MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

8—28

A Knowledge-Based Registration System using Radio-Frequency Identification

Mustapha Boukraa and Shigeru Ando[1]
Graduate School of Engineering,
Department of Mathematical Engineering and Information Physics,
University of Tokyo

Abstract

Attaching visual tags such as barcodes to some elements in a scene allows applications such as navigation, handling industrial objects, or augmented reality. In the past few years, a new type of non-visual tags has been the object of intensive developments: Radio-Frequency (RF) Tags. In this paper we present the architecture of a machine vision system that uses RF Tag identification (RFID). RFID enables detecting the presence of an object, which polyhedral model and properties are remotely retrieved from a knowledge network database. This model is used for object registration by projective invariant matching. Our system not only facilitates the problem of object identification and registration to a model, but also allows new Tag-Based applications to be built around the concept of ubiquitous objects.

1 Introduction

For human observers, the world is full of items that play the role of tags: brand names and logos, road signs, symbols, gestures and poses, advertisement jingles, etc… Each of these tags can be interpreted according to one's experience and behavior characteristics.

Autonomous machine vision systems have been designed that handle visual tags [1]. Such systems ease the process of knowledge-based 3D scene analysis and object recognition, since each visual tag is a key to a specific object model. However, when relying on visual tags, the system has to accurately detect remote tags, which is subject to occlusion and errors (due to noise, degradation of the tag's aspect, etc...). In an IT-based approach, visual tags are not of very practical use because they are not designed to be re-writable. Moreover, there is no unique standard for product encoding.

A new tagging technology has emerged in recent years and is being adopted by a growing number of product manufacturers, security companies, banks and many others (recently individuals have been tagged using this technology): RF Tags, also known as Electronic Tags.

RF Tags are small devices (transponders) that contain an antenna and can communicate with a tag interrogator (or "reader") using electromagnetic coupling or propagation. They contain a few bits to 1Mb of data that is whether read-only or readable-writable [2,9]. There are efforts to create standard that aim at making RF Tag-based systems interoperable [2,6]. Advantages of RF Tags over classic visual tags such as barcodes include the fact that RF Tags can be embedded into the object whether barcodes are usually put on the object's package and modify its appearance, they come in tiny sizes that make them discreet and applicable to small objects, and they can operate in tough environments (water, chemicals,…). Also, for some applications where active sensing is needed they can come with a power supply (capacitor).

All sorts of applications can be thought of with such a technology, leading to autonomous systems able of evolving in an ubiquitous environment of tagged objects, dynamically referring to or modifying the common knowledge about those objects and taking the appropriate course of actions.

For a computer vision system the first stage is naturally to estimate each tagged object's location and pose. The work presented in this paper is original in that it tackles the problem of 3D object identification and recognition to 2D images where the identification step is performed by the highly reliable RFID technology. Recognition is then a matter of retrieving the object's model and registering it to the image data.

Fig.1 shows the general architecture of our RF Tag-based system, comprising two local sensing channels, a remote model retrieval module and the processing core.

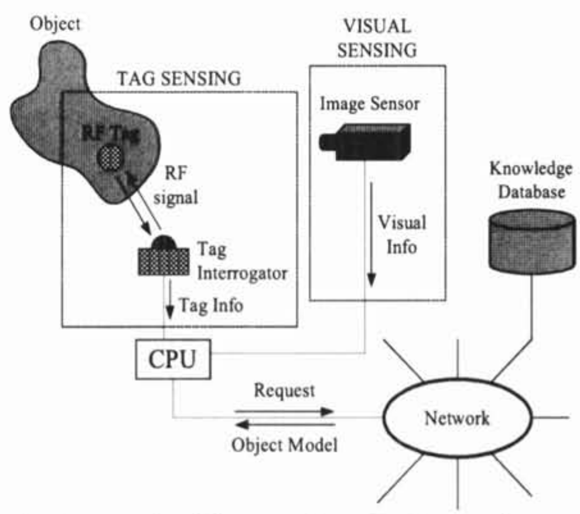


Fig. 1: general architecture of our RF Tag-based computer vision system. Tag sensing and information retrieval, and visual sensing, are combined at the CPU level to perform knowledge-based 3D registration of the detected objects.

In the following section we describe the 3 elements of our system, namely RFID, the building up of object models, and image processing for object registration. Section 3 deals with the practical implementation of our architecture, presenting some experiments of object registration.

Address: Hongo 7-3-1, Bunkyo-ku, Tokyo 133-8656 Japan. E-mail: moose@alab.t.u-tokyo.ac.jp

2 A Tag-Based Vision System

RFID 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 [2], [9]:

. frequency range: typical RFID systems use frequencies in the VHF, UHF and up to the microwave 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 bit rates.

. 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 to rigid rings of alloy or flexible laminates, etc... allowing to adapt them to objects of various sizes and shapes.

In our application of object registration, we use smart card shaped, 2.45GHz, passive RF tags with chips that support R/W and multiple tag detection. Each tag contains only a unique identifying code, which means that the entire object related data management is handled through the network database. The operational distance of the system is about 1m when facing the antenna.

Our application is original in that it resorts to RFID to assist image analysis. We argue that RFID is particularly suited for such problem, since the detection of RF tags is unambiguous and error-free itself. Whereas visual tag detection is subject to noise and orientation error, RF detection makes it obsolete (except for the cases where reflections on metallic surfaces or absorption by water create perturbations of the RF signal). On the other hand, because RF tag detection is pose-independent, no prior information is available concerning object location/orientation.

Logic Models for Physical Objects In this study, object description is the process of creating a logic model of a physical object. It is a complex problem, and a general endeavor is bound to be unsatisfactory for specific applications [6]. In our case, discriminative and manageable data to represent object geometry are needed, whereas other existing standards focus on other aspects such as logistics and trade. Among standards for structured data exchange on the Internet, we can cite EDI, UDDI, ebML and DCMI, each being supported by different economic partners. Fig. 2 shows the general classification we have chosen for object description. We focus on object geometry and location, leaving other elements unspecified at this stage, as they are not determinant for our application.

We use XML as a description language for object model data, making use of its ease of editing and parsing, and its popularity as a language for data sharing on the Internet.

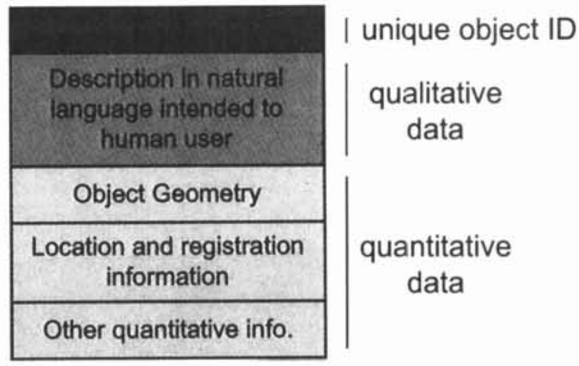


Fig. 2: overview of our object model description. The tag ID is the primary key to the object data. Qualitative information is aimed at providing a human user an understanding of the object (for applications such as augmented reality). Qualitative information is to be processed by an autonomous system. In our computer vision system we focus on the description of object geometry and localization. Each type of information can be constant or variable.

Object registration using projective invariance Our registration algorithm relies on point-set matching with projective invariants under full projective perspective. The basic result we make use of is taken from [7].

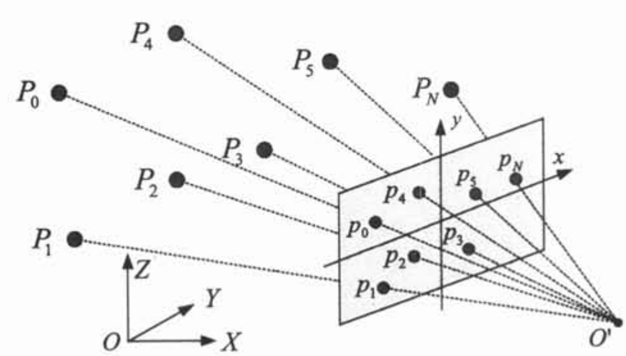


Fig. 3: two sets of N points in correspondence by pinhole camera projection. The absolute 3D affine referential is (O, X, Y, Z) and O' is the camera center.

Consider a set of N 3D vertices describing a given object. We can extract a subset of 6 vertices $\{P_i, 0 \le i \le 5\}$. Using the canonical injection of the 3D real space into the 3D projective space, each of these vertices can be represented by 4 homogeneous coordinates (X_i, Y_i, Z_i, T_i). 3D projective invariants of the P_i remain constant for any transformation of the 3D projective space undergone by the P_i. For the subset $\{P_i\}$, three independent such invariants can be formed as follows:

$$I_1 = \frac{M_1 M_2'}{M_1' M_2}, I_2 = \frac{M_1 M_3'}{M_1' M_3}, I_3 = \frac{M_1 M_4'}{M_1' M_4} \quad (1)$$

where the following 4x4 determinants appear:

$$\begin{cases} M_k = \left|P_0 \ldots P_{k-2} P_k \ldots P_3 P_4\right| \\ M_k' = \left|P_0 \ldots P_{k-2} P_k \ldots P_3 P_5\right| \end{cases} \quad (2)$$

We note p_i the projections of the P_i onto the 2D projective space representing the image plane. Each of the p_i

has homogeneous coordinates (x_i, y_i, t_i). Thus we get the following 2D projective invariant cross-ratios [7]:

$$i_1 = \frac{m'_{12}m_{14}}{m_{12}m'_{14}}, i_2 = \frac{m'_{12}m_{35}}{m_{12}m'_{13}}$$
$$i_3 = \frac{m'_{12}m_{13}}{m_{12}m'_{13}}, i_4 = \frac{m'_{12}m_{45}}{m'_{14}m_{25}} \tag{3}$$

where 3x3 determinants are computed as:

$$\begin{cases} m_{jk} = \left| p_0 \cdots p_{j-2} p_j \cdots p_{k-2} p_k \cdots p_3 p_4 \right| \\ m'_{jk} = \left| p_0 \cdots p_{j-2} p_j \cdots p_{k-2} p_k \cdots p_3 p_5 \right| \end{cases} \tag{4}$$

With these notations, and the only hypothesis that there is 3D-2D projective correspondence between the 2 subsets of points, we have the necessary condition [7]:

$$\varepsilon = (i_3 - i_2)I_1I_2 + (i_1i_2 - i_3i_4)I_2I_3 + (i_4 - i_1)I_3I_1$$
$$+ (i_2 - i_4)I_1 + (i_4 - 1)i_3I_2 - (i_2 - 1)i_1I_3 = 0 \tag{5}$$

Null determinants create degenerate cases in the relationship. Therefore these cases, corresponding to coplanar points in 3D, must be avoided. Eq.(5) can be written as a quadric equation in a 4D vector space:

$$\varepsilon = I^T \begin{bmatrix} 0 & i_3 - i_2 & i_4 - i_1 & i_2 - i_4 \\ i_3 - i_2 & 0 & i_1i_2 - i_3i_4 & (i_4 - 1)i_3 \\ i_4 - i_1 & i_1i_2 - i_3i_4 & 0 & -(i_2 - 1)i_1 \\ i_2 - i_4 & (i_4 - 1)i_3 & -(i_2 - 1)i_1 & 0 \end{bmatrix} I$$

and $\varepsilon = 0$ (6)

with $I = (I_1, I_2, I_3, I_4 = 1)^T$

Now, replacing P_5 with any of the remaining $N-6$ object vertices P_i, and provided we know its 2D projection p_i, we can form a system of $N-5$ terms ε_n as in eq.(5), where n is the index of the 6-point subset. All the 6-point subsets have points $P_0, ..., P_4$ in common: those points form a projective basis and the 3 invariants are in fact the projective coordinates of the sixth point in this basis. Keeping only those terms corresponding to non-degenerate cases and visible points, the following 3D-2D necessary condition of projective consistency can be expressed as:

$$\|\mathrm{E}\|^2 \xrightarrow{\min} 0 \tag{6}$$

where E is a vector of global inconsistency. If K is the number of non-degenerate equations as in (5) that we can form with those object vertices which are not occluded, we have:

$$\mathrm{E} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \ldots & \varepsilon_K \end{bmatrix}^T \tag{7}$$

Eq.(5) expresses that the term noted ε is a measure of projective inconsistency of a 6-point subset. Therefore E defined in eq.(7) represents a term of global projective inconsistency of the set of relevant points.

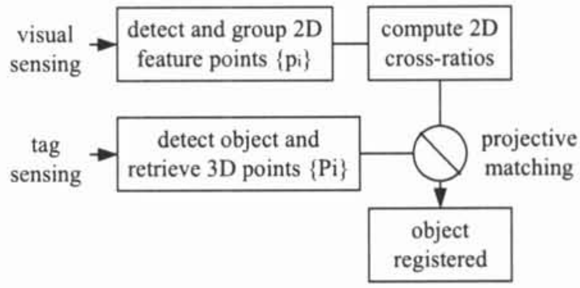


Fig. 4: chart of our registration algorithm. The 2D cross-ratios of eq.(3) are computed after extraction of the 2D image feature points. On the RFID and knowledge database side, the object model is retrieved in parallel. Projective matching then consists in determining the 3D points that best satisfy the condition (6).

Making use of this observation, we devise the registration algorithm shown in Fig. 4. The two sensing channels (visual and tag) simultaneously provide with data are combined after extraction of the visual features. The next section deals with the practical problems encountered when implementing this method.

3 Practical Implementation

Our system, which three main components we presented above, will eventually be integrated in an autonomous computer vision system. At this stage we seek to validate its feasibility. To that end, we focus on the most critical element, namely the object model registration to the image.

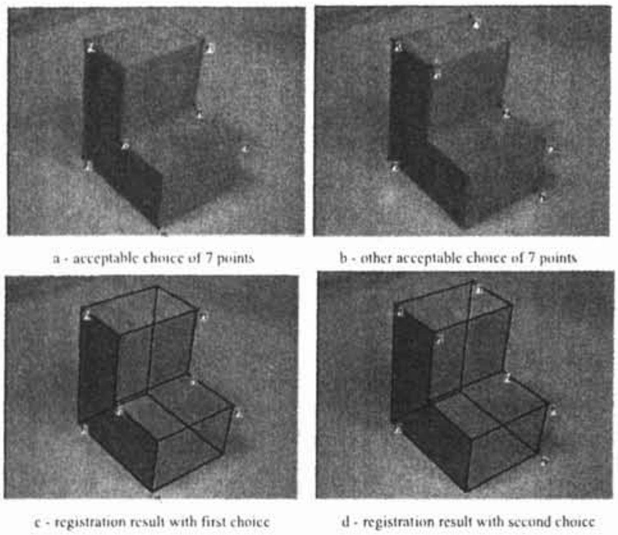
The performance of the registration method critically relies on a robust 2D feature extraction and grouping. Ideally, we should be able to detect all the salient feature points (corners) belonging only to the object to register. We are currently performing experiments and calibrating our system using Ando's OMNIVAR corner detection tools [1]. However, in order to increase the robustness of the object feature detection and grouping, region-based segmentation (on texture, color or gray similarity) is desirable to isolate object points and eliminate spurious points.

Moreover, in order to avoid degenerate cases when computing the projective inconsistency between a set of 2D feature points and a given set of 3D vertices, we need to ensure that:

✓ $p_0, ..., p_4$ correspond to, i.e. they are projections of, object points $P_0, ..., P_4$ (resp.) that form a projective basis. This is realized if no subset of four points among the five object points is coplanar in the 3D affine space. When considering polyhedral objects, we can have a key to the 3D configuration by looking at the 2D edge connection of their projections. In general however, we have no information *a priori*, but we can assume that the more complex the object is, the less likely we are to be faced to degenerate sets, and the more simple it is, the easier it is to detect coplanarity of points and discard degenerate point sets.

✓ an additional sixth image feature point is not the projection of a point P_5 that is coplanar with any four points from the projective basis $P_0, ..., P_4$.

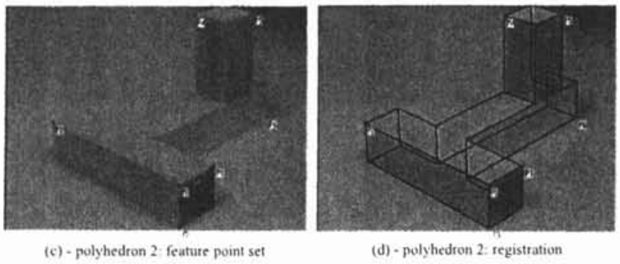
Fig. 5 shows a registration test performed on a L-shaped polyhedron. The polyhedron's wireframe model is back-projected on the image, which permits to qualitatively assert the accuracy of the registration.



a - acceptable choice of 7 points
b - other acceptable choice of 7 points
c - registration result with first choice
d - registration result with second choice

Fig. 5: registration test on simple polyhedral shape. Feature points are picked manually. The registration (c) with points (marked with white labels) as chosen in (a) gives $||E|| = 8.70E-3$. In (d) we have $||E|| = 8.08E-4$

We have carried out experiments with objects of polyhedral shapes. Each of them is attributed a RF tag with a specific Tag ID, allowing to retrieve its wireframe model stored on a remote XML database server. For the sake of experimentation tags are waved before the antenna and not attached directly to objects. The successful registration results are presented in Fig. 6 for 2 object models. On each view the feature points used for projective matching are marked with white labels. Those features are picked manually from the set of points computed by corner detection [1]. Indeed, we are still working on devising a robust feature grouping algorithm suitable to our method.



(a) - polyhedron 1: feature point set
(b) - polyhedron 1: registration
(c) - polyhedron 2: feature point set
(d) - polyhedron 2: registration

Fig. 6: experiments of object registration for validation of projective matching algorithm. Registration is performed for 2 polyhedral models, each with a set of set of 7 matching points.

Conclusion

We have presented an architecture for knowledge-based object registration where efficient object detection and knowledge retrieval is achieved thanks to the promising RFID technology. Such architecture aims at bridging physical objects and their quantitative and qualitative representations. The object registration algorithm we have devised uses projective matching of image feature points with 3D object data. It has been validated experimentally on some test images, and needs to be integrated with an efficient low-level feature extraction tool and be adapted to deal with special degenerate cases, subjects currently under investigation.

We believe our demo application to open new perspectives in domains where automatic systems need to perform fast and accurate object registration, and more generally when an automated system is expected to acquire a "deep" understanding of physical objects.

References

[1] S. Ando, "Image Field Categorization and Edge/Corner Detection from Gradient Covariance", *Trans. IEEE PAMI*, vol. 22, no. 3, pp.179-190, Feb 2000.

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

[3] 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.

[4] D. Lowe, "Three-Dimensional Object Recognition from Single Two-Dimensional Images", *Artificial Intelligence*, vol. 31, no. 3, pp. 355-395, March 1987.

[5] R. Mohr, B. Triggs, "Projective Geometry for Image Analysis", Tutorial given at ISPRS, Vienna, July 1996.

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

[7] 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.

[8] 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.

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