [17]. BBC Primer [7], a real-time web-based shot framing system hopes to utilize 4K unmanned cameras set-up to cover a wide shot. Enabling directors to change the shot simply by framing out a new segment. Given all the work in this space however, these scenarios are still limited by the availability of reliable and detailed metadata to drive each production process.

Metadata Driven New Consumption Scenarios

Media with associated metadata has historically been used to drive a wide range of consumption scenarios, from some of the first usages with 'Interactive TV' services such as 'red button' services [19] to newer 'second screen' experiences. A classic example, are interactive and branching narrative experiences that require carefully constructed and labelled media sets to enable a rich and meaningful experience. Ursu *et al.* [48,49] initially explored the possibilities of this genre for interactive TV and produced 'Shape Shifting TV'. The authors explored atomisation of content, enabling the media to be reordered with a lower impact on the narrative and explored reconfigurable media within documentary production [50]. However, these workflows required a large amount of additional pre-production work from that of traditional TV. Zsombori *et al.* [53] continued this work by attempting to leverage UGC content and narrative building to circumvent such limitations, however the style and production value of such content is limited. 'TryFilm' [6] approached the problem by allowing the cast and crew of an interactive narrative production to interact with the content on location. The intention was to bootstrap the post-production process by creating useful metadata during the production, however this proved difficult in practice.

More recently we have seen a concerted effort by broadcasters to deliver detailed and immersive second screen experiences. Indeed, Pablo *et al.* & Obrist *et al.* explore these interactive and second screen experiences [9,10,33] envisaging the second screen as a device that compliments the viewing experience both in terms of providing control and extra information about the current content as well as social media integration.

Indeed, many consumers are moving away from traditional forms of media consumption with severe declines in set-top boxes in homes and the consumption of Live TV [46,52], instead preferring more convenient and discoverable media. On-demand services like Netflix and Amazon Prime Instant Video allow for a much more flexible and accessible means to enjoy media. The delivery of these services via the Internet also opens up new levels of bandwidth and control that enable Object-Based Broadcasting [12,43]. The wide availability of consumer devices such as Google Chromecast and Amazon Fire TV also enable a cheap and convenient way to provide On-Demand and potentially interactive services to existing TV consumers thus providing a practical means to create the immersive media experiences made possible by OBB.

Sensing in the Environment & Internet of Things

Inherently, vision based metadata generation solutions lack the ability to describe actions and processes that are not visually captured in the media. With the emergence of IoT as an approach to achieving ubiquitous sensing environments, it is within reach to envision consumption scenarios which make use of IoT devices as part of the viewing experience. By considering the production studio as an embedded sensing environment, the viability of IoT for capture as well as output is immediately made clear, however current applications of IoT have largely been confined to two main areas:

- Smart-Homes: Connecting hardware within the home to the internet enabling new levels of control and automation of devices such as thermostats [31], lighting [38], locks [2] and appliances [44].
- Industrial Sensing: Deploying cheap, reliable sensing equipment for real-time reporting and analysis of production lines [29], manufacturing infrastructure [14] and asset monitoring [8].

The latter application, while lesser known, is far more prevalent. Previously manual processes such as meter reading, and parcel tracking, for example can now be automated and validated by embedded technologies in the environment. Such sensors embedded within the environment could be used to provide the detailed metadata required to achieve Object Based Media. Since many filming scenarios are based around the same constraints as consumer IoT (i.e. the home), the approach of smart environments provides a natural symmetry to smart production environments. In a practical sense, this symmetry grounds reconfigurable media in what is possible for consumers while at the same time potentially reducing production costs.

In recent work, Churnside *et al.* and Cox *et al.* [11,13] discuss the possibility of IoT for consuming responsive media, where the playback system responds to actions from the consumer to adjust what is played. Whilst not producing media, the Ambient Kitchen by Olivier *et al.* explored using a sensor instrumented kitchen to gain context from user

actions in a kitchen and aid patients with dementia doing tasks [20]. This project drew inspiration from Ambient Intelligence proposed by Tscheligi *et al.* [47], a new concept of user centred design. Olivier *et al.* and Chong & Olivier [34,35] were able to perform stage detection using the sensors within the environment helping dementia patients to complete tasks. They did however struggle with detecting true task completion and action qualities and the work required excessive instrumentation of the environment to achieve this. Further work based in the Ambient Kitchen investigated the application of such technology to assist in language learning [21,39,41] through cooking. The system cued participants with ingredients or tasks in a new language which had to be identified/used before the recipe continued. These works demonstrate the power offered by connected environments with many of the application scenarios immediately applicable to media consumption in a connected environment. Instrumentation of production studios and capturing using IoT technologies provides a cheap readily available source of sensor data to exploit existing production data sources (camera positioning, props, appliances etc.) as well as new potential data sources (positional information and interaction data) to describe the created media.

APPLICATION SCENARIOS

OBB (object based broadcasting) opens a wealth of opportunities for novel media consumption experiences. To help define the requirements for point-of-capture metadata generation, we can think about a number of key production scenarios that can be supported through the production of descriptive and context based metadata.

New Consumption Scenarios

Cookery is a prime example of a rich consumption scenario with possibilities for OBM production. The kitchen is a constrained scenario that suits sensing in both production and consumption, as such, instrumentation and reconfiguration possibilities are broad and more feasible.

Task Allocation for Multi-User Playback

Playback of a cookery show is typically a linear film with a set of tasks to be completed. In a multi-user playback scenario, each user could be assigned a set of tasks to complete to provide a collaborative cooking effort. To be able to adequately understand where tasks start and finish and which tasks are dependent upon the completion of each other we must have rich metadata describing each asset. This rich metadata enables the media to be segmented and assigned to each user, leading to more efficient or playful collaborative playback of the media.

Metadata required: location of actor, position and usage of ingredients, use of appliances and utensils, blocking points in production (i.e. put into oven).

Adaptive & Customizable Playback Experience

Structured and labelled media affords new opportunities to adapt media to environmental, physical or other constraints, as well as offering the users means to customize their experience. A cookery show, where precedence and dependencies have been determined allows for a show with multiple components to have sections removed and added dynamically. Simplistic changes such as swapping ingredients and adapting to available equipment in the environment is also possible. The production can adapt to the available screen real-estate, user configuration and other environmental and physical factors, offering a more compelling and appropriate consumption experience for a wider range of devices and environments.

Metadata required: Segmentation of tasks, description of shot types and identification of ingredients in scene.

New Production Workflows

Semantic labelling of raw media assets at the point of capture affords new opportunities for existing production workflows. We outline a few examples of the application of this context metadata to the post production workflow.

Sensed Editor

Structure, labelling and organization of media produced from productions is currently limited to those provided by the production team and camera operators. Using metadata generated during capture we can provide richer semantic labelling on media, facilitating more meaningful queries on corpus of media. For example, rather than manually searching by thumbnail or preview, a search could be performed using actions or content e.g. “Filleting of the fish.”, “Washing vegetables in sink.”. This offers streamlining of the post-production editing process, enabling quick swapping of shots and the creation of powerful tools for editors to use more natural language querying of media sets.

Metadata required: segmentation and labelling of individual shots

Quick Editor

The quick editor takes large overwhelming sections of footage from a production and makes simplistic section breaks in order to produce a more approachable and cleaner editing workflow, reducing the ingest time. Initially the system uses the metadata to identify the sections where it deems the production was paused, removing breaks in the shoot. Then dynamically generating sections from the sensor data based upon continued usage of certain objects labelled sub-sections. This is presented in a chronological order with generated labels describing the utensils, appliances and general area of interest. Editors can preview the generated content from each camera angle, select their preferred shots from each section. Once each desired shot is selected the editor can finalize the edit and the production is cut together.

Metadata required: media labeled by camera angle, segmentation and content of clip.

MEDIA OF THINGS

In response to requirements of these example scenarios we developed Media of Things (MoT), a production workflow and supporting infrastructure that is designed to flexibly support the creation of context based metadata at the point of

capture by embedding multiple IoT sensors in a production environment. MoT is designed to capture raw sensor data, in a film shoot context to enable the subsequent creation of OBM as shown in Figure 2.

MoT applies the principles of ambient sensing [47] in which sensors are embedded within the production environment in an unobtrusive manner. This is key to video production where any visual change to the set would not be acceptable. MoT consists of a ***'capture'*** system for recording raw sensor data from IoT sensors in a production environment, and a ***'post production'*** workflow which segments and contextualizes this data alongside any video shot on location.

The nature of OBB being the segmenting and description of objects from raw data streams (such as video, audio and text) means that it is important to capture data in as granular form as possible, so that inferences on the data can be made at a later time as new uses and forms of objects emerge. As such, MoT's primary purpose is to capture time-based raw sensor readings from multiple inputs. Inferences on sensor readings are then made on the data during post production.

MoT is designed to be agnostic to sensor type and volume, and produce segmented metadata without prior training of specific actions as a starting point for media object creation.

Capture System

MoT supports many levels of data abstraction, and the capture system is designed to reliably record and store native sensor readings from multiple sensors in real-time. This approach to collection of metadata is in line with other research in metadata capture. Project Orpheus from BBC R&D [60] is a project looking at rich metadata in radio, in this example the raw input and output from the mixing desk is captured for later processing. This approach of maintaining the original data allows for 'objects' to inferred from the data during post production.

Running hundreds of sensors continuously and reliably within a production scenario where shooting may not be able to be repeated, with all data correctly time-synced requires a robust system architecture. The system architecture for supporting the continuous streaming of the sensor data was largely informed by current best practices with regards to large IoT sensor deployments as well as similar architecture to the one deployed for Oliver *et al.'s* Ambient Kitchen [34]. A message queue server is at the heart of the system to handle the sheer volume of messages from the sensors. All sensor data is stacked into structured queues to allow listeners to retrieve data for set of sensors or a specific sensor. The data storage listens on all data sets and stores all data with timestamps for replay. The current system architecture is flexible in that it will accept additional sensors or new sensor types.

A RabbitMQ message queue server running on Ubuntu 16.04 sits at the heart of the Media of Things. A MongoDB instance runs on a separate Ubuntu 16.04 instance with a capture script running to save all data running through the message queue. All sensor nodes were running Universal Windows Apps running on Windows 10 IoT on Raspberry Pi 2.

Inference System

MoT records raw sensor data at a variety of rates from multiple types of sensors. To provide useful inferences for OBB usage, the data is passed through three processing steps.

i) Raw sensor data is converted into consistent discrete time windows and common data format
ii) this sensor data is then interpreted into usage information, monitoring when each implement and object is used and when they were last seen.
iii) this usage information is mapped against user-defined areas of interest (physical areas of the film-set) generating a probability for each time window of where the focus of the production should be.

The inferences, detailed below, are obtained from the usually one-to-one relationship of sensor to object, rather than explicit activity recognition. Many of the sensors were embedded in the cookware and utensils directly so deeper activity recognition would have yielded only marginally more information. A knife has a limited set of uses, within this context, as does a peeler, therefore basic usage information suffices. These inferences enable us to build a logic model for selection of media, a query able dataset synced to the relevant media assets can provide editors with a

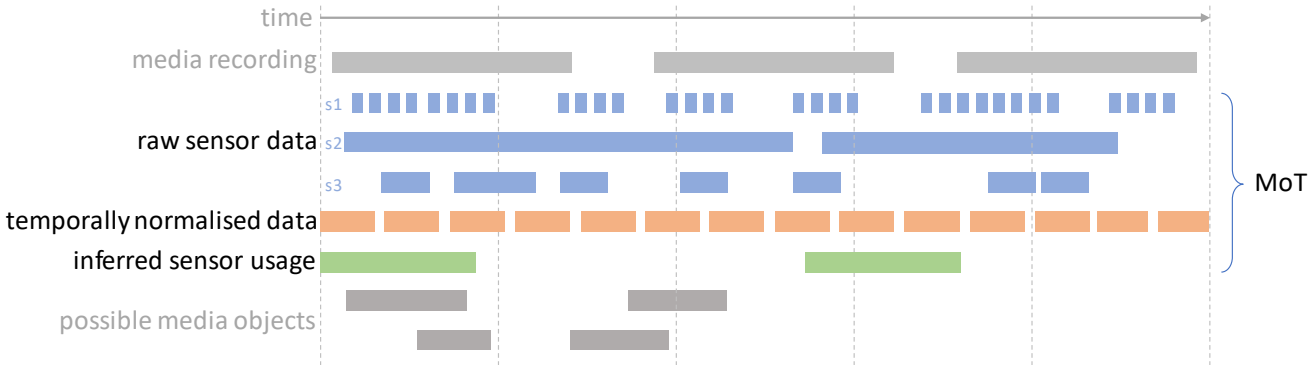

Figure 2. MoT captures and stores time based raw sensor data producing inferences that can be used to create Media Objects

powerful tool in post-production.

Object Usage Information

Data from sensors are combined into a set of inferences about object usage, including last usage time and duration of current usage. These higher-level inferences provide another level of metadata from footage, cleaning up noisy raw signal data. This information can be further decomposed into a taxonomy of actions related to each object. As an example, the raw signal from an accelerometer in a knife is constructed into a time-sequenced representation of when the object was last seen, and duration of last/current usage.

Area of Interest Classification

Object usage information from the previous step is combined to produce a probability value for each area of interest in each time window. This value represents the likelihood that this area should become the focus of the production. Degrading significance of pre-configured areas of the production studio are used to maintain the classifications. Sensors have been attributed different significance values, RFID transponders moving or appearing are assigned high significance if they can be located to specific area. Sensors without inherent location (i.e. not RFID) such as utensils are used to maintain the current significance level across the available areas. Appliance usage when located to an area with an RFID transponder is attributed a high significance. This classification could be used to, for example, make decisions on which shot is appropriate in a shot framing system.

Future Activity Recognition

The multi-layered storage of sensor data and inferences lends itself to future work conducting more detailed activity recognition, no data is lost. One immediate challenge however, in introducing deeper activity recognition beyond basic usage information is this increases the specificity of Media of Things. As such, balancing the improvements in gathered sensor data while maintaining our vision of the platform being sensor agnostic is an interesting area to explore. The recognizer architecture adopted in the MoT with sensor data read off the message queue in real-time processed and all data piped back in enables a wealth of possibilities for many layers of recognition with the message queue the point of scale. However, each new sensor type introduced and application scenario envisioned requires some bespoke development of activity recognizers to support the generation of new inferences.

THE PRODUCTION

As demonstrated by our examples scenario, the cookery show is a well-defined scenario in which to test such a production tool. Alongside multiple rich new consumption scenarios, a kitchen is a physically constrained scenario with many opportunities for embedded sensing. In addition, cookery shows maintain a distinct segmentation related to actions on screen, allowing us to test the applicability of our MoT implementation more easily.

In partnership with a major broadcaster, we commissioned a cooking programme and hired a team of producers, script writers, lighting, sound and camera crew. The production was designed around Object-based Broadcasting as its intended output and as such the recipe selections were made that would enable reconfiguration with a range of complex technical tasks and utilizing a range of appliances. The production was run with a single chef cooking 4 recipes, chosen as ones with a wide range of technical tasks. Requiring chopping, filleting, baking, frying and Bain Marie melting.

MoT for Cooking

To deploy sensors and IoT devices within the shoot environment which were discrete and largely invisible, we constructed a custom, fully working kitchen island. Although outwardly this looks like a traditional film-set, internally it contains many embedded sensors. To explore which types of sensors would generate adequate information about context, we selected a number of key technologies to offers both redundancy and variety:

Accelerometers

OpenMovement WAX3 [32] wireless streaming accelerometers were chosen to capture usage information for

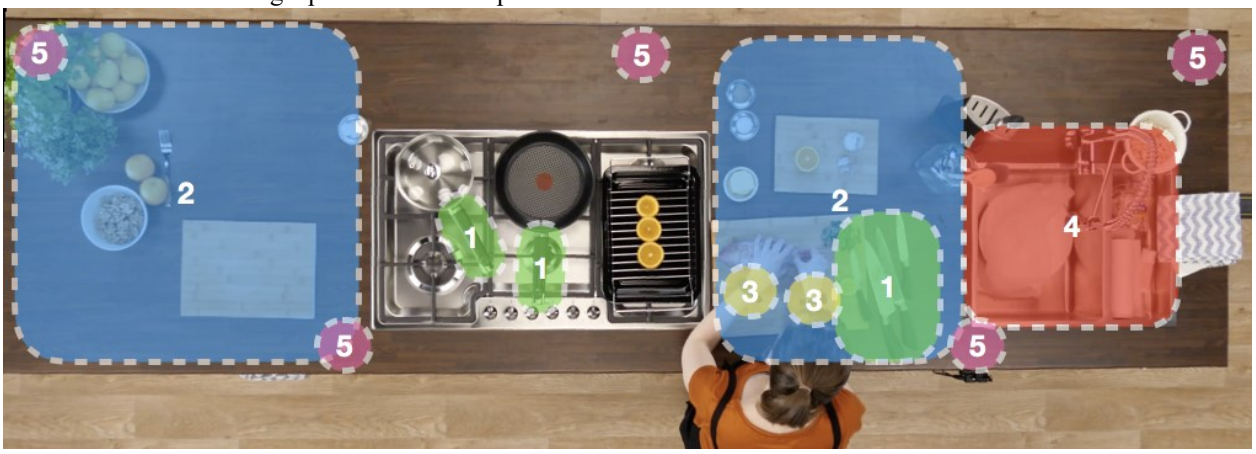

Figure 3. MoT Kitchen Sensing: (1) Accelerometers in utensils, pots and pans; 2. RFID positional information on ingredients and cookware; (3) IMU on chef's wrists; (4) Water flow meter on sink tap; (5) Bluetooth beacons for position of chef.

utensils, pots and pans. These sensors are small enough to embed without being seen in objects. Accelerometers are a key technology in the fabric of IoT, and provide basic usage information, what utensil or equipment is being moved at any given time. Through a post-production process, this data has the potential to provide rich activity information such as whisking, chopping etc. as demonstrated by Pham and Olivier [37].

In addition, OpenMovment WAX9 accelerometer and gyroscopes packages were worn by the chef on each wrist and deployed on the refrigerator and oven doors. In production use, these could be replaced by consumer Smart Watch or Fitness bands. BLE and Zigbee receivers for these sensors were placed inside the kitchen unit and connected via IP to MoT.

BLE Location

Indoor positional information was captured by a LG G Watch [61] placed in the pocket of the chef. This device continuously measured the RSSI of 6 Estimote [62] beacons located strategically around the set.

RFID

Benchtop appliances, bowls, plates, and other large items were tagged with RFID labels. Embedded beneath the kitchen worktop were 6 FEIG ($0.5m^2$) RFID pads [63]. Objects were continuously tracked when on the bench and identifiable from the transponder IDs referenced from a production configuration. RFID transponders are cheap to deploy and come in a range of shapes and sizes to accommodate as many assets as possible. These were placed underneath workbenches, with transponders placed on appliances, cookware and ingredients bowls.

Water Flow Monitor

Water flow meters [64] were placed on the pipework in the sink measuring exact amounts of water dispensed. Since the water system was self-contained (waste and fresh water), energy monitors were also placed on the pump to record usage.

Appliance Usage

Energy signature monitors are widely available on the consumer market. We deployed 2 Smappee [65] devices to monitor usage on plug sockets within the kitchen island to monitor benchtop appliance usage. Combined with RFID positional information of the appliance, this provides a rich picture of appliance usage.

The Film Shoot

The production took place over 3 days in a black box studio location in 2016, in collaboration with the BBC. The crew consisted of ~20 professionals including lighting, set, electrician, camera crew, director and producer and talent who were hired for the production. A TV chef was employed to perform the preparation of 6 meals on set, which were captured by both static and manned cameras over the production. In total, we captured ~12 hours of sensor data using MoT and continuous video footage from 4 DMC-GH4 [66] using BBC Primer [7] for time synchronization and capture from the production over 2 days of shooting. To aid with data validation and to enable development of future tools, we deployed an un-manned 4K static camera system consisting of 4 wide-shot cameras around the set. These were capturing continuously during the shoot. MoT was deployed to capture raw sensor data during this same period. Around 72000 sensor data points were captured alongside 4 camera angles. Three members of our research team joined the production team to manage the equipment, data capture and physical components of the augmented set.

PRODUCTION REPORT

The production was an overall success in terms of data and video captured, however a number of incidents and specific failures during the shoot offer us a chance to reflect on how MoT performed within this context.

Whilst the production was planned with the whole production team being aware of the focus on OBB, the production workflow failed to adequately adapt. The need for sensors within the set formed a limited part of the production planning resulting in utensils and cooking equipment being hard to augment with sensors once on location. RFID transponders provided reliable object location information throughout the production, however due to the equipment being washed between shooting sequences, RFID tags were sometimes accidentally removed from devices, requiring the researchers to re-apply and record tags. Additionally, the material of some objects (such as glass bowls) prevented tags being attached at all.

During the production, the WAX9 sensors (to be used for local positioning of the chef) unfortunately failed to function during the production due to a data transfer issue, resulting in malformed data. Additionally, a laptop being used as a data capture relay overheated under the studio lights resulting in the data being lost for a take early in the shoot and before the start of the production the water flow meter failed, however the energy signature monitor offered us redundancy for tap usage data.

The success of the data collection largely vindicated our choice of sensors for this type of media production. In particular, the WAX3 accelerometer data produced a consistent and useable dataset.

Very quickly, our researcher operating the MoT system became integrated into the production team's workflow, and roll-call. Temporary sensor failures and resets became analogous with lighting adjustments and camera focus issues to the production team, and the researcher was considered an equal member of the production team who was able to call 'halt' when required for adjustments. This enabled many temporary issues to be addressed immediately resulting in a more consistent dataset. However, understandably we were able to use this relationship in a limited fashion, as maintaining the production schedule was of primary importance.

In total, data from 25 kitchen implements (i.e. utensils, pots and pans) was recorded, alongside RFID position information for 28 objects (i.e. plates, bowls and ingredients), and power usage data from 3 appliances. After performing MoT's post production inference process, the data consisted of ~450 discrete movement actions, and ~200 changes in positional information for objects on the kitchen unit. The output of which was time-based metadata related to individual tagged items on set, to be mapped against the production notes made by the team to determine which utensil or item was used.

DESIGN RECOMMENDATIONS

The deployment of MoT integrated within an existing production workflow provides us with invaluable insights into the suitability of existing production. Through self-reflection by the research team and observational notes of the shoot on our deployment and the challenges and successes of integrating MoT within a film production workflow, we can draw out three key design recommendations for OBB context capture.

Introduction of the 'Sensor Operator'

The introduction of new technologies and workflows into film is a slow and difficult process, due in part to the inherent link between the experience of film crews and the technology they use. The industry's response to the introduction of new workflows has historically been to apply a new team role, as seen with the introduction of digital film media (the 'DT' role), and computer graphics (the 'CG' role) who fits within the well understood practices of the team. It is this standard practice that enabled our researcher to build legitimacy within the film crew as a member of the production team, thus providing the ability to request pauses of the shoot (through negative confirmation when 'ready sensors' was shouted) to reconfigure or fix emergent problems. Being physically located on set also increased recognition of the process to the rest of the crew, which led to easier integration of production tasks such as resetting objects on set for continuity. Given that MoT can provide live feedback of sensor inferences and instant recall of time based data, we envision this tool could be used by the production team for more efficient continuity adjustments on set.

Embedded Sensors in Production Environments

Our deployment clearly highlights the importance of redundancy in recording sensor data in production scenarios. However, the selection of sensors used to capture is more nuanced and there are multiple factors which contribute to this selection: visibility on camera; ability to embed in the environment; amount of pre-configuration required; access for adjustment and the quality of data captured.

In the context of a kitchen, where the set is physically constrained and a limited, known set of manipulability objects exists, accelerometers proved to be a good balance between visibility and rich data, and provided useful information. We felt however that the use of BLE indoor localization technology did not justify the potential results. To obtain sufficient granularity of position, the sensor technician had to perform a lengthy (~2 hour) calibration of the sensor before it could be used, and the resulting data was only sufficient to identify which end of the set the actor was located.

In many cases (such as with raw food), objects are unable to be directly tagged, so proxy tags are used (such as the packet the food came from), but this has limitations, particularly when dealing with fresh food. Utilizing customized embedded vision approaches such as FoodBoard [36] would be one way of obtaining this data.

Given the expensive and temporally compressed nature of film production however it is important to be resilient to sensor failure, as there may not be a second chance for capture. In our scenario for example, RFID pad coverage overlapped and water flow information was available through pump usage. Therefore, MoT could not only cross-validate sensor data but support multiple points of failure.

Key in supporting this is the Sensor Operator's ability to watch a live feed of all sensors' data being recorded by MoT allowing them to respond appropriately when issues occur, in the same way that the sound operator monitors his recording.

Other Production Types

Media of Things is a flexible concept that should accommodate most studio based productions, because it is a sensor agnostic platform we can accept input from more bespoke sensing units as appropriate for each production type. Non-studio productions are obviously more unpredictable environments with more dynamic meta-data capture needs, detailed action information and descriptive information is therefore more complex to gather. However, there are further opportunities for metrics in sports, character labelling and more.

Configuration as a Key Pre-Production Task

Although MoT is both sensor agnostic and context agnostic and thus does not require pre-training or ground truth data in order to infer metadata, each type of sensor may require configuration on location. In particular, sensors need to be recorded against a production reference that will make sense in post-production (e.g. 'the spoon'). Neglecting to sufficiently plan the film shoot with these tasks and constraints in mind led to confusion during setup of the shoot, as set design, production managers, and sensor operators struggled to find a solution for problems that should have been identified earlier.

Each sensor required careful labelling, and each transponder needed to be mapped to an object. The process of this configuration was very time consuming as a record must be kept of which sensor is assigned to which object, while this can be recovered manually post-production from the footage. This experience demonstrates that sensing needs to be considered at all points of the pre-production pipeline. The type, location and variety of sensors should be considered alongside the design of the set, and production timings

should be adjusted to include sensor configuration and setup. In hindsight, a shared codebook of set objects should have been shared amongst the production team, enabling a shared reference for labelling objects during production changes. Future productions should also consider the importance of labelling individual objects. Each new object added to the production as a labelled artifact or member increases the complexity of the configuration process.

Production Crew Reflections

The impact of MoT on the crew was important to capture. In general, their response was positive, the director commented that the "sensors were seamless" and they went "unnoticed". However, at the outset there was some resistance to deploying them within the environment. The chef was asked to wear sensors on their wrists to enable post-production activity recognition. Initially however the director found this to be too visually invasive, and a compromise was reached regarding costume. The value placed on running sensor capture also raised tensions, with some member of the crew placing more value of video and audio data, and willing to continue shooting without sensors. Future sensor driven productions should consider the impact on costume design, set design and aesthetics as well as the trade-off between sensing importance at each stage of the production including the potential for blackouts.

MAKING USE OF MOT INFERENCES

MoT is designed to facilitate the production of sensor agnostic metadata that can be used to infer specific context throughout the production and consumption pipeline. The raw data and basic inferences about object usage within the shooting environment can potentially be used to drive new workflows, as envisioned earlier in the paper. This generic and fundamentally reusable approach is a shift from existing work in OBM production which requires the output scenario to be pre-defined. We believe this to be an important step towards enabling a sustainable workflow for metadata generation which can be applied without re-configuring well-defined professional workflows.

Although useful to broadcasters and top-down media production, we propose that MoT opens up important opportunities outside of professional production. Much of current IoT development is consumer product driven, thus by approaching metadata generation with IoT as a possible solution we have demonstrated that products already within the consumer environment (i.e. the home) could form a key part of the media consumption and production experience. Indeed, one could imagine how IoT appliances could be used to record rich metadata whilst recording home YouTube cooking videos, or to enrich the viewing of immersive content by adjusting ambient lighting or audio devices.

The workflows and processes we have developed around MoT go towards generating raw and low-level inferred metadata, however we acknowledge that this data without context applied by the production team during post-production is not sufficient to define fully formed context-based objects that could be used directly by a consumption scenario. We would argue that all such media is defined by the creative production processes that directors, producers and designers apply during the rest of the creation pipeline.

CONCLUSION

In this paper, we present Media of Things, our implementation of a point-of-capture metadata capture system which integrates into the existing production workflow. By supporting the capture of raw, time-based sensor data from multiple sources in real-time, MoT provides a solid base for creating rich Object-based Media based experiences in the future. We highlight the symmetry in using IoT based technologies with 'invisible' sensors which can record actions and object use on a film set which responds to the need to capture such contextual and descriptive information from broadcasts in nearly-live scenarios and post production. Through the deployment of Media of Things in a real production scenario, we explore how metadata capture can become part of the film production workflow.

Our primary contribution is the validation of sensor-based metadata capture as a sustainable and flexible process that can be included reasonably within the existing constraints of the media production pipeline. Specifically, we recommend that the 'Sensor Role' should be recognized as a distinct and valued member of the production team, in line with the roles allocated to 'Sound' and 'Lighting' professionals, facilitating a smooth transition into the professional workflow. When planning for MoT style productions however we caution that the selection of sensors is a nuanced and often difficult trade-off between granularity of data and operational requirements for configuration. This is an area that we acknowledge would benefit from more work, trialing different sensors for the collection of data.

Although our deployment was within the situated and constrained environment of a cookery show, the structured and standardized nature of film production means that MoT has been used against the inherent situational factors which define such workflows. Combined with the flexibility offered by our sensor agnostic approach, we envision that MoT could be used as part of the normal production workflow for any production scenario, when appropriate sensing is applied. In summary, we encourage the community to consider how production tools such as MoT which leverage existing trends in IoT can be integrated into existing production pipelines now, to more quickly enable the rich, immersive and indeed exciting experiences that Object-based Media can deliver.

ACKNOWLEDGMENTS

This work was funded by EPSRC awards EP/L505560/1 (BBC iCASE 2014) and EP/M023001/1 (Digital Economy Research Centre), and the British Broadcasting Corporation.

REFERENCES

1. J. Assfalg, M. Bertini, C. Colombo, A.D. Bimbo, and W. Nunziati. Automatic extraction and annotation of

soccer video highlights. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, II-527-30. https://doi.org/10.1109/ICIP.2003.1246733

2. August. August Smart Lock. Retrieved from http://august.com/products/august-smart-lock/
3. Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei, and Michael I. Jordan. 2003. Matching Words and Pictures. *Journal of Machine Learning Research* 3, Feb: 1107–1135.
4. Tom Bartindale, Daniel Jackson, Karim Ladha, Sebastian Mellor, Patrick Olivier, and Peter Wright. 2014. RedTag. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video - TVX '14*, 19–22. https://doi.org/10.1145/2602299.2602303
5. Tom Bartindale, Guy Schofield, and Peter Wright. 2016. Scaffolding Community Documentary Film Making using Commissioning Templates. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2705–2716. https://doi.org/10.1145/2858036.2858102
6. Tom Bartindale, Guy Schofield, Peter C Wright, and Peter Wright. 2016. TryFilm: Situated Support for Interactive Media Productions. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 1410–1420. https://doi.org/10.1145/2818048.2819929
7. BBC R&D. Nearly Live Production. Retrieved from http://www.bbc.co.uk/rd/projects/nearly-live-production
8. BlackBerry IoT. Internet of Things – Asset Tracking. Retrieved from http://uk.blackberry.com/internet-of-things/platform/application-modules/vertical-applications/asset-tracking.html
9. Pablo Cesar, Dick C A Bulterman, and A. J. Jansen. 2008. Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 168–177. https://doi.org/10.1007/978-3-540-69478-6_22
10. Pablo Cesar and Konstantinos Chorianopoulos. 2007. The Evolution of TV Systems, Content, and Users Toward Interactivity. *Foundations and Trends® in Human-Computer Interaction* 2, 4: 373–95. https://doi.org/10.1561/1100000008
11. A. Churnside, M.E.F. Melchior, M. Armstrong, M. Shotton, and M. Brooks. 2014. Object-based broadcasting - curation, responsiveness and user experience. In *International Broadcasting Convention (IBC) 2014 Conference*, 12.2-12.2. https://doi.org/10.1049/ib.2014.0038
12. Tony Churnside. Object-Based Broadcasting - BBC R&D. Retrieved from http://www.bbc.co.uk/rd/blog/2013/05/object-based-approach-to-broadcasting
13. Jasmine Cox, Rhianne Jones, Chris Northwood, Jonathan Tutcher, and Ben Robinson. 2017. Object-Based Production. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17 Adjunct*, 79–80. https://doi.org/10.1145/3084289.3089912
14. Deloitte University Press. The Internet of Things in the oil and gas industry. Retrieved from http://dupress.deloitte.com/dup-us-en/focus/internet-of-things/iot-in-oil-and-gas-industry.html
15. Boon-lock Yeo, Minerva Yeung, Bede Liu Dennis Yow. Analysis And Presentation Of Soccer Highlights From Digital Video. Retrieved June 2, 2016 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.24
16. P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. 2002. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. . Springer Berlin Heidelberg, 97–112. https://doi.org/10.1007/3-540-47979-1_7
17. Facebook. Facebook Live. Retrieved from https://live.fb.com/
18. Google Cloud Platform. Vision API - Image Content Analysis. Retrieved from https://cloud.google.com/vision/
19. Pam Hanley and Rachel Viney. 2001. Pressing the red button: Consumers and digital television. *Cultural Trends* 11, 43–44: 35–60. https://doi.org/10.1080/09548960109365165
20. Jesse Hoey, Thomas Plötz, Dan Jackson, Andrew Monk, Cuong Pham, and Patrick Olivier. 2011. Rapid specification and automated generation of prompting systems to assist people with dementia. *Pervasive and Mobile Computing* 7, 3: 299–318. https://doi.org/10.1016/j.pmcj.2010.11.007
21. Clare J. Hooper, Patrick Olivier, Anne Preston, Madeline Balaam, Paul Seedhouse, Daniel Jackson, Cuong Pham, Cassim Ladha, Karim Ladha, and Thomas Plötz. 2012. The french kitchen. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 193. https://doi.org/10.1145/2370216.2370246
22. Chuanping Hu, Zheng Xu, Yunhuai Liu, and Lin Mei. 2015. Video structural description technology for the new generation video surveillance systems. *Frontiers of Computer Science* 9, 6: 980–989. https://doi.org/10.1007/s11704-015-3482-x

23. Rene Kaiser, Marcus Thaler, Andreas Kriechbaum, Hannes Fassold, Werner Bailer, and Jakub Rosner. 2011. Real-time Person Tracking in High-resolution Panoramic Video for Automated Broadcast Production. In *2011 Conference for Visual Media Production*, 21–29. https://doi.org/10.1109/CVMP.2011.9
24. Rene Kaiser, Marian F. Ursu, Manolis Falelakis, and Andras Horti. 2015. Enabling Distributed Theatre Performances through Multi-Camera Telepresence. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences - ImmersiveME '15*, 21–26. https://doi.org/10.1145/2814347.2814351
25. Rene Kaiser, Wolfgang Weiss, Manolis Falelakis, Spiros Michalakopoulos, and Marian F. Ursu. 2012. A Rule-Based Virtual Director Enhancing Group Communication. In *2012 IEEE International Conference on Multimedia and Expo Workshops*, 187–192. https://doi.org/10.1109/ICMEW.2012.39
26. Girish Kulkarni, Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, and Tamara L Berg. 2011. Baby talk: Understanding and generating image descriptions. *PROCEEDINGS OF THE 24TH CVPR*. Retrieved September 21, 2016 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.232.7714
27. Livestream. Mevo Camera. Retrieved from https://getmevo.com/
28. Britta Meixner, Maxine Glancy, Matt Rogers, Caroline Ward, Thomas Röggla, and Pablo Cesar. 2017. Multi-Screen Director. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17 Adjunct*, 57–62. https://doi.org/10.1145/3084289.3089924
29. Microsoft. KUKA creates a connected factory with IoT. Retrieved from https://www.microsoft.com/en-us/cloud-platform/customer-stories-kuka-robotics
30. Margaret Mitchell, Xufeng Han, Jesse Dodge, Alyssa Mensch, Amit Goyal, Alex Berg, Kota Yamaguchi, Tamara Berg, Karl Stratos, and Hal Daumé III. 2012. Midge: generating image descriptions from computer vision detections. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 747–756.
31. Nest. Nest Learning Thermostat. Retrieved from https://nest.com/uk/thermostat/meet-nest-thermostat/
32. Newcastle University. OpenMovement. Retrieved from https://github.com/digitalinteraction/openmovement
33. Marianna Obrist, Pablo Cesar, David Geerts, Tom Bartindale, and Elizabeth F. Churchill. 2015. Online video and interactive TV experiences. *interactions* 22, 5: 32–37. https://doi.org/10.1145/2799629
34. Patrick Olivier, Guangyou Xu, Andrew Monk, and Jesse Hoey. 2009. Ambient kitchen. In *Proceedings of the 2nd International Conference on PErvsive Technologies Related to Assistive Environments - PETRA '09*, 1–7. https://doi.org/10.1145/1579114.1579161
35. Cuong Pham, Clare Hooper, Stephen Lindsay, Dan Jackson, John Shearer, Jurgen Wagner, Cassim Ladha, Karim Ladha, Thomas Plötz, and Patrick Olivier. 2012. The ambient kitchen: a pervasive sensing environment for situated services. Retrieved December 14, 2015 from http://eprints.lincoln.ac.uk/15375/1/DIS2012demo.pdf
36. Cuong Pham, Daniel Jackson, Johannes Schoening, Tom Bartindale, Thomas Ploetz, and Patrick Olivier. 2013. FoodBoard. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*, 749. https://doi.org/10.1145/2493432.2493522
37. Cuong Pham and Patrick Olivier. 2009. Slice&Dice: Recognizing food preparation activities using embedded accelerometers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 34–43. https://doi.org/10.1007/978-3-642-05408-2_4
38. Philips. Philips Hue. Retrieved from http://www2.meethue.com/en-gb/
39. Anne Preston, Madeline Balaam, Paul Seedhouse, Salla Kurhila, Lari Kotilainen, Ashur Rafiev, Daniel Jackson, and Patrick Olivier. 2015. Can a kitchen teach languages? Linking theory and practice in the design of context-aware language learning environments. *Smart Learning Environments* 2, 1: 9. https://doi.org/10.1186/s40561-015-0016-9
40. Guy Schofield, Tom Bartindale, and Peter Wright. 2015. Bootlegger. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 767–776. https://doi.org/10.1145/2702123.2702229
41. Paul Seedhouse, Anne Preston, Patrick Olivier, Dan Jackson, Philip Heslop, Thomas Plötz, Madeline Balaam, and Saandia Ali. 2013. The French Digital Kitchen: implementing task-based language teaching beyond the classroom. *International Journal of Computer-Assisted Language Learning and Teaching* 3, 1: 50–72. https://doi.org/10.4018/ijcallt.2013010104
42. Chris Shallue. Research Blog: Show and Tell: image captioning open sourced in TensorFlow. Retrieved from https://research.googleblog.com/2016/09/show-and-tell-image-captioning-open.html
43. Matthew Shotton, Frank Melchior, Michael Evans, Tony Churnside, Matthew Brooks, and Mike Armstrong. Object-based broadcasting - curation,

responsiveness and user experience - BBC R&D. Retrieved March 20, 2016 from http://www.bbc.co.uk/rd/publications/whitepaper285

44. Smarter. Smarter - Kitchen Appliances. Retrieved from http://smarter.am/#
45. TensorFlow. TensorFlow — an Open Source Software Library for Machine Intelligence. Retrieved from https://www.tensorflow.org/
46. TIME. Fewer People Than Ever Are Watching TV. Retrieved from http://time.com/3615387/tv-viewership-declining-nielsen/
47. Manfred Tscheligi, Boris de Ruyter, Panos Markopoulus, Reiner Wichert, Thomas Mirlacher, Alexander Meschterjakov, and Wolfgang Reitberger (eds.). 2009. *Ambient Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-05408-2
48. Marian F. Ursu, Ian C. Kegel, Doug Williams, Maureen Thomas, Harald Mayer, Vilmos Zsombori, Mika L. Tuomola, Henrik Larsson, and John Wyver. 2008. ShapeShifting TV: interactive screen media narratives. *Multimedia Systems* 14, 2: 115–132. https://doi.org/10.1007/s00530-008-0119-z
49. Marian F. Ursu, Julia Sussner, Ulf Myrestam, Nina Hall, Maureen Thomas, Ian Kegel, Doug Williams, Mika Tuomola, Inger Lindstedt, Terence Wright, Andra Leurdijk, and Vilmos Zsombori. 2008. Interactive TV narratives. *ACM Transactions on Multimedia Computing, Communications, and Applications* 4, 4: 1–39. https://doi.org/10.1145/1412196.1412198
50. Marian F. Ursu, Vilmos Zsombori, John Wyver, Lucie Conrad, Ian Kegel, and Doug Williams. 2009. Interactive documentaries. *Computers in Entertainment* 7, 3: 1. https://doi.org/10.1145/1594943.1594953
51. Gerard Wilkinson. Media of Things. Retrieved from https://github.com/digitalinteraction/MediaOfThings
52. Yahoo Finance. Cisco's Set-Top Box Business Continues to Decline. Retrieved from http://finance.yahoo.com/news/cisco-set-top-box-business-170635200.html
53. Vilmos Zsombori, Michael Frantzis, Rodrigo Laiola Guimaraes, Marian Florin Ursu, Pablo Cesar, Ian Kegel, Roland Craigie, and Dick C.A. Bulterman. 2011. Automatic generation of video narratives from shared UGC. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia - HT '11*, 325. https://doi.org/10.1145/1995966.1996009
54. EBU Technology & Innovation - Growing network of metadata developers makes MDN 2017 the biggest yet. Retrieved August 16, 2017 from https://tech.ebu.ch/news/2017/07/growing-network-of-metadata-developers-makes-mdn-2017-the-biggest-yet
55. 2-IMMERSE. Retrieved August 16, 2017 from https://2immerse.eu/
56. Opta Premier League. Retrieved September 5, 2017 from http://www.optasports.com/events/premier-league.aspx
57. Hawk-Eye Innovations Official Website. Retrieved September 5, 2017 from https://www.hawkeyeinnovations.com/products
58. LockitScript - the script supervising app on the App Store. Retrieved September 5, 2017 from https://itunes.apple.com/gb/app/lockitscript-the-script-supervising-app/id842485792?mt=8
59. Script Evolution on the App Store. Retrieved September 5, 2017 from https://itunes.apple.com/gb/app/script-evolution/id824223248?mt=8
60. ORPHEUS - BBC R&D. Retrieved April 20, 2017 from http://www.bbc.co.uk/rd/projects/orpheus
61. LG G Watch W100 Mobile Phone Accessories - LG Electronics UK. Retrieved June 26, 2017 from http://www.lg.com/uk/smart-watches/lg-W100
62. Estimote Beacons — real world context for your apps. Retrieved June 26, 2017 from https://estimote.com/
63. UHF Mid Range Reader - ID ISC.MRU102 - Identification - Products - FEIG ELECTRONIC. Retrieved June 26, 2017 from http://www.feig.de/en/products/identification/product/id-iscmru102/
64. Liquid Flow Meter. Retrieved July 25, 2017 from https://www.adafruit.com/product/833
65. Smappee Energy Montior. Retrieved June 26, 2017 from http://www.smappee.com/uk/energy-monitor
66. DMC-GH4 4K Compact System Camera | Panasonic UK & Ireland. Retrieved June 26, 2017 from http://www.panasonic.com/uk/consumer/cameras-camcorders/lumix-g-compact-system-cameras/dmc-gh4rh.html