Integration of RFID and Computer Vision for Remote Object Perception for Individuals Who Are Blind

Troy L. McDaniel Kanav Kahol Daniel Villanueva Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing, Arizona State University
699 S. Mill Avenue, 3rd Floor
Tempe, Arizona, 85281
480-694-0021
{troy.mcdaniel, kanav.kahol, daniel.villanueva, panch}@asu.edu

ABSTRACT
Over the last few years, Radio-Frequency Identification (RFID) technology has gained popularity for use in assistive technology for individuals who are blind. Recently, RFID-based wearable assistive devices have been developed for individuals who are blind to assist with navigation or remote object perception. However, RFID-based assistive technology suffers from two major drawbacks: (1) information overload in environments with many tagged objects, and (2) usability issues in untagged environments. In this paper, we propose a framework for integrating RFID and computer vision in assistive devices for remote object perception to overcome the aforementioned limitations. Computer vision enables content selection to help prevent information overload and provide users with only relevant information found through RFID. Moreover, computer vision can be used to learn a mapping between visual data and object features as acquired through tags, which will enable computer vision to replace RFID in untagged environments.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *artificial, augmented, and virtual realities.* H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, human information processing.* I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *color, object recognition.*

General Terms
Algorithms, Design, Reliability, Human Factors.

Keywords
RFID, computer vision, distal object perception, assistive technology, haptic user interfaces.

HAS'08, Feb 11–14, 2008, Quebec City, Quebec, Canada.

1. INTRODUCTION
Wearable or handheld assistive devices for remote object perception for individuals who are blind convey to users information about distal objects in the environment. Typically, such systems are either vision-based [1-4] or tag-based [5-8]. Vision-based systems use computer vision algorithms to extract object features, e.g., shape, size, texture, etc., from images of objects in the environment. However, these systems are limited as computer vision is challenging in uncontrolled, real-world environments due to lighting changes, scale changes, pose changes, motion blur, video noise, etc. [9].

On the other hand, tag-based systems for remote object perception rely on identifying tags to extract object features. Tags may be visual in which computer vision algorithms find and identify tags, or embedded, in which radio frequency signals are used to identify tags. The latter is referred to as Radio-Frequency Identification (RFID) technology. Tag-based systems are limited by two major drawbacks: (1) information overload and (2) usability issues in untagged environments. For example, consider a scenario where a user enters a store where merchandise is equipped with RFID tags. As the user explores the store, he or she will be overwhelmed with information about tagged objects as the RFID reader (or interrogator) reads the tags of possibly many surrounding objects. Instead, a methodology must be developed wherein content selection may be used to convey to the user only relevant information.

Another important drawback of tag-based assistive technology is that it is not usable in untagged environments. A vision-based learning module may be used wherein ground truth provided by tagged objects will enable computer vision algorithms to continually learn the mapping between visual data (images captured by the wearable system) and object features (stored on the tags). This is in contrast to typical training of computer vision algorithms where algorithms are trained from limited datasets and then expected to perform well in a variety of real-world conditions that may significantly differ from those seen during training.

In this paper, we propose a framework that integrates RFID and computer vision for remote object perception by taking advantage of the strengths of each technology to overcome their individual

limitations. In our proposed framework, computer vision is utilized in two respects: first, it enables content selection to help prevent information overload, and secondly, it can help develop a learning module to improve system usability in untagged environments. Moreover, to combat the problem of the effects of real-world conditions on the performance of computer vision algorithms, we propose the use of *reliability measures* [10]. Reliability measures assess current environmental conditions such as illumination or motion blur, and provide the user with (1) an assessment of each condition, and (2) an overall measure of system reliability. Assessment of environmental conditions is achieved through comparison of *real working conditions*, i.e., the current environmental conditions, and the *optimal working conditions*, i.e., the environmental conditions algorithms were trained in. To ensure reliability measures are intuitive and enable real-time perception, it is recommended they operate at a perceptual level, i.e., each reliability measure provides a classification of an environmental condition into one of several pre-determined categories. By understanding environmental conditions and the reliability of system-made decisions, users can act on the environment to improve reliability and therefore the usability of assistive devices. Hence, this is a collaborative framework wherein human computation is taken advantage of to solve tasks that the system alone cannot.

In the next section, we provide a survey of related work that combines RFID and computer vision. In Section 3, our proposed conceptual framework is presented. In Section 4, we introduce our experimental methodology. And finally, Section 5 concludes and presents possible directions of future work.

2. BACKGROUND AND RELATED WORK

In this section, we review assistive technology for remote object perception. This brief review divides approaches into vision-based technology, tag-based technology and hybrid approaches that combine both vision and RFID technology. Vision-based techniques may be divided into those approaches that work at a physical or perceptual level [2], and tag-based techniques may be divided into visual and RFID tagging systems (see Figure 1).

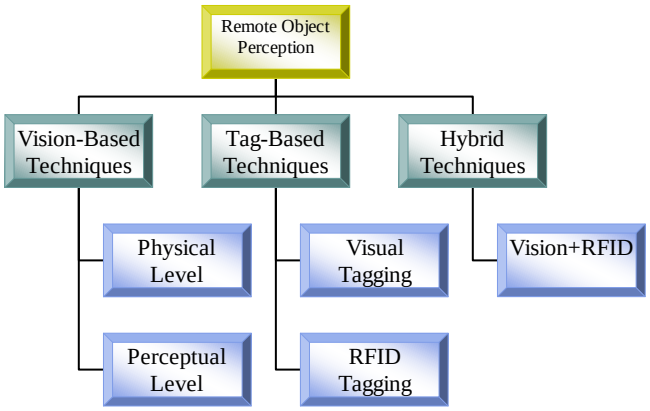

Figure 1. Approaches for remote object perception for assistive technology.

2.1 Vision-Based Techniques

VIDET [1], a vision-based wearable assistive device for remote object perception, enables users to feel objects from a distance. Physical 3D models of objects are first extracted through stereo vision. Users may then feel the 3D models through a wire-actuated haptic user interface wherein the movement of the user's finger is constrained by a wire as the user's finger moves across the 3D model. Unfortunately, experimentation revealed that users have a difficult time recognizing objects through the haptic exploration provided by VIDET.

Given the difficulty of perceiving physical objects through haptic user interfaces, researchers have taken other approaches to remote object perception, specifically, the presentation of information at a perceptual level [2-4]. Object features may be classified into pre-determined perceptual categories, and then subsequently conveyed to users to invoke mental concepts [2]. This method is much easier compared to working with physical representations, and can enable users to perceive remote objects in real-time.

A few examples of approaches that work at a perceptual level include a suite of visio-haptic transfer algorithms [2] that extract haptic features, specifically haptic shape, size, texture and material, from an object's visual image; a vision-based wearable assistive device for landmark recognition [3]; and a handheld vision-based device for object feature detection including color and size [4]. In the next section, we review tag-based approaches, both visual and RFID, for remote object perception.

2.2 Tag-Based Techniques

Badge3D [5] is a relatively recent wearable assistive device for object recognition and obstacle avoidance designed for individuals who are blind. Computer vision algorithms are used to find and extract visual bar codes, which help identify objects, as well as estimate object position and pose. Untagged objects lying on the ground are identified through ultrasonic signals to help users avoid them. Another visual tagging and extraction system is CyberCode [6] developed by Rekimoto and Ayatsuka. In their work, they present a new visual tagging system based on 2D visual patterns, which are easy for off-the-shelf camera's to recognize, and can be used for a variety of tasks including object recognition and pose estimation.

Unfortunately, visual tags have several drawbacks. First, visual tags must be in the line-of-sight of the camera otherwise they cannot be detected. Secondly, visual tags alter the appearance of objects whereas RFID tags may be embedded inside objects. And finally, information stored in visual tags is more difficult to alter compared to RFID tags. These limitations warrant research into alternatives to visual tags, specifically, the use of RFID tags for remote object perception.

An example of the use of RFID tags for navigation and remote object perception is RoboCart [7]. RoboCart is a robotic shopping assistant for individuals who are blind or visually impaired. RFID tags are placed at key locations in a grocery store, and help guide users to desired products. Although the system is designed for navigation, it can be easily extended to remote object perception. As another example, Willis and Helal [8] developed a framework for a RFID information grid to assist with navigation and remote object perception. In this framework, RFID tags are placed in a

grid formation both indoor and outdoor to assist with navigation. Moreover, tags may be placed at entrances to inform users about objects within rooms such as tables, chairs, etc., and their relative locations. Next, we review hybrid approaches for remote object perception that combine RFID and computer vision. Although these approaches have been limited to the field of robotics to improve object recognition and registration, in understanding these techniques, we can learn a lot about how these may be applied to assistive technology for individuals who are blind.

2.3 Hybrid Approaches: Vision+RFID

The integration of RFID and computer vision has been limited to the field of robotics with the purpose of simplifying object recognition, object localization, object tracking and task planning [11-16]. One of the earliest approaches to tag-based visual object recognition was developed by Mae et al. [11]. An object's appearance model, extracted from its tag, is used to recognize the object in the scene. If no appearance model, contained in the tag, matches the object, then a new appearance model is accumulated and stored in the tag. After enough models with different poses are accumulated for an object, a robot will be capable of recognizing the object without RFID. One shortcoming of this approach is that it might be difficult to accumulate new appearance models given the difficulty of segmentation in real-world conditions.

In [12], Boukraa and Ando attempt to solve the problem of vision-based object recognition and registration through the use of RFID. A polyhedral object model is extracted from an object's tag and is used for object registration via projective geometry. Similar to [11], this approach solves the problem of real-time model-based object recognition as only a subset of object models need to be considered during object recognition.

One issue of [12], as identified by Takemura et al. in [13], is that for users, object models are time consuming and costly to create. Hence, in [13], when a tag is encountered, its CAD model, designed by manufacturers, is downloaded and used for analysis. Obviously one drawback of this strategy is that not every object will have a CAD model available from the manufacturer.

In [14], Hontani et al. developed an object tracking system that uses a CAD model obtained from an object's tag. In this framework, visual tags are used rather than RFID tags, and hence, tags may not be identified if they are not in the camera's view.

Chong et al. [15] have developed a framework for combining RFID and computer vision for task planning in robotics. They developed an experimental setup wherein a robot's task is to clear dishes from a table after customers have finished eating. First, the robot obtains CAD models via RFID tags placed on each object. These models are then used to find the position of each object. This information helps the robot learn how to grasp objects so that it may pick up and remove each tagged object from the table.

Finally, in a more recent approach, Kim et al. [16] developed a robotic system for object recognition and localization. The authors proposed the use of smart tags which have an active landmark (IRED) and a data structure consisting of geometrical, physical and semantic information. When a tagged object is read, its IRED is activated, and the robot searches the scene for the active landmark, which will be a flickering light. When the light is found, stereo vision on a pan-tilt mechanism is used to find the object's depth, size and pose. This information may then be used to grasp the object. The major disadvantage of this system is that smart tags might be too costly and bulky to use for most applications.

Similar to those approaches just described, we are interested in tag-assisted visual object recognition. However, rather than a model-based approach, we propose an approach based on invariant local features. Moreover, as our framework is for remote object perception for wearable assistive technology, tag-assisted visual object recognition is used to prevent information overload and enable usability in untagged environments, as previously discussed. In the next section, our proposed conceptual framework is presented.

3. Conceptual Framework

We propose a novel framework that integrates RFID and computer vision to enable remote object perception in wearable assistive devices for individuals who are blind or visually impaired. With respect to remote object perception, we are mainly interested in conveying perceptual haptic (i.e., tangible) features, which will enable users to feel objects from a distance, but at a perceptual level rather than a physical level. As discussed in Section 2, existing wearable systems to accomplish this task utilize visio-haptic transfer algorithms to convert visual information into haptic features, either at a physical level or a perceptual level. However, by working at a perceptual level, we enable real-time perception and improve system usability.

Our proposed framework is depicted in Figure 2, and consists of the following main components: (1) visio-haptic ground truth collection; (2) RFID sensing; (3) content selection; (4) learning; and (5) reliability measures. The proposed framework for reliability measures is depicted in more detail in Figure 3. Each component is further discussed in the following subsections.

3.1 Visio-Haptic Ground Truth Collection

In [17], we proposed a methodology to collect reliable ground truth for training and testing visio-haptic transfer algorithms and visual-to-tactile image translation systems. The framework was used to create the Visio-Haptic Object Database (VHOD). Ground truth is in the form of physical and perceptual visual ground truth, and physical and perceptual haptic ground truth. Only perceptual haptic ground truth is covered here as it is of most relevance. We refer the reader to [17] for more information regarding the other aforementioned types of ground truth.

Reliable ground truth is critical as perceptual classifications are often subjective. Hence, a large sample population of perceptual classifications is required to achieve good ground truth. First, objects are collected such that within each object category, there are four subsets of objects that each vary along one of four dimensions: haptic shape, size, texture and material.

Perceptual haptic ground truth is collected through a controlled capture session involving blind-folded participants. A participant is handed each object, which is randomly selected, to haptically explore one time. Ground truth collection is completed once each object is explored at least five times (preferably more) over the course of all capture sessions.

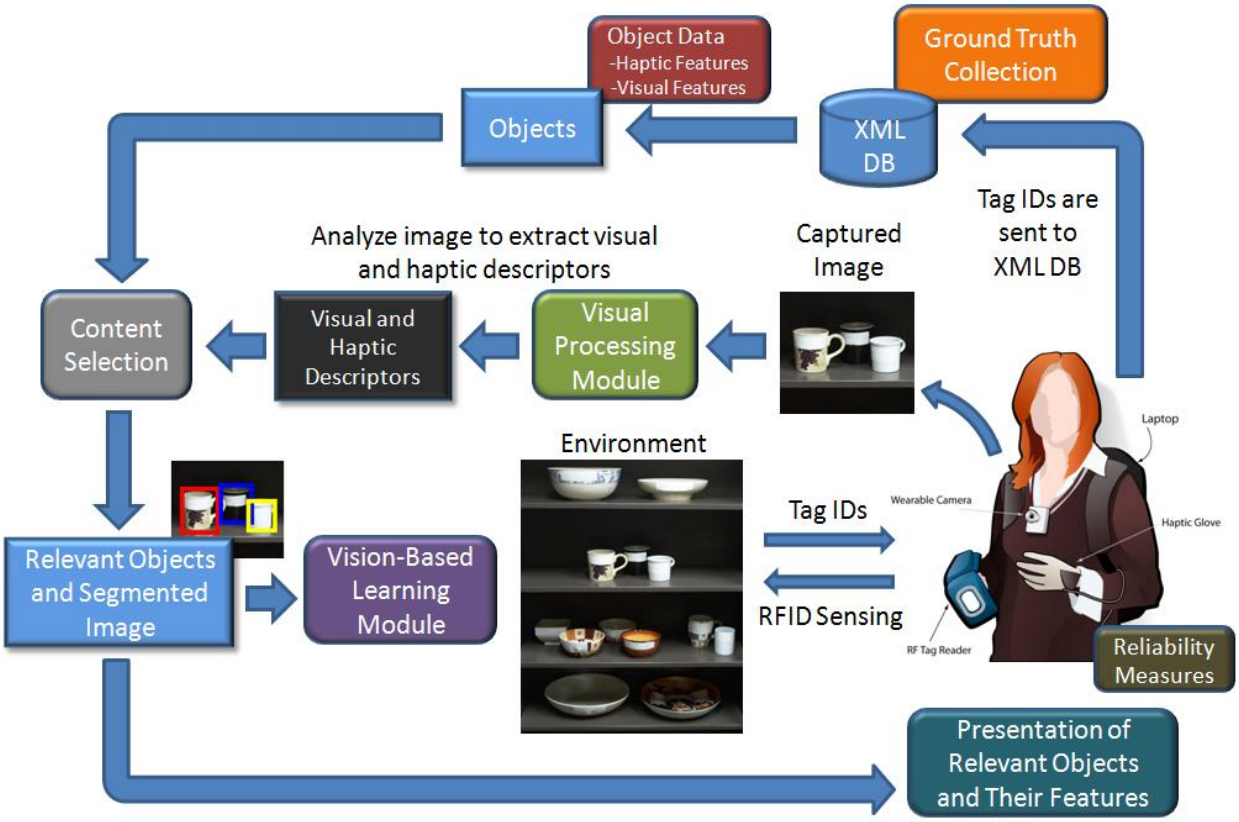

Figure 2. Proposed conceptual framework for integrating RFID and computer vision for the task of remote object perception for individuals who are blind.

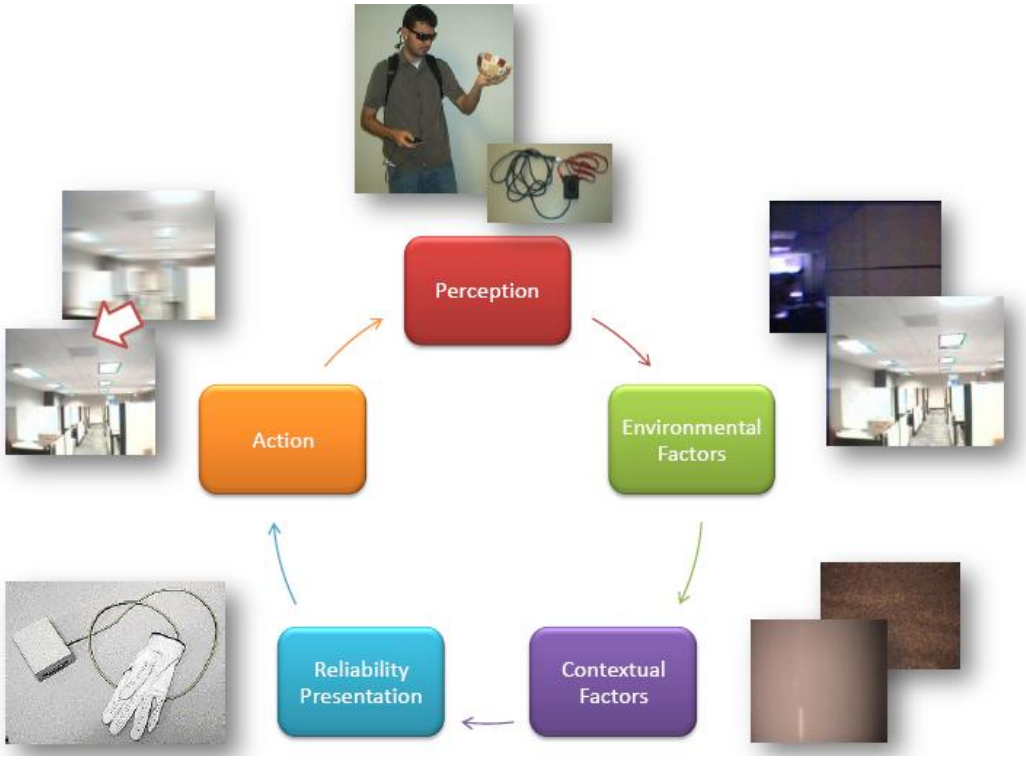

Figure 3. Proposed conceptual framework for reliability measures.

```
<Item ID="HE004447306010000">
    <Haptic>
      <OverallShape>Cup</OverallShape>
      <OverallSize>Large</OverallSize>
      <OverallTexture>Smooth</OverallTexture>
      <OverallMaterial>Ceramic</OverallMaterial>
      <TallerThanWider>Yes</TallerThanWider>
      <BaseLargerThanRim>No</BaseLargerThanRim>
      <RimShape>Round</RimShape>
      <RimSize>Large</RimSize>
      <RimTexture>Smooth</RimTexture>
      <RimMaterial>Ceramic</RimMaterial>
      <BaseShape>Round</BaseShape>
      <BaseSize>Medium</BaseSize>
      <BaseTexture>Smooth</BaseTexture>
      <BaseMaterial>Ceramic</BaseMaterial>
      <UVRShape>Low</UVRShape>
      <UVRSize>Large</UVRSize>
      <UVRTexture>Smooth</UVRTexture>
      <UVRMaterial>Ceramic</UVRMaterial>
      <MVRShape>Low</MVRShape>
      <MVRSize>Medium</MVRSize>
      <MVRTexture>Smooth</MVRTexture>
      <MVRMaterial>Ceramic</MVRMaterial>
      <LVRShape>Low</LVRShape>
      <LVRSize>Small</LVRSize>
      <LVRTexture>Smooth</LVRTexture>
      <LVRMaterial>Ceramic</LVRMaterial>
    </Haptic>
    <Visual>
      <Color>green</Color>
      <Color>purple</Color>
      <Color>white</Color>
    </Visual>
  </Item>
```

Figure 4. XML object sample.

Participants are asked to haptically explore each object without seeing it to answer a series of questions. An object is divided into six regions: the rim, base, inside, lower vertical region (LV), middle vertical region (MV), and the upper vertical region (UV). Participants answer a total of 28 questions, four questions for the overall object and each object region except the inside region which has only two questions. Furthermore, two questions are asked pertaining to auxiliary information. The multiple choice questions for each region or overall object pertain to that particular region's shape, size, texture and material. Given a lack of space, only questions pertaining to the overall object are listed here, but the complete list of questions can be obtained online as part of the database [17]. The following questions were used to acquire some of the ground truth for VHOD. These questions will of course differ depending upon the reader's object categories and desired perceptual classifications. Questions about the overall object for VHOD include (1) Is the overall shape of the object (a) a bowl; (b) a cup; (c) a glass; or (d) irregular; (2) Is the overall size of the object (a) small; (b) medium; or (c) large; (3) Is the overall texture of the object (a) smooth; (b) medium; or (c) rough; (4) Is the overall material of the object (a) plastic; (b) metal; (c) ceramic; (d) glass; (e) cloth; or (f) other; (5) Is the object taller than it is wider (yes/no); (6) Is the base larger than the rim (yes/no). The majority answer is taken as the consensus for each question.

The ground truth just described provides reliable perceptual classifications of haptic features for objects, which is stored in an XML database. To assist with content selection, described in Section 3.3, ground truth in the form of visual features may also be stored. In the next section, we describe how this information is stored for objects and how it can be accessed using an RFID reader.

3.2 RFID Sensing

An object's overall haptic features and haptic features for each region are stored in an XML database under its tag ID (see Figure 4 for a sample XML object). Moreover, we may store visual features, which are used for content selection, described in the next section. When an RFID reader comes in range of an object's RFID tag, its tag ID is retrieved and used to obtain its information (haptic features, visual features, etc.) from the XML database. The database, which may reside on a local server, is accessed wirelessly via 802.11.

Upon sensing and gathering an object's haptic features from the XML database, these features are conveyed to the user through an audio or haptic interface based on user preference. Users may decide what type of information to receive such as overall features or more detailed information about the object. The audio interface describes the object to the user, e.g., "The object is a bowl, its size is large, its texture is smooth and its material is ceramic," whereas the haptic interface may convey information efficiently through the use of tactile cues [2]. That is, object features are conveyed as patterns of short and long vibrations, and are delivered to different parts of the hand based on the neurophysiological layout of tactile sensors. Extensive experimentation has revealed fast learning rates and ease-of-use. Moreover, by conveying haptic features to users at a perceptual level, we may invoke mental concepts, and hence, users can achieve efficient and accurate perception.

When a user encounters multiple tagged objects, RFID readers typically resolve any collisions, and eventually retrieve each tag ID. In our system, as tags are retrieved, object information is conveyed to the user. This presents a problem in situations where there might be ten, fifteen or more tagged objects in close proximity, such as tagged merchandise in a store. Hence, some method of deciding which tags are relevant and which are not must be developed in order to avoid overloading the user with useless information. In the next section, we present a novel approach wherein information overload is prevented in RFID systems through the use of computer vision.

3.3 Content Selection

We propose the use of computer vision to help select relevant incoming information sensed through RFID. The goal is to discard any information that is not relevant to the user. Here we assume that objects directly in front of the user are of relevance, as well as objects of interest as determined by user preference. Using visual features (e.g., color, visual texture, etc.) and/or visio-haptic features (shape, size, texture, material, etc.) found through RFID sensing, the respective objects may (or may not) be recognized in the scene in front of the user as captured by a wearable camera. That is, as RFID sensing is continuously occurring, image capture is occurring simultaneously, and in turn, tagged objects are recognized, and those that are found within the image, are conveyed to the user. RFID tags that are sensed but belong to those objects not found in the captured image are not conveyed to the user. However, objects may be conveyed to the user based on

user preference, even if the object is not recognized in the captured image. For example, if the user is in a store and he or she is interested in buying a jacket, all RFID tags belonging to jackets could be conveyed to the user, or at least the user can be notified of surrounding tagged jackets.

Tag-assisted visual object recognition has largely been accomplished through model-based object recognition, as previously discussed in Section 2.3. Instead, we propose tag-assisted visual object recognition through invariant local features. In this work, we use color information, which is invariant to many environmental variations, although it is limited by illumination variations. As part of future work, we hope to investigate the use of SIFT features for this application. Next, we describe how color information can be stored and utilized in the task of tag-assisted visual object recognition.

In the XML structure depicted in Figure 4, we include another entry for color information. Rather than a single color, a list of colors may be added. It's important to note that only an object's significant colors, i.e., those a camera will be able to capture from a distance, should be added. It may be useful to include and utilize visual textures, i.e., patterns created by color variations, but we save this for future work.

After reading all detected tags through RFID, an image is captured and color segmented through the following process. Given an image, each pixel is first classified independently of its neighboring pixels using Bayesian classification. A pixel is classified as the color category that maximizes the posterior probably conditioned on the pixel value:

$$P(C_i \mid x) = \frac{p(x \mid C_i)P(C_i)}{\sum_{j=1}^{n} p(x \mid C_j)P(C_j)} \quad (1)$$

where C_i is the i^{th} color category, x is the pixel value and n is the number of color categories. As shown in (1), the posterior probability is equal to the likelihood of C_i times the prior probability of C_i divided by a normalization factor, which can be ignored for the task of classification. The prior probability is the number of occurrences of a certain color category divided by the total number of pixels in the training set. The likelihood of C_i can be estimated using Maximum Likelihood Estimation (MLE). Assuming the densities are Gaussian, MLE is achieved by computing the mean and covariance matrix of each color category.

Given that vision-based wearable systems are (1) usually equipped with low-cost off-the-shelf video equipment, and (2) must operate in real world conditions with possibly extreme environmental variations, point-based color classification often misclassifies pixels, resulting in noisy segmentation results. Instead, we can take into account a pixel's neighborhood to improve segmentation. In our framework, we use the methodology of [18], which uses the Iterated Conditional Modes (ICM) algorithm [19] to maximize a pixel's conditional probability based on its neighborhood. As in [18], we assume that the classes of neighbors are known, and each color category is treated as an independent process, modeled by the first order Gibbs-Markov random field:

$$P(C_i \mid N) = \frac{1}{Z} e^{\lambda \frac{N_i}{N}} \quad (2)$$

where N is the neighborhood, N_i is the number of pixels in the neighborhood that fall into color category C_i, Z is the normalization factor, which can be ignored since it is constant across posterior probabilities, and λ is the clique potential, which determines the dependence of a pixel on its neighborhood. As λ increases, a pixel's dependence on its neighborhood strengthens. In the ICM algorithm, (2) is applied to the image multiple times until a stopping criterion is met.

The percentage of each color found in the image is computed, and colors with very small percentages (e.g., 3% or less) are ignored as these represent noise. Background subtraction is performed to obtain only colors belonging to foreground objects. (Although background subtraction is difficult in general, if we know some prior information about the current environment, much better results can be achieved. For example, when the user enters a new environment, through RFID, he or she could receive background images of walls, floors, ceilings, etc., which could assist the system in background subtraction.) The remaining colors are grouped into color profiles based on their proximities, and compared with the color profiles of objects found through RFID. The objects with matching color profiles are then conveyed to the user.

3.4 Learning

In untagged environments, computer vision may be used in place of RFID for extraction of haptic features from visual data, i.e., visio-haptic transfer. To learn the mapping between visual and haptic information, two pieces of information are required: haptic ground truth and segmented images. Haptic ground truth is obtained through the sensing of RFID tags, whereas segmented images are gathered through content selection, described in Section 3.3. Given a captured image, visual and haptic descriptors are extracted by the visual processing module of Figure 2 for use in content selection. At the stage of content selection, we are performing tag-assisted visual object recognition; hence, the location of each tagged object is found within the captured image, which provides us with a segmented image consisting of tagged objects (foreground) and the remaining background.

In this paper, computer vision algorithms for learning visio-haptic transfer are part of future work, and will be discussed in more detail in Section 5. In the next section, we discuss reliability measures.

3.5 Reliability Measures

The proposed framework to incorporate reliability measures into wearable assistive devices is depicted in Figure 3. The framework consists of a cycle of five phases: (1) Perception; (2) Environmental Factors; (3) Contextual Factors; (4) Presentation; and (5) Action. The cycle begins with the system performing

analysis on input from the environment to complete a task requested by the user, another system or the system itself. In our case, reliability measures will be used in combination with tag-assisted visual object recognition, where reliability is determined through environmental and contextual factors, described next.

Environmental variations are assessed by the system, and based on the current working conditions, each reliability measure provides a penalty or reward, which help derive the reliability of the current system-made decision. Penalties and rewards should be determined based upon the effect environmental variations have on an algorithm's performance. For example, as illumination conditions worsen, the penalty, as provided by a reliability measure for illumination, should become more significant. Continuing with the same example, if illumination conditions are optimal, e.g., the illuminant matches that of the illuminant used during the training of a vision-based algorithm, then the reliability measure may provide a reward to the reliability value. Penalties and rewards are issued to a reliability value that either begins at 100%, for deterministic recognition algorithms, or below 100%, for stochastic recognition algorithms. Rewards should be taken into account after penalties, and reliability should never increase above 100%. Possible environmental factors for vision-based wearable assistive devices include illumination conditions, motion blur, video noise, etc. Here, we present novel reliability measures for illumination and motion blur, described next.

The illuminant of a scene can have adverse effects on vision-based systems as many computer vision algorithms are sensitive to illumination changes. Often, algorithms tend to fail when required to perform under an illuminant different from that used during training. Moreover, changes in illumination can create shadows or specularities on an object's surface, which is problematic for many computer vision algorithms [9]. Hence, these issues provide motivation for a reliability measure for illumination wherein the illuminant of a scene is classified and conveyed to the user. Illumination classification is defined as matching the illuminant of a scene to one of several illuminant models. In this work, five coarse illumination classes are proposed including *poor (too dark)*, *good (a bit dark)*, *great*, *good (a bit bright)* or *poor (too bright)*. Such a categorization will enable users to efficiently perceive current environmental conditions, and take action to improve results. Whereas past approaches are limited by requiring object segmentation prior to illumination classification [20], we propose a novel approach for real-time illumination classification that does not require scene segmentation.

First, an image is captured and converted to grayscale. The mean grayscale value is then used to classify the illuminant of the scene as one of the five aforementioned illuminant classes. Classification is accomplished through Bayesian classification wherein the mean grayscale value is classified as class i that has the maximum posterior probability. The class-conditional density of each illuminant class is assumed to be Gaussian, and the mean and variance is estimated from the sample mean and sample variance of the training data. Here, training data may be labeled images of outdoor and indoor environments with varying illumination conditions. Prior probabilities reflect how often each illuminant class occurs in the training data.

Another important environmental factor for vision-based wearable assistive devices is motion blur caused by user movement and/or object movement. Motion blur in vision-based assistive technology is an important issue since wearable systems are mobile and hence captured images are subject to extreme motion blur. Motion blur can have a significant effect on the accuracy of computer vision algorithms, especially those that rely on edge information as motion blur blurs edges, making edge detection and extraction very difficult. Hence, there is a need to develop reliability measures for motion blur wherein the motion blur as estimated from an image is classified as one of several degrees. In this work, four classes of motion blur are proposed including *no motion blur*, *small motion blur*, *large motion blur* and *extreme motion blur*.

One approach for motion blur classification is through the use of a no-reference, perceptual blur metric developed by Marziliano et al. [21]. This algorithm measures edge spread through edge detection followed by computing the average edge width, where edge width is defined as the local extrema locations closest to the edge. In the context of feedback mechanisms, it is important to know why reliability has been penalized as this communicates to the user what must be done to improve reliability. Unfortunately, the approach of [21] is sensitive to any type of blur, and thus is not useful in this context. The approach proposed here is a real-time algorithm for classifying the amount of motion blur contained in an image, which requires no reference image and is sensitive to only motion blur. This algorithm aims to classify the overall motion blur present in an image, whether this blur is caused by user and/or object movement. Further, an indirect approach is taken in that it doesn't directly classify motion blur, but overall movement, which is a good indicator of the amount of motion blur in an image for low-cost, off-the-shelf camera equipment for wearable systems. The algorithm is described next.

First, two frames are sequentially grabbed from the video stream, and the difference image of these two images is computed by subtracting them. If a pixel value remains the same between the previous and current frames, it will have a value of zero in the difference image; otherwise, it will have a value greater than zero. A binary threshold is then performed on the difference image, and the image is scanned both horizontally and vertically to detect vertical and horizontal lines, respectively. (An adaptive threshold is not used as the difference image is being worked with directly, and hence, it is not required.) The average width of vertical and horizontal lines is computed. A line is vertical if its height is more than its width, and a line is horizontal if its width is more than its height. The greater of these two averages is then classified as one of four motion blur classes previously described. The range of average line widths is divided into four sub-ranges representing the four classes. These sub-ranges are determined though experimentation such that motion classifications predict the corresponding motion blur levels.

Often the validity of system decisions can be furthered assessed by taking into account context. Recognition algorithms are often greatly assisted by context, but here, the goal is to verify algorithmic output using context. A few examples of features that may be used as context include the following: an object's material, shape, size, texture, or color profile, objects nearby the object of

interest, background or surrounding locations, etc. For clarity, consider the following example. The user has asked the system to determine what object lies on the table in front of him or her. The system responds and informs the user that the object is a bowl. However, based on a contextual cue that the object is made of cloth, this is not very likely, and hence the system reports a reliability of, e.g., 50%. Of course for this approach to work successfully, the step of context recognition, in this example, material recognition, would have to be achieved with high reliability. In the current work, the focus is on environmental factors rather than contextual factors. However, future work will focus on contextual factors as a powerful aid for reliable and usable wearable assistive devices.

As previously discussed, each reliability measure, e.g., illumination, motion blur, etc., issues a penalty or reward, based on how factors affect the performance of recognition algorithms, and ultimately, the final output is a reliability value that represents the reliability of the system-made decision. Conveyed to the user, either through audio or haptic output, is the system-made decision, along with the reliability of that decision. In understanding a system's reliability and causes of varying reliability levels, users may act upon the environment to help improve reliability, thereby improving system usability. That is, the user and system collaborate to solve challenging problems that the system may not be able to solve alone. For many environmental variations such as motion blur, strategies for improving reliability are straightforward. For example, to reduce motion blur, the user should reduce his or her movement. However, for other reliability measures, the best strategy for improving reliability depends on the environment and users' preferences. For example, illumination conditions can change dramatically between locations. Users may know of specific locations where the lights are usually turned off and need to be switched on upon entry, e.g., a workplace environment or home; or, locations that have poor lighting, such as a parking garage, in which case users may prefer to use a wearable illumination source to improve lighting.

4. Experimental Methodology

A set of 35 objects (14 bowls, 14 cups and 7 irregular objects made from LEGO® building blocks) were collected. Objects varied in shape, size, texture and material. Data capture sessions were held were ten participants each haptically explored all objects in the set to classify overall haptic features and haptic features of each region, as outlined in Section 3.1. The majority vote was used to decide final classifications for each object feature. An XML database was built using the object format shown in Figure 4.

A wearable assistive device was built, which consisted of a portable laptop, RFID reader, wearable camera, USB keypad, and headphones and/or haptic glove. We used Intermec's IP4 portable RFID reader and their 700 series mobile computer. Passive RFID tags were used, each tuned to the material it was placed on to achieve maximum read range. An existing application on Windows CE for RFID sensing was provided by Intermec for our use. This application was modified to enable the portable reader to communicate with a server wirelessly through 802.11. The server, which resides on the portable laptop, provides (1) a means for the user to interact with the reader; (2) enables information to be gathered from the reader; (3) access to the XML database; and (4) enables content selection and reliability measurement. The following briefly describes the operation of our proposed system when RFID tags are read (ignoring content selection and reliability measurement for the moment).

When RFID tags are read by the reader (the reader is set up to continuously read), unique tag ID's are stored in a resizable array. This array is periodically checked by the server for new entries. When a new tag ID is found, the server grabs the ID, looks it up in the XML database, retrieves the object's information and conveys it to the user. The user interacts with the system via a USB keypad; the user can specify audio or haptic output, specify which object features are to be conveyed, turn reliability measures on or off, etc. A wearable camera is worn around the user's neck, and in the current implementation, has two purposes: content selection and reliability measurement, described next.

The proposed framework for content selection using color information was implemented using Intel's Open Source Computer Vision library. When a number of tags are read at once, the wearable camera captures a 320x240 image of the scene, and this image is subsequently analyzed to determine which tagged object is in the image through matching color profiles. For preliminary testing, we used a blue background to aid object segmentation. Training data consisted of manually segmented color images taken from the COREL color image database. Our color categories, and respective pixel counts for training, included white (1600), gray (300), black (1072), red (1100), light red (400), dark red (400), green (600), light green (100), dark green (300), blue (1100), light blue (400), dark blue (400), orange (500), purple (600), yellow (500) and brown (600). Priors and class-conditional probabilities were estimated and used in (1). We assumed Gaussian densities, and used MLE to estimate the means and covariance matrices. Through experimentation, we estimated parameters for (2). We found a neighborhood size of 3 and a clique potential of 0.1 to work well. Our stopping criterion was when the number of classification updates is below a threshold, which is recommended by [18]; we found a threshold of 1000 to work well. As a preprocessing step, noise is reduced using a median filter before point-based classification. Some segmentation results are shown in Figure 5.

Algorithms for illumination and motion blur classification were implemented and integrated into our system. Users now had the option of using reliability measures, which have been shown to make wearable assistive technology more reliable in real-world environments [10]. When the system is used to remotely perceive objects from a distance, after object features have been communicated to the user, the reliability of these system-made decisions, as well as environmental variation classifications, are conveyed to the user. Motion blur categories *no motion* and *small motion* did not penalize reliability, but *large motion* and *extreme motion* each generated a penalty of 30%. Both poor illumination categories generated a penalty of 20%, and both good illumination categories generated a penalty of 10%. Illumination category *great* did not penalize reliability. In our system, the initial reliability begins at 100%. The reliability measure based on illumination was

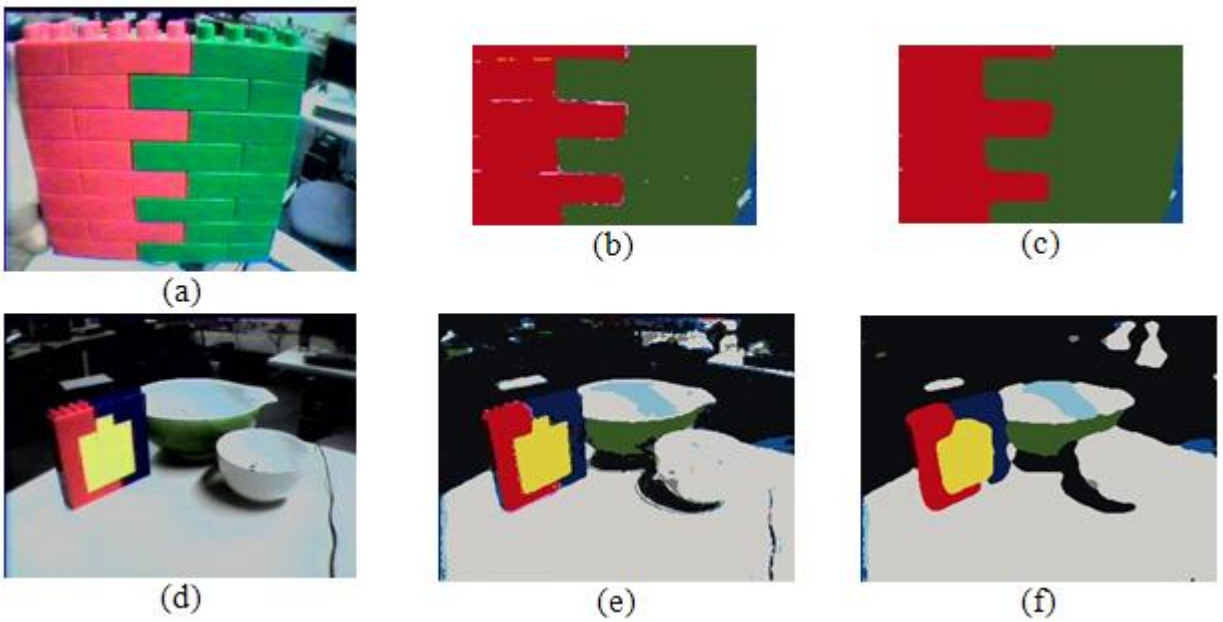

Figure 5. Segmentation results: (a) and (d) are the original images, (b) and (e) are point-based classified images and (c) and (f) are the final, segmented images.

trained using 1200 images of indoor and outdoor environments, captured from the wearable system. Each image was manually labeled as one of the five illuminant classes. Ten-fold cross-validation was used to evaluate the algorithm, which provided a classification accuracy of 96%. To train the reliability measure based on motion blur, we manually adjusted the decision boundaries until the four motion levels corresponded with the four motion blur classes. To test the algorithm, we collected 400 image pairs from video recorded from the wearable system. These image pairs were manually labeled, and then classified by our algorithm with an accuracy of 95%.

Formal usability testing will be performed as part of future work. Up until this point, we have conducted some preliminary usability tests, which have revealed the usefulness and usability of our wearable assistive device for remote object perception.

5. Conclusion and Future Work

In this paper, we've presented a conceptual framework wherein RFID and computer vision are integrated for the task of remote object perception in a wearable system for individuals who are blind or visually impaired. Computer vision aids RFID in that it enables content selection to prevent information overload, and enables the system to be used in untagged environments. Moreover, we've presented a framework to integrate reliability measures into vision-based wearable assistive technology to help make systems that utilize computer vision more reliable in real-world environments. This is accomplished by providing a collaborative framework in which the user and system work together to help solve challenging problems that alone the system cannot solve.

As part of future work, extensive usability testing will be conducted to ensure that (1) the system is useful to individuals who are blind; (2) it is usable in real-world environments; and (3) it is easy to use. Comments and feedback will be taken into account to improve the design of the system. SIFT features will be evaluated for their use in tag-assisted visual object recognition. SIFT features are desirable as they are robust to many environmental variations including illumination changes, scale changes, and to some extent, pose changes. Moreover, contextual factors will be investigated as a powerful measure of reliability for assistive technology. And finally, a vision-based learning module will be proposed and built as part of future work; visio-haptic transfer algorithms for haptic shape, size, texture and material will be implemented to simulate intermodal transfer from vision to touch.

6. References

[1] Stefano, L.D., and Mattoccia, S. 2002. Real-Time stereo within the VIDET project. In Proceedings of Real-Time Imaging, 8, 439-453.

[2] Kahol, K. 2006. Distal object perception through haptic user interfaces. Doctoral Thesis. Arizona State University.

[3] Luo, A., Zhang, X.F., Tao, W., and Burkhardt, H. 1999. Recognition of artificial 3-D landmarks from depths and color: a first prototype of electronic glasses for blind people. In Proceedings of SCIA99.

[4] Hub, A., Diepstraten, J., and Ertl, T. 2004. Design and development of an indoor navigation and object identification system for the blind. In Proceedings of ACM Sigaccess Accessibility and Computing, 77-78, 147-152.

[5] Iannizzotto, G., Costanzo, C., Lanzafame, P., and La Rosa, F. 2005. Badge3D for visually impaired. In Proceedings of IEEE

Computer Society Conference on Computer Vision and Pattern Recognition, 3, 29-36.

[6] Rekimoto, J. and Ayatsuka, Y. 2000. CyberCode: designing augmented reality environments with visual tags. In Proceedings of Designing Augmented Reality Environments (DARE 2000).

[7] Kulyukin, V., Gharpure, C., and Nicholson, J. 2005. RoboCart: toward robot-assisted navigation of grocery stores by the visually impaired. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2845-2850.

[8] Willis, S. and Helal, S. 2005. RFID information grid for blind navigation and wayfinding. In Proceedings of IEEE International Symposium on Wearable Computers, 34-37.

[9] Dana, K.J., van Ginneken, B., Nayar, S.K., and Koenderink, J.J. 1997. Reflectance and texture of real-world surfaces. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 151.

[10] McDaniel, T., Kahol, K. and Panchanathan, S. 2007. An interactive wearable assistive device for individuals who are blind for color perception. In Proceedings of HCI International, 751-760.

[11] Mae, Y., Umetani, T., Arai, T. and Inoue, K. 2000. Object recognition using appearance models accumulated into environment. In Proceedings of International Conference on Pattern Recognition, 4, 845-848.

[12] Boukraa, M., and Ando, S. 2002. A computer vision system for knowledge-based 3D scene analysis using radio-frequency tags. In Proceedings of International Conference on Multimedia and Expo, 2, 245-248.

[13] Takemura, K., Ohara, K., Ohba, K., Chong, N.Y., Hirai, S., and Tanie, K. 2004. Knowledge distributed tag-based vision system. In Proceedings of the 1st International Workshop on Networked Sensing Systems.

[14] Hontani, H., Baba, K., Kugimiya, T., Sato, K., and Nakagawa, M. 2003. Visual tracking system using an ID-tag and the network. In Proceedings of SICE 2003 Annual Conference, 3, 2375-2380.

[15] Chong, N.Y., Hongu, H., Miyazaki, M., Takemura, K., Ohara, K., Ohba, K., Hirai, S., and Tanie, K. 2004. Robots on self-organizing knowledge networks. In Proceedings of International Conference on Robotics and Automation, 4, 3494-3499.

[16] Kim, J.Y., Im, C.J., Lee, S.W., and Lee, H.G. 2005. Object recognition using smart tags and stereo vision system on pan-tilt mechanism. In Proceedings of ICCAS2005.

[17] McDaniel, T.L., Kahol, K., Tripathi, P., Smith, D.P., Bratton, L., Atreya, R., and Panchanathan, S. 2005. A methodology to establish ground truth for computer vision algorithms to estimate haptic features from visual images. In Proceedings of IEEE International Workshop on Haptic Audio Visual Environments and their Applications, 95-100.

[18] Abdel-hakim, A.E., and Farag, A.A. 2005. Color segmentation using an Eigen color representation. In Proceedings of 8th International Conference on Information Fusion, 25-29.

[19] Besag, J. 1986. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society, 48, 3, 259-302.

[20] Hel-or, H.Z. and Wandell, B.A. Object-based illumination classification. Pattern Recognition, 35, 1723-1732.

[21] Marziliano, P., Dufaux, F., Winkler, S., and Ebrahimi, T. 2004. Perceptual blur and ringing metrics: application to JPEG2000. Signal Processing: Image Communication, 19, 163-172.