IEEE Conference on Emerging Technologies for a Smarter World, Seoul, Korea, September, 2010

REANA: An RFID-Enabled Environment-Aware Navigation System for the Visually Impaired

Klaus Mueller and Samir Das

Computer Science Department, Stony Brook University, USA

Abstract **– Nearly 5% of the US population is visually impaired, and nearly half of this population is within 'working age' (18-69). In humans, the visual sense is the most important of the five senses, and a lack of it poses serious limitations, both in daily life activities and also in job placement. The most basic task enabled by the visual system is the navigation of spatial environments. More advanced tasks are the perception, recognition, and localization of objects and people in the scene, which also includes visual search. Recent developments in several technologies with the wireless network in the centerpiece can aid the visually impaired in many of these tasks. For example, GPS, infrared laser scanner, and RFID tags can help in spatial navigation that we exploit in our REANA system. In addition to basic navigation, REANA aims to provide capabilities for more advanced visual tasks, such as object and scene perception, identification, localization, and search. REANA has the potential to make the life of the visually impaired considerably more similar to that of a seeing person, giving him/her substantially improved quality of life.**

Keywords – RFID, navigation, visually impaired, wireless

I. INTRODUCTION

Just in the United States alone, an estimated 14 million people over the age of 45 are visually impaired, implying that they are either legally blind or have serious difficulty seeing that cannot be corrected with ordinary glasses. This amounts to almost 5% of the total US population. Hence, this population represents not just a niche, but a significant faction of our general population. Among the five senses (vision, sound, touch, smell, taste), vision is the most dominant and important sense in humans. The visual cortex in the brain performs cognition, but cognitive processes also occur in the visually impaired, in the same brain real estate. The main difference is that these cognitive processes are fed by less efficient perceptive sensors, that is, the four remaining senses and the weak visual channel. This proposal aims to boost the efficiency of these channels via assistive computing, using pervasive sensor and information network systems.

The system we have begun to build, REANA, represents a novel, unique and holistic approach to making the life of a visually impaired person significantly more similar to that of a seeing person. It is an integrated sensor-driven system that provides the visually impaired with an immersed experience of the environment, a significant improvement over the current traditional passive devices such as the white cane. REANA non-trivially takes advantage of the immense growth of a number of major technologies: (i) sensors such as Radio Frequency ID (RFID) tags, 2D and 3D cameras, and GPS, (ii) wireless communication with on-line databases, (iii) video eyewear, (iv) commodity high-performance computing, and (v) natural language processing – all tied together with a personalized stimulus-alert and guidance framework, implemented as a wearable embedded computer system.

The structure of our paper is as follows. Section 2 illustrates the overall concept of our system, Section 3 describes related work, Section 4 follows with a concrete application example, Section 5 describes our methods, and Section 6 finishes with conclusions and future work.

II. OVERALL CONCEPT

RFID tags have frequently been hailed as the enabler of the “”internet of things” – attaching a unique identifier to any object produced, allowing one to retrieve further information about this object over the internet. The development of RFID tags has been vigorously pursued by large corporations for inventory control. We are utilizing this passive infrastructure, currently untapped within a social context, to create a pervasive on-demand information infrastructure and technology offering electronic access to associated object semantics, history, and even computational behavior stored on internet databases and indexed by the RFID. The general concept of REANA is outlined in Figure 1. With REANA, an RFID tag essentially transforms a passive object into an active information node, connecting it to the internet which then allows it to directly communicate with the user. Such an infrastructure is particularly valuable and critically needed for the visually impaired since this population group has no means to gain access to this information via visual communication, to spawn subsequent (sub)-conscious processes in the higher-level brain cortex. This motivates our choice of the visually impaired as our main target group, but our technology is also helpful to the seeing population.

While the RFID-based communication framework delivers information about object presence and semantics, it is inherently too imprecise to enable sufficiently accurate location information beyond a range of 2 ft. We therefore have chosen, in on-going work, to complement it with a computer vision based system that will acquire and process visual scene information completely autonomously. This

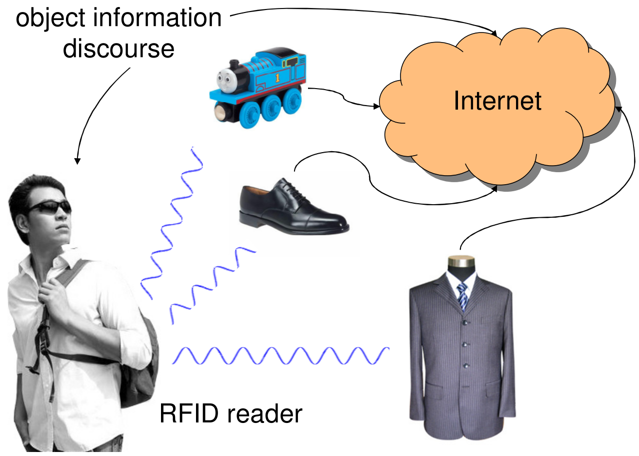

Figure 1: Using the object's RFID tag to "connect" the object to the internet, and in turn to the user, allowing the object to communicate with the user and back. In the case pictured here, the user may have been looking for a suit, dress shoes, and a toy for his little nephew's birthday. With instances of these objects of possible interest in reach, he may want to examine their features, history, user reviews, etc more closely. REANA establishes a 2-way information discourse, with the internet providing the communication medium. Given closer interest, the system would then guide the user to the object, using REANA's RFID-coupled computer vision system.

integration of RFID-based communication with computer vision creates a synergy in which each (wireless) modality complements the weakness of the other, creating a system in which the overall value is much greater than the sum of its parts. More concretely, while generalized object recognition via computer vision is prohibitively expensive, it is quite feasible with RFIDs. On the other hand, while precise known-object localization is infeasible with RFID technology, it is within the reach of computer vision based techniques, especially when accelerated on commodity graphics hardware chips (GPUs).

III. RELATED WORK

For years, visually impaired people have used canes and guide dogs to navigate the environment. While a variety of more advanced sensory devices, such as the SonicGuide [8] and the Laser Cane [1] have been developed over the years, these were eventually not successful as commercial products. On the other hand, computer-based products have become quite popular among the visually impaired. Apart from Blind-ready PCs, PDAs have also become available, mostly based on HPs iPAQ platform. HumanWare [9] (formerly VisuAide) markets the *Maestro,* which has a text-to-speech interface and a special tactile keyboard that sits on top of the iPAQ touch-screen. They also developed *Victor*, a digital audio book, while Sendoro Group [15] offers *Trekker*, a GPS-based orientation and navigation device. Freedom Scientific [6] offers the *ScanTalker* which scans the UPC code of a product label and reads its info to the user.

A variety of public services have also been conceived to aid the visually impaired in their daily activities. Some cities, at selected locations, have chirping sounds at traffic light mediated intersections to *sonify* (audio) the 'stop/go' visual signals, but much more frequent is the inexpensive braille labeling at public sites. A 'shared' technology are the audio-based tour guides often available at museums and tourist sites, which are also used by visually impaired persons and thus require no additional investment. Likewise, our proposed system is also to a large extent a shared technology, exploiting infrastructure and devices that already are or soon will be ubiquitously available.

IV. APPLICATION EXAMPLE

Figure 2 illustrates the use and capabilities of system via a sufficiently complex example application scenario: helping a visually impaired person to navigate and experience a shopping mall (see caption for a detailed description). Note that the traditional cane or guide dog would still provide direct navigation support – the REANA system is meant only as an added utility, not as a replacement of the devices the visually impaired user has relied on for many years.

V. METHODS

In our REANA system, a wearable RFID reader constantly scans the surrounding scene, detecting objects in the vicinity by their RFID tags. A wearable custom mini-computer then decodes the RFID code and retrieves, over wireless internet, the description of these objects. In the future, a wearable actuated camera will scan the environment for other environmental cues. Further, due to the potentially massive number of RFID-tags and information present in the scene, the computer first compares the delivered object descriptions with the current user preference parameters stored on the computer. Only in the case of a match, which can occur at many levels of refinement will the user be alerted. Should the user then decide to approach the object, the actuated 2D/3D camera system will aid him/her in localizing the object, and an audio- visual interface consisting of a lightweight video eyewear unit, LED display, and speech will provide the guidance for grasping it. Our system also logs all identified objects – environmental and merchandise – and build a map of the traversed environment to aid in a later navigation to the object, should interest in the arise.

V.1 Dealing with the massive number of objects

Returning to our shopping mall stroll example, it is not unrealistic to assume that a (sighted) person would casually <u>encounter</u> 1,000s of objects (in a mall of 100,000 objects), of which 100s he would <u>notice</u> and <u>remember</u>, while a dozen he would eventually <u>grasp</u>. This requires (i) rapid object recognition (the encountering), (ii) information filtering based on personal interest (the noticing), (iii) localization and storage of these interesting objects (the remembering), and (iv) micro-navigation to a desired object

Figure 2: Shopping mall scenario: The visually impaired person is casually strolling along his usual guide dog or using a cane (both are shown here but in practice only one or the other is used). He has a backpack (with computer, RFID reader, compass, and GPS inside), actuated shoulder-worn 2D / 3D camera, navigation eye-wear, and a headset, consisting of an earphone and a microphone.

(**Left**): Initiated by our system, the RFID reader reads all the RFID tags in the vicinity and also takes images of the scene around (in the figure, these objects are drawn at full intensity). Descriptive information about these objects is retrieved from on-line databases and logged in the system, along with their spatial location. An environment-map is also being built, using the GPS and RFID information.

(**Right Top**): Only certain objects (drawn at full intensity) match the user's preference profile. The system generates alerts for these objects, and upon request, guides the user to the desired object.

(**Right Bottom**): The user instead wishes to take a rest, and thus, changes the user-preference to "rest." The system guides the user to the near-by bench, using the environment-map built.

(the grasping). People with normal vision easily master these tasks. We now describe how REANA will assist a visually impaired person to accomplish them as well.

It is tempting to model the visually impaired person simply as a robot and then use the standard set of techniques from robot localization and navigation [16] to achieve these goals, however, the problem here is conceptually very different. First, while robots can be measured and controlled very precisely, the kinematics and geometry of humans are hard to assess and even more difficult to instruct. Second, our aim is to facilitate a *semantic* recognition of the scene, and not just the detection and recognition of strong landmark features based on low-level saliency measures.

On the other hand, for the recognition task above, if one were to apply a standard computer vision based technique to substitute for the lack of good vision, one would use an indexed database of objects each encoded as a set of low-level features, and then try to match those with equivalent features extracted from the scene. An large number of recognition methods, using SIFT [10] feature encoding and others, have been described (and many are quite robust and accurate [2][12]). But the abundance of scene objects will make the use of standard computer vision impractical. We therefore resorted to RFIDs to manage this task. The following section details our approach.

V.2 System description

As mentioned in the introduction, we propose to use information derived from RFID tags to quickly recognize nearby objects and retrieve semantic information about them from internet databases and general websites using standard wireless communication. Herein it is assumed that all consumer products are tagged in such a way. Facilities, such as benches, might also be tagged and identified in this manner. The 2D/3D camera system will be used to guide and supplement this process. While the recognition is aimed to be exact, the localization of the objects at this stage will be approximate. A second stage, based on the imaging system, will make this approximate localization exact. Finally, a third stage, using both the camera system and the audio-visual feedback system, will then micro-navigate the user to the object for grasping.

RFID-Based Recognition: An RFID tag returns a 96-bit EPC (Electronic Product Code) unique for each instance of a product, which is then routed by an ONS (Object Name Service) to an online EPC-IS (Information Service) database to retrieve, over a WI-FI or cellular connection, specific information on the RFID-tagged object [3][4]. We will assume that these online databases will store, for any EPC code, a vendor's standard object descriptions, such as dimensions, category, and one or more images in a public area (a secured area will hold more proprietary information). We further assume that privacy rights will be preserved by moving all object information into a secure area once the product has been sold. It can be expected that these web services will grow in utility and scale, given the fact that the cost of an RFID-tag is projected to fall below 5 cents, making it sufficiently economical. Further, due to its convenience for inventory tracking, RFID technology is being heavily pursued by large corporations, such as Wal-Mart. Many articles predict that RFIDs will be the new UPC bar code [13] and they are often called the *internet of things* [14].

Filtering: The number of tags read on an RFID-read burst can be vast. Filtering of these occurs on three levels. First, in order to minimize network traffic for tag information we filter out duplicates of multiple instances of the same object. This can be determined from the EPC code. But even then, there still can be a large number of objects. Thus, a second filtering stage uses the approximate spatial clustering of the tags and the taxonomy derived from the EPC code to produce a summary of the scene objects. This can be presented to and refined by the user on request. The third level produces the sparsest set of objects. For this we propose to use machine learning techniques to learn a user's preference profile, such as style and size of clothing, taste of music, preferences of food, budget, and so on. All this can be matched with the categorization of the EPC code, and other categorization hierarchies, such as WordNet [18]. The learning of hierarchical object classifiers has become a heavily researched area in recent years, spawned by the need for content-based image retrieval from large online collections, such as Flickr [5]. Established techniques are support vector machines [17], neural networks, or AdaBoost [7]. A facility allows the user to focus his profile to his current short-term interest. The profile itself is set up in an initial training session and then updated during system use.

Use modes: We provide two basic modes of operation, to be altered at any time: 'cruise' and 'explore'. The 'cruise' mode is activated when the user seeks to engage in a casual stroll, only to be alerted when an object (at a selected detail) matches his profile, while in 'explore' the user performs a focused search. But even without an alert, information is logged for possible future recall, given proper match with the general profile (see below).

VI. CONCLUSIONS AND FUTURE WORK

REANA is a unique assembly of sensor, wearable computing, wireless communication/networking and HCI technologies to deliver a holistic system with a significant societal value, here for the visually impaired population. Current work also extends REANA's capabilities to using GPS as well. Via GPS it can provide visually impaired, but also seeing, shoppers with directions, automatic alerts, and silent storage for later recall, to products of possible interest when strolling through a neighborhood or shopping mall. These actions will occur on the basis of the types of products that the shopper has previously expressed an interest in, the purchases he has made so far, products he may be in need of or may complement items he already calls his own. Further, the shopper may tune this profile before the stroll, and he may extend it to also include the profiles and needs of friends, family, or business. The system we propose will use intelligent planning technology to compute a policy that optimizes the expected utility of a shopper's walk through the shopping mall, taking into account uncertainty about (i) whether the shopper will actually find a suitable product in a given location and (ii) the time required for each purchase.

REFERENCES

[1] D. Bolgiano, E. Meeks. Laser cane for the blind IEEE Journal of Quantum Electronics, 3(6):268, 1967.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In Proc 9th European Conf on Computer Vision. pp. 404–417, 2006.

[3] http://www.epcglobalinc.org/

[4] http://www.verisign.com.sg/guide/epc/epc_architecture.pdf

[5] R. Fergus, L. Fei-Fei, P. Perona and A. Zisserman. Learning Object Categories from Google's Image Search, Proc. of the 10th Inter. Conf. on Computer Vision, ICCV 2005.

[6] http://www.freedomscientific.com/

[7] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1):119–139, August 1997.

[8] B. Goldstein, W. Wiener. Acoustic analysis of the Sonic Guide. The Journal of the Acoustical Society of America 70(2): 313-320, 1981.

[9] http://www.humanware.ca

[10] D. G. Lowe: Object Recognition from Local Scale-Invariant Features. ICCV 1999: 1150-1157

[11] Masciari. RFID data management for effective objects tracking. Proceedings of the 2007 ACM Symposium on Applied Computing pp. 457-461, 2007.

[12] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," IEEE Trans. Pattern Anal. Mach. Intell., 27(10), pages 1615-1630, 2005.

[13] http://www.idii.com/wp/MhaRfid.pdf

[14] http://magazine.digitalidworld.com/Nov03/Page66.pdf

[15] http://www.humanware.ca/web/en/p_DA_Trekker.asp

[16] S. Thrun. Robotic Mapping: A Survey. In Exploring Artificial Intelligence in the New Millenium. G. Lakemeyer and B. Nebel,eds., Morgan Kaufman, 2002.

[17] V. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, 1995.

[18] http://wordnet.princeton.edu/