An Invariance-Based Object Registration Architecture using Electronic Tags and XML

M. Boukraa and S. Ando

Abstract-- We present an architecture for knowledge-based object registration that relies on radio-frequency tagging to detect and identify objects. When a tagged object is detected, the network resources are searched for the object’s logic model, and when this model is retrieved various knowledge-based treatments can be performed. The three main components of our system of which we present an early implementation in this paper are the electronic tagging technology, the object logic representation in XML standard, and the object registration method relying on invariance matching. We believe that such architecture will allow new applications in computer vision.

Index Terms-- knowledge-based, computer vision, invariance, logic representation, radio-frequency tags.

I. INTRODUCTION

WHEN handling and analysing an object, human observers relies on their senses to directly perceive different properties of the object, as well as on their proper experience and knowledge, that provide them with a model of the physical world. In monocular image analysis, many limitations are inherent to 3D-2D projective image formation, and some information is impossible to sense without a model. Thus the optical flow constraint is underdetermined, and 3D object recognition is affected by occlusion and the loss of depth information.

In machine vision systems, a popular way to create the conditions of knowledge-based analysis, by triggering a contextual call to a knowledge pool, is to implement a tag detection module; such module interrogates a tagged object to retrieve its unique identifier (the Tag ID), and then uses this Tag ID as a key to some relevant information about the object previously stored in a local or remote database. Barcodes constitute an example of visual tags widely used for automatic object handling and logistics [1]. Proprietary visual tags are also often used for camera self-calibration, robot navigation and object recognition [7]. Electronic tags, also known as radio-frequency (RF) tags, have been introduced a few years ago as a new alternative [2], [13]. RF tags are tiny devices that communicate energy and data when a propagative/inductive coupling is initiated with an appropriate RF antenna. They can be attached or embedded to usual objects.

They combine some advantages over classic tags, one of which being that they require no processing effort for unambiguous remote detection. The present study considers the problem of 3D object registration from 2D images. This problem, challenging to the image processing community though easily carried out in human vision [6], has been tackled by using range images, multiple images, or active vision techniques such as structured light [3], [8]; 3D recognition from 2D image methods needs resort to certain assumptions about object classes, which limit their generality [6], [11].

Object representation consists in storing a model of the object that best describes it in a given “space of information”. This representation is an important part of a system that implements tag detection, and it must follow some consistency rules, so as to allow network data exchange and automatic processing. In order to make models available on the Internet and accommodate for automated parsing, we choose the popular XML standard as a description language for object models. Among object properties, the most critical part for our computer vision system is the description of its geometry.

The description of object geometry is determined by the recognition method, which is itself conditioned by the features available at the sensing level. Range finders provide with a distance-to-obstacle suited to matching with CAD models, where object surface are often expressed following IGES specifications [4]. Features available from a single image can be local (brightness intensity, edges, vertices, corners,...) or global (moments, directions, region segmentation by texture or color,....). In our application, the object’s wireframe model is registered to 2D local edges and vertices, for the sake of simplicity. The method used relies on projective matching with geometric invariants [10]-[12].

In the following section we briefly present the RF tag technology and RF identification (RFID), which made our system feasible. Section III deals with the choices we made for the representation of physical objects and the constitution of our network database. Finally in section IV we describe the layout of our computer vision system that integrates the elements introduced above with a polyhedral object registration algorithm. We conclude by mentioning some key perspectives allowed by “electronic tag and XML”-based systems.

II. RADIO-FREQUENCY IDENTIFICATION

A. Overview of RF tags

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], [13]:

. 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.

B. RFID and Machine Vision

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.

III. SPECIFICATIONS FOR OBJECT MODELS

A. Object Classification

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 [9]. 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. 1 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.

<table>
<thead>
<tr>
<th>unique object ID</th>
<th>qualitative data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag ID</td>
<td>Description in natural language intended to human user</td>
</tr>
<tr>
<td>Object Geometry</td>
<td>Location and registration information</td>
</tr>
<tr>
<td>Other quantitative info.</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. General 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.

Whatever the chosen classification is, setting a structure that enables automated treatment and sharing is a primordial concern. This is why use XML as a description language for object model data, making use of its popularity as a language for data sharing on the Internet.

B. Object Geometry

Various formats are available to describe a 3D solid shape. Proprietary formats are used in CAD design, though most systems follow the recommendations of IGES [4]. Some formats are designed for object visualization in computer graphics, such as OpenGL, VRML and its evolution X3D, the last two aiming at transmission of 3D models on the Internet.

```xml
<Shape type="polyhedral">
  <VertexNumber number of vertices in the model>
    <X unit="mm">X</X>
    <Y unit="mm">Y</Y>
    <Z unit="mm">Z</Z>
  </VertexNumber>
  <ConnectedVertexIndices number="cvnb">
    <Index index="index1">
      <Index index="index2">
        ...
      </Index>
    </Index>
  </ConnectedVertexIndices>
</Shape>
```

Fig. 2. XML markup tags for polyhedral object description. VertexNumber is the total number of object vertices. Each 3D vertex is expressed by an index, 3D coordinates (in the object’s local 3D referential) and the indices of connected neighbor vertices. The object’s optimum registration method being dependent on its geometry, this information is attached after the shape info. The “Invariant Matching” method is explained in the following.
There is a tradeoff between the physical accuracy of a model and its ease of use in image processing and network sharing. Moreover, in the general case, 3D characteristic object features can be lost in camera projection, and inversely, some image features do not correspond to particular 3D features. For example, a 2D object boundary does not correspond to any 3D curve on the object surface in general. Restricting our application to polyhedral objects allows us to consider that we keep this practical 3D-2D feature correspondence (3D vertices (resp. segments) project onto 2D vertices (resp. vertices)). Fig. 2 shows our markup tags for storing a polyhedral shape as a mesh of connected vertices.

C. Object Location and Registration

The purpose of our system is to determine the pose of an object of known shape, relatively to the observing camera. To realize this, we set a `<Location>` tag that is aimed at accommodating for absolute location in a global referential, which we have not specified at this point, and relative location `<RelativeLocation>`.

The latter consists of the registration parameters of the object, relative to a given “viewing device”, which is in our system a unique CCD camera, referenced by its own `<TagID>` and `<AbsoluteLocation>`. The registration parameters are the 12 coefficients of the camera projection matrix. It is an 11-degree of freedom set, since a scaling does not affect location and pose [5]. Some of the markup tags used in practice are given in Fig. 3.

Note that instead of a “viewing device”, we could store the translation and rotation parameters of the object relative to another reference object. However, this would be impractical in the context of highly mobile objects, and needs the supervision of a camera to ensure the consistency of the relative location data.

D. Sharing Object Models

Our XML model database is designed to store object models and location information, so that it is available to all the nodes on the network.

In practice, we have a client application that manages RFID and image analysis, and communicates with a database server that handles object models. In order to keep a non-specific architecture, the communication is established with a UDP socket. Further developments will involve setting a naming service for XML data files to efficiently be retrieved on the network.

IV. APPLICATION: KNOWLEDGE-BASED REGISTRATION OF POLYHEDRAL OBJECTS

A. System Implementation.

The architecture of our system, integrating the elements presented above, is summarized in Fig. 5. The central CPU integrates the RFID module and the vision module, and sends requests to the network for logic object models every time a physical object is detected.

B. Invariance-Based Polyhedron Recognition.

The recognition algorithm is that of an invariant matching with 6 points, under full perspective, as introduced in [10]. This algorithm uses a relationship between 3D invariant and 2D invariant cross-ratios in order to match a set of 6 3D points to a set of 6 vertices extracted from the image.

Invariants of the projective space for a set of $n$ points are any value of those points that remain constant for any transformation of the projective space. Fig. 4 shows two subsets of 6 points in projective correspondence: $P_i$ are 3D points that project onto the image points $p_i$. For our registration algorithm, the geometry of edge connection is of particular importance.

We focus on configurations where 3 points $P_2, P_3, P_4$ are connected to a trihedral point $P_1$ to form an affine basis. Two points $P_5, P_6$ are chosen such that they are not connected to any of the points $P_2, P_3, P_4$ as explained in the following.

Fig. 3. Object location information. AbsoluteLocation describes the object’s location in some predefined global referential. This information is updated by a reference viewing device (a camera in general) identified by its Tag ID. The object’s relative location consists in the calibration parameters relative to the viewing device’s referential.

Fig. 4: two 6-point subsets in correspondence by pinhole camera projection. The absolute 3D affine referential is $(O,X,Y,Z)$ and the camera center is $O'$. The optical axis, not drawn on the figure, is $(O',z)$, perpendicular to the image plane.
Fig. 5. General architecture of the tag-based vision system. The central CPU commands the RF Tag identification system and the visual sensor, and is connected to the network. RF Tags are attached to physical objects, and each tag contains a unique “Tag ID”. In the space of logic objects, designed as XML data files, the object’s Tag ID is the key to its properties.

This relation shows that in the general non-degenerate case, \((I_1, I_2, I_3)\) in the affine space \(A^3\) belongs to a quadric determined by the parameters \((i_1, i_2, i_3, i_4)\). Equation (5) theoretically enables to register a polyhedral object with one characteristic 6-point subset. In practice there are 2 degenerate cases that induce spurious results:

1) \(\exists j \in \{2, 3, 4\} / M_j = 0 \) or \(M'_j = 0\). This causes the 3D invariants \(I_j\) to be null or infinite. To prevent this case, we select \(P_5\) and \(P_6\) not connected to any of the \(P_1, P_3, P_4\). This way we minimize the chance of getting four 3D coplanar points.

2) \(\forall (j, k), i_j \approx i_k \approx 1\). This case adversely affects accuracy since it creates spurious minima for \(|\varepsilon|\). When grouping feature points, we reject subsets for which this is verified.

To further constraint the necessary condition expressed in (5), we make use of edge connection configuration, by searching for features \(p_i\) that follow the pattern of Fig.4 \((P_5\) and \(P_6\) connected to each other). Such configurations are present in general non-trivial polyhedrons. Moreover, using more feature points will provide additional equations and will further constraint the projective matching. Our knowledge-based registration method consists in performing the steps described in Fig.6.
We perform feature extraction with a Hough transform. This choice is driven by the fact that we consider polyhedral objects, for which lines and edges are easy to detect characteristic features. Moreover, linear features are natural cues to directions and points at infinity in the image. However, the regular Hough transform is a global feature extraction technique that relies on predefined parameters (size, resolution, threshold, ... of accumulator) that need to be optimized.

The term “projective inconsistency” in step 5 refers to $\mathcal{E}$ in (5). The main problem, which needs improvements, is to eliminate spurious minima of $\mathcal{E}$. This could be done by direct comparison of image edges and backprojected model at step 5.

Some results for the registration of a simple polyhedron are given in Fig. 7. Accuracy can be assessed visually, as the backprojected polyhedral model matches the borders and vertices of the image of the object on each 2D view.

V. CONCLUSION

In this article we presented an application that aims at bridging physical objects and their quantitative and qualitative representations. These representations follow the XML specifications in order to make them easy to share and retrieve on a network and parse by an autonomous system. We gave some specifications for two important types of data making up this representation: object geometry, expressed with a polyhedral model, and object localization, defined as the registration parameters relative to the reference camera.

As a practical application of this system, made possible thanks to the tremendous progress of the RFID technology, we introduced a basic scheme for model-based polyhedral object registration, using some recent result of projective geometry.

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