Here and Now: Reality-Based Information Retrieval

Wolfgang Büschel
Interactive Media Lab
Technische Universität Dresden
bueschel@acm.org

Annett Mitschick
Interactive Media Lab
Technische Universität Dresden
annett.mitschick@tu-dresden.de

Raimund Dachselt
Interactive Media Lab
Technische Universität Dresden
dachselt@acm.org

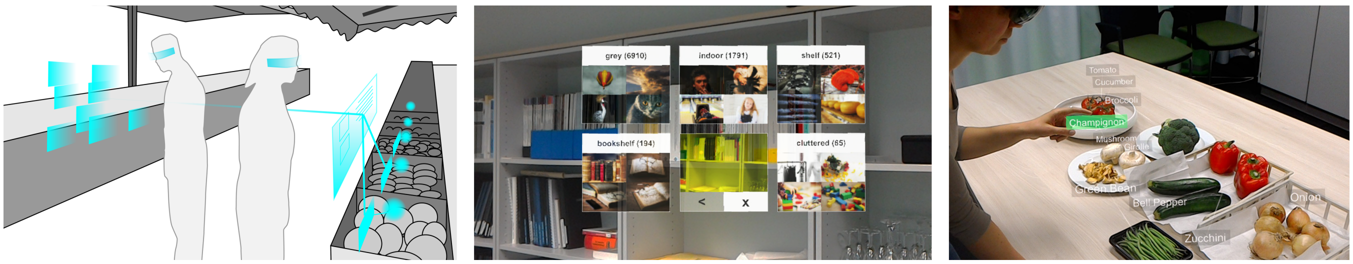

Figure 1: Left: Basic concept of RBIR illustrated by the envisioned scenario of grocery shopping. Center: Our situated photograph Image Retrieval prototype, showing a photograph taken by the user (yellow, in the middle) and extracted tags. Result pictures retrieved from pixabay.com. Right: Recipe search prototype, showing tagged groceries as input for a recipe search.

ABSTRACT

Today, the widespread use of mobile devices allows users to search information "on the go", whenever and wherever they want, no longer confining Information Retrieval to classic desktop interfaces. We believe that technical advances in Augmented Reality will allow Information Retrieval to go even further, making use of both the users' surroundings and their abilities to interact with the physical world. In this paper, we present the fundamental concept of Reality-Based Information Retrieval, which combines the classic Information Retrieval process with Augmented Reality technologies to provide context-dependent search cues and situated visualizations of the query and the results. With information needs often stemming from real-world experiences, this novel combination has the potential to better support both Just-in-time Information Retrieval and serendipity. Based on extensive literature research, we propose a conceptual framework for Reality-Based Information Retrieval. We illustrate and discuss this framework and present two prototypical implementations, which we tested in small user studies. They demonstrate the feasibility of our concepts and inspired our discussion of notable challenges for further research in this novel and promising area.

CCS CONCEPTS

• **Information systems → Search interfaces**; • **Human-centered computing → Mixed / augmented reality**;

CHIIR '18, March 11–15, 2018, New Brunswick, NJ, USA

ACM ISBN 978-1-4503-4925-3/18/03…$15.00
https://doi.org/10.1145/3176349.3176384

KEYWORDS

Reality-based Information Retrieval, Augmented Reality, Spatial User Interface, Immersive Visualization, In Situ Visual Analytics

ACM Reference Format:
Wolfgang Büschel, Annett Mitschick, and Raimund Dachselt. 2018. Here and Now: Reality-Based Information Retrieval. In *CHIIR '18: 2018 Conference on Human Information Interaction & Retrieval, March 11–15, 2018, New Brunswick, NJ, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3176349.3176384

1 INTRODUCTION & BACKGROUND

Searching for information plays a dominant role in the everyday life of people of the so-called "information age". We rely on search engines to fulfill our information needs, which might be more or less abstract or specific, highly depending on the real-world context or completely in the digital domain, with a practical impact or just out of sheer curiosity. Originally, information retrieval was exerted in the physical world, for example in a library, by asking other people, and reading public notifications. Later, it has long been (and is still) performed on desktop computers, fixed at a single location as a gateway to a virtual environment containing the digital information sources. These traditional search interfaces force the user to abstract information needs to specify the query and to "translate" the results back to match his or her real-world demands. Today, with the dissemination of powerful mobile devices and appropriate bandwidth, search has gone mobile. Statistics show that today more web searches take place on mobile devices than on computers[1].

1.1 Understanding Mobile Search & Information Needs

Since the beginning of the smartphone era, a large number of short- and long-term studies have been conducted to investigate mobile

[1] Google Blog, May 2015: https://adwords.googleblog.com/2015/05/building-for-next-moment.html

search behavior and mobile information needs. The most comprehensive one to date is a 3-month-study of daily information needs with 100 users by Church et al. [14]. It shows that *"daily information needs are highly varied, intricate, and dynamic"* and highly influenced by contextual factors. A preceding study [15] showed that most information needs occurred when participants were mobile, i. e., away from desk, traveling, on-the-go. They distinguished *geographical* needs, needs related to *Personal Information Management (PIM)*, like schedules, contacts, etc., and *informational* needs, that is, *"focused on the goal of obtaining information about a topic"*. The majority of information needs (mobile and non-mobile) were *informational*, and 64% of informational intents arose in a mobile context [15]. Still, a large extend of research on mobile search interfaces is focused on geographical or PIM-related needs, incorporating the user's context (location, time, activity, social context, etc.) to adapt the search interface or visualize results, e. g., map-based interfaces [31], and informational needs are addressed to a lesser extend. Approaches that deal with the various informational needs (e. g., how-to's, facts, explanations, advice) either focus on the advancement of input modalities and interaction techniques for complex, explicit queries, like [4, 32, 45], or they try to limit the extent of required input as much as possible. For example, they use reasoning to derive implicit information from background knowledge [37], provide automatic query-term extraction from Web content [55], or use question-answering systems [45] to extent given keywords and provide natural language answers. Of course, both strategies are driven by the limited capabilities of mobile devices for query formulation, e. g., text input by touch keyboard or speech input, or visual input by camera [22]. Other research is focused on situation-aware filtering or ranking of search results [8] taking into account the limited display size of mobile devices.

1.2 The Potential of AR-based Information Retrieval

We believe that in the future, Information Retrieval (IR) will not only be location-based but actually *return to happen in the real world*. It will be tightly interwoven with the physical world itself in what we call *Reality-based Information Retrieval* (RBIR). With the help of Augmented Reality (AR) technology, we aim to decrease the gulfs of execution and evaluation [27] for a broad range of search applications by providing Natural User Interfaces based on the situated visualization of search stimuli, queries, and results. In a few years, AR glasses are likely to become mainstream and to be a common form factor in a "post-smartphone era". Search facilities will be a vital part of such versatile devices, not only because of the constant need for information in mobile contexts but also because of the novel opportunities to satisfy it much better. One of these opportunities is the concept of Just-in-time Information Retrieval (JITIR) [48], proactively retrieving information "based on a person's local context in an easily accessible yet non-intrusive manner", which is very similar to the idea of "finding what you need with zero query terms (or less)" envisioned in [2]. JITIR heavily relies on modeling the user's context (e. g., location, time, application usage, individual preferences and interests, and nearby objects) and situationally matching it to the environment. So far, it has seen limited use in mobile AR settings (e. g., [1]).

In addition to that, a focus of current research in Human-computer Information Retrieval concentrates on the understanding of *serendipity, "an unexpected experience prompted by an individual's valuable interaction with ideas, information, objects, or phenomena"* [43], leading to unforeseeable, but much more valuable insights. According to [41], some strategies can stimulate serendipity, like *"going out and about"* to experience new things one might not have come across otherwise, or *"keeping [...] eyes and ears open to things happening"* in order to recognize or to be receptive for connections. More than any other digital environment, AR is able to support serendipity in information access by fusing virtual and physical information artifacts, suggesting contextual information [1], and assisting the user to *"follow up on potentially valuable opportunities"* [41]. Furthermore, studies investigating mobile search activities showed that *"interactions with the material world tend to create more information needs and information seeking behaviors than virtual interactions"* [12]. Therefore, we believe that examining how to combine IR and AR to better integrate Information Retrieval into the physical world is a promising field of research.

In our endeavor to pave the way for a new generation of immersive, in-the-wild IR systems, we seek to close the gap between both research fields, Augmented Reality and Information Retrieval. Thus, our main contributions in this perspective paper are as follows:

(1) We present the general and novel concept of *Reality-based Information Retrieval* (RBIR),
(2) and suggest a conceptual framework for the design of future RBIR systems by integrating Natural Interaction and Situated Augmentations into the classical information retrieval model.
(3) We report on two implemented prototypes which we tested in small-scale user studies that demonstrate the feasibility of our ideas, and
(4) derive and discuss future challenges of RBIR.

2 ENVISIONED SCENARIOS AND BASIC CONCEPT

We illustrate the basic concept of Reality-based Information Retrieval with the following scenario:

Alice and Bob are at the market place to buy groceries. Although they have a broad idea on what they plan to cook, they are not sure about the dessert. They use the new Reality-based Information Retrieval system with its lightweight AR glasses (see Figure 1, left). As they look at the different fruits and vegetables, small indicators light up, showing that the system has additional information ready for them. Going nearer, virtual tags appear floating above exotic fruits in addition to the physical price tags. The tags tell the couple that these are kumquat and provide information about their origin and the supply chain. Bob and Alice select the fruit's tag to search for kumquat recipes. The recipes appear scattered in front of them, showing the most relevant results closer to them. Organically, clusters of related recipes form. Bob selects a recipe with an interesting preview picture and gets nutrition and allergy information. He notices that a graph of other required ingredients is visualized, connecting the kumquats with beets and avocados at a different booth. Surprised that there are recipes with both kumquats and beet, Bob wants to find other such recipes. He adds the beets to the search query; most recipes fade out while some light up. While looking at other fruits and vegetables, Alice *remembers that they had a delicious avocado dessert last week. She wants to share the recipe with others. She connects it to the avocados, allowing future market visitors to find the dish and get inspired just like her.*

We also envision other use cases for RBIR, e. g.,

- Image retrieval: *Anne searches for pictures to decorate her living room by virtually placing relevant results from an image database on her wall. The RBIR application extracts visual features from the surrounding for content-based image retrieval to filter the result set of relevant images, e. g., regarding dominant colors.*
- Literature search: *Alex draws a book from a bookshelf of his housemate and searches for related publications of the author. The RBIR application identifies the book and its metadata. The writing on the book cover is augmented by virtual controls as overlays which Alex physically touches to start a search. Results are presented in the bookshelf in front of him.*
- Video retrieval: *Currently watching a soccer game, Phil searches for soccer videos with scenes showing similar positions of the players for comparison. The RBIR application analyses the relative positions and motion vectors of the identified players on the field and suggests similar recordings.*

The scenario and short examples illustrate our concept of *Reality-based Information Retrieval* from the user's point of view. Locations and objects in the physical world regularly trigger information needs in our daily lives. Thus, the fusion of virtual and physical information artifacts and the application of JITIR could lead to unexpected, but valuable information. Addressing this, information about the physical world, as well as associated abstract content, is digitally connected to these objects or locations and accessed by the means of AR technology. Users wear mixed reality headsets or smart glasses. They get context-specific suggestions for search terms, e. g., in form of tags virtually attached to real world objects. The basis for these tags can be manifold: Low-level (visual) features from an image analysis, higher-level meaning derived from object recognition, or even information explicitly attached by other users or supplied by smart things. Making use of these sources, the users build search queries addressing their information needs. Retrieved documents are then presented, can be browsed, and allow for relevance feedback. To seamlessly integrate the system into the users' routines, the visualizations of tags, queries, and results should all be designed to prevent visual overload or the occlusion of important real-world information. Furthermore, all interaction steps are to be supported by natural interaction techniques that allow users to benefit from, e.g., proprioception, and minimize cognitive load.

3 RELATED WORK

Although some application-specific approaches do exist, as of yet little work has been done to address issues in the intersection of the fields of IR and AR. Ajanki et al. [1] developed a prototype system which retrieves information relevant to contextual cues, i. e., people identified using face recognition techniques and objects with attached markers, to be presented with AR techniques on a handheld or head-mounted display. A similar approach was presented in [30]: face recognition techniques were used to retrieve video snippets from a personal lifelog to immediately show previous encounters with the person the user is currently looking at. General visual cues as input parameters were addressed in work from the field of Mobile Visual Search [22], showing the feasibility of content-based query term extraction from photos taken of objects in the environment [21, 59]. Other perceptive and contextual cues have been addressed, with Mobile Audio Search (involving technologies like speech recognition, audio fingerprinting, query-by-humming) and Location-based Mobile Search leading the way [58]. On the other side, ongoing research deals with the challenges of data integration and provision, e. g., using linked data and Semantic Web technologies for Augmented Reality [42, 61], to exploit the huge amount of publicly available interlinked information.

Another interesting field of related work is the realization of information retrieval in a virtual reality (VR) setting, involving 3D information exploration and browsing. Work like [13, 46, 62] informs the design of AR interaction and visualization techniques for query and result interaction even if it lacks the aspect of registration in the real world.

Placing information and labels in AR according to their connection to the real world has been subject to extensive research published in the last few years [23, 29, 39, 40, 47] and the concepts of *In Situ Visual Analytics* [20] and *Situated Analytics* [17] bring visual analytics into AR environments. The tightness of the coupling between virtual and physical world characterizes the different strategies: the spectrum ranges from a very weak coupling like in the concept of 2D information spaces in 3D mixed reality environments [19, 20], to a very tight coupling like in the concept of *embedded data representations* [57].

The design space of AR applications and the extent of research in the field of AR (cf. [50] for an overview) forms the basis of our conceptual framework for the novel paradigm of *Reality-based Information Retrieval*.

4 A CONCEPTUAL FRAMEWORK FOR REALITY-BASED IR

In the following we present and discuss our conceptual framework for *Reality-Based Information Retrieval* (RBIR). We base this framework on the general model of the Information Retrieval process as described in various forms [10, 25, 49]. Figure 2 shows our concept as an adaptation of the system-oriented IR model described in [25]. In contrast to other models, e. g., those based on cognitive IR theory [28], it is ideally suited to emphasize and delineate the two aspects *Interaction* and *Representation*. In that model, a user's *Information Need* is formulated as a machine-understandable *Query* that is matched against the internal representation of the source documents (*Indexed documents*) which form the database. The result, a usually sorted set of *Retrieved documents*, then either satisfies the information need or leads to a reformulated or new query or an abort. The processes of *Query formulation* and *Feedback*[2] form the *Interaction* side of the front-end to the IR system, the *Query* itself and the set of *Retrieved documents* form the *Representation* side. To integrate the aspects of Augmented Reality to form a conceptual model of *Reality-Based Information Retrieval*, we adopted the three

[2]Of course, user feedback in terms of Relevance Feedback can also be used to adapt matching parameters or even indexing parameters within a dynamic IR system. This is not in the scope of this paper.

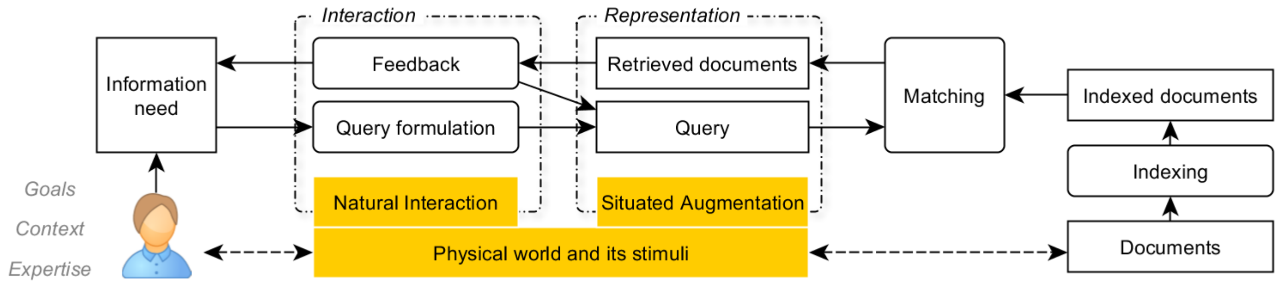

Figure 2: Conceptual model of Reality-based Information Retrieval. We extend the IR model from [25] by adding the user and three aspects of AR: the physical world, situated augmentations, and natural interaction.

elements of an AR application described by Billinghurst et al. [7]: real physical objects, virtual elements and interaction metaphor. These three components are represented in Figure 2 as *Physical world* (i. e., real physical objects), *Situated augmentation* (i. e., virtual elements), and *Natural interaction* (i. e., interaction metaphor).

The user, depicted on the left side, is an integral part of the physical world and can be described with parameters such as overall *Goals* (from which a certain information need arise), *Context* (which both may trigger or influence an information need), and *Expertise* (which affects the user's ability to identify and specify an information need). These parameters are influenced by past or present real world experiences. On the other side, the content of documents (or information objects in general) which the user wants to retrieve represent parts of the physical world or are at least associated with them. The *Indexing* process involves an analysis of these associations and representations in combination with the (predicted) information need of the user.

4.1 The Physical World

Embedding the IR process into the physical world is one of the key aspects of Reality-Based Information Retrieval. Sensory input from the real world, mainly visual and auditory stimuli, are one of the main triggers for information needs: We want to know which song is playing on the radio, are interested in the name of an actor on a movie poster, or require nutrition information for a product in the super market. We focus mainly on visual and auditory input as these are best supported by today's hardware and also have the highest bandwidth. However, in the future, other input channels (e. g., smell) could also come into play.

4.1.1 Physical World Stimuli. Both specific real-world objects and the environment in general provide the contextual cue and input for the users' queries [1]. We differentiate several classes of *physical world stimuli* (inpired by commonly used abstraction levels in Content-based Image Retrieval [16]):

(1) *Low-level* features that can be directly extracted from the input stream. Visual examples are dominant colors and textures, acoustic examples are loudness or pitch.
(2) *Mid-level features* that are usually based on classification or pattern detection processes, such as object classes or materials, based on their visual appearance like texture, shape, surface reflectivity, etc. Acoustic examples are instrumental features and distinguishing music from spoken content.
(3) *High-level concepts* that include identities derived from low- and mid-level features and background knowledge, e. g., specific people or devices, a specific piece of music, speaker recognition (who is it) and recognized spoken content (what is said).
(4) *Associated data and services* that are either provided by a (smart) object itself or are externally hosted and logically connected to an object, e. g., user-generated content connected to an object or location, social media channels about or related to the object, etc.

These classes also show a progression from physical to human-defined properties. As such, the first two and sometimes the third class are openly observable features while the latter needs instrumentation (i. e., active beacons) and/or external databases.

4.1.2 Contextual Cues. The interpretation of the scene and its objects can be dependent on the context. We differentiate between context free cues, which are usually low-level features (e. g., color), cues with a weak context (e. g., knowledge about general settings such as "office" helps to identify object classes), and those with a strong context (including, e. g., the exact location, beacons, or QR codes). A deep contextual knowledge can facilitate sensemaking and is also the gate to attached data and services, making for a more powerful system than achievable by context free data only. However, information about the context may be missing, limited, or misinterpreted. As such, Reality-based Information Retrieval systems should at least support a fall back to context-free cues if the use case allows for it.

4.1.3 Output. The physical world is not only the source for information needs and query input. By appropriation of existing displays or *embedded physicalizations* [57] (i. e., physical objects used for in-situ data representation), the physical world can also be used to visualize queries or result sets, or even give acoustic feedback. Furthermore, using the physical world as output enables collaborative search scenarios, making it easier to share and discuss information or results, or direct users to particular artifacts. We see embedded physicalizations as an optional feature for Reality-based Information Retrieval to enhance user experience, but its realization should be considered carefully regarding privacy issues.

4.2 Situated Augmentation for Reality-based IR

The registration of content in 3D forms the connection between the physical and virtual world. Aligning virtual objects with related physical objects (locally) or the physical world (globally) makes them intuitively understandable and easier to interpret. For content that is semantically related to the environment, White and Feiner introduced the term *Situated Visualization* [56]. In this case, the connection between presented information and the location is characterized by the congruence or intersection of their meaning. Here, we use the term *Situated Augmentation* to emphasize that it should not be restricted to visual representation. Within our framework for RBIR we propose three concepts for Situated Augmentation as described in the following.

(1) Situated Stimulus. We define a *Situated Stimulus* as the representation of real-world object properties (see above) as input for a search query. Ideally, but not necessarily, they would correspond to the psychological stimuli from the physical world that trigger an information need. The simplest form of visualization for situated stimuli are labels placed in the scene. It is important that the user mentally associates them to the objects or environments that they belong to. Such a connection can be visually supported by placing the labels near the corresponding object or linking them with lines (cf. for example Figure 1, right). Additional visual variables can be used to encode, e. g., the type of the underlying property. More advanced visualizations include example pictures or highlighting of the corresponding physical object.

There are two strategies to dynamically create situated stimuli in an AR scenario: (a) processing the whole scene (either 2D or 3D) in front of the user to find Regions-of-interest (ROI) and identify visible physical objects in the camera's field-of-view [24] or (b) analyzing the users gaze points to detect and extract ROI [54] and identify and label objects in the visual focus. Using 2D or 3D information to solve these tasks depends on the technical abilities of the used AR device. The second strategy has the advantage of significantly reducing the complexity of object detection by narrowing down the search area according to users interests. Nevertheless, the first strategy seems better suitable to support serendipity.

(2) Situated Query Representation. The visualization of the queries depends on their complexity and the requirements of the use case. In a very simple case where single situated or only a few equally weighted stimuli are used as search input, no further query visualization may be needed except a feedback which stimuli are used to form the query (e. g., by highlighting the label).

For more complex queries consisting of several terms which are, e. g., individually weighted and/or combined with logical operators, a 2D virtual user interface [19] showing the selected terms and their relation can be used to provide an overview of the request and also allow for editing it. In Figure 3 we illustrate a conceivable example for a 2D query representation based on the *Filter/Flow* metaphor by Young and Shneiderman [60] to visualize boolean queries. Other visualization techniques developed for classical 2D representation would be suitable as well. Spatial visualization of a query in 3D provides the opportunity to give weight to the connection between query components and the physical world (e. g., as a network [6]), but regarding data overload and visual cluttering it should be considered carefully.

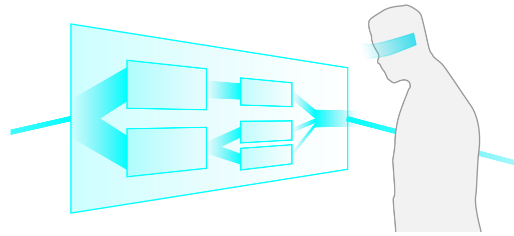

Figure 3: Example illustration of an AR application using the Filter/Flow model from [60] for boolean query formulation.

(3) Situated Result Representation. Of course, the representation of search results plays a very important role in a RBIR application. AR provides a number of possibilities determined by

(1) the intrinsic order/structure of the result set(s) (e. g., ordered by relevance, hierarchy, certain properties like timestamp, etc.)
(2) the reference to the real-world (e. g., placing visualizations relative to the user, i. e., body-referenced, or relative to the environment by augmenting vertical or horizontal surfaces) [50], and
(3) the mapping of the result space (e. g., using metaphors like a bookshelf, 3D shapes in free-space [62], or local coordinate systems like the space above a table [11]).

Although the results may also be presented on a 2D virtual user interface with its "advantages in efficiency, speed, precision and reduction of clutter" [19], the opportunity to map parameters to the physical world, e. g., using distance to the user to represent relevance of individual results , and the ability to move within or relative to it has proven benefits (e. g., [5, 11]).

4.3 Natural Interaction for Reality-based IR

Various interaction techniques have been proposed for AR [7], ranging from traditional input devices, such as keyboard, mouse, and touch screen to advanced 3D interaction methods and natural interaction techniques. Regarding our envisioned application scenarios, the use of traditional input devices such as keyboard and mouse is not suitable. However, an additional touch-enabled device such as a smartphone could simply provide text input or serve as a handy device for pointing, selecting or data transfer [38]. Furthermore, several ideas of purpose-built devices for interaction in AR have been proposed [34, 53]. Other interaction modalities applicable in an AR environment are voice input, free-hand gestures, tangible input, body motion, and gaze, in summary also referred to as *natural interaction techniques*. Gaze is by far the most unobtrusive and discreet input modality when we think of use cases in public space. Based on the elementary interaction tasks in the IR process (see Figure 2), we distinguish four Natural Interaction tasks within RBIR as described in the following.

(1) Natural Query Specification. A query can be submitted to an IR system in the form of

(a) text, like keywords, tags, natural language, artificial query language, commands, etc.
(b) key-value pairs, e. g., for property-based or faceted search,

(c) one or more examples, in case of Query-by-Example / similarity-based search, or
(d) via associations, e. g., in browsing scenarios.

Of course, the most natural and efficient input modality for text is voice input [26], but it has its drawbacks regarding privacy issues, social acceptance, and in noisy environments. This is to a great extend also the case for free-hand gestures [35, 36] and body motion [51] as interaction modalities. We believe that gaze input in combination with additional hardware like a smartphone or smartwatch for confirmation [52] and manipulation is most applicable for the above mentioned query modalities (b)-(d), so long as query formulation can be down-scaled to a selection task. This in turn means that the user should be able to select from a range of items, i. e., a set of possible values for properties or facets (b), a set of examples for similarity-based search (c) and a set of linked items for exploration/browsing (d). These items can be virtual, but also and more particularly physical (physical world stimuli), e. g., picking a color from a surface, capturing a part of the scene as visual example (a query-by-photograph metaphor [3]) or browsing through a taxonomy of recognized higher concepts. Certainly, this only works if the system is able to provide items or the physical world contains according stimuli.

(2) Natural Result Exploration and Interaction. Depending on the concrete reference and mapping of the result representation (see above), we envision *spatial interaction techniques* as proposed in [33] or gaze-supported multimodal interaction like [52] for natural result exploration. Spatial interaction involves a much more intense immersion. Locally or globally registered information spaces can be explored by physical movement of the user, e. g., determining the level of detail by the distance and the type of information by the angle to the object [33]. We also imagine unobtrustive wearable input devices like smartwatches, ring devices [18] or other novel devices [34] for casual interaction with even large result sets using zooming, panning, sorting, and filtering techniques.

(3) Natural User Feedback. Interacting with a result set is of course an implicit form of giving relevance feedback. Thus, any above mentioned modality could serve as input for user feedback, especially gaze [63] and spatial interaction [33], e. g., coming closer to an item gives it more weight, literally turning one's back on an item excludes it from the relevant results. Additionally, we imagine to rearrange results as a form of relevance feedback using motion and gestures or additional devices. We propose the idea of *metaphorical relevance feedback* using or establishing relations between results and the physical world, e. g., by explicitly placing results onto or near physical-world objects that symbolize a certain usage like a trash basket (removal), a notice board (keeping), or a personal device (taking along).

(4) Natural Annotation. Beyond the rather implicit user feedback described above, we also envision a natural way for the user to annotate physical stimuli, situated stimuli as well as retrieved results, resp. the connection between stimuli and results. Of course, the user benefits from his/her personal annotations. They support sensemaking but also serendipity in the sense of "remembering and drawing on previous experiences" as one of the key strategies [41]. Furthermore, like in the envisioned scenario in Section 2 we also imagine social interaction or loose collaboration. User annotations left in AR may inspire other users or help them discover new aspects or connections.

Creating and leaving annotations in AR necessitates interaction techniques that allow the user to register them to the environment, in either a specific or abstract way. Precisely placing virtual artifacts in the real world is typically a very cumbersome task. In order to achieve a rather casual interaction, placing annotations and thus connecting them to physical stimuli should be supported by semantic and contextual cues, e. g., annotations could "snap" to their location according to their meaning and the user's context.

5 EXPERIMENTS WITH REALITY-BASED IR

To show the general feasibility of our proposed concept for Reality-based Information Retrieval, we implemented two prototypes. Both implementations described below realize different information retrieval concepts in AR for the Microsoft HoloLens device[3], using its build-in functions for image processing, augmentation and interaction. We also tested these prototypes in short, preliminary studies to collect feedback for further design iterations.

5.1 Case Study I: Situated Photograph Image Retrieval

Our first prototype realizes the basic concept of situated photograph queries in AR for image retrieval. It incorporates the usage of a photograph metaphor based on [3] to extract visual information from the environment as query parameters. The implementation is quite simple: performing an air-tap gesture the user takes a photograph of the center of the visible scene, which is displayed as a query object in the middle of a 2D canvas located in free space in front of the user (see Figure 1, center). This implements the three main aspects of RBIR: *Natural interaction techniques* to formulate a query based on some *physical, real-world stimuli* and *situated augmentations* enriching the user's environment. The picture is sent to the Microsoft Computer Vision REST API[4] to retrieve automatically assigned tags. The tags are then used to retrieve preview images from the Pixabay API[5]. The retrieved images are displayed as preview images around the initial query image and labeled with the corresponding tags. The actual query, i. e., searching for images with specific tags, is executed by tapping on a tag, thus making use of the same techniques to interact with digital and physical objects. A second 2D canvas serves as result visualization, which can be freely placed on available surfaces by the user using gaze direction and air tap. This result canvas shows the sorted set of images retrieved from the Pixabay API using one or more selected tags (see Figure 4). The user can also take multiple photos and thus create multiples query canvases to combine tags corresponding to different objects. Selecting one of the preview images replaces the original photo with the chosen image which is then used as query source, allowing the user to iteratively browse through pictures.

[3] Microsoft HoloLens: https://www.microsoft.com/hololens
[4] Microsoft Cognitive Services: https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/home
[5] RESTful interface for searching and retrieving free high-quality images from Pixabay, see https://pixabay.com/api/docs/.

Figure 4: Aggregation of the tags "red" (left) and "keyboard" (right) from different photographs to build a combined query (middle). Result pictures retrieved from pixabay.com.

Evaluation: We conducted a qualitative user study to evaluate the usability of situated photograph queries as a basic interaction technique for RBIR. Seven male students from the local university, aged between 22 and 26 years, took part. A series of tasks was chosen to provoke interaction between the user and the application. Every task consisted of one or two keywords that the user should search for. The task was considered complete when the keyword(s) were active on the result canvas. Simple tasks were set at the beginning to facilitate getting started. In order to follow and later reproduce the participants' actions during usability testing, all discrete user activities (e. g., taking pictures, tapping a preview image for browsing) were logged. Users were asked to "think aloud" and give statements about the application whenever they encountered a special situation or problem. These statements were recorded using a camcorder. In a follow-up questionnaire we assessed the application using the System Usability Scale (SUS) [9].

During the study we could observe two opposing ways of interacting with the prototype, which was confirmed by an analysis of the logs: (1) using the photograph technique only as an initial query and browsing through tags by selecting preview images (*browsing*) and (2) using the photograph technique repeatedly to refine a query (*non-browsing*). While non-browsing participants performed fewer interactions per minute (average of 4.8 versus 5.4), the average time spent using the application was also lower (26 min.) compared to browsing participants (38 min.). In total, non-browsing participants performed fewer interactions (average of 124 versus 209) to solve the tasks. This possibly results from the fact that browsing often incorporates rather quick jumping from one picture to the next, which involves comparatively many tapping interactions in a short period of time. The browsing strategy was more time consuming and involved more interactions compared to a non-browsing behavior. Additionally, the analysis of the questionnaire also indicates that browsing participants needed more effort (2.3 versus 1.7), especially for "getting the right keywords", and felt less pleasure (3.4 versus 4.3) when using the application. The perception of usefulness varies regarding the provided features, but the average differences are small (3.9 versus 3.6). Although differences in the SUS score between the groups of participants are generally not big (72.5 for browsing versus 79.4 for non-browsing), it is worth noting that browsing participants rated the application lower in 7 of the 10 statements compared to non-browsing participants.

Although general implications should not be made, we observed that browsing participants were less aware of the room and used it less during their interaction with the application. Participants who recognized their environment and its possibilities were not only more effective but also more confident. This insight encouraged our work on a succeeding experiment (described in the following) where we decided to provide visual guidance to the user in order to support the awareness of the environment and physical stimuli.

5.2 Case Study II: Recipe Search

For our second prototype we were inspired by the food search scenario presented earlier. The concept behind this prototype is that food, e. g., fruits and vegetables, is augmented with labels. The user then selects such tags to form a search query. As results, a set of recipes containing the selected ingredients are presented. In contrast to the first prototype, we opted against the user explicitly taking photographs. Instead, the concept is to have the current camera image automatically analyzed in the background, showing tags as they are generated and thus better supporting serendipity. In reference to our framework, the tags are examples for *mid-level features* that represent object classes, either specifically (e. g., "apple") or more generally (e. g., "meat"). For improved reliability and stable results necessary for the evaluation, we decided to pre-generate the tags for the prototype instead of using an actual computer vision API. The recipe search is implemented using the Food2Fork web API[6], which returns a list of recipes with ingredients, preview images, and other metadata.

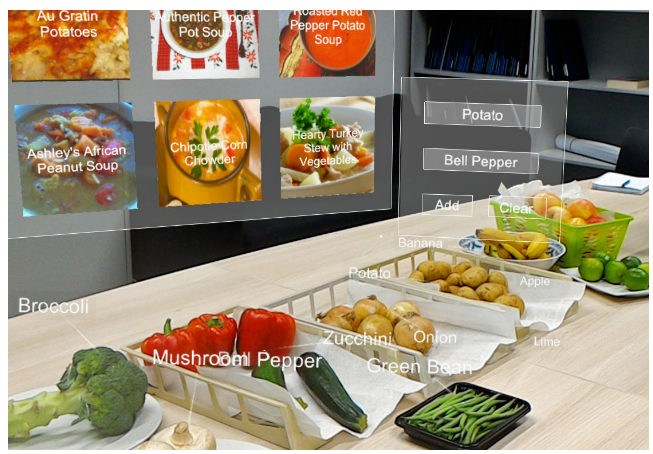

Figure 5: First-person mixed reality capture of the Recipe Search prototype showing the tags in the foreground, the result visualization (left) and the query (right) in the background. Result data retrieved from Food2Fork.com.

The visualization is graph oriented, with tags being connected to the corresponding real-world location and also to the query that they are a part of. This helps to show the spatial relations between keywords and results. Extending this, it would be feasible to also visually link result ingredients to the search query. The query as well as the results are shown on spatially positioned 2D canvases that automatically turn towards the user. Similar to the first prototype, interaction is based on gaze pointing and the air tap gesture of the HoloLens. Users can select tags to add them to

[6] Food2Fork: https://food2fork.com/about/api

the query or, when tag and canvas are next to each other, simply drag them to the canvas.

Evaluation: We tested our second prototype with four usability experts (three male, one female, aged 26 to 34). To this end we prepared a large table in our lab with 13 different fresh vegetables and fruits, their arrangement resembling a real market place. All goods where digitally annotated with their type. After signing informed consent, the participants received a short introduction to the use case and the prototype. They then had the chance to freely explore the system for approximately 10 minutes, looking up recipes based on selections of the augmented ingredients presented on the table. As in the first study, the participants were encouraged to voice their opinion and talk about positive and negative aspects as they encountered them (*think-aloud*). Afterwards, they filled out a questionnaire consisting of nine 5-point Likert items on usability and five items on serendipity, based on the *Serendipitous Digital Environment scale* (SDE) [44].

The experts reported a high usability: For instance, learnability was rated with an average score of 4.5 out of 5, and confidence in using the application was rated 4 out of 5. Our participants reported a medium level of serendipity, e. g., partly agreeing that it invites examination of its content (3.5 out of 5). We believe that the study setting and its limited scope negatively affected serendipity. In their comments, the experts provided valuable feedback on several potential problems of the prototype for real world use. This feedback, in combination with the results of the first study, informs the challenges that we present in the following section.

6 CHALLENGES

In this section we identify open research challenges, informed both by our review of related literature and our own experiences with our prototypes. We believe that addressing these challenges will be the key to future successful RBIR.

A main challenge is to clarify if existing retrieval models are suitable for RBIR. While we believe that current models can be applied, incorporating extended context models and considering the imprecise nature of the user input will be beneficial. Subsequently, new evaluation methodologies should be examined to suitably measure effectiveness while considering the serendipitous nature of the interface. Approaches like the *Serendipitous Digital Environment scale* [44] might be promising starting points. Additionally, there are some technical challenges, which are not specific to RBIR but still need to be addressed. They include the weight and battery life of the devices and the availability of a reliable, low-latency communication infrastructure. Also, object detection and registration of virtual content in the scene is an ongoing research challenge.

In the following we would like to highlight some challenges we think are crucial for the success of RBIR:

Reliable Stimulus Detection & Identification: Currently, there is no system that accurately classifies arbitrary objects and their properties in real time, let alone on a mobile device. We believe that local pre-computation, e. g., finding regions of interest or determining the scene context, combined with cloud-based object classification could be promising for the use case of RBIR. However, stimulus detection is not only a question of computer vision. It is just as important to assess the user's attention to objects or environmental features as a form of relevance feedback. In this context, having a reliable user model allows to estimate what a user might be interested in and how to correctly interpret stimuli. On the other hand, we believe that unexpected (i. e., wrong) results could actually improve serendipity by presenting new and surprising opportunities, as long as the system as a whole still feels reliable.

Visibility & Subtlety: Designing suitable and effective visualizations for the queries and results in 3D is a major challenge. In particular, there are two conflicting goals regarding the choice of visualization. On one hand, the aim is to design augmentations that stand out well enough to be recognized even in cluttered environments. A specific goal has to be supporting the mental model of the user, in this case the association between real objects and virtual content. Here, in contrast to many other AR applications, not only spatial cues but also, e. g., a similar color could be used to link stimuli to objects. On the other hand, augmentations need to be seamlessly integrated as to not obfuscate or occlude important features of the physical world. This is especially important in RBIR, assuming a system that runs in the background for extended periods of time. There has been a lot of research on, e. g., label placement in AR, however, achieving not only optimal visibility but also a visually calm presentation that does not lead to visual overload is still challenging.

Limited Interaction Capabilities: Working with our prototypes, not all users were able to reliably use the air tap gesture. Also, the concept of a gaze directed pointer, similar to a traditional mouse pointer, was not always clear. While these and similar problems are general challenges of spatial input for AR, a specific question for RBIR is how to achieve the expressiveness and flexibility of traditional, desktop based searches. For instance, we do not believe that a system could solely rely on contextual information taken directly from the environment. Instead, the explicit input of search strings will also sometimes be necessary. Speech-to-text is one approach for text input with AR headsets, however especially with IR systems, questions of privacy remain (see below). Another challenge are complex, facet-based searches that usually require complex menus. For such information needs, a successful RBIR system should not be significantly harder to use than regular systems.

Perspicuity & User Control: Typically, IR systems are a black box for the user. In the future, when they affect decisions in our daily lives even more directly, an understanding of the factors leading to the results is even more important. Thus, it is a challenge to design Reality-based IR in such a way that the users feel in control. Support for easy relevance feedback is one particular way to keep the *human in the loop*. Building on that, it is important to clearly visualize the system state and how the user's input (e. g., their relevance feedback) influences the IR process.

Accessibility & Inclusion: Often, it is assumed that every user can interact with a designed system, when in reality accessibility is lacking. For example, in AR there is a very strong reliance on visuals with little regard for alternative forms of output, excluding visually impaired users. Similarly, RBIR assumes physical navigation and spatial input that may not be feasible for users with motor impairments. Being locked out from using future general purpose RBIR systems could become a serious problem. However, exclusion can also happen on other levels, through the language or the metaphors that the system uses or by the type and form of presentation of the results. Thus, it is an important research challenge to come up with a system design that is inclusive for all user groups, independent of factors such as age, cultural background, or abilities.

Privacy: A person's search history allows to derive detailed knowledge of their interests and is as such highly sensitive information. Privacy in RBIR is an even bigger concern because of the combination with real-time location tracking, gaze information, and other metadata. However, another challenge is to support privacy not only against search providers but also bystanders and shoulder-surfing attacks. Very explicit forms of interaction, e. g., voice input or gestures, may leak information about the search interests of a user. AR headsets, while single user, may also allow others to perceive parts of their content and, in comparison to smartphones, can not be easily shielded by the user.

Filter Bubbles & Discriminating Algorithms: The risk of any user-model is to only show things that the user knows, expects, or wants. This stands in direct opposition to our goal of serendipity. While problematic even in today's systems, it could become even worse in AR. There, these Filter Bubbles would extend to the real world with differing opinions being hidden from the user. Furthermore, recent developments show that many IT companies do not shy away from using their influence on their users to shape public opinion. Thus, even explicit discrimination or the suppression of information in the physical world by search engine providers would become a scary possibility in the future. Similarly, there have been examples of algorithms discriminating against groups of people, be it by malice or ignorance of the developers. How to support the creation of open and transparent platforms that are resilient to these effects is another research challenge for RBIR.

User-generated Annotations. User-generated annotations in AR may refer to a physical stimulus in its specific representation (e. g., the physical book in the shelf) or the abstract concept of it (e. g., avocados in general). Automatically detecting the reference of an annotation is a particular challenge which might profit from other users' annotations, machine learning techniques, and extensive domain knowledge. Furthermore, the physical world is subject to changes, artifacts move or exist only temporally. And finally, user-generated content shared in a virtual community also involves the questions of how to create, filter and manage them, including the danger of misuse and malpractice, a complex challenge requiring synergies from multiple disciplines.

7 CONCLUSION

In this perspective paper, we introduced the concept of Reality-based Information Retrieval (RBIR). Since information needs are often stimulated by the user's surroundings, we combined the Information Retrieval process with Augmented Reality by extending the IR model with the notions of the physical world, Natural Interaction, and Situated Augmentations. With intuitive user interfaces for spatial, in-situ visualizations of search stimuli, queries, and results, RBIR has the potential to support serendipitous Just-in-time Information Retrieval. We presented a conceptual framework for the design of RBIR systems and reported on two implemented prototypes, which we tested in small-scale user studies. These studies show the feasibility of our ideas and allowed us to derive challenges specific to RBIR. We aim to address these challenges and hope that our work can spark a discussion about future IR systems interwoven with the physical world.

8 ACKNOWLEDGMENTS

We would like to thank Philip Manja for his contributions, including his work on the first prototype. This work was funded in part by grant no. 03ZZ0514C of the German Federal Ministry of Education and Research (measure Twenty20 – Partnership for Innovation, project Fast).

REFERENCES

[1] Antti Ajanki, Mark Billinghurst, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, Teemu Ruokolainen, et al. 2010. Contextual information access with augmented reality. In *Proc. 2010 IEEE Int. Workshop MLSP*. IEEE, 95–100.

[2] James Allan, Bruce Croft, Alistair Moffat, and Mark Sanderson. 2012. Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012 the Second Strategic Workshop on Information Retrieval in Lorne. *SIGIR Forum* 46, 1 (May 2012), 2–32. https://doi.org/10.1145/2215676.2215678

[3] Jürgen Assfalg, Alberto Del Bimbo, and Pietro Pala. 2002. Three-Dimensional Interfaces for Querying by Example in Content-Based Image Retrieval. *IEEE Trans. Vis. Comput. Graphics* 8, 4 (Oct. 2002), 305–318. https://doi.org/10.1109/TVCG.2002.1044517

[4] Juhee Bae, Vidya Setlur, and Benjamin Watson. 2015. GraphTiles: A Visual Interface Supporting Browsing and Imprecise Mobile Search. In *Proc. MobileHCI '15*. ACM, New York, NY, USA, 63–70. https://doi.org/10.1145/2785830.2785872

[5] Robert Ball, Chris North, and Doug A. Bowman. 2007. Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays. In *Proc. CHI '07*. ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/1240624.1240656

[6] Daniel Belcher, Mark Billinghurst, SE Hayes, and Randy Stiles. 2003. Using Augmented Reality for Visualizing Complex Graphs in Three Dimensions. In *Proc. ISMAR '03*. IEEE CS, Washington, DC, USA, 84–93. https://doi.org/10.1109/ISMAR.2003.1240691

[7] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Found. Trends Hum.-Comput. Interact.* 8, 2-3 (March 2015), 73–272. https://doi.org/10.1561/1100000049

[8] Ourdia Bouidghaghen, Lynda Tamine-Lechani, and Mohand Boughanem. 2009. Dynamically Personalizing Search Results for Mobile Users. In *Proc. FQAS '09*. Springer-Verlag, Berlin, Heidelberg, 99–110. https://doi.org/10.1007/978-3-642-04957-6_9

[9] John Brooke. 1996. SUS: a "quick and dirty" usability scale. In *Usability Evaluation in Industry*. Taylor and Francis.

[10] Michael K. Buckland and Christian Plaunt. 2000. On the Construction of Selection Systems. *Library Hi Tech* 12 (06 2000).

[11] Wolfgang Büschel, Patrick Reipschläger, Ricardo Langner, and Raimund Dachselt. 2017. Investigating the Use of Spatial Interaction for 3D Data Visualization on Mobile Devices. In *Proc. ISS '17*. ACM, New York, NY, USA, 62–71. https://doi.org/10.1145/3132272.3134125

[12] Juan Pablo Carrascal and Karen Church. 2015. An In-Situ Study of Mobile App & Mobile Search Interactions. In *Proc. CHI '15*. ACM, New York, NY, USA, 2739–2748. https://doi.org/10.1145/2702123.2702486

[13] Patrick Chiu, Andreas Girgensohn, Surapong Lertsithichai, Wolf Polak, and Frank Shipman. 2005. MediaMetro: Browsing Multimedia Document Collections with a 3D City Metaphor. In *Proc. MULTIMEDIA '05*. ACM, New York, NY, USA, 213–214. https://doi.org/10.1145/1101149.1101182

[14] Karen Church, Mauro Cherubini, and Nuria Oliver. 2014. A Large-scale Study of Daily Information Needs Captured in Situ. *ACM Trans. Comput.-Hum. Interact.* 21, 2, Article 10 (Feb. 2014), 46 pages. https://doi.org/10.1145/2552193

[15] Karen Church and Barry Smyth. 2009. Understanding the Intent Behind Mobile Information Needs. In *Proc. IUI '09*. ACM, New York, NY, USA, 247–256. https://doi.org/10.1145/1502650.1502686

[16] John Eakins, Margaret Graham, and Tom Franklin. 1999. Content-based Image Retrieval. In *Library and Information Briefings*.

[17] Neven ElSayed, Bruce Thomas, Kim Marriott, Julia Piantadosi, and Ross Smith. 2015. Situated Analytics. In *2015 Big Data Visual Analytics (BDVA)*. 1–8. https://doi.org/10.1109/BDVA.2015.7314302

[18] Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In *Proc. SUI '16*. ACM, New York, NY, USA, 99–102. https://doi.org/10.1145/2983310.2985757

[19] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. 2014. Ethereal Planes: A Design Framework for 2D Information Space in 3D Mixed Reality Environments. In *Proc. SUI '14*. ACM, New York, NY, USA, 2–12. https://doi.org/10.1145/2659766.2659769

[20] Barrett Ens and Pourang Irani. 2017. Spatial Analytic Interfaces: Spatial User Interfaces for In Situ Visual Analytics. *IEEE Comp. Graph. Appl.* 37, 2 (2017), 66–79. https://doi.org/10.1109/MCG.2016.38

[21] Xin Fan, Xing Xie, Zhiwei Li, Mingjing Li, and Wei-Ying Ma. 2005. Photo-to-search: Using Multimodal Queries to Search the Web from Mobile Devices. In *Proc. MIR '05*. ACM, New York, NY, USA, 143–150. https://doi.org/10.1145/1101826.1101851

[22] Bernd Girod, Vijay Chandrasekhar, David M Chen, Ngai-Man Cheung, Radek Grzeszczuk, Yuriy Reznik, Gabriel Takacs, Sam S Tsai, and Ramakrishna Vedantham. 2011. Mobile visual search. *IEEE Signal Process. Mag.* 28, 4 (2011), 61–76.

[23] Raphael Grasset, Tobias Langlotz, Denis Kalkofen, Markus Tatzgern, and Dieter Schmalstieg. 2012. Image-driven View Management for Augmented Reality Browsers. In *Proc. ISMAR '12*. IEEE Computer Society, Washington, DC, USA, 177–186. https://doi.org/10.1109/ISMAR.2012.6402555

[24] Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, and Jianwei Wan. 2014. 3D Object Recognition in Cluttered Scenes with Local Surface Features: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 11 (Nov 2014), 2270–2287. https://doi.org/10.1109/TPAMI.2014.2316828

[25] Djoerd Hiemstra. 2009. Information Retrieval Models. In *Information Retrieval*. John Wiley & Sons, Ltd, 1–19. https://doi.org/10.1002/9780470033647.ch1

[26] Ronghang Hu, Huazhe Xu, Marcus Rohrbach, Jiashi Feng, Kate Saenko, and Trevor Darrell. 2016. Natural language object retrieval. In *Proc. CVPR '16*. 4555–4564. https://doi.org/10.1109/CVPR.2016.493

[27] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct Manipulation Interfaces. *Hum.-Comput. Interact.* 1, 4 (Dec. 1985), 311–338. https://doi.org/10.1207/s15327051hci0104_2

[28] Peter Ingwersen and Kalervo Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA. https://doi.org/10.1007/1-4020-3851-8

[29] Daisuke Iwai, Tatsunori Yabiki, and Kosuke Sato. 2013. View management of projected labels on nonplanar and textured surfaces. *IEEE Trans. Vis. Comput. Graphics* 19, 8 (2013), 1415–1424.

[30] Masakazu Iwamura, Kai Kunze, Yuya Kato, Yuzuko Utsumi, and Koichi Kise. 2014. Haven'T We Met Before?: A Realistic Memory Assistance System to Remind You of the Person in Front of You. In *Proc. AH '14*. ACM, New York, NY, USA, Article 32, 4 pages. https://doi.org/10.1145/2582051.2582083

[31] Matt Jones, George Buchanan, Richard Harper, and Pierre-Louis Xech. 2007. Questions Not Answers: A Novel Mobile Search Technique. In *Proc. CHI '07*. ACM, New York, NY, USA, 155–158. https://doi.org/10.1145/1240624.1240648

[32] Amy K. Karlson, George G. Robertson, Daniel C. Robbins, Mary P. Czerwinski, and Greg R. Smith. 2006. FaThumb: A Facet-based Interface for Mobile Search. In *Proc. CHI '06*. ACM, New York, NY, USA, 711–720. https://doi.org/10.1145/1124772.1124878

[33] Jens Keil, Michael Zoellner, Timo Engelke, Folker Wientapper, and Michael Schmitt. 2013. *Controlling and Filtering Information Density with Spatial Interaction Techniques via Handheld Augmented Reality*. Springer, Berlin, Heidelberg, 49–57. https://doi.org/10.1007/978-3-642-39405-8_6

[34] Konstantin Klamka and Raimund Dachselt. 2015. Elasticcon: Elastic Controllers for Casual Interaction. In *Proc. MobileHCI '15*. ACM, NY, USA, 410–419. https://doi.org/10.1145/2785830.2785849

[35] Jarrod Knibbe, Sue Ann Seah, and Mike Fraser. 2014. VideoHandles: Replicating Gestures to Search Through Action-camera Video. In *Proc. SUI '14*. ACM, New York, NY, USA, 50–53. https://doi.org/10.1145/2659766.2659784

[36] ByoungChul Ko and Hyeran Byun. 2002. Query-by-Gesture: An Alternative Content-Based Image Retrieval Query Scheme. *J. Vis. Lang. Comput.* 13 (2002), 375–390.

[37] Jihoon Ko, Sangjin Shin, Sungkwang Eom, Minjae Song, Dong-Hoon Shin, Kyong-Ho Lee, and Yongil Jang. 2014. Semantically Enhanced Keyword Search for Smartphones. In *Proc. WWW '14 Companion*. ACM, New York, NY, USA, 327–328. https://doi.org/10.1145/2567948.2577331

[38] Ricardo Langner, Ulrich von Zadow, Tom Horak, Annett Mitschick, and Raimund Dachselt. 2016. *Content Sharing Between Spatially-Aware Mobile Phones and Large Vertical Displays Supporting Collaborative Work*. Springer International Publishing, 75–96. https://doi.org/10.1007/978-3-319-45853-3_5

[39] Gang Li, Yue Liu, and Yongtian Wang. 2017. Evaluation of Labelling Layout Methods in Augmented Reality. In *Proc. VR '17*. 351–352. https://doi.org/10.1109/VR.2017.7892321 00000.

[40] Jacob Boesen Madsen, Markus Tatzqern, Claus B Madsen, Dieter Schmalstieg, and Denis Kalkofen. 2016. Temporal Coherence Strategies for Augmented Reality Labeling. *IEEE Trans. Vis. Comput. Graphics* 22, 4 (April 2016), 1415–1423. https://doi.org/10.1109/TVCG.2016.2518318

[41] Stephann Makri, Ann Blandford, Mel Woods, Sarah Sharples, and Deborah Maxwell. 2014. "Making my own luck": Serendipity strategies and how to support them in digital information environments. *J. Assn. Inf. Sci. Tec.* 65, 11 (2014), 2179–2194. https://doi.org/10.1002/asi.23200

[42] Tamás Matuszka. 2015. The Design and Implementation of Semantic Web-Based Architecture for Augmented Reality Browser. In *The Semantic Web. Latest Advances and New Domains*. Springer, 731–739. https://doi.org/10.1007/978-3-319-18818-8_46

[43] Lori McCay-Peet and Elaine G. Toms. 2015. Investigating serendipity: How it unfolds and what may influence it. *J. Assn. Inf. Sci. Tec.* 66, 7 (2015), 1463–1476. https://doi.org/10.1002/asi.23273

[44] Lori McCay-Peet, Elaine G. Toms, and E. Kevin Kelloway. 2015. Examination of relationships among serendipity, the environment, and individual differences. *Information Processing & Management* 51, 4 (2015), 391 – 412. https://doi.org/10.1016/j.ipm.2015.02.004

[45] Taniya Mishra and Srinivas Bangalore. 2010. Qme!: A Speech-based Question-answering System on Mobile Devices. In *Proc. HLT '10*. Association for Computational Linguistics, Stroudsburg, PA, USA, 55–63. http://dl.acm.org/citation.cfm?id=1857999.1858006

[46] Munehiro Nakazato and Thomas S Huang. 2001. 3D MARS: immersive virtual reality for content-based image retrieval. In *Proc. ICME '01*. 44–47. https://doi.org/10.1109/ICME.2001.1237651

[47] Jason Orlosky, Kiyoshi Kiyokawa, Takumi Toyama, and Daniel Sonntag. 2015. Halo Content: Context-aware Viewspace Management for Non-invasive Augmented Reality. In *Proc. IUI '15*. ACM, New York, NY, USA, 369–373. https://doi.org/10.1145/2678025.2701375

[48] Bradley J Rhodes and Pattie Maes. 2000. Just-in-time Information Retrieval Agents. *IBM Syst. J.* 39, 3-4 (July 2000), 685–704. https://doi.org/10.1147/sj.393.0685

[49] Cornelis J. Van Rijsbergen. 1979. *Information Retrieval* (2nd ed.). Butterworth-Heinemann, Newton, MA, USA.

[50] Dieter Schmalstieg and Tobias Hollerer. 2016. *Augmented reality: principles and practice*. Addison-Wesley Professional.

[51] Kimiaki Shirahama and Kuniaki Uehara. 2011. Query by Virtual Example: Video Retrieval Using Example Shots Created by Virtual Reality Techniques. In *Int. Conf. on Image and Graphics '11*. 829–834. https://doi.org/10.1109/ICIG.2011.158

[52] Sophie Stellmach, Sebastian Stober, Andreas Nürnberger, and Raimund Dachselt. 2011. Designing Gaze-supported Multimodal Interactions for the Exploration of Large Image Collections. In *Proc. NGCA '11*. ACM, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/1983302.1983303

[53] Zsolt Szalavári and Michael Gervautz. 1997. The Personal Interaction Panel – a Two-Handed Interface for Augmented Reality. *Computer Graphics Forum* 16, 3 (1997), C335–C346. https://doi.org/10.1111/1467-8659.00137

[54] Norimichi Ukita, Tomohisa Ono, and Masatsugu Kidode. 2005. Region Extraction of a Gaze Object Using the Gaze Point and View Image Sequences. In *Proc. ICMI '05*. ACM, NY, USA, 129–136. https://doi.org/10.1145/1088463.1088487 00005.

[55] Nayuko Watanabe, Masayuki Okamoto, Masaaki Kikuchi, Takayuki Iida, Kenta Sasaki, Kensuke Horiuchi, and Masanori Hattori. 2011. *Designing Mobile Search Interface with Query Term Extraction*. Springer, Berlin, Heidelberg, 67–76. https://doi.org/10.1007/978-3-642-23854-3_8

[56] Sean White and Steven Feiner. 2009. SiteLens: Situated Visualization Techniques for Urban Site Visits. In *Proc. CHI '09*. ACM, New York, NY, USA, 1117–1120. https://doi.org/10.1145/1518701.1518871

[57] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. 2017. Embedded Data Representations. *IEEE Trans. Vis. Comput. Graphics* 23, 1 (Jan 2017), 461–470. https://doi.org/10.1109/TVCG.2016.2598608

[58] Xing Xie, Lie Lu, Menglei Jia, Hua Li, Frank Seide, and Wei-Ying Ma. 2008. Mobile Search With Multimodal Queries. *Proc. of the IEEE* 96, 4 (April 2008), 589–601. https://doi.org/10.1109/JPROC.2008.916351

[59] Tom Yeh, John J. Lee, and Trevor Darrell. 2008. Photo-based Question Answering. In *Proc. MM '08*. ACM, New York, NY, USA, 389–398. https://doi.org/10.1145/1459359.1459412

[60] Degi Young and Ben Shneiderman. 1993. A graphical filter/flow representation of Boolean queries: A prototype implementation and evaluation. *J. Am. Soc. Inf. Sci.* 44, 6 (1993), 327–339. https://doi.org/10.1002/(SICI)1097-4571(199307)44:6<327::AID-ASI3>3.0.CO;2-J

[61] Stefan Zander, Chris Chiu, and Gerhard Sageder. 2012. A Computational Model for the Integration of Linked Data in Mobile Augmented Reality Applications. In *Proc. I-SEMANTICS '12*. ACM, New York, USA, 133–140. https://doi.org/10.1145/2362499.2362518

[62] Qi Zhang, Simon Zaaijer, Song Wu, and Michael S. Lew. 2014. 3D Image Browsing: The Planets. In *Proc. ICMR '14*. ACM, New York, USA, Article 511, 3 pages. https://doi.org/10.1145/2578726.2582613

[63] Yun Zhang, Hong Fu, Zhen Liang, Zheru Chi, and Dagan Feng. 2010. Eye Movement As an Interaction Mechanism for Relevance Feedback in a Content-based Image Retrieval System. In *Proc. ETRA '10*. ACM, New York, NY, USA, 37–40. https://doi.org/10.1145/1743666.1743674