Tag-Based Vision: Assisting 3D Scene Analysis with Radio-Frequency Tags

Mustapha Boukraa and Shigeru Ando
Department of Information Physics and Computing
Tokyo University,
Bunkyo-ku, Hongo 7-3-1
Tokyo 133-0031, Japan
moose@alab.t.u-tokyo.ac.jp

Abstract - *In image sensing and processing, ambiguities arise when only one source of information is used. Thus, 3D object recognition and localization is a difficult task when using intensity image as single input. This paper presents a machine vision system that uses a Radio Frequency (RF) Tag device to identify objects prior to locating them visually. The tag system consists of a tag reader that can interrogate, and receive radio signals from, tags attached on objects and characterizing them. Laying the basis of an object model database shared on a network, we perform a knowledge-based recognition task where the information retrieved from the database query serves as a prior knowledge. The recognition algorithm used is a matching with projective invariants. We describe how this system can be used for efficient object registration and how the concept of integrated tag based systems can provide new insights in image processing and machine vision.*

Keywords: machine vision, tags, model-based vision, sensor fusion.

1 Introduction

The problem of object recognition and localization in a 3D context is a fundamental task for autonomous vision systems. It is the preliminary step to autonomous navigation and object manipulation. 3D object recognition requires constructing and efficiently storing models of the 3D world, and matching features extracted from sensed information with these models [2,4]. Recognition based on 2D single intensity images is very hard without prior models or assumptions that make up for the lack of depth information in monocular vision[2,5,8,14]. Such additional information can be provided by an auxiliary sensing channel, through which the system can improve its understanding of the surrounding world [7,10].

Now, an increasing amount of information is available through the Internet, and efforts are made to structure it for improved retrieval. Namely, the XML standard is a promising and appealing technology for sharing information organized according to its semantic content. Thanks to this nature, XML appears as a powerful tool when designing a system expected to dynamically acquire knowledge about an object or infer the best course of action to undertake in a given context.

Based on these observations, some researchers implemented systems where visual tags are attached on objects, allowing to characterize objects in a visual scene[9,12]. However, inconvenients of using visual tags are that, among others, line of sight is required, visual tags can wear off over time, they are not easily rewritable, they alter the aspect of the object, and they can be hard to recognize unambiguously themselves. Example of a visual tag of widespread use is provided by barcodes used to register merchandise goods. In this article we present the architecture of a vision system that uses RF tags that can be attached to or embedded in objects. Such tags contain a unique identifier, the Tag ID, and can be read by a tag interrogator coupled to the camera.

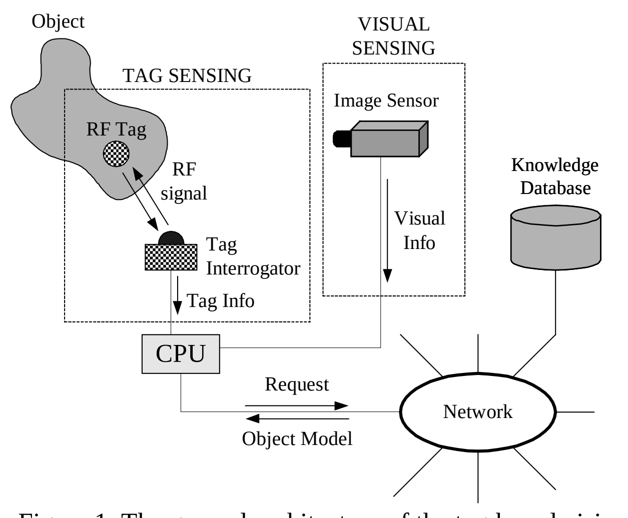


Figure 1. The general architecture of the tag-based vision system. The central CPU commands the RF tag identification system and retrieves information from the network, knowing the object's Tag ID. Tag sensing and visual sensing are then integrated to perform a high-level vision task.

The main technical drawback of using RF Tags instead of visual tags is that we don't have direct access to the object location. Fig.1 shows the architecture of a RF Tag-based vision system.

When an object equipped with a RF Tag enters the operational field of the RF receiver, the Tag Interrogator retrieves the object's unique Tag ID, and sends a request to the network to ask for the object model. This model is then used to perform a model-based vision task.

In the next section, we describe the principles that motivate our vision system, section 3 will briefly summarize the paradigm of RF tagging, and in section 4 we present a practical implementation for 3D scene analysis. In the final section we present some results obtained with our object recognition system still under construction. They suggest that a RF tag identification module can expand the capabilities of a machine vision system in brand new directions, providing ways to design more "intelligent" image analysis devices and systems.

2 Principles of a Tag-Based Vision System

2.1 Added value of using tags

The key point in using identifying RF tags is that, since the detection of the tag can be considered reliable, we have access through the network to a model representing the object at any different level of description that we choose. Thus, the system not only can retrieve a physical model of the object, but also can get a high level functional "understanding" of the object. In this sense, and in the perspective of information technology, RF tagging is much more than just an alternative to conventional barcode tagging.

Let us draw some key elements that constitute the added value of a tag based vision module compared to classical systems for visual analysis:

. **accuracy:** the preliminary tag identification process ensures the presence of the object which tag ID has been detected by the system. Therefore, the recognition step is posed as a rigorously constrained knowledge based matching of the database information with the visual information directly available.

. **complementary information:** the RF Tag can contain - or give a key to - information that cannot be retrieved only from the visual content. This includes physical and chemical properties (such as 3D configuration and real dimensions, weight, tactile info., acoustic info., odor, etc...) as well as higher level information about the function and purpose of the object. This property is essential in the design of autonomous systems or systems aimed at interacting with and providing output to humans.

. **capacity of evolution:** the system doesn't have to be trained with a predefined set of objects, but rather can acquire "awareness and knowledge" of new objects each time it performs a database query knowing the object's ID.

As a type of widely used identification tags, we can cite barcodes, for which some systems of visual recognition have been succesfully developed, as in [17]. However, in comparison to barcodes, RF Tags are more easily and unamibiguously detected, they do not require line of sight, they do not wear off over time, they can be embedded into objects to make them invisible, and they can be rewritable to update/modify their information.

2.2 RF Tags and Machine Vision

As possible applications of the concept of Tag-Based image analysis, we can think of systems that recognize objects/places/people in a tagged environment, and detect the degree of correspondence with the model, for inspection, diagnosis, automatic repairing, etc… when accuracy is needed, or when the objects encountered can be very versatile. In augmented reality applications, the system can display useful information about the detected objects. The possible applications of RF Tag –based vision systems are really infinite, all the more since RF Tags and the information in the database can be modified, allowing several systems to cooperate and update the information of the database according to the result of their observation/action. In a typical example, the system updates the information concerning the location of the object that is detected, and shares this information on the network. In the same open manner, one can think of unlimited "Tag-Based Applications", for such areas as augmented reality, virtual reality, entertainment, automated inspection and repairing of objects, and ultimately intelligent autonomous robots.

As a very basic illustration of this concept, and in order to present a practical implementation, we devise in this paper a recognition system for polyhedral objects from single view. We make use of some recent results derived from considerations of projective geometry [14].

3 RF Tag Identification

RF Tag Identification (RFID) consists in an architecture where information contained in a small tag is remotely read by a reference system, using radio frequencies. This technology represents a fast grwoing market and is constantly attracting more and more economic actors. In this section we will present a brief survey on RF tagging technologies and issues.

3.1 An overview of RFID

The principle of RFID was invented in 1948 (transmission of energy and data by inductive/propagative coupling), but it really developed as a business only about two decades ago[18]. According to market analysts Frost&Sullivan, it is expected to grow fastly and reach a US$2b global business by 2006. RFID was originally driven by applications concerning Electronic Article Surveillance (EAS) and security (checking luggage, granting access to facilities,…). In the last few years RFID has much diversified, and RFID manufacturers are striving to provide with an ever wider choice of hardware options.

RF Tags can be categorized according to the following parameters, among others[19]:

. **frequency range:** typical RFID systems use frequencies in the VHF, UHF and up to the micro-wave band. Much effort is made for practical implementations of the 125-135kHz range and the 13.56MHz and 2.45GHz frequencies. Higher frequencies naturally allow for higher bitrates.

. **active/passive:** whether the tag contains a power source or not. Active tags allow bigger operational range, whereas passive tags cost less and have longer lifetime.

. **chip/chipless:** whether there is an embedded integrated circuit (IC) in the tag. Microchips allow greater functionalities (R/W, on-tag processing).

. **conventional/low cost:** the industry is pushing towards the achievement of low cost tags that will overcome limitations due to the cost of integrating RF Tags.

Besides, RF Tags come in all types of shapes, ranging from tiny devices (Hitachi recently released the smallest chip RF Tag, the "µ-chip", that can be embedded in paper) to rigid rings of alloy or flexible laminates, etc….

3.2 The future of RF Tags

Industrial applications of RFID are concerned with gains in productivity security, and anti-counterfeiting. Thus, typical examples of developing applications are: assembly line monitoring, product storage and tracking, access control, EAS. Also, the European Central Bank is planning to put RF Tags in Euro banknotes by 2005.

On the other hand, some different and ambitious directions are being taken. Of particular interest is the MIT's vision of the "Internet of Things"[20] that seeks to create the global network of physical objects equipped with low cost RF Tags. For this purpose, they devise a standard for Tag IDs that they call Electronic Product Code (EPC).

As global standards are being developed by industrial partners and low cost RF Tags are achieved, RFID applications are due to be more and more popular.

4 Implementation

The chartflow of our application is presented in fig.(2). Our purpose is to realize an efficient integration of the information processing and image processing modules. Tag sensing is performed by the system described in 4.1. In the information processing stage, we design an object model database as in 4.2. The image processing stage is described in 4.3.

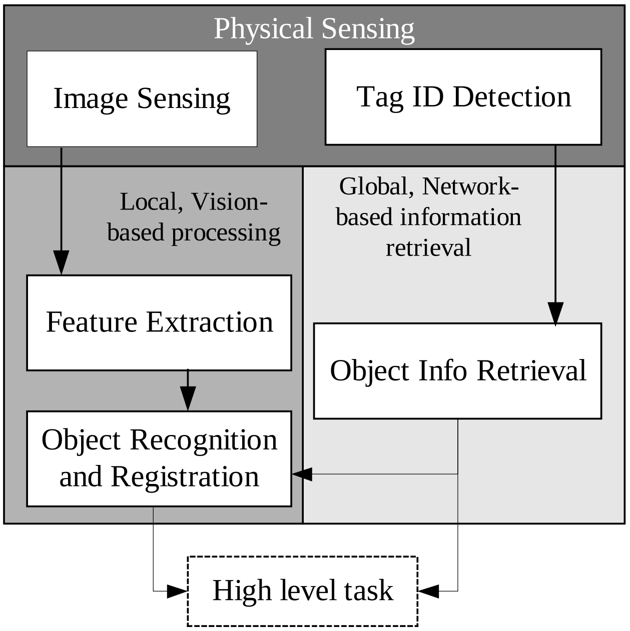


Figure 2. The chartflow of the system, splitted in 3 functional domains. The lower "High level task" is mentioned as an ultimate goal and is not tackled here.

4.1 The RF Tag system

The RF Tag system we use in practice consists of an industrial Tag Interrogator device and its corresponding smart card-shaped tags that contain 1024bits of rewritable data each, which allows storing an amount of object information directly in the Tag. The operational frequency is the 2.45GHz standard, for a bandwidth of 16kbs, and the detection range is limited to 1.2m. Our system belongs to the family of passive tags with microchips. Of interest is the fact that 2.45GHz radiations are not absorbed by water and the human body: the system we are using was originally designed to be used for granting access to people to facilities and the RF Tags were supposed to be worn as badges. Due to vector components of the RF field, the reliability of Tag detection is sensitive to the position and orientation of the tags.

For our purpose, each Tag contains an identifier made of 6 characters, much simpler than MIT's EPC and proprietary, corresponding to the primary key of the model entry in the model database.

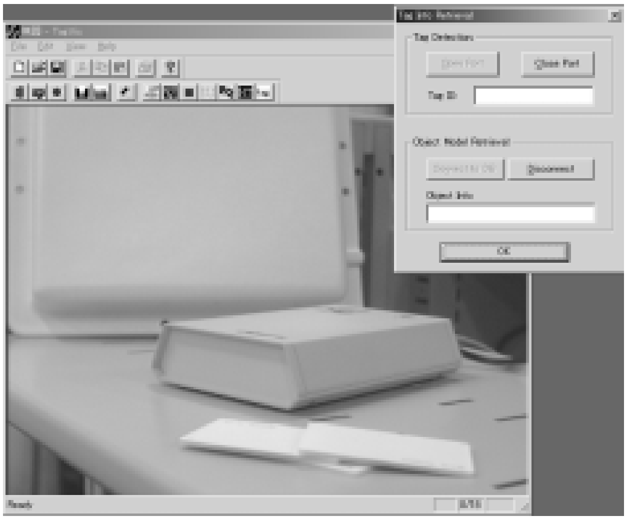


Figure 3. The main window of our image analysis application, showing the interface for Tag detection/object info retrieval, and the video capture window, where the Tag Interrogator (central box) and some card-shaped Tags are visible in the foreground, and the RF antenna in the background.

4.2 The object model database

For an automated system to efficiently analyse an object and dynamically undertake an « intelligent » course of action, it is very useful to store information according to its content. This is why we choose to build the object model in the XML fashion. Our vision system is aimed at identifying and locating industrial objects, of which a simple and natural representation is a wire-frame model formed with edges connected by vertices. This representation is inspired by the IGES norm for CAD modelization of wire-frame models[6]. We devise our XML network database including this model and appending textual information that describes the object at a high level. Fig.4 shows the listing of our rudimentary Document Type Definition(DTD).

The vertices of the object are stored in homogeneous coordinates, with each vertex containing the indices of its neighbour points. Thus the physical model of the object is transfered to the vision system as a 3D vertex network. We intend to handle more complex models such as NURBS in the future.

4.3 Model-based recognition

Projective geometry is a powerful tool to account for the representation of the 3D world on the image plane. It has been widely used for camera calibration and multiple view feature correspondence. The shape of the object is stored as 3D homogeneous coordinates, and although no direct correspondence is available from P^3 to P^2, various algorithms allow recognition of polyhedrons with projective invariants [14,15,16].

```
<?xml version='1.0' encoding='utf-8'?>
<!-- DTD for the sample object model database -->

<!ELEMENT ObjectList(object+)>

<!ELEMENT object(tagID,name,maker*,description*,
shape,color*,texture*)>
<!ATTLIST objectclass CDATA "undefined">
<!ELEMENT tagID (#PCDATA)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT maker (#PCDATA)>
<!ELEMENT description (#PCDATA)>

<!ELEMENT shape (point+)>
<!ATTLIST shape type CDATA "polyhedron">

<!ELEMENT point (X,Y,Z,T,connected)>
<!ATTLIST point index ID #REQUIRED>
<!ATTLIST point unit (mm|cm|dm|m) "mm">

<!ELEMENT X (#PCDATA)>
<!ELEMENT Y (#PCDATA)>
<!ELEMENT Z (#PCDATA)>
<!ELEMENT T (#PCDATA)>
<!ELEMENT connected (#PCDATA)>
```

Figure 4. Our DTD for wire-frame object models. Each object is nested inside an <object> node, referenced by its unique Tag ID, and contains structured information (name, maker, description, shape parameters…). The shape is stored as connected vertices.

Consider the projection of a 3D world point $P(X,Y,Z)$ onto the image point $p(x,y)$, as performed by the pinhole camera model. This is expressed:

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (1)$$

The 12 coefficients c_{ij} are the camera calibration parameters that need to be determined when locating the projection of an identified object in the image. In this paper, we wish to registrate a given known shape to the image sensed by the camera. We claim that using a RFID system greatly speeds up the registration process. Indeed, we want to compute the 3x4 camera calibration parameters of eq.(1), which is often done by solving the linear system of projection equations of n points in correspondence from

3D to 2D, as in eq.(2). In this equation, x_i and y_i are the image coordinates of a point which world coordinates are X_i, Y_i and Z_i.

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1Y_1 & -y_1Z_1 & -y_1 \\ \dots & & & & \dots & & & & \dots & & & \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_nX_n & -x_nY_n & -x_nZ_n & -x_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_nX_n & -y_nY_n & -y_nZ_n & -y_n \end{bmatrix} . \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ \dots \\ c_{31} \\ c_{32} \\ c_{33} \\ c_{34} \end{bmatrix} = 0 \quad (2)$$

We use the result proved by Weiss and Ray[14] in their theorem 2, giving a formal relation between the 3D projective invariants of 2 subsets of 6 points in P^3 and their images by a projective transformation. Eq.(2) needs at least 6 points in order to determine those parameters. Solving for each set of points being computationaly expensive, we use Weiss and Ray's result to check for the correspondence of 2 sets of 6 points, one from the object model and one from the image.

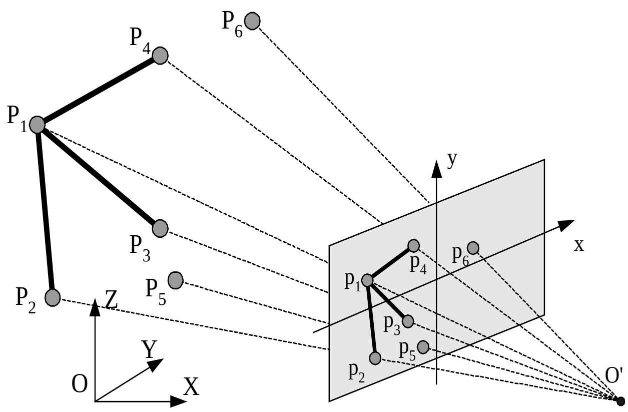


Figure 5. Two sets of 6 points in correspondence by projective projection. The points P_i are 3D world points projected into the points p_i in the image plane. The point p_1 in the image is a trihedral point because it is connected to the 3 points p_2, p_3 and p_4.

From the set of 3D points P_i such as in fig.5, we can form a set of three 3D projective invariants as follows:

$$I_1 = \frac{M_1 M_2'}{M_1' M_2}, I_2 = \frac{M_1 M_3'}{M_1' M_3}, I_3 = \frac{M_1 M_4'}{M_1' M_4} \quad (3)$$

where :

$$\begin{cases} M_k = \det(P_1 \dots P_{k-1} P_{k+1} \dots P_4 P_5) \\ M_k' = \det(P_1 \dots P_{k-1} P_{k+1} \dots P_4 P_6) \end{cases} \quad (4)$$

From the projections of the P_i onto the image plane, they define the following 2D projective invariant cross-ratios :

$$\begin{aligned} i_1 &= \frac{m_{12}' m_{14}}{m_{12} m_{14}'}, i_2 = \frac{m_{12}' m_{35}}{m_{12} m_{13}'} \\ i_3 &= \frac{m_{12}' m_{13}}{m_{12} m_{13}'}, i_4 = \frac{m_{12}' m_{45}}{m_{14}' m_{25}} \end{aligned} \quad (5)$$

where :

$$\begin{cases} m_{jk} = \det(p_1 \dots p_{j-1} p_{j+1} \dots p_{k-1} p_{k+1} \dots p_4 p_5) \\ m_{jk}' = \det(p_1 \dots p_{j-1} p_{j+1} \dots p_{k-1} p_{k+1} \dots p_4 p_6) \end{cases} \quad (6)$$

With these notations, 2 subsets of poins in correspondence as in fig.4 verify the relation :

$$\begin{aligned} \varepsilon = I_3(I_2-1)i_1 i_2 - I_3(I_1-1)i_1 - I_1(I_2-1)i_2 \\ -I_2(I_3-1)i_3 i_4 - I_2(I_1-1)i_3 - I_1(I_3-1)i_4 = 0 \end{aligned} \quad (7)$$

Eq.(7) expresses a necessary condition that we constarint by considerations of edge relationship: if 2 points from the object model P_i and P_j are connected by a 3D edge, their images p_i and p_j should be linked by an edge.

The way the sets are formed is as follows: from the image, we extract 6 points (intersections of Hough lines detected in the image) connected to each other by edges. From the object, we just group points knowing their connection information previously stored in the model. When the best match is found over the whole image, corresponding to the set that best verifies eq.(7). we solve for eq.(2) with the single value decomposition (SVD) method. This means finding the eigenvector of $({}^tA.A)^{-1}.{}^tA$ of least eigenvalue, where A is the left-hand side matrix of eq.(2) built with the corresponding points $P_i(X_i, Y_i, Z_i)$ and $p_i(x_i, y_i)$.

5 Experiments

Our system is developed on a desktop PC driving the Tag Interrogator and the image sensor (CD-711 camera by Sony), and connected to a server PC that manages the object model database. The Tag-Based Vision algorithm falls into 4 parts:

1)read Tag info and retrieve object model from database.

2)detect lines with Hough transform and extract image vertices as their intersections, if the local gradient intensity is above a gradient threshold. When a new valid intersection point is found, it is rejected if it lies on an edge segment joining 2 points that we already found. This way we keep edge segments of maximum length, but we cannot deal with T-junctions.

3)form subsets of six 2D feature points, and for each subset, compute the invariant relation of eq.(7) with all the 3D subsets of 6 points from the object vertices which have a compatible connection configuration. This gives a criterion of confidence of the two subsets being in correspondence by a pinhole camera projection. The 6-point subset from the image is deemed relevant when:

- it contains a trihedral point such as p_1 in fig.5 (point of intersection of 3 edges). Thus we expect the 3D points $P_1, P_2, P_3,$ and P_4 to form an affine basis.

- feature points p_5 and p_6 (resp.) are not connected to any of the points p_2, p_3 and p_4. Thus we minimize the chance of having coplanar 3D points that cause the determinants M_i and M_i' to become null, which cannot be dealt with by our algorithm. Here again we use a gradient threshold to determine the likeliness of $[p_j, p_k]$ for being an edge.

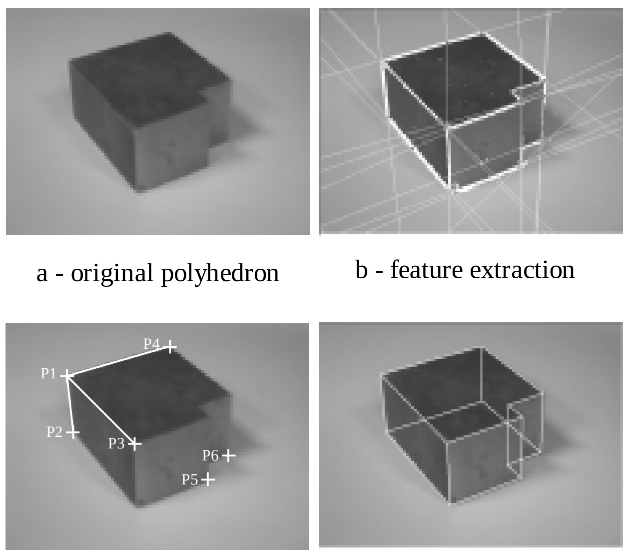
4)for the image point subset that gives the best match, i.e. the subset pair for wich ε in eq.(7) is minimum, use this optimum pair of 2x6 points to compute the camera calibration parameters with eq.(2).

5)knowing the camera calibration parameters, backproject the object model onto the image and display object information. According to a vocabulary commonly used in computer vision, the system has formed its *scene*, meaning its representation of the surrounding context consistent with a model.

When performing step 2), we speed up the computation of the classical Hough transform by using consistent gradient operators without smoothing the image sequence[1], and relying on their approximation of the local direction of edges, allowing to accumulate the Hough parameter space only in a narrow [θ-δθ, θ+δθ] area. Feature points in the image are found as intersections of Hough lines. This allows finding principal directions and vanishing points more easily, and also suggests that non-polyhedral objects can be described with directions or points other than vertices.

When forming subsets of connected points in step 3), we start by selecting trihedral points in the image. Thus, the computing complexity of step 3) is of $O(T.N^2n^6)$ where: T is the number of trihedral feature points, N is the average number of connected neighbours for each image feature point, and n is the average number of neighbours for each model vertex (n=3 if only trihedral corners are present on the object).

Fig. 6 shows the window of our main application for image analysis. The object we use for our test demo is a simple polyhedron with 12 vertices. The RF Tag (not shown) containing the object's model is presented to the Tag interrogator when the object enters the camera's field of vision.



a - original polyhedron

b - feature extraction

c - extraction of 6-point subset

d - object model registered

Figure 6. Example of recognition for a simple polyhedron with our image analysis software. In (b) lines and edges are detected by Hough transform, and a discriminating 6-point subset is extracted from those feature points (c). The camera is calibrated and the objectis registered in (d).

6 Conclusion

In this paper, we introduced a novel architecture for a vision system that integrates a cooperative strategy for image analysis and acknowledges the ever-growing availability of structured information on the computer networks. The integration is performed thanks to a RF Tag object identification system, providing a key to retrieve a complete model of the detected object through a network knowledge database, designed in XML. This system enables to simplify any object recognition task to a problem of registering the object model to the image, and the recognition part is independent of the number of models in the database. We use wire-frame object models. A more elaborate model, implementing general components from the IGES standard as well as other structural object description elements will be needed to handle general objects. Moreover, we still need to work on the robustness of our system. Those points, as well as the application of our architecture to various scenarios in machine vision, will be the objects of future developments.

References

[1] S. Ando, "Consistent Gradient Operators", *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 22, no. 3, pp. 252-264, Mar. 2000.

[2] P. Besl and R. Jain, "Three dimensional object recognition", *ACM Comput. Surveys*, vol. 17, pp. 75-145, 1985.

[3] R. Campbell and R. Flynn, "A Survey of Free-Form Object Representation and Recognition Techniques", *Comput. Vis. and Im. Understand.*, vol. 81, pp.166-210, 2001.

[4] R. Chin and C. Dyer, "Model-based recognition in robot vision", *ACM Comput. Surveys*, vol. 18, pp. 67-108, 1986.

[5] O.D. Faugeras and M. Hebert, "The representation, recognition, and location, of 3D objects", *Intern. J. Robotics Res.*, vol. 5, pp. 27-52, 1986.

[6] P.J. Flynn and A.K. Jain, "CAD-Based Computer Vision: From CAD Models to Relational Graphs", *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, no. 2, pp. 114-132, Feb. 1991.

[7] J. Hackett and M. Shah, "Multi-sensor fusion: a perspective", in *Trends in Optical Engineering*, 1993.

[8] D. Lowe, "Three-dimensional object recognition from two-dimensional single images", *Artificial Intelligence*, vol. 31, pp. 155-195, 1987.

[9] J. Matas, L.M. Soh, J. Kittler, "Object Recognition using a Tag", *Proc. of Int. Conf. on Image Processing*, 1997.

[10] N. Oswald and P. Levi, "Cooperative Object Recognition", *Pattern Recognition Letters*, vol. 22, pp. 1273-1282, 2001.

[11] A. Pope, "Model-based object recognition – A survey of recent research", *Technical Report TR-94-04*, 1994.

[12] J. Rekimoto and Y. Ayatsuka, "CyberCode: Designing Augmented Reality Environments with Visual Tags", *Proceedings of DARE*, 2000.

[13] F. Sandakly and G. Giraudon, "3D scene interpretation for a mobile robot", *Robotics and Autonomous Systems*, vol. 21, pp. 399-414, 1997.

[14] I. Weiss and M. Ray, "Model-Based Recognition of 3D Objects from Single Images", *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 2, 2001.

[15] Y. Zhu, L. Seneviratne and S. Earles, "New algorithm for calculating an invariant of 3D point sets from a single view", *Im. and Vis. Comput.*, vol. 14, pp. 179-188, 1996.

[16] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu and N. Pillow, "3D object recognition using invariance", *Artificial Intelligence*, vol. 78, no.1-2, pp. 239-288, 1995.

[17] S. Ando and H. Hontani, "Automatic Visual Searching and Reading of Barcodes in 3D Scene", Proc. IEEE Intl. Vehicle Electronics Conf. 2001, Tottori, pp. 49-54, 2001.

[18] "Draft Paper on the Chracteristics of RFID-Systems", www.aimglobal.org, AIM Inc. Frequency Forum White Paper, July 2000.

[19] R. Das, "RFID explained", IDTechEx White Paper, www.idtechex.com, 2002.

[20] S. Sarma, K. Ashton, D. Brock, "The Networked Physical World", MIT Auto-ID Center White Paper,1999.