Beyond RFID: Tagging Integrated with Image-based Object Tracking

Stanley Yuen, Ph.D., Patrick Chen, Michael P. Lenaghan & Thomas Tsao, Ph.D.
CompuSensor Technology Corp.
352-C Christopher Ave., Gaithersburg, MD 20879
syuen@cstcorp.com, cchen@cstcorp.com, mlenaghan@cstcorp.com ttsao@cstcorp.com

I. Abstract

In this paper we describe the synergistic integration of RFID tagging with image-based object tracking technology. Coupling these two technologies will have a profound impact on the competitiveness and efficiency of the retail and consumer goods industries. The adoption of RFID technology is inevitable and its enormous transformational promise is clear. We will explore the possibilities of enhancing RFID with visual validation.

The relevance of object detection and tracking, also known as Automatic Target Recognition (ATR) and Automatic Target Tracking (ATT) within the defense industry, to the ongoing RFID revolution has not been touched on in the open literature, until now. Specifically, a scenario in connection with loss prevention and anti-sabotage and another scenario focusing on enhancing consumers-retailers interaction are described. Important issues and the latest advancements of ATR / ATT, along with concepts on how these capabilities can be coupled with an RFID infrastructure, are mentioned. The above ideas and anticipated impacts are applicable to a number of other industries and commercial sectors as well.

II. Introduction

Radio frequency identification (RFID) technology has been around for decades, and is already used in a number of applications, from hazardous waste management, livestock tracking and feeding, electronic toll collection, fleet management, inventory management and logistics, etc. These applications are generally vertical, industry and task specific, each using its own standard. For Example: RFID tags have been imbedded into Compact Disks (CDs) packaging for years, yet the task and make up is specifically for Loss Prevention at Point-of-Sale Retail locations.

<u>RETAIL / CONSUMER INDUSTRIES</u>

Wal-Mart, DoD and Target, and many others, are currently mandating RFID compliance. It is no longer a hype that the first decade of the 21st century may see intelligent devices being embedded in essentially all products of all sorts in the retail industry such as consumer goods and clothing. These electronically ID-tagged and "smart" objects and products will increasingly be networked via the wireless Internet.

Another crucial milestone in speeding up the RFID revolution is the establishment of the Electronic Product Code (EPC), which is the emerging standard for RFID applications in the retail supply chain. It represents an industry consensus on the best technological approach to successful implementation of RFID. The overall EPC concept is designed to work in a range of applications – from pallet and carton tracking in the “backroom” or “cross dock”, to tagging individual items in the “selling floor.” EPC emerged from the Auto-ID Center, a partnership between almost 100 companies and five of the world’s leading research universities including the Massachusetts Institute of Technology.

While the RFID revolution is increasingly gaining momentum, another important development is that widely deployed wireless networks have made it possible to transmit images and video sequences to and from mobile devices. Images and video sequences can now be captured, displayed, and transmitted using mobile telephones or personal digital assistants (PDAs). It is clear that the “next big wave” is already underway, propelled by the integration of the Internet, sophisticated functionalities of cell phones and PDAs, the constant expansion of wireless broadband networks, and “the RFID train” *which has already left the station*. But there is an inconspicuously missing ingredient in the making of this new technological revolution. It is **Computer Vision**.

COMPUTER VISION

Computer Vision can be viewed as a branch of artificial intelligence and image processing. It typically requires a combination of *low-level* image processing to enhance image quality (e.g., removing noise, increasing contrast, etc.) and *higher-level* pattern recognition and image understanding to recognize features present in the image.

It is an enormously challenging research topic; any major breakthrough would require an accurate modeling at the detailed computational level of certain aspects of human vision. Computational investigation into the human representation and processing of visual information has been pursued over the past few decades. Due to the multi-disciplinary nature and the general lack of effective mathematics models for analyzing the human nervous system, overall progress had been slow.

The good news is that some of the abstract mathematics behind human visual perception that is consistent with various neuro-physiological findings have been established over the past few years [1-4].

In parallel with the R&D efforts on Computer Vision within the academic circle and commercial sector, the defense industry over the past 30 years or so has actively been pursuing Automatic Target Recognition (ATR) and Automatic Target Tracking (ATT), which is concerned with the detection, tracking and recognition of small targets using input data and images. Hence the differences in definition among Computer Vision, Image Understanding, and ATR / ATT are quite blurry. To simplify the terminology in our discussion, the term ATR / ATT is used for the rest of the paper since it is more descriptive.

This paper is atypical in the sense that it bridges the gap between RFID related system concepts and the latest breakthroughs in ATR / ATT. We are motivated by the following considerations:

1) From a consumer's perspective, he / she is increasingly relying on his / her cell-phone with a built-in camera to capture images at any time and from anywhere for a wide spectrum of purposes and reasons. Naturally there will be a strong need to carry out accurate and robust *image matching*. From a consumer's standpoint, transmitting an image of a product of interest will increasingly be a routine. On the other hand, the transmitted image somehow needs to be matched against related images stored in various retailers' databases. Given this unstoppable trend, accurate image matching is an important task in the consumer-retailer interaction loop, and the consequence of that is that RFID tagging and image matching will be tightly coupled in the electronic buying process. This is further elaborated via Scenario A in Section III. (The linkage between image matching and ATR/ATT is explained in Section IV.)

2) The demand for video monitoring systems using data received by RFID readers as positional information is undoubtedly on the rise. For instance, by inserting an RFID tag to the nametag of a kindergarten child and with several cameras installed in the kindergarten, we can automatically select images from particular cameras based on RFID tags that identify individual persons or objects. To put it another way, such a system can automatically switch the transmitted images according to the child's movements so that the parents can watch their own child from home or the office. This basic concept can be extended to include even more sophisticated applications involving precision tracking of objects and persons. And this precision tracking capability can be incorporated based on the latest advances in ATR / ATT. It is directly aimed at loss prevention and anti-sabotage measures in a selling floor environment. This is further elaborated via Scenario B in Section III.

3) The defense industry has invested billions of dollars on ATR / ATT related R&D over the past few decades. US Department of Defense is in fact strongly encouraging commercialization of these R&D results, e.g., through the Small Business Innovation Research (SBIR) Commercialization Assistance Program of the US Navy [1]. Inserting these sophisticated and cost-effective ATR / ATT capabilities into the future RFID infrastructure of the retail and consumer industries and other sectors will take place. It is therefore constructive and educational to inform the business sectors as well as taxpayers of what can be expected in terms of commercialization opportunities.

4) Scenarios A & B mentioned above reflect new business and technology development opportunities within the retail industry that is undergoing the RFID transformation. Given the ongoing momentum associated with the RFID infrastructure rollout and the impetus to be generated by the marriage of RFID and ATR / ATT technologies, this paper can provide

insights into the distinct possibility of *commercial-to-military conversion* down the road.

The paper is organized as follows. In Section III, the two aforementioned Scenarios A & B are graphically shown to establish a clear mental picture, emphasizing the essence and key points involved. Due to the multi- and inter-disciplinary nature of this article, background information of and the interrelationships among GPS, RFID, sensor technologies, and the challenges surrounding ATR / ATT R&D are summarized in Section IV so as to make this paper self-contained. In addition, important scientific and engineering issues involved in ATR / ATT R&D and their relevance to the two specific scenarios are also briefly explained with the help of simple diagrams. A summary of the key issues covered in this article as well as some closing remarks are then given in Section V.

III. Two Application Scenarios

Scenario A: enhancing interaction efficiency between consumers and retailers

According to a forecast conducted by the market research firm IDC, 800 million camera-phones will be in use worldwide by 2007. Images are very inexpensive to send. Each new cell phone currently being sold in Japan has a built-in camera. Obviously this trend has important implications in terms of creation of new business opportunities in the retail industry as a whole. A simple example of shopping-with-an-image is illustrated below:

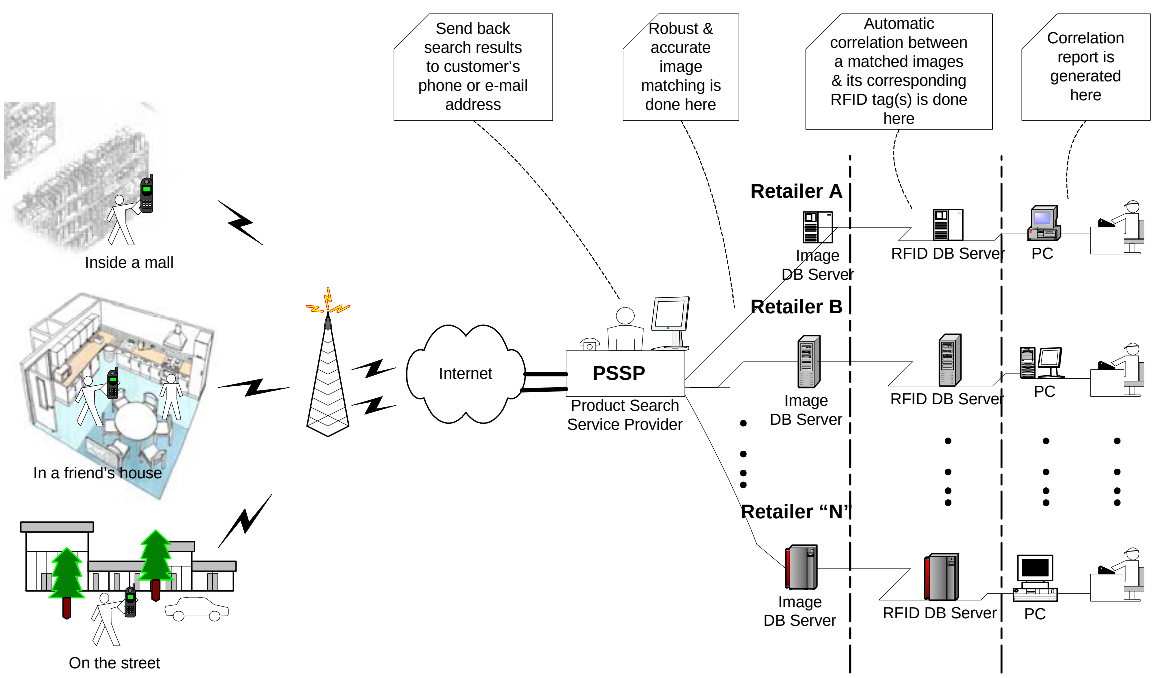

Figure 1

Figure 1 is self-explanatory but the following points are shared to emphasize a few key comments:

1) Besides Internet Service Providers (ISPs) and Contents Providers (CPs), it is highly likely that another category of service providers designated as Product Search Service Providers (PSSPs) above will appear in the retail sector.

2) In order to make the above scheme feasible, some sort of small service fee may have to be paid to the PSSP by the image-sender using a cell-phone. In addition, the PSSP may have to split this small fee with the retailers as an incentive for them to cooperate and provide support.

3) Upon receiving the search results from the PSSP, the image-sender (i.e., the potential buyer) has the options of not taking any action, purchasing the product directly from a store of his/her liking, or going through the same PSSP to place in the order.

4) It would be an efficient and competitive process to all parties concerned if the order is eventually placed through the same PSSP since the PSSP has all the relevant information and knowledge about the potential transaction. In this case, the consumer would save a lot of time, and the PSSP would also receive certain amount of commission from the retailer to facilitate the subsequent delivery of the product. All in all, it is a "win-win-win" situation.

Scenario B: loss prevention, anti-sabotage and tracking hackers in real time

The benefits and productivity improvements associated with establishing an RFID infrastructure is clear. It is also clear that this type of infrastructure could be abused by hackers and tech-savvy shoplifters. Such a system could allow thieves to fool retailers by altering the identity of goods. When RFID reader hardware and software for reprogramming radio tags become widely available, hackers could use a homemade handheld device and the associated software to mark expensive goods as cheaper items and walk out through the self-checkout exit. Furthermore, pranksters could create confusion and chaos by randomly swapping tags.

Does your RFID infrastructure have loss prevention? What prevents disabling, changing, decoding, switching, unbundling, duplicating, demagnetizing, or defrauding a device or reader? The potential for illegal tag intervention a threat to the system. **We call this RACKING** or **R**fid h**ACKING**.

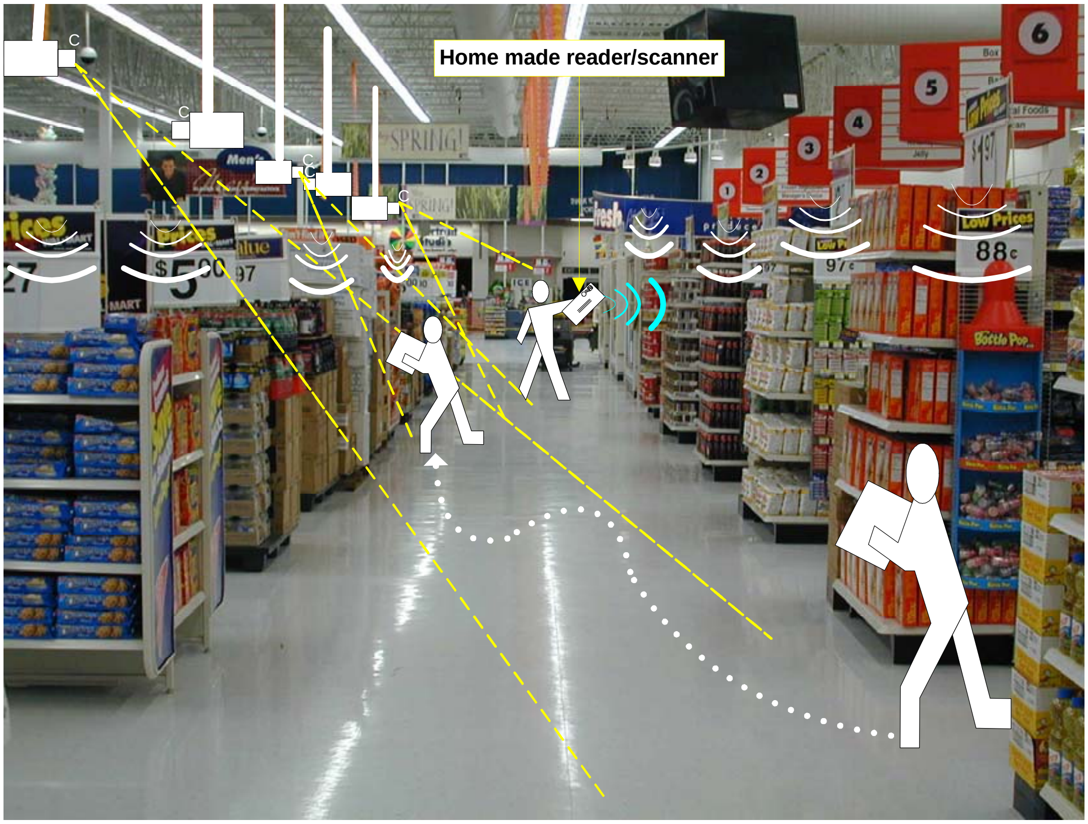

Figure 2

Over the past few years, a number of R&D firms and organizations have proposed to develop an automated security surveillance system that combines closed circuit video cameras, RFID technology, and computer modeling of human behaviors, with the goal of achieving rapid detection of suspicious events warranting the attention of security personnel. The ultimate objective is to develop search and retrieval tools for rapid compilation and presentation of relevant video clips.

This type of top-down approach of combining sensor technologies and developing computer reasoning modules, while ambiguous and logical, often totally overlooks the crucial challenges and subtle issues surrounding the detection of events of interest as well as the required tracking of humans and objects in a clustered environment. As discussed in the next section, achieving such kind of noble goals involves "*true* artificial intelligence" and "true AI" requires in-depth knowledge in perceptual psychology, neuro-physiology, advanced mathematics, so on and so forth. Without resolving the fundamental and crucial issues of automatic image matching, object segmentation (from its background), tracking of partially occluded humans and objects, etc., the conventional top-down approach could only lead to *ad hoc* solutions at best, which means that there is no solid scientific foundation for a sustainable development.

COMPUSENSOR TECHNOLOGY

CompuSensor has been pursuing advanced ATR / ATT R&D mainly for defense related applications over the past ten years. In particular, the core technology it has developed is based on accurately modeling biological vision using advanced mathematics and knowledge from perceptual psychology, image processing and parallel computing. And these R&D outputs are consistent with the findings from neuro-physiological research on the visual cortex of the brain.

Given this unique background, CompuSensor is in a unique leadership position to transfer these ATR / ATT related knowledge, experience and technical know-how to the upcoming technological transformation of the retail industry involving RFID tagging. Next section provides additional information on the background of CompuSensor's technology development and other relevant issues.

IV. Interrelated Concepts and Important Technical Issues

POSITIONAL INFORMATION

It is common knowledge that the Global Positioning System (GPS) provides positional information and is used in a wide spectrum of navigation applications. RFID technology, on the other hand, is generally not viewed as something related to the extraction of positional data; is short ranged, limited and specific to the product. However, since we normally would know where the RFID tag readers are in a given application, the position of a particular tagged object could therefore be approximated in a dynamic situation where the tagged object moves around in an environment covered by a distributed set of tag readers. In this sense, RFID technology is linked to the concept of localization of objects, at least in a coarse sense.

Currently passive RFID tags can only be used in close proximity of within 1 meter to the reader, while active RFID tags (i.e., with an internal battery) can be used within about 10 meters in diameter. These distance limits will be improved in the near future and are tied to the technological advancements of readers. Another reality is that GPS cannot be used indoors. Consistent with the above idea and principle, a few companies are now developing video monitoring systems using positional information extracted from an existing infrastructure of RFID tags and readers.

TARGET TRACKING

We can go one step further from the above-mentioned video monitoring scenario that is linked with RFID tagging. Retailers and other end users in the near future will demand for more sophisticated systems that can perform image-based tracking of objects, as well as shoplifting suspects and vandals who intend to sabotage an existing RFID infrastructure. (This is graphically shown as Scenario B in the previous section.) It should be emphasized that the goal of performing precision tracking of objects and humans using simply video sequences from cameras is a great technical challenge. This is especially true for situations involving moving cameras or *image handoff* from one camera to the next. Image-based precision tracking has been an important component of defense R&D over the past three decades. CompuSensor has been actively pursuing this type of R&D for a wide range of defense and security applications since 1993, and has established a few critical milestones [2,3] in this field and accordingly received a number of professional recognitions [4-7]. Commercialization of its revolutionary tracking and missile homing technology is also supported and encouraged by the Navy [1]. To provide a sense of realism and the type of research efforts involved, a few examples on image-based precision tracking/homing are illustrated below.

Figure 3 (below) shows the situation in which *a particular pixel point* on a tank is being homed-in during the terminal homing phase. This example is associated with the Navy-sponsored commercialization initiative in 2003 [1].

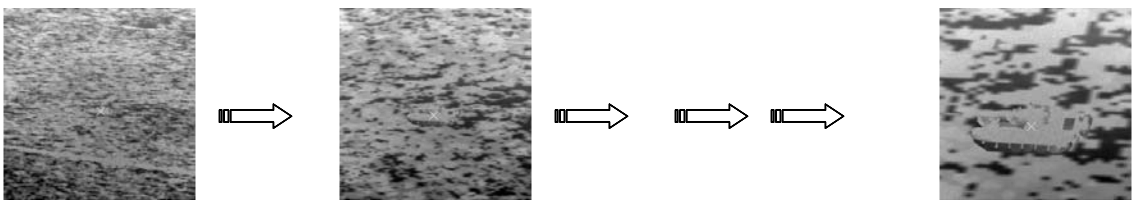

Figure 3: In a SBIR contract funded by the Army Missile Lab., simulated anti-tank infrared image sequences provided by the U.S. Army were used to test missile homing accuracy. Each sequence has more than 1600 consecutive image frames. CompuSensor's biologically inspired tracking solution was successful in tracking any given pixel point on the target throughout the entire 1600-plus frames. This long-duration precision tracking capability involving *sub-pixel accuracy* from frame-to-frame is shown here. This established an important milestone for the DoD in terms of developing a robust tracking/homing capability in an environment of strong background clutter. CompuSensor's unique capability is that it is the first and perhaps the only company that has demonstrated such a capability in such an environment.

It is appropriate at this point to clarify the relationship between image matching and ATR / ATT research. This would also bring out the multi-disciplinary nature of ATR / ATT. As can be seen in Figure 3 above, the object being homed-in in this case is a tank and its appearance changes from frame to frame. This frame-to-frame image alteration can be modeled accurately using a mathematical concept known as *affine transformation*. In essence, an affine transformation corresponds to a *6-dimensional change* in a geometric sense, involving translation, rotation, scaling and shear as shown in **Figure 4** below:

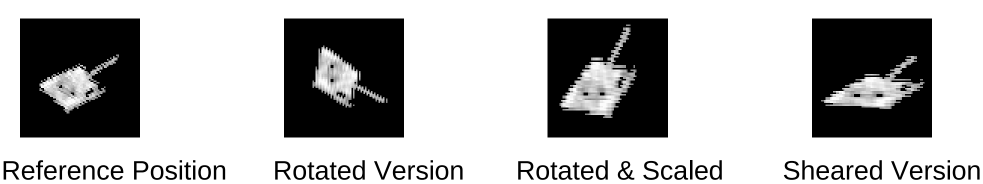

Figure 4: concept of an affine transformation

To relate image matching with ATR / ATT, all we need to realize is that any frame to frame distortion must be computed and compensated if we were to track an object in an image sequence successfully. Hence image matching, also known as Image Registration, is a core component in ATR / ATT research.

Again, matching an image with its affine-transformed version involves a 6-D match and this has been an enormous challenge to engineers and scientists in many technical fields. The computational difficulty involved can be appreciated as follows: if there are10 discrete possibilities (i.e., "positions") for each of the 6 dimensions, the total number of matches one has to go through would be 10^6. If there are 100 discrete "positions" per dimension, the total number of required matches would be 10^{12} !!!   On the other hand, if we are interested in tracking the movement of *any particular pixel* associated with a target or object, we must be resolve the above *computational nightmare*.

PHYSICAL VISION

Fortunately, we know from our own experiences that our vision system is extremely efficient in dealing with affine transformations, which is analogous to carrying out a set of "mental alignments." In our everyday experience, we continuously perform "mental alignments" and they can in fact be done instantaneously. Interestingly enough, accurate and robust image matching is directly related to the mathematics behind visual perception and the concept known as Psychological Constancy. That is, objects in the visual field are recognized as *what they are* no matter how their appearance may be distorted by viewing conditions: near or far, right or left, rotated or not, etc. This is termed the *correspondence* problem in psychology, and is equivalent to the alignment task in image matching.   CompuSensor determined over 10 years ago that in order to establish a solid scientific footing in our ATR/ATT R&D efforts, one must first understand the underlying *bio-mathematics* associated with biological vision. Consequently it took CompuSensor a decade of efforts in formulating and validating the underlying bio-mathematics [4].   It is clear now that the seemingly simple subject of image matching is in fact directly tied to the monumental research efforts in perceptual psychology, neuro-physiology, advanced mathematics, and defense-related ATR/ATT.

Figures 5 and 6 illustrate CompuSensor's current activities in the area of tracking multiple moving objects from an Unmanned Aerial Vehicle (UAV) platform under very challenging conditions, including poor video quality, frame-to-frame instability, large shadowing effects, significant occlusion of objects being tracked, etc. It should be emphasized again that CompuSensor's proprietary image tracking technology is inspired by and consistent with human visual perception.

Figure 5: tracking multiple targets simultaneously from an UAV platform.

Figure 6: tracking partially occluded objects under shadowing and other difficult conditions (also from an UAV platform).

PASSIVE VS ACTIVE SENSING

A few words should be mentioned about passive vs. active sensors. There are many types of sensors and sensor technology. Within the context of commercial applications using ATR/ATT or Computer Vision techniques, we mainly focus on the use of regular cameras and infrared sensors. These are known as *passive* sensors since they do not transmit any electromagnetic energy and subsequently process the return echoes. Using an *active* sensor gives us the advantage of being able to better predict, manage and process the return echoes. This “extra luxury” of course comes with a price and that is the cost of the sensor. Imagine the difference in price between a regular digital camera versus an advanced laser or radar capable of transmitting electromagnetic pulses with high precision.

At this point, we should point out the key reason behind the enormous challenge of conducting image-based (i.e., “passive”) ATR/ATT research. Since the main thrust of ATR/ATT research is to intelligently process whatever images the sensor “sees,” the most effective and efficient processing solution naturally correspond to the computational mechanism between the eyes’ retinas and the part of the brain known as the visual cortex. This is “true AI” which requires in-depth understanding and inter-disciplinary knowledge from perceptual psychology, neuro-physiology, advanced applied mathematics, and so forth. Being the most complex piece of machinery in this planet, the human brain obviously employs a lot of structured, abstract and extremely sophisticated mathematics, and understanding its processing capabilities at the mathematical level is still at an infancy stage. Even a concise summary of the ongoing efforts by a few dozens of the most prominent research institutions engaging in visual perception research is way beyond the scope of this paper. Interested readers are referred to [3,4] for the background information on CompuSensor’s breakthrough on biologically-inspired ATR/ATT works relevant to this paper.

V. Summary

In closing, while an RFID tag contains a lot of useful data about a given product, the actual *image* of that product is still a necessary piece of information that must be seen and available to the consumer or potential buyer. Hence a proper *marriage* between these two pieces of information is a must if we were to enhance, enlarge and speed up the interactive consumers-retailers business process. The two scenarios illustrated in Section III reflect many key points on how to create additional business through the shopping-with-an-image arrangement and improve security for the retail industry.

While emphasizing the importance of this marriage, this paper also points out the scientific and multi-disciplinary challenges, as well as some recent breakthroughs, in the field of Automatic Target Recognition and Automatic Target Tracking which has been heavily invested on by the defense industry. The future RFID infrastructure rollout in the retail industry can and should take advantage of this huge investment by the government.

VI. References

[1] Navy Opportunity Forum (May, 2003); see www.dawnbreaker.com

[2] Thomas Tsao, Laveen Kanel, "Scene Registration Method based on a Dynamical Receptive Field Model of Biological Vision", in *Pattern Recognition in Practice VI*, Vlieland, The Netherlands, June 1999.

[3] Thomas Tsao, "Image-based Target Tracking through rapid sensor orientation change, in *Optical Engineering*, March 2002, Volume 41, Number 3. (Winner of 2003 SPIE Kingslake Medal & Prize).

[4] www.oemagazine.com/fromTheMagazine/sep03/spieworld.html

[5] Thomas Tsao, "Method and Apparatus for Determining Image Affine Flow Using Artificial Neural System with Simple Cells and Lie Germs" U.S. Patent No. 5535302, July 1996.

[6] Thomas Tsao, "Method and Apparatus for Determining Binocular Affine Disparity and Affine Invariant Distance Between Two Image Patterns" U.S. Patent No.5911035, June 1999.

[7] Thomas Tsao, Stanley Yuen, "Affine Transformation Analysis System and Method for Image Matching" U.S. Patent Application; Application No.: 10/377,649; Filing.

VII. Footnote

The concept of matching the two technologies was formulated by Mike Lenaghan of Michael P. Lenaghan LLC, an expert in the field of Supply Chain and Finance, in conjunction with Dr. Stanley Yuen and Dr. Thomas Tsao. This paper is collaboration between them and includes the valuable input of Patrick Chen.