Open Research Online

The Open University's repository of research publications and other research outputs

Object analysis with visual sensors and RFID

Book Section

How to cite:

Karmakar, Gour C.; Dooley, Laurence S.; Karmakar, Nemai C. and Kamruzzaman, Joarder (2012). Object analysis with visual sensors and RFID. In: Karmakar, Nemai Chandra ed. Chipless and Conventional Radio Frequency Identification: Systems for Ubiquitous Tagging. Hershey, PA, USA: IGI Global, pp. 234–250.

For guidance on citations see [FAQs](#).

© 2012 IGI Global

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.4018/978-1-4666-1616-5>

<http://www.igi-global.com/book/chipless-conventional-radio-frequency-identification/61620>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Chiplless and Conventional Radio Frequency Identification: Systems for Ubiquitous Tagging

Nemai Chandra Karmakar
Monash University, Australia

Managing Director:	Lindsay Johnston
Senior Editorial Director:	Heather A. Probst
Book Production Manager:	Sean Woznicki
Development Manager:	Joel Gamon
Development Editor:	Hannah Abelbeck
Acquisitions Editor:	Erika Gallagher
Typesetter:	Jennifer Romanchak
Cover Design:	Nick Newcomer, Lisandro Gonzalez

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: <http://www.igi-global.com>

Copyright © 2012 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Chipless and conventional radio frequency identification: systems for ubiquitous tagging / Nemaï Chandra Karmakar, editor.
p. cm.

Includes bibliographical references and index.

Summary: "This book explores the use of conventional RFID technology as well as chipless RFID technology, which provides a cheaper method of implementation, opening many doors for a variety of applications and industries"-- Provided by publisher.

ISBN 978-1-4666-1616-5 (hardcover) -- ISBN 978-1-4666-1617-2 (ebook) -- ISBN 978-1-4666-1618-9 (print & perpetual access) 1. Radio frequency identification systems. 2. Radio frequency--Identification. 3. Wireless communication systems. 4. Radio--Antennas. I. Karmakar, Nemaï Chandra, 1963-
TK6570.I34C45 2012
621.3841'92--dc23

2011050428

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Chapter 12

Object Analysis with Visual Sensors and RFID

Gour C. Karmakar
Monash University, Australia

Laurence S. Dooley
The Open University, UK

Nemai C. Karmakar
Monash University, Australia

Joarder Kamruzzaman
Monash University, Australia

ABSTRACT

Object analysis using visual sensors is one of the most important and challenging issues in computer vision research due principally to difficulties in object representation, segmentation, and recognition within a general framework. This has motivated researchers to investigate exploiting the potential identification capability of RFID (radio frequency identification) technology for object analysis. RFID however, has a number of fundamental limitations including a short sensing range, missing tag detection, not working for all objects, and some items being just too small to be tagged. This has meant applying RFID alone has not been entirely effective in computer vision applications. To address these restrictions, object analysis approaches based on a combination of visual sensors and RFID have recently been successfully introduced. This chapter presents a contemporary review on these object analysis techniques for localisation, tracking, and object and activity recognition, together with some future research directions in this burgeoning field.

INTRODUCTION

The innate desire for humans to have their computers and machines do exactly whatever they are able to do has, to date remained largely unfilled. While humans can effortlessly analyse and recognise an

object and its activity or event, computer-based automatic systems still remain stubbornly in an embryonic state. This has lead to computer vision being one the most challenging research topic areas over the past few decades. The application of computer vision is now however, rapidly in-

creasing due to the advancement of software and hardware technologies, including RFID (*radio frequency identification*), digital image analysis, artificial intelligent and robotic vision systems.

Computer vision has been extensively applied in a wide range of diverse domains. These include, but are not limited to: robotic vision (Hsiao et al., 2008), autonomous car and mobile robot (Milanés et al., 2010), event detection, surveillance and tracking (Yilmaz et al., 2006; Lee et al., 2007), airport identification from aerial photographs, object-based image identification and retrieval (Veltkamp & Tanase, 2000), object and activity recognition (Smith et al., 2005; Wu et al., 2007; Bashir et al. 2007), behaviour prediction (Skinner, 1953; Pentland, 1999; Chen et al., 2004), object-based image and video coding (ISO/IEC 14496-2, 1999), criminal investigation using video footage analysis, computer graphic, and medical diagnoses (cancerous cell detection, segmentation of brain images, skin treatment, and intrathoracic airway trees) (Paul & Paul, 1993; Pham & Prince, 1999; Liu et al., 1997). An essential preprocessing step in all the aforementioned real-world examples is object detection and separation (segmentation), which both involve the core element of image analysis to retrieve and understand an event's information in real-time (Computer vision, 2011). Image analysis, typically includes image segmentation, feature extraction, and object classification (Baxes, 1994).

Image segmentation is the process of separating mutually exclusive homogeneous regions of interest (i.e., having similar pixel intensities) from other regions in an image. In reality, most natural objects are not homogeneous and comprise numerous objects in a hierarchical order. Furthermore, there are generally a huge number of objects and a myriad of variations amongst them. This makes object segmentation in an image an interesting if intractable task as it somewhat contradicts the above definition of object-based image segmentation. This is because there is no universally accepted standard definition of im-

age segmentation. The properties of an object to be segmented depend on its applications and human perception, so segmenting an object is in essence an ad hoc process, which depends on the emphasis given to the set of desired properties and a trade-off among them (Paul & Paul, 1993; and Karmakar et al., 2001). To reduce the complexity of detecting and separating an object and capturing its dynamic movement pattern (i.e., motion), computer vision researchers are progressively exploiting temporal information from visual sensors such as digital video cameras (Song & Fan, 2006; Ahmed et al., 2007).

Even if an object has been successfully separated, there still remains the question of *how can the object be automatically identified or recognized?* Automatic object identification is regarded as the most difficult problem in computer vision because of being unable to define a unified feature set to represent generic natural objects. Any identification system requires an accurate representation and supervised machine learning of an object, its interaction with other objects and a physical context. Many techniques (Lou et al., 2002; Yu et al. 2002) have evolved though their performance is limited. In addition, the development of a suitable training set requires considerable manual intervention and so is time consuming (Shirasaka et al. 2006). This has prompted the adoption of RFID technology in object analysis, which has been already proved as an effective approach in application domains like object tracking in automated assembling lines (Wang, 2007), workflow optimization (Faschinger et al., 2007) and inventory control (Goodrum et al., 2006; Ko, et al., 2007).

The main drawbacks of RFID are very short sensing range, missed detection, sensitivity to environmental noise and the inherent fact that the technology does not work for certain materials like metals and food items, while some items are simply too small to have a tag attached. For these reasons, RFID alone is often ineffectual, which motivated the investigation of alternative approaches based on combining RFID and com-

puter vision techniques (Wu et al., 2007). This chapter presents an overview of contemporary approaches towards applying RFID in combination with computer vision for object analysis, involving location estimation, tracking and, object and activity recognition.

OBJECT ANALYSIS USING RFID

Object analysis applications have tended to evolve using either vision sensors or RFID tags separately rather than in combination. The latter is now gaining traction because as alluded above, the main benefit of RFID tags is they provide a general framework for object identification, which addresses the main shortcoming of visual sensor-based systems. This section presents an exploration of some contemporary research projects involving object analysis and which have exploited the potential of combining RFID and visual sensors.

Object Localisation

For many applications that require multiple targets detection such as, activity recognition and robotic vision in home environments to either monitor or track particular objects, localisation of the object is the important initial step. Based on the type of signal analysis, object localisation techniques can be classified into two categories: *i)* visual and *ii)* acoustic sensor. The latter based approaches are mainly applied to determine the location of a passive source such as the object producing a sound (Sheng et al., 2005; Rahman et al., 2005), and are normally sensitive to environmental noise. Conversely, object localisation using visual information is a classical problem for the computer vision research community due to its high accuracy requirement for location estimation, with much literature available (Park et al., 2006). Moreover, the number of approaches combining RFID with visual sensors (Kamol et al., 2007) is expanding due to the potential identification capability

of RFID, which addresses the key issue facing visual analysis techniques, namely approximate location estimation. A RFID tag strategy with both mobile and static RFID antennas together with some ceiling cameras has been proposed to determine object location in an indoor scenario (Kamol et al., 2007). The main processing steps involved are shown in Figure 1, where the location is estimated from the information derived by recognition. Colour histogram features of an object are stored in the tag attached to it. Initially RFID tags and RFID antennas are used to determine the area (approximate location) of an object which is nearby an RFID antenna. Since mobile robots attached with RFID antennas continuously roam the floor, they can detect those objects which are not detected by the static RFID antennas.

The features obtained from both the RFID tags and images captured by ceiling cameras are used to recognise an object by creating the back project of the colour histogram features, details of which are given in (Kamol et al., 2007). Following object recognition, its precise location or *location vector* is estimated by particle filtering, which used the multiple views obtained from the ceiling cameras. The authors conducted an experiment in an indoor environment comprising 10 different objects. The results revealed that both recognition accuracy and time overheads using RFID tags significantly improved compared to not using RFID tags, with a 34.5% improvement in recognition accuracy achieved allied with a reduction in computational time of around half an hour. They also showed the average localization error was around 0.96%, i.e., 79cm of the room diameter.

Object Tracking

Object tracking is another very challenging research problem, with numerous diverse real-world applications including the military, hospital, manufacturing lines and mining. Tracking an object of interest using the *global positioning*

Figure 1. Processing steps for localisation using the method articulated in (Kamol et al., 2007)

system (GPS) (Lee, 2009) is popular and widely adopted provided the object to be tracked is known *a priori*. Since GPS determines the object location using information obtained from satellites, it is highly sensitive to shadowing effects caused by surrounding structures such as buildings, bridges, hills and streets for example (Hosrt, 2007). Depending on the weather conditions, noise, radio signal, satellite position and shadowing effect, the location estimation error for GPS can be in the range of 1 to 10 metres¹, and for this reason, researchers are attempting instead to track objects using their attributes, like sound and appearance.

Tracking passive sources have been used in many applications. These applications primarily use acoustic sensors which are normally low cost and easy to deploy, though to determine the location of any object requires knowledge of the location of at least three sensors (Rahman et al., 2005), which can be affected by environmental noise so compromising the location estimate. Object tracking using only RFID tag is widely applied in automatic manufacturing lines (Wang et al., 2007) however this is very susceptible to the prevailing environment and often produces inaccurate object positions and movement paths.

To solve this problem, visual sensors have been proposed as they afford both reliable and consistent location estimation (Yilmaz et al., 2006). As mentioned previously, their main drawback however, is the requirement for either object identification using an appearance-based feature or coarse localisation. In order to provide coarse location information about an object, a new paradigm is emerging which jointly exploits RFID coverage and identification facility together with the reliability and location estimation accuracy of visual sensors. Lee et al. (2007) introduced one such model consisting of the following two steps:

1. Coarse location estimation with RFID tags
2. Refinement of the coarse location using visual sensors

These will now be respectively discussed.

Coarse Location Estimation with RFID Tags

While the accuracy and reliability of visual sensors-based location estimation is high (Lee et al., 2007), their disadvantage is requiring a

Figure 2. Overlapping regions ($S_1, \dots, S_7$) amongst the coverage of RFID readers ($R_1, \dots, R_3$)

priori information in terms of an approximate location estimate of an observed object. Coarse location estimation using RFID readers potentially bridges this hiatus. A viewing area is covered by a number of RFID readers, where the coverage of a RFID reader depends on its range. This inevitably leads to there being a number of overlapping and non-overlapping RFID coverage regions. To utilise these regions for coarse localization, the concept of *virtual sensing nodes* is introduced. As an example, three RFID readers (and their corresponding virtual sensing nodes (are displayed in Figure 2.

The maximum number of visual sensing nodes is represented by all possible combinations of RFID readers, which in this example is 7 (${}^3C_1 + {}^3C_2 + {}^3C_3$), with the linkage between RFID coverage and the virtual sensing node being shown in Figure 3. For instance, readers R_1 , R_2 and R_3 respectively cover the set of virtual sensing nodes: $\{S_2, S_3, S_5, S_6\}$, $\{S_1, S_2, S_3, S_4\}$, and $\{S_3, S_4, S_6, S_7\}$.

Each virtual sensing node is represented by a reference point, which is the centre of its rough sensing range. The reference point of each virtual sensing node in Figure 4 is shown by $\odot$. The coarse localisation of an object is the reference point of a virtual sensing node covering that object. The actual object location ($\odot$) at different time slots from t_1 to t_5 and its real movement path are indicated by the solid lines in Figure 4, while the estimated tracking trajectory with the coarse localisations are displayed by dotted lines.

Refinement of the Coarse Location Using Visual Sensors

As shown in Figure 4, the course localisation error depends on the sensing range of a visual sensing node with in the worst case, it being equal to the RFID reader coverage range. To reduce this error, two parallel projection cameras are used to take snapshots of the object to be tracked. The virtual viewable plane of each camera is the image frame generated by the parallel projection of an object. The object plane passes through the actual position of an object and is parallel to its corresponding virtual viewable plane i.e. P_y^o is parallel to the virtual viewable plane for the first camera, while P_x^o is for the second. Let P_y^c and P_x^c be the parallel planes passing through the estimated coarse location (x_c, y_c) . The location (x_r, y_r) approximated with the visual sensors (two cameras) is then defined as:

$$x_r = x_c \pm \Delta x \quad (1)$$

$$y_r = y_c \pm \Delta y \quad (2)$$

where Δx and Δy are the distance between P_y^o and P_y^c , and P_x^o and P_x^c respectively. The sign $\pm$ reflects the relative positions of an object with respect to its estimated coarse location (x_c, y_c) .

Figure 3. RFID reader coverage for virtual sensing node regions

Figure 4. The reference points of virtual sensor nodes shown by ©. Dotted line - tracking trajectory based on coarse estimation; solid line - the actual moving path.

Simulation Results for Object Tracking

The authors conducted two experiments for a single object tracking inside a room and considered both one and two RFID readers, together

with two visual sensors – Camera 1 (C_1) and Camera 2 (C_2) located on two walls of the room. The simulation setup is shown in Figure 5. An example of the object trajectory for two RFID readers, namely R_1 and R_2 is also presented in Figure 5 where the positions of the two RFID readers are represented by ©.

The simulation results showed that the usage of visual sensors improved the location estimation accuracy and does not depend on the number of RFID readers i.e., the tracking accuracy was almost the same for both one and two RFID readers, though the more RFID readers were used, the better the coarse localization accuracy achieved.

The proposed method is entirely dependent on parallel projection cameras which are generally not available (Chai & Shum, 2000). Moreover, it raises many questions over the effectiveness of this method in practical field deployments for both single and multiple objects tracking. For instance, how will the camera automatically detect an object from a cluttered and complex background scene? And, since RFID reader coverage for the passive tags is generally between 4cm and a few metres, how effective will this method be for location estimation in larger field deployments?

Activity Recognition

There are many vision-based techniques in the literature that can semantically interpret an object based on activity patterns analysis (Lou et al., 2002), perform behaviour-based similarity measurement using motion (Shechtman et al., 2007) and learn to recognize human action sequences using eye and head movements (Yu et al. 2002). All these methods are constrained however, to a specific action for a particular object (especially for humans) and only a fixed environment is ever considered. In most techniques, objects are represented by spatio-temporal video features such as colour/brightness distribution, height and width. However, these are insufficient to represent all object types since typical video sequences usu-

Figure 5. Object (★) trajectory determined using two RFID readers

ally have a large number of perceptual objects and a multitude of variations amongst them. In most cases, temporal information is considered with either *Hidden Markov Model* (HMM) (Yu et al. 2002) or *Dynamic Bayesian Networks* (DBN) (Murphy, 2002), both of which use supervised machine learning techniques. To improve the recognition of action, the relationship between an object and other objects that exist in its context is exploited (Han et al., 2009). Even with the rapid advances in computer vision, activity recognition remains extremely difficult in terms of object separation and recognition and it is very much in this context that the alternative paradigm of using RFID technology for activity and object recognition has evolved.

RFID cannot alone produce a robust recognition system because of the many limitations previously delineated, such as the reader being too far away or wrongly sensing a tag if the reader is close to another object. In addition, there are many objects which are either so small that it is

not possible to attach tag on them or where RFID does not work such as for metallic and food objects. This provides the backdrop for exploiting visual information in conjunction with RFID and *common-sense* knowledge, to directly solve the problems created by incorrect and missing tag detections and losing objects without tags attached.

Any robust recognition system should be able to operate in different environments having varying degrees of complexity over a range of objects and activities. This led to the development of learning-based models, which ultimately restricts scalability, because in such models, objects used in an activity belong to a training set which needs to be identified, separated and manually labelled, which is extremely time consuming. It is clearly insuperable to do this for all images as a typical hour long video with 25fps, contains 90000 images.

To recognise an activity with a smaller number of related actions, Wu et al. (2007) introduced a scalable and robust activity recognition system,

which uses object identification information obtained by sensing its RFID tag to automatically develop the learning model. RFID tags are attached on certain objects and the object (i.e., the person) which performs an action wears an RFID reader (bracelet on their hand). When the person performs an action, the RFID reader reads and wirelessly transmits the tag information associated with an object, to the activity processing server. The recognition system then uses this tag information to identify the object involved with that action. As a representative example of an experiment, the recognition of an activity within a kitchen environment is shown in Figure 6, where the person is holding and manipulating a water jug, with other objects also having RFID tags.

A DBN model is then designed which synergistically combines RFID detections, objects usage, common-sense knowledge and video sequence information. The complete model is shown in Figure 7(a), where common-sense knowledge is encoded along the edge between an activity and object, and A_t , O_t , R_t and I_t respectively denote activity, object, RFID and the frame at time t .

The various DBN parameters are defined as follows:

$P(A_t)$ is the *a priori* distribution;

$P(O_t|A_t)$ and $P(O_{t+1}|O_t, A_{t+1})$ are the observation probabilities;

$P(A_{t+1}|A_t)$ is the state transition probability; while $P(R_t|O_t)$ and $P(I_t|O_t)$ are the output probabilities.

$P(I_t|O_t)$ is the only parameter automatically learned via RFID detection during the DBN learning process. During this process, objects are separated from their backgrounds using a change detection technique and are represented by texture features, namely Scale Invariant Feature Transform (SIFT) (Lowe, 2004). The values of all the other parameters are heuristically determined from domain knowledge (Wu et al., 2007).

The authors conducted a series of experiments for each of the models shown in Figure 7(a) to 7(e), by considering 16 kitchen activities including boiling water, making tea, coffee and cereal, packed lunch, drinking juice, taking medicine, making salad which involved in total, 33 different objects. The experiments were conducted using three different video sequences, *Video_1*, *Video_2* and *Video_3*. The recognition results of various testing scenarios using the model learned with common-sense knowledge shown in Figure 7(a) for *Video_1* are presented in Table 1.

The test data were constructed i.e., the activities and objects to be predicted were labelled using their respective models represented in Table 1. The corresponding results reveal that due to the incorporation of common-sense knowledge, the combination RFID+visual sensor produced activity and object recognition rates of 80.67% and 72.36%, respectively, compared with 80.97% and 73.30% when only the visual sensors were applied. This vindicates that after learning the model, extra information from RFID detection of the test set does not provide any further useful information irrespective of the common-sense knowledge. In contrast, RFID only recognised 64.31% and 63% of activities and objects accurately, from which it can be concluded that even learned with RFID detection, visual sensor information and common-sense knowledge, RFID alone in a test set produces inferior recognition. Similar observations can be made for the results for both *Video_2* and *Video_3*, though *Video_3* contained mainly texture-less objects that were not detected by visual sensor alone due to the fact that in DBN, objects are only represented by texture features, namely SIFT.

The validation test of the learned model was also performed using *Video_1* and *Video_2*, with the model learning with the former and tested by the latter sequence. RFID only had activity and object recognition rates of 80.33% and 66.54%,

Figure 6. A sample kitchen environment. The user wearing a RFID bracelet (blue rectangle) and holding water jug (red rectangle). Note some other objects with RFID tags (green rectangle) attached.

respectively compared with 73.37% and 71.02% produced by RFID+visual sensor. The lower activity recognition rate achieved by RFID+visual sensor was directly due to its inability to classify texture-less objects such as the kettle and lunch bag.

A further series of experiments were conducted to test the robustness of the model in regard to either missing tag detections or objects that do not have tag attached. Results plotted recognition rate versus number of missing tag for both activity and object. The recognition rates for RFID+visual sensor were significantly higher than those of RFID only for both activity and objects, though in both cases, the recognition rates decrease when the number of missing tag detections increased.

While this research project presented fundamental work in activity recognition and will help to develop many practical applications, its main limitation is the use of only texture features in object representation. Key primitive object features such as shape, which is one of the most perceptual characteristics of an object, were not considered (Ahmed et al., 2007). The actions of a particular object are represented by the movement patterns of different body parts (e.g., head and eye

movement are tightly coupled with actions (Yu et al., 2002)), gesture and interactions with other objects and/or physical context, so an elementary action is reflected by a pattern of underlying motion. Furthermore, an activity has a particular sequence of actions and the interactions among objects and their movement patterns that have not been explicitly considered in this project.

Training Set Construction for Object Detection

As mention previously, most vision sensor based applications exploit the object analysis techniques and thus require the detection and recognition of an object in advance. These techniques generally use supervised learning and need to develop training sets consisting of either geometric or appearance-based object features. Creating such training data sets is laborious and incurs considerable time expenditure. As it is difficult to observe which tag ID represents which object due to the inability of RFID readers to accurately detect all tags, this mandates that vision sensors and/or cameras are able to detect and recognise objects. However, their potential is limited and varies with the complex-

Figure 7. Bayesian network structures for activity and object recognition

ity of objects and their surrounding environment. This motivated Shirasaka et al. (2006) to introduce an approach to construct a training set using the appearance features of objects, where RFID tags are used as supervisors. The various processing steps involved in this approach are presented in Figure 8.

Obtaining Target Information from RFID Tags

All the objects to be detected in an environment have RFID tags attached to them, which contains identification information. The RFID readers shown in Figure 8 detect these tags and using the identification information to determine the relevant object. To apply supervised learning in this scenario, it is essential to derive two different types of information – the appearance-based features of an object and its corresponding target

information. This latter information can be derived directly from its RFID tag, while the former can be calculated using any suitable feature approximation technique such as in (Karmakar et al., 1999). In this approach however, spatial resolution reduction together with kernel *principal component analysis* (PCA) for dimensionality reduction (Schölkopf et al., 1999) has been applied.

Object Detection and Separation

The most challenging part of this particular process is to detect and separate an object from the image captured by vision sensors. Since it is assumed the model is being deployed in a controlled indoor environment, the background complexity is generally limited and almost static. For this reason an image difference approach has been adopted in Figure 8 to separate the object from its background, though if the background changes are not insig-

Table 1. Recognition rates using different test scenarios

Common-sense knowledge used	Testing Scenarios	Activity (%)	Object (%)
Yes	RFID only (model in Figure 7 (c))	64.31	63.00
Yes	RFID+Visual Sensors (model in Figure 7 (a))	80.67	72.36
	Visual Sensor only (model in Figure 7 (d))	80.97	73.30
No	RFID+Visual Sensor (model in Figure 7 (b))	60.84	74.68
No	Visual Sensors only	62.76	74.72

nificant, the image difference approach becomes ineffectual. While the performance is satisfactory for a single object, if the image contains multiple objects, it is then necessary to determine which RFID tag corresponds to which object. How this can be achieved is not presented in (Shirasaka et al., 2006).

Storing Object Images and Corresponding ID in a Database

The RGB colour images containing the objects are obtained by the differential technique described above and then spatially reduced to 30 *pixels* square. All background pixels, except the object pixels are set equal to zero to eliminate this information and then these object images and

their corresponding RFID tags used to construct a training set which is stored in a database (see Figure 8).

Feature Dimensionality Reduction

The reduction of feature set dimensionality is a major issue as identifying the optimal set of attributes of a feature set, which perform equally well for all datasets, is extremely intractable. If the dimension of a feature set exceeds a certain limit, the feature set loses its distinguishing capability and hence compromises the detection accuracy. In addition, the higher the dimension, the greater the computational complexity, so to make the system suitable for real-time applications, judiciously applying dimensionality reduction is essential.

Figure 8. The processing steps for training set construction (Shirasaka et al., 2006)

While there are many techniques available for feature dimensionality reduction (Fodor, 2002), the authors chose the most widely adopted, namely kernel PCA (Schölkopf et al., 1999). This decreases the original dimensions of the feature set, from $30 \times 30 \times 3$ to n dimensions, where the feature set derived by kernel PCA is represented by $x = \{x_1, x_2, \dots, x_n\}$.

Object Classification Using the EM Algorithm

The overall objective is to design an object classification model using the constructed training set, i.e., to determine the *a posteriori* probabilistic function $p(id/x)$ from a training set, which can automatically detect any object (i.e., estimate the object id) from a given appearance-based feature set x . Depending on the location and contextual object, RFID readers may fail to detect the tag of an object. The *expectation maximisation* (EM) algorithm is often used in the Gaussian mixture model to estimate the maximum likelihood of missing data and concomitantly improve the computational complexity. The EM algorithm seeks to maximise the *likelihood function* $l(x / \theta_{id})$ where, θ_{id} represents the set of parameters (the mean and covariance matrix) of a normal distribution function for an arbitrary object referenced by its id.

Experiments have been conducted with three scenarios, which had one, two and three objects respectively. All objects and a RFID reader were placed both on and under a table, respectively. Vision sensors focused only upon the objects on the table. The mean classification accuracy, which is the ratio between the number of features for which the classifier classified accurately and the total number of feature used in the experiment, was a very commendable 98.7%.

The fundamental concept introduced in (Shirasaka et al., 2006) is the development of a semi-automatically learning model which avoids

significant manual processing. It advances many supervised learning-based research questions for numerous computer vision applications including tracking and surveillance. The classification accuracy of this method has been shown to be high because the experimental environment was restricted and simple. If the deployment context is more complex such as a room consisting of many different static and mobile background objects, and the objects are moving from one place to another, it is anticipated the classification accuracy of this method will decrease significantly. This is mainly due to the poor method adopted to separate objects from their background and not in considering multiple object movements, which may force a RFID reader to detect multiple objects. This confusion will lead the recognition system to match which object relates to which RFID tag. This can be resolved by storing the main object features into the RFID tags. What type of features, how many of them are required for accurate object identification; and how can they be seamlessly added to a RFID tag, remains a major research question? Some other future research directions are considered in the following section.

FUTURE RESEARCH DIRECTIONS

The potential to store colour histogram features of an object into a RFID tag to improve recognition has already been considered (Kamol et al., 2007), however colour information alone is not sufficient to recognise all possible natural objects as it cannot differentiate between objects with similar colour distributions. This mandates object representations based more on structural representations such as shape and texture (Karmakar et al., 2003). It is probable that in the future a large number of cutting edge applications will require RFID tags to store various types of object features. A myriad of natural objects and huge variations among them have made the representation of an object in a

generic form very complicated. Moreover, the perception and definition of an object is application dependent so it will be a challenging task to determine what type and which set of features should be stored in RFID tags to effectively recognise a wide variety of objects for a particular application context. This firstly requires designing RFID tags which are capable of storing essential features along with their tag id.

An activity is represented by a number of actions, and the objects and context that are involved in that action, most literature has mainly considered isolated actions. However, without identification of an object, and interpretation and interaction with other objects and context, the activity recognition performance will be limited as the same action may represent a number of different activities for different objects and contexts. For example, consider a chef's cutting action i.e., vegetable, fruits, fish and meat, with a knife in a kitchen. This can represent the cooking activity, while for a doctor in an operation theatre, performing exactly the same action is interpreted as surgical activity. There are a myriad number of activities and an activity comprises a number of hierarchically organised elementary actions distributed in different levels that make the activity recognition challenging. The various temporal sequences of the same action set may represent different activities, which makes activity recognition even more complicated. Without a proper understanding of the hierarchical relationships and temporal sequences amongst actions in conjunction with object interactions with domain or context, the potential of any activity recognition system will remain limited.

To understand and semantically interpret the behaviour of a particular object of interest from a video sequences remains a major research problem to both the artificial intelligence and multimedia technology research communities, as it covers a very wide range of applications from national security to medical diagnostics. In all these cases, the users know what types of objects they need to

detect and can then predict the object's behaviour. This raises a question on how to detect and predict the behaviour of a moving object from its given descriptor. The answer has defiantly remained elusive due to the huge number of objects and the myriad of variations amongst them. This makes the efficacy of the existing detection techniques very limited and difficult to segment any object based solely on pixel properties. There is no existing approach to both detect and separate an object for a given descriptor except a rather rudimentary technique proposed in (Ahmed et al., 2007). The descriptor could be developed automatically using RFID technology and then applied to a method similar to (Wu et al., 2007) for improving recognition rates. The activity and objects of interest can then be recognised using such descriptor and RFID technologies.

Contextual relationships are essential in identifying particular events such as a car-bomb attack. This can for example, involve the making and distributing of the bomb, identification of the bomb maker, abnormal behaviour of a person, psychological profiling, and unattended bags in an airport. If we can understand and consider the social context in which an event is initiated, developed, and finally occurs, it will add another valuable modality to the decision making. For example, which events or acts or contexts make a child hyperactive or a mentally ill patient violent? The inclusion of this information means a better likelihood that the prediction of an object's behaviour will be more accurate, and where it is currently impossible to detect suspicious events, it may make them at least feasible. This will open a new avenue of research in for example, child growth and mentally ill patient analysis based on applying RFID technology within a controlled environment. McFate (2005) analysed the social environment in which *improvised explosive devices* originated, and were subsequently produced, distributed and applied. He discovered that social network analysis provided a valuable tool to locate

insurgents within the tribal networks of Iraq, which ultimately led to the capture of Saddam Hussain.

Chen et al. (2004) have analysed social interaction patterns and provided reports of activities and behaviour of patients in nursing home environments. In their research, the labelling of doors, walls, ceilings and floors were performed manually, so this is not suitable for dynamic environments consisting of complex objects, especially for arbitrarily shaped objects and where their locations are not *a priori* known. To address this issue, the identification objects (e.g., human, contextual information - doors, walls, ceiling and floors) could each have RFID tags attached to them and their various features stored to give impetus to the computer vision field in managing things automatically, just like we humans can do in any given context.

CONCLUSION

Object analysis is a fundamental processing step in many leading-edge applications involving location estimation, tracking and monitoring, activity recognition, event detection, and behaviour analysis and prediction. Either vision sensors or RFID technology alone have been shown to be inadequate in terms of their performance which has promoted the introduction and investigation by the computer vision community of hybrid approaches based on a combination of vision sensors and RFID. This chapter has presented a contemporary review of these evolving hybrid approaches. The advantages and drawbacks of each approach have been highlighted along with some insightful future research directions. This has revealed key aspects, such as the interactions with other objects, deployment contexts and social interactions which are deemed essential to consider in the analysis and prediction of event behaviour. This chapter has also offered some perceptive insights into the future research scope that may help to improve the human condition.

REFERENCES

- Ahmed, R., Karmakar, G. C., & Dooley, L. S. (2007). Probabilistic spatio-temporal video object segmentation using a priori shape descriptor. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-07)*, 1, (pp. 15-20).
- Bashir, F. I., Khokhar, A. A., & Schonfeld, D. (2007). Object trajectory-based activity classification and recognition using hidden Markov models. *IEEE Transactions on Image Processing*, 16(7), 1912–1919. doi:10.1109/TIP.2007.898960
- Baxes, G. A. (1994). *Digital image processing: Principles and applications*. New York, NY: John Wiley & Sons, Inc.
- Chai, J., & Shum, H. (2000). Parallel projections for stereo reconstruction. *IEEE Conference on Computer Vision and Pattern Recognition*, 2, (pp. 493-500).
- Chen, D., Yang, J., & Wactlar, H. D. (2004). Towards automatic analysis of social interaction patterns in a nursing home environment from video. *ACM SIGMM International Workshop on Multimedia Information Retrieval*, (pp. 283-290).
- Faschinger, M., Sastry, C. R., Patel, A. H., & Tas, N. C. (2007). An RFID and wireless sensor network-based implementation of workflow optimization. *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM07)*, (pp. 1-8).
- Fodor, I. K. (2002). *A survey of dimension reduction techniques*. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, Center for Applied Scientific Computing, USA.
- Goodrum, P., McLaren, M., & Durfee, A. (2006). The application of active radio frequency identification technology for tool tracking on construction job sites. *Automation in Construction*, 15(3), 292–302. doi:10.1016/j.autcon.2005.06.004

- Han, D., Bo, L., & Sminchisescu, C. (2009). Selection and context for action recognition. *IEEE 12th International Conference on Computer Vision (ICCV09)*, (pp. 1933-1940).
- Hosrt (2007, August 19). *GPS accuracy—Myth and truth—Part 1—DGPS/Shadowing effect*. Sensors-GPS, <http://autocom.wordpress.com/2007/08/19/genaues-gps-mythos-und-wahrheit-teil-1-dgps-abschattung/>
- Hsiao, K., Vosoughi, S., Tellex, S., Kubat, R., & Roy, D. (2008). Object schemas for responsive robotic language use. In the *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, (pp. 233-240).
- ISO/IEC 14496-2. (1999). *Information technology—Coding of audio-visual objects. Part 2: Visual*.
- Kamol, P., Nikolaidis, S., Ueda, R., & Arai, T. (2007). RFID based object localization system using ceiling cameras with particle filter. In the *Proceedings of the Future Generation Communication and Networking (FGCN07)*, 2, (pp. 37–42).
- Karmakar, G. C. (2003). *An integrated fuzzy rule based image segmentation framework*. Australia: Gippsland School of Information Technology, Monash University.
- Karmakar, G. C., Dooley, L., & Rahman, S. M. (2001). Review on fuzzy image segmentation technique. In Rahman, S. M. (Ed.), *Design and management of multimedia information systems: Opportunities and Challenges* (pp. 282–313). Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-930708-00-6.ch014
- Karmakar, G. C., Rahman, S. M., & Bignall, R. J. (1999). Composite features extraction and object ranking. In the *Proceedings of the International Conference on Computational Intelligence for Modeling, Control and Automation (CIMCA'99)*, Vienna, Austria, (pp. 134 – 139).
- Ko, Y., Roy, S., Smith, J. R., Lee, H., & Cho, C. (2007). An enhanced RFID multiple access protocol for fast inventory. *IEEE Global Telecommunications Conference (GLOBECOM07)*, (pp. 4575 – 4579).
- Lee, J. (2009). Global positioning/GPS. In Kitchen, R., & Thrift, N. (Eds.), *International encyclopedia of human geography* (pp. 548–555). doi:10.1016/B978-008044910-4.00035-3
- Lee, J., Park, K., Hong, S., & Cho, W. (2007). Object tracking based on RFID coverage visual compensation in wireless sensor network. *IEEE International Symposium on Circuits and Systems (ISCAS)*, (pp. 1597 – 1600).
- Liu, J., Bowyer, K., Goldgof, D., & Sarkar, S. (1997). *A comparative study of texture measures for human skin treatment*. Presented at International Conference on Information, Communications and Signal Processing (ICICS '97), Singapore.
- Lou, J., Liu, Q., Tan, T., & Hu, W. (2002). Semantic interpretation of object activities in a surveillance system. *International Conference on Pattern Recognition*, 3, 777 – 780.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94
- McFate, M. (2005). Iraq: The social context of IEDs. *Military Review*, (May-June): 2005.
- Milanés, V., Llorca, D. F., Vinagre, B. M., González, C., & Sotelo, M. A. (2010). Clavileño: Evolution of an autonomous car. *International IEEE Annual Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, (pp. 1129-1134).
- Murphy, K. P. (2002). *Dynamic Bayesian networks: Representation, inference and learning*. Berkeley, USA: University of California.

- Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26, 1277–1294. doi:10.1016/0031-3203(93)90135-J
- Park, S., Kim, K., Park, S., & Park, M. (2006). Object entity-based global localization in indoor environment with stereo camera. *Proceedings of SICE*, (pp. 2681-2686).
- Pentland, A. (1999). Modelling and prediction of human behaviour. *Neural Computation*, 11, 229–242. doi:10.1162/089976699300016890
- Pham, D. L., & Prince, J. L. (1999). An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recognition Letters*, 20, 57–68. doi:10.1016/S0167-8655(98)00121-4
- Rahman, M. Z., Karmakar, G. C., & Dooley, L. S. (2005). Passive source localization using power spectral analysis and decision fusion in wireless distributed sensor networks. *International Conference on Information Technology: Coding and Computing (ITCC-05)*, Las Vegas, USA, (pp. 260–264).
- Scholkopf, B., Smola, A., & Müller, K. (1999). Kernel principal component analysis. In Scholkopf, B., Burges, C. J. C., & Smola, A. J. (Eds.), *Advances in kernel methods—Support vector learning* (pp. 327–352). Cambridge, MA: MIT Press.
- Shechtman, E., & Irani, M. (2007). Space-time behaviour based correlation or How to tell if two underlying motion fields are similar without computing them? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11), 2045–2056. doi:10.1109/TPAMI.2007.1119
- Sheng, X., & Yu-Hen Hu, Y. (2005). Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Transactions on Image Processing*, 53(1), 44–53.
- Shirasaka, Y., Yairi, T., Kanazaki, H., Junichi Shibata, J., & Machida, K. (2006). Supervised learning for object classification from image and FDID data. *SICE-ICASE International Joint Conference*, Bexco, Busan, Korea, (pp. 5940-5944).
- Skinner, B. F. (1953). *Science and human behavior*. New York, NY: The Free Press.
- Smith, J. R., Fishkin, K. P., Jiang, W. B., Philipose, M., Rea, A. D., Roy, S., & Rajan, K. S. (2005). RFID-based techniques for human-activity detection. *Communications of the ACM*, 48(9), 39–44. doi:10.1145/1081992.1082018
- Song, X., & Fan, G. (2006). Joint key-frame extraction and object segmentation for content-based video analysis. *IEEE Transactions on CSVT*, 16(7), 904–914.
- Veltkamp, R. C., & Tanase, M. (2000). *Content-based image retrieval systems: A survey*. Technical Report UU-CS-2000-34, Department of Computing Science, Utrecht University.
- Wang, J., Luo, Z., Wong, E. C., & Tan, C. J. (2007). RFID assisted object tracking for automating manufacturing assembly lines. *IEEE International Conference on e-Business Engineering*, (pp. 48-53).
- Wikipedia. (2011). *Computer vision*. Retrieved October 1, 2011, http://en.wikipedia.org/wiki/Computer_vision
- Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., & Rehg, J. M. (2007). A scalable approach to activity recognition based on object use. *IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, (pp. 1–8).
- Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys*, 38(4), 1–45. doi:10.1145/1177352.1177355

Yu, C., & Ballard, D. H. (2002). Learning to recognize human action sequences. *International Conference on Development and Learning*, (pp. 28-33).

KEY TERMS AND DEFINITIONS

Computer Vision: A field of computer science that uses image analysis for retrieving and understanding the information of an event in real-time.

GPS: The global positioning system which determines the object location using information obtained from satellites.

Image Analysis: An image processing technique which includes image segmentation, feature extraction, and object classification.

Object Localisation: The process of locating an object.

Recognition: To identify an object or event.

RFID: Radio Frequency Identification technology which uses radio waves to read the data stored in an electronic tag.

Segmentation: The process of separating mutually exclusive homogeneous regions or objects of interest.

SIFT (Scale Invariant Feature Transform): A feature approximation technique which is invariant to scale, rotation, and translation.

Training Set: A data set used for learning of a machine.

Visual Sensor: A digital electronic device which is used to capture images

ENDNOTE

- ¹ <http://www.maps-gps-info.com/gps-accuracy.html>