2008

Modeling and performance analysis of a UAV-based sensor network for improved ATR

Xiaohan Chen
West Virginia University

MODELING AND PERFORMANCE ANALYSIS OF A UAV-BASED SENSOR NETWORK FOR IMPROVED ATR

XIAOHAN CHEN

Dissertation submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Electrical Engineering

Natalia A. Schmid, Ph.D., Chair
Matthew C. Valenti, Ph.D.
Xin Li, Ph.D.
Powsiri Klinkhachorn, Ph.D.
Erdogan Gunel, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2008

Keywords: Automatic Target Recognition, Sensor Network, PCA, Channel Capacity,
Error Exponent, Probability of Outage

ABSTRACT

MODELING AND PERFORMANCE ANALYSIS OF A UAV-BASED SENSOR NETWORK FOR IMPROVED ATR

Xiaohan Chen

Automatic Target Recognition (ATR) is computer processing of images or signals acquired by sensors with the purpose to identify objects of interest (targets). This technology is a critical element for surveillance missions. Over the past several years there has been an increasing trend towards fielding swarms of unattended aerial vehicles (UAVs) operating as sensor networks in the air. This trend offers opportunities of integration ATR systems with a UAV-based sensor network to improve the recognition performance. This dissertation addresses some of design issues of ATR systems, explores recognition capabilities of sensor networks in the presence of various distortions and analyzes the limiting recognition performance of sensor networks.

We assume that each UAV is equipped with an optical camera. A model based recognition method for single and multiple frames is introduced. A complete ATR system, including detection, segmentation, recognition and clutter rejection, is designed and tested using synthetic and realistic images. The effects of environmental conditions on target recognition are also investigated.

To analyze and predict ATR performance of a recognition sensor network, a general methodology from information theory view point is used. Given the encoding method, the recognition system is analyzed using a recognition channel. The concepts of recognition capacity, error exponents and probability of outage are defined and derived for a PCA-based ATR system. Both the case of a single encoded image and the case of encoded correlated multiple frames are analyzed. Numerical evaluations are performed. Finally we discuss the joint recognition and communication problems. Three scenarios of a two node recognition sensor network are analyzed. The communication and recognition performances for each scenario are evaluated numerically.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Dr. Natalia A. Schmid, for guiding me deeply and brightly during my four-year research and study in Morgantown, WV. Her perseverance, extraordinary vision and love of life have inspired me. The thesis is impossible without her invaluable guidance, generosity and patience. She has devoted so much time and effort to teaching me high standards in doing and communicating my research.

I am also grateful to all the faculty members in the Lane Department of Computer Science and Electrical Engineering. Special thanks go to Dr. Matthew C. Valenti, Dr. Xin Li, Dr. Tim McGraw and Dr. Tim Menzies for their excellent teaching efforts on the courses that I took during my Ph.D. career. I want to thanks Dr. Powsiri Klinkhachorn and Dr. Erdogan Gunel for accepting to be my committee members, taking the time to read my thesis and providing me with helpful feedbacks.

I would like to thank all colleagues and friends for their friendship and support. My special thanks go to Jinyu Zuo, Francesco Nicolo, Nathan Kalka and Shanshan Gong in the Statistical Signal Analysis Lab. Exchanges of ideas and useful discussions with them are always a big source for my work. I also want to thank Shi Cheng for his great help in communication related work.

I would also like to thank Augusta Systems Inc. for supporting a part of my research.

Finally, I want to thank my parents for their endless love and support, and my husband, Jinwen Xi, for his love and encouragement. The sunshine smiles on their face are the source of my happiness.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

OVERVIEW AND OBJECTIVES

Automated Target Recognition (ATR) is a very specific field of study within the general scope of image processing and image understanding. This technology is a critical element of surveillance missions. A variety of sensors and techniques have been developed to approach the problem. Recent researches in ATR suggest that two or more sensors may significantly improve the overall system performance. Over the past several years, swarms of Unmanned Aerial Vehicles (UAVs) equipped with cameras operated as large-scale sensor networks in the air. This provides opportunities of integration ATR systems with UAV-based sensor networks to enhance and augment the surveillance and reconnaissance abilities.

In the past, the extensive research related to UAV networks has been mostly focused on designing and evaluating communication protocols, describing strategies to control UAVs, evaluating collision detection capabilities, etc. With respect to ATR protocols, a vast literature describing various strategies and approaches is available. The designed algorithms range from purely deterministic structural approaches to complex neural networks recognizing 3D objects or complex stochastic models. In spite of this vast research related to UAVs and their control and a large number of ATR protocols with claimed exclusive recognition performance available, the problem of designing recognition protocols utilizing optical imagery acquired by a sensor network (UAV-based recognition network) and the problem of quantifying limits and limitations of recognition sensor networks in ideal environment and highly complex practical environment have not been solved. Without understanding limits and capabilities of using sensor network in unconstrained practical environment, reliable identification of objects will not be possible. Without understanding how the environment influence the performance of recognition network, the optimized designs compensating distortions in images due to the environment will also not be possible.

This thesis approaches these problems and proposes a number of solutions to them. The main objective of this thesis is to design a complete object recognition system that operates on long-range optical imagery, explore the capabilities of the designed system

to recognize objects under a variety of environmental and camera effects, which are typical features in practical situations, and evaluate large scale recognition capabilities of designed system. The remainder of the chapter is organized as follows. The structure of ATR systems is introduced in Section 1.1. The challenges to achieve high performance are also discussed in this section. In Section 1.2, we briefly describe a swarm of UAVs. Finally, the organization of the thesis and the main contribution are summarized in Section 1.3 and Section 1.4.

1.1 Automatic Target Recognition

Automatic Target Recognition (ATR) is the computer processing of images or signals (data), acquired by optical, radar, infrared, or other imaging sensors with the purpose to identify objects of interest (targets) based on information contained in the images or signals [2]. Human operators cannot reliably identify targets on a continuous basis due to the need for rapid reaction times, difficulty in interpreting observations from all sensors, and general cognitive overload [41]. These limits are the origin of the concept of Automatic (or aided) Target Recognition. The application of ATR technology is a critical element of the future success of Intelligence, Surveillance and Reconnaissance (ISR) missions.

Fig. 1.1 shows a block-diagram of a traditional ATR system. An ATR system can be decomposed into three major subsystems. The first subsystem performs preprocessing of acquired data. Preprocessing may involve denoising, contrast adjustment, normalization, etc. The second subsystem is to obtain a set of subregions that indicate what pixels in the original image are likely to belong to targets. By locating the subregions, we can filter out a large amount of background clutter from the terrain scene, making object recognition feasible for large data sets. At the last step, the preprocessed signals in each subregion are utilized to compute the values of a set of attributes or feature vectors. And the classifier takes the feature vectors and returns the target's identity by comparing query data against a target library. Other information such as a priori probability of occurrence and position coordinates from Global Positioning System (GPS) can also be incorporated with the information of targets to improve recognition performance. Some systems may have additional preprocessing blocks such as a segmentation subsystem that can be placed between the detection and classification blocks in order to extract the targets from the background as accurately as possible and thus reduce redundancy in the data.

Figure 1.1. Structure of an ATR system

The topic of the ATR has been researched for decades. A large number of systems utilizing a variety of sensory data have been designed. However, the recognition capabilities of most of systems remain unsatisfactory for using these fully automatic ATR systems in practice. A particular problem is high false alarm rates [44]. The main reasons for unsatisfactory performance of ATR systems in practice is in using idealized data for estimating parameters of the systems (training) and evaluating performance (testing). The uncontrollable imaging conditions, complex and unknown background, obscuration of targets and low signal-to-noise ratios characterizing practical data are the challenges that most ATR systems face. A practical solution to this problem is to use more information about targets and background, which can be achieved by fusing data. This has been proved to result in performance improvement. However, establishing a theoretical foundation for ATR is the key point to understand and predict ATR performance, since it is impossible to involve all situations in training. In this thesis, we will focus on the classification part and on the analysis of the recognition performance.

1.2 Unmanned Aerial Vehicles

Unmanned Aerial Vehicles (UAVs) are remotely piloted or self-piloted aircrafts that can carry sensors, communication equipments or other payloads. UAVs not only decrease the risks confronted by military personnel, but also have long flight durations and large flight heights because they are not burdened with the physiological limitations of human pilots. By taking different equipments, UAVs have the capability to perform multiple missions. Over the last several years, UAVs have served to enhance and augment the surveillance and reconnaissance abilities of the military.

One of the primary concerns about UAV applications is “gold plating”. The fear is that good designs will become loaded up with more sensors and more missions until they become too expensive to build or too valuable to use [8]. One of the practical solutions is to build an UAV system composed of several UAVs with indispensable equipments. The other problem for most early UAVs is that they lacked autonomy and had no on-board ATR capability. UAVs acquire sensory data and send the data to a central location such as a base station, where potential targets are identified using image analysis algorithms

[21]. However, this centralized model for ATR possesses a number of drawbacks such as scalability with the number of UAVs and network delays in communicating with the central location.

In developing the next generation of UAVs, one of the ideas is to utilize reactive agents and the associated swarming behavior as part of the command and control system for a group of UAVs functioning cooperatively and independently from ground control. Previous works [7, 14] demonstrate that this technique provides a suitable mechanism to assimilate the capabilities of individual UAVs into a group of coordinating UAVs that perform ATR in a distributed manner. In a swarmed system, multiple mobile entities are directed to converge on a single point of interest, disperse and regroup again. To achieve distributed ATR using swarming, each UAV individually searches for potential targets within an area of interest using its image sensor. As soon as the image of an object is sensed to be a possible target by a UAV, other UAVs cooperate with it by swarming towards the potential target to collectively perform ATR in a data fusion manner and confirm the object as a target. One of swarm intelligence techniques, digital pheromone is utilized to coordinate the movements of multiple UAVs based on a computational analog of pheromone dynamics. In contrast to a centralized model for ATR, a distributed ATR model possesses a potential to provide minimal user intervention, a high level of robustness and largely autonomous operation. In this thesis, we are focusing on the integration of an ATR capability into the UAV system exhibiting swarmed behavior in order to maximize ATR effectiveness for the UAV system.

1.3 Thesis Overview

In this chapter, we reviewed the concepts and current situations of ATR and UAV-based networks. We will then review the related works in Chapter 2, including sensor selection and literature research on state-of-the-art ATR methods and performance analysis. We motivate use of optical sensors. In Chapter 3, two databases, a simulated database generated using a 3D Optical ATR Tool provided by Augusta Systems, Inc and a real image database, COIL-100, are described. A set of distorted data, including distortions due to illumination, contrast, noise, motion blur and defocus blur, are also simulated.

This thesis contains two parts. Chapter 4-6 form the first part, where we focus on building an ATR system and exploring the capabilities of the designed system to recognize objects under different environmental and camera effects. The second part of the thesis is

mainly on recognition performance analysis, including Chapter 7-10. Fig. 1.2 illustrates relationships between chapters.

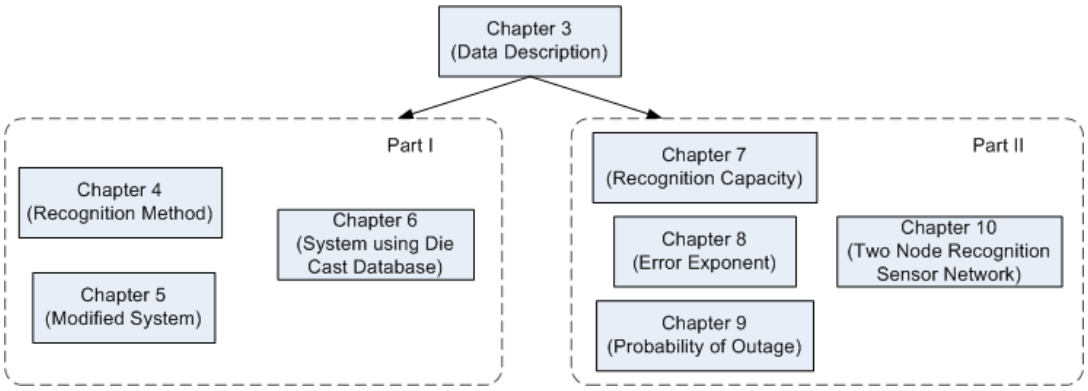


Figure 1.2. Structure of the thesis

In the first part, we will first introduce a model based recognition method, Bessel K forms-based method in Chapter 4 and describe a complete ATR system in Chapter 5, including a Haar-like feature based detector, a region-based segmentation method using B-splines and a post processing to reject non-targets. The recognition performances under different environmental conditions and camera effects are further evaluated using synthetic images. In Chapter 6 a real image database generated using 6 die cast models is introduced . Distorted images are taken by adjusting the lighting conditions and the camera parameters. The recognition method designed in Chapter 4 is then applied. The effect on recognition performance using distorted images is also evaluated. The overall system performance is then tested using the images taken in the simple and complex environments.

In the second part, we focus on a PCA-based recognition system. In Chapter 7, a model-based approach is applied to encoded sensory data, and a concept of recognition capacity is introduced. The expression for the capacity of a recognition system under the constraint of PCA-based encoding is further derived. The joint recognition capacity of multiple images is also defined, and the expression for the multi-frame capacity is obtained. We also define the recognition rate and analyze the empirical mutual information rate as a function of the recognition rate. The empirical capacity is estimated from the sequence of empirical mutual information rates. Chapter 8 describes the empirical recognition reliability function using PCA encoding. The exponents of the random coding lower bound and the space partitioning upper bound are further derived for both single- and multiple- frame recognition channels. In Chapter 9, the concept of recognition proba-

bility of outage is introduced. Two types of outage are defined for PCA-based recognition system. The numerical results are evaluated using images at different distortion levels.

The joint recognition and communication problem is discussed in Chapter 10. The performance measures introduced in the previous three chapters are used to characterize the recognition channel with the constraint of wireless communication channel undergoing Rayleigh block fading. The system probability of outage is the recognition probability of outage using the transmitted PCA codes. Three scenarios of operation of a two node network are described and a comprehensive performance evaluation is performed by changing the parameters of the communication channel.

Finally, in Chapter 11, we conclude the thesis and discuss the potential directions for future research.

1.4 Contributions

The main contributions of the thesis are listed below.

1. A model-based recognition method using Bessel K forms is introduced for target recognition. A combination algorithm is designed to balance accuracy and speed. A recognition method using two images from the same target is designed based on multivariate Bessel K forms and its performance evaluated.

2. A complete ATR system including detection, segmentation, recognition and clutter rejection is proposed and tested using both simulated 3D dataset and a dataset of real objects.

3. The recognition performance of designed systems is evaluated varying environmental conditions.

4. The concept of recognition channel and recognition capacity are introduced. For an ATR system, the PCA encoded data from both a single image or correlated multiple frames are statistically modeled. The PCA constrained recognition capacity is derived and evaluated numerically. The recognition reliability function and the recognition probability of outage of a PCA-based ATR system are further defined and analyzed.

5. Performance of a small wireless sensor network for object recognition is evaluated numerically. The system probability of outage is defined to measure the overall system performance. Three protocols of operation of a two node sensor network are analyzed.

CHAPTER 2

RELATED WORKS

In this chapter, we will briefly introduce the diverse sensors used in ATR systems and select one sensor source as the research object. The state-of-the-art ATR methods and performance analysis methods are further presented based on the selected sensor.

2.1 Sensor Selection

Here we present a brief review of sensor sources utilized in ATR systems. For more detailed discussions see [5, 23, 44]. Sensors can be divided into two main classes: Active, which require the generation of a signal to scan the scene, and Passive, which rely in the energy provided by the scene. Table 2.1 summarizes the various advantages and critical issues for sensor options.

Generally, there is no single type of sensor that clearly dominates the others with respect to all characteristics. Some sensors, such as Forward-looking infrared (FLIR) and Synthetic aperture radar (SAR), are suitable for detection, since they have a wider all-weather capability, can penetrate vegetation to some extent and detect targets at ambient temperatures. But they may not perform well in recognition or maybe too bulk or expensive. Target recognition generally requires more information and greater resolution than that required for detection. FLIR and SAR signatures are variable and ambiguous [45]. Researchers in [1, 44] have demonstrated that two or more sensors can significantly improve overall performance of ATR systems.

The selection of sensors also depends on the whole system requirements (weight, size, capabilities). In this research, the sensors are installed on the UAVs. Since the purpose is to build a cheap system, the UAVs traditionally used by military for surveillance purpose are relatively small in size and light in weight. Thus, a relative light and inexpensive sensor is needed. Here only Electro-Optical (EO) sensors, such as near infrared sensor and Charge Coupled Device (CCD) camera, are involved. In our research, we use optical images as the sensor observations and assume that each UAV has a camera on-board.

Table 2.1. Performance Tradeoffs for ATR Sensor Options (Adapted From [23])

Sensor	Advantages	Critical Issues
Forward-looking infrared (FLIR)	High target to background contrast	False alarm from background clutter, vegetation, & animals
Passive	Day & night operation	Range uncertainty
	Penetrates fog, haze & dust	Occlusions from terrain & vegetation
		Target signature variability
		Aspect angle dependence
Millimeter Wave (MMW) Radar	All weather	False alarms from background clutter, rocks, isolated buildings, & metal structures
Active	Day & night operation	Terrain occlusions
		Target signature varies with aspect angle
Synthetic Aperture Radar (SAR)	All weather	False alarms from background clutter,rocks, isolated buildings, and metal structures
Active	Day & night operation	Terrain occlusions
	Large target to background contrast	Target signature varies with aspect angle
Laser Radar	Penetrates fog, haze & dust	Target signatur varies with aspect angle
Active	Potentially high level of discrimination	Complex technology
	for range map signatures	Power requirements
	Vibration signature show promise	Requires long dwell time on target
	Doppler laser radar for moving targets	Very precise tracking & stabilization required
Electro-Optical (EO)	Light weight	Relatively low target to background contrast
Passive	Inexpensive	No night or all weather capability
	High Resolution	
	Reliable	

2.2 State-of-the-art ATR Methods based on Optical Images

An ATR system first detects targets and then identifies them. A general block-diagram was displayed earlier in Fig. 1.1. It is important to contrast detection with the problem of recognition. A target detection system knows how to differentiate targets from everything else, while a target recognition system knows the difference between a target A and a target B. The recognition methods vary as the signals/images from different sensors and the size of objects. In this section, we will focus on the recognition system with targets imaged using EO sensors. An efficient automatic classifier should be able to select those informative features (geometric, topological, spectral, etc.), which maximize the similarity of objects from the same class and minimize the similarity of objects from different classes. Similar to most recognition systems, ATR system based on optical images are broadly classified into 3 categories (based on encoded information or features). The categories are: (1) Shape-based method; (2) Appearance-based method and (3) Computer Aided Design (CAD)-based method.

In shape-based recognition ([4, 35]), the contour or silhouette of the object is extracted, then the shape templates are used to match the extracted contour. Hausdorff distance is used as matching criterion. The recognition performance of these algorithms is sensitive to the accuracy of an edge detection method. In appearance-based (or view-based) approach, the 2D intensity templates of 3D target acquired from different viewpoints are stored as a model. Some view-based methods ([46, 66]) use statistical techniques to analyze the distribution of the target image vectors in the vector space, and derive an effective representation (feature space) according to different applications. Other methods ([33, 43]) design distortion-invariant filters to perform a correlation matching between the model view and the input image. In practice, the number of available training images covering different target poses is small, which limits the performance of appearance-based target recognition systems. In CAD-based ATR ([40, 56]), an explicit 3D model of a target is generated and subsequently used in target recognition employing imagery acquired by a variety of sensors. Target models coupled with environmental affects models presumably can represent any state in which the target can occur. Then, the images that the sensor produces are compared to the library models until a match occurs with some level of confidence. The main step of CAD-based ATR is to estimate the pose of the CAD model so that the projection of 3D model matches with the query image. However, it is not possible to acquire CAD models of all targets. In [65], a multi-view morphing algorithm is generated to provide 3D model using several images.

Some classification schemes that have been used for target recognition are K-nearest neighbor, linear and quadratic discriminators, tree-based classifiers, multi-layer neural networks, Support Vector Machine (SVM), etc.

There are also some other methods, such as morphological detection [47] and anomaly detection [10] for small target (less than 20×20 pixels). In our research, the target size is from 100×100 pixels to 200×200 pixels and CAD models of the targets are not available. Thus, we will focus on the appearance-based approach and try to develop an ATR system using images from other UAVs to improve the recognition performance and reliability.

We introduce a Bessel K based recognition method, which is a stochastic model for capturing image variability. The direct modeling of image is difficult due to the large dimensionality of the images. A popular idea is to first reduce dimensions using purely numerical considerations and then impose probability models on the reduced data. Principal components, independent components, Fisher's discriminant, etc. are all instances of this idea. The main advantage of such representations is the computational efficiency. But a lack of physical or contextual information leads to a limited performance, especially in challenging situations.

Instead of modeling the image itself, we may decompose images into their spectral components using a family of bandpass filters. Thus, the definition of a probability model on images is through its spectral representation. Low dimensional statistics of these filtered components are used as reduced data to represent images. To build probability models, non-parametric or parametric estimators are researched. Bessel K forms are derived to state the probabilities in a convenient form, and the resulting analysis can be simplified considerably compared the analysis performance on the full non-parametric forms. The relationship between Bessel parameters and certain physical characteristics of the images objects are discussed in [55, 58].

2.3 Performance Analysis of ATR Systems

There is a large amount of literature devoted to analysis of ATR systems including both numerical and analytical approaches. Most of works use traditional performance measures such as Signal-to-Noise Ratio (SNR), probability of detection, probability of false alarm, Receiver-Operator-Curve (ROC), average probability of recognition error and confusion matrix. These measures lead to analysis of ATR systems in real time, but do not allow to predict performance. A traditional approach to performance predictions is to use bounds, approximations and limits. For example, the work in [58] suggests use of Laplace

approximation method for solving an integral for Bayesian probability. The asymptotic analysis relies on vanishing value of noise variance, which leads to overoptimistic results. Some references on application of bounds and approximations can be found in [22,28,37], etc.

In [53], the problem is to recognize CAD models at arbitrary orientations observed via the projective transformation on a sensor with additional noise. Rate-distortion theory is applied to establish the bounds on codebook size. The more general methodologies for predicting performance of pattern (biometric) recognition systems was done by Schmid and O'Sullivan in [50] from channel capacity viewpoint. It was further analyzed by Westover and O'Sullivan [64]. Similar to a communication channel, a recognition channel is characterized by its capacity, with the difference being recognition capacity. In an ATR problem, recognition capacity can be thought as being the maximum number of targets that can be successfully recognized with probability close to zero when the number of informative samples gets large. While the recognition capacity gives a comprehensive measure of ATR system capabilities, it does not provide us with the value of probability of error, an important characteristic of ATR systems. In [64], Westover etc. considered the tradeoff between the amount of resources devoted to data representation and the complexity of the environment. A general model for recognition systems subject to resource constraints was described and the resource-complexity tradeoff was characterized in terms of three rates.

2.4 Summary

This chapter reviewed sensors used in ATR systems. The CCD camera is selected as the sensor to be used in a network of UAVs because it is small in size and light in weight. The state-of-the-art recognition methods based on optical images are further researched. In this work, we will focus on the appearance based approach since no CAD models of targets are available. The recognition performance analysis is also reviewed. Traditional performance measures can lead to real time analysis of systems, but do not predict performance. A general methodology for prediction of recognition system performance from channel capacity view point will be used.

CHAPTER 3

DATA DESCRIPTION

In this section, we will introduce two databases that are used in this work. The first database is generated using an ATR Training Tool provided by Augusta Systems, Inc. The generated data (images) are assumed to mimic data acquired by optical cameras mounted on board of a network of UAVs. The second database is collected by Columbia University using a fixed color camera. A set of distorted images are further simulated from each clear image to imitate the camera and weather effects.

3.1 Baseline "Clear" Data

An ATR Training Tool provided by Augusta Systems, Inc. was used to build a simulated database. The tool is capable of generating prospective projections of 15 distinct objects projected at different orientation and elevation angles and sampled at distinct resolutions. Fig. 3.1 is the illustration of the parameters, where θ is the elevation angle, α is the orientation angle and d is the distance from the object to the camera. The objects can be manually superimposed onto a background to simulate various ground conditions. The camera parameters such as position, azimuth, declination and distance can be varied to simulate an UAV flight. The resolution of captured images can be adjusted from 512×384 to 1152×864. A snapshot of the Graphical User Interface (GUI) of the tool is shown in Fig. 3.2. Every image generated by the 3D optical tool is first processed by a target detector and then fed into a recognition system. Prior to recognition, a potential target is located and placed in a canonical (or object-centered) reference frame suitable for recognition. In our experiments, we use three target types: tank, truck, and tractor. Sample images of targets used for recognition are shown in Fig. 3.3. We built a dataset by projecting each 3D target into a 2D plane at discrete orientation angles spaced 5 degree apart and elevation angles from 0 to 75 spaced 15 degree apart.

The second object dataset used in this work is a subset of Columbia Object Image Library (COIL-100). The COIL-100 database consists of color images of 100 objects, 72

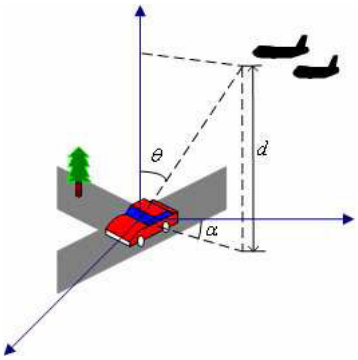


Figure 3.1. The illustration of camera parameters.

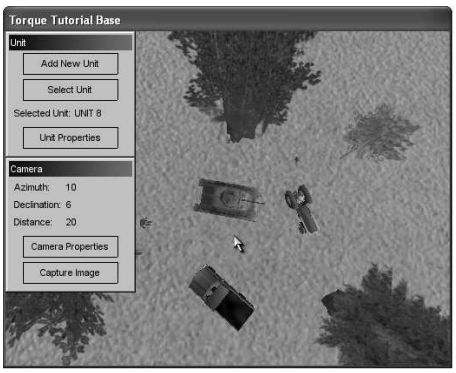


Figure 3.2. The GUI of the ATR training tool.

Figure 3.3. Top view images of tank, truck and tractor for recognition from simulated ATR database.

images per object taken uniformly 5 degrees apart in orientation (see [25] for the detailed description). In our experiments, we select 11 visually similar objects (toy cars) shown in Fig. 3.4 from the database. We use gray-level images instead of the color images.

3.2 Simulated Environmental and Camera Effects

Apart from generated images of objects as described in the previous section, we expand the dataset by adding six distorted versions of each original image. The involved distortions mimic various camera and environmental effects. The types of distortions that we impose include Gaussian noise, Poison noise, illumination effect, effect of contrast change, motion blur and defocus blur. The details of generation procedures are summarized below.

1. Images contaminated by Gaussian noise contain additive white noise with zero mean and variance σ^2.

Figure 3.4. Images of 11 objects from COIL-100 database selected for our experiment.

2. Shot noise from Charge Coupled Device (CCD) camera is modeled by Poisson process. ([54] provides physical descriptions.) The intensity of the Poisson noise depends on the intensity of the underlying data. The mean of the Poisson process is equal to the square root of the image intensity.

3. The images are brightened or darkened by increasing or decreasing the intensities [29]. This procedure simulates illumination effect. Denote by β ($\beta \in (-1, 1)$) the parameter that controls the level of illumination. We first normalize image intensities to $(0, 1)$, then brighten images by raising to the power of a number less than one, that is, $(1 - \beta,\ \beta \in (0, 1))$ or darken images by raising to the power of a number larger than one, that is, $(\frac{1}{\beta+1},\ \beta \in (-1, 0))$.

4. We model contrast change by linear mapping the normalized histogram to a new one [29]. If the histogram is "squeezed," then the new image will have low contrast. The more compression, the lower the contrast is. The range is determined by parameter $1-2LF$, where LF specifies the fraction of the image to saturate at low intensities.

5. A linear relative motion of an optical camera or an object is simulated by convolving images with a two parameter point spread function (PSF) [48]. Length L in pixels and angle θ in degrees correspond to motion in specific direction with predefined camera velocity. The parameter θ follows uniform distribution on $[0, 360°]$.

6. The images are filtered by a two-dimensional circular averaging filter to generate defocus blur [48]. Defocus level corresponds to the radius r of the averaging filter.

By controlling the value of the parameters, different levels of noise in images can be generated. Table 3.1 lists the parameter values used to achieve various levels of distortions

in our experiments. Distortions increase from Level 1 to Level 5. The samples of distorted images of the tractor are displayed in Fig. 3.5.

Table 3.1. Parameters used to simulate various distortion levels.

Effects	Parameter	Level 1	Level 2	Level 3	Level 4	Level 5
Gaussian Noise	σ^2	0.005	0.01	0.015	0.02	0.025
Illumination (dark)	β	-0.1	-0.2	-0.3	-0.4	-0.5
Contrast	LF	0.15	0.2	0.25	0.3	0.35
Motion Blur	L	2	4	6	8	10
Defocus Blur	r	2	4	6	8	10

Figure 3.5. Distorted images of the tractor. From the top left to the bottom right: the image with additive Gaussian noise, the image distorted by Poisson noise, the image characterized by a low illumination, a low contrast image, motion-blurred image and defocused image.

3.3 Summary

In this chapter, a simulated database and a real database are introduced. Images from different rotation and declination angles for each target are generated. 6 distorted images for each clear image are further added into the database to simulated the camera and environmental effects, which include Gaussian noise, Poison noise, illumination effect, effect of contrast change, motion blur and defocus blur. By controlling the parameters in the simulation, 5 levels of distorted data are formed.

CHAPTER 4

RECOGNITION METHOD BASED ON BESSEL K FORMS

In this work, we select an Electro-Optical sensor, a CCD camera, to obtain the information from the field of interest with the benefit of low cost and a small size. We assume that the original images taken by the UAVs have been processed through a proper detection method. The object of interest is located and placed in a canonical (or object-centered) reference frame suitable for recognition.

Assuming that the CAD models of the targets are not available during the testing procedure and considering the complex structure of targets, we focus on the appearance-based or view-based approach. In this chapter, the recognition method based on Bessel K forms is introduced. The training and testing results using single image and two images are then discussed based on the simulated data. We will also analyze the influence of environmental and camera effects on recognition performance.

4.1 Recognition based on Single Frame

Comprehensive studies [55], [57] of natural scenes have shown that the distributions of pixel intensities in linearly filtered images are described by a family of Bessel K distribution functions. This constitutes a basis for the implemented recognition algorithm.

Bessel K forms is a stochastic model that can be used to measure image variability. This parametric family is applied to model lower order probability densities of pixel values resulting from bandpass filtering of images. The main idea of the recognition algorithm based on Bessel K forms is to select the critical features of each object class by passing an image through a bank of linear filters and then analyzing statistics of the filtered images. As shown by Grenander and Srivastava [57], Bessel K forms parameterized by only two parameters: (1) the shape parameter p, $p > 0$, and (2) the scale parameter c, $c > 0$, may provide a good statistical fit to empirical histogram distributions of filtered images.

Denote by I an image and by $\mathcal{F}$ a filter, then the filtered image $\mathcal{I} = I * \mathcal{F}$, where $*$ denotes the 2-dimensional convolution operation. Under the conditions stated in [55], the probability density function of the random variable $\mathcal{I}(\cdot)$ can be approximated by

$$f_K(x; p, c) = \frac{2}{Z(p,c)}|x|^{p-0.5}K_{(p-0.5)}(\sqrt{\frac{2}{c}}|x|), \tag{4.1}$$

where $K_\nu(x)$ is the modified Bessel function of the second kind, and $Z(p, c)$ is the normalization given by

$$Z(p,c) = \sqrt{\pi}\Gamma(p)(2c)^{0.5p+0.25}.$$

Given J filters, the image I can be represented using $2J$ Bessel parameters.

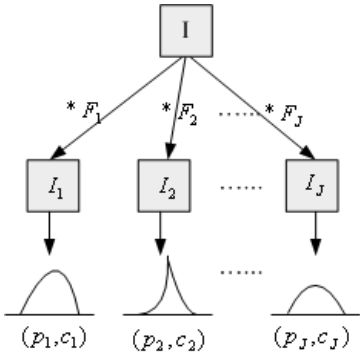


Figure 4.1. Representation of an image I using $2J$ Bessel parameters.

To approximate the empirical density of the filtered image by a Bessel K form, the parameters p and c are estimated from the observed data using

$$\hat{p} = \frac{3}{SK(\mathcal{I}) - 3} \text{ and } \hat{c} = \frac{SV(\mathcal{I})}{\hat{p}}, \tag{4.2}$$

where SK is the sample kurtosis and SV is the sample variance of the pixel values in $\mathcal{I}$. Since the moment-based estimate of p in (4.2) is sensitive with respect to outliers, in our computations we replace it with an estimate based on empirical quartiles given by

$$\hat{p} = \frac{3}{\hat{SK}(\mathcal{I}) - 3}, \text{ with } \hat{SK}(\mathcal{I}) = \frac{q_{0.995}(\mathcal{I}) - q_{0.005}(\mathcal{I})}{q_{0.75}(\mathcal{I}) - q_{0.25}(\mathcal{I})},$$

where $q(\cdot)$ is the quartile function that returns the x quartile of a set of samples. More information on quartile estimates can be found referred in the work by Freund [19]. This method provides reasonable fit. As shown in Fig. 4.2, the histogram (dashed line) of

images filtered by Gabor filters [27] closely follows the estimated Bessel K forms (solid line). To quantify the difference between two filtered images based on their distributions,

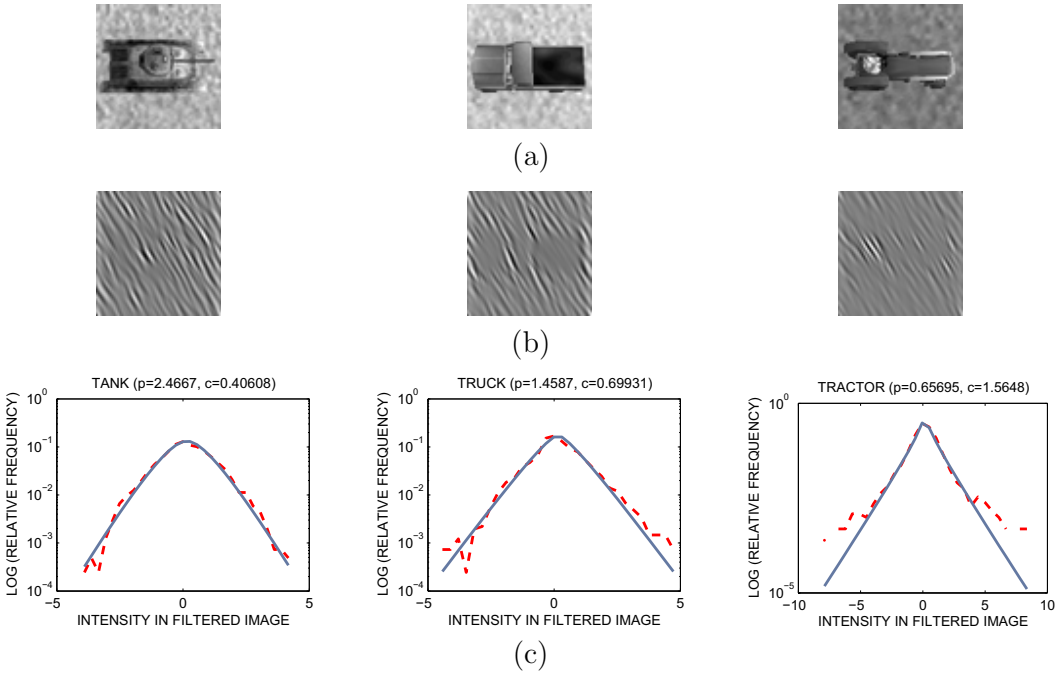


Figure 4.2. (a) Images, (b) Gabor components of images in (a), and (c) the marginal densities using targets in ATR dataset. The empirical histogram distributions are marked in dashed line. The Bessel K form approximations are shown in solid lines.

two distance measures: (1) a pseudo-metric introduced by Srivastava [55] and (2) the K-measure [36] between two Bessel K forms, are used. The pseudo-metric is defined as

$$d_I(\mathcal{I}_1, \mathcal{I}_2) = \sqrt{\int_{-\infty}^{+\infty} \left(f_K(x; p_1, c_1) - f_K(x; p_2, c_2)\right)^2}. \tag{4.3}$$

The closed form of $d_I(\mathcal{I}_1, \mathcal{I}_2)$ for the case of $p_1, p_2 > 0.25, c_1, c_2 > 0$ is given by:

$$d_I(\mathcal{I}_1, \mathcal{I}_2) = \left[\frac{\Gamma(0.5)}{2\sqrt{2\pi}}\left(\frac{\mathcal{G}(2p_1)}{\sqrt{c_1}} + \frac{\mathcal{G}(2p_2)}{\sqrt{c_2}} - \frac{2\mathcal{G}(p_1+p_2)}{\sqrt{c_1}}\left(\frac{c_1}{c_2}\right) p_2 \mathcal{H}\right)\right]^{\frac{1}{2}},$$

where $\mathcal{G}(p) = \frac{\Gamma(p-0.5)}{\Gamma(p)}$ and $\mathcal{H} = H\left(p_1 + p_2 - 0.5, p_2; p_1 + p_2; 1 - \frac{c_1}{c_2}\right)$. The function H is the hypergeometric function. In cases where $\hat{p} < 0.25$ for an image-filter combination, we compute the pseudo-metric numerically using the quadrature integration.

The K-measure is defined as

$$d_{KL}(\mathcal{I}_1, \mathcal{I}_2) = D\left(f_K(x; p_1, c_1) \| f_K(x; p_2, c_2)\right) + D\left(f_K(x; p_2, c_2) \| f_K(x; p_1, c_1)\right), \quad (4.4)$$

where $D\left(f_K(x; p_1, c_1) \| f_K(x; p_2, c_2)\right)$ is the relative entropy between two distribution functions $f_K(x; p_1, c_1)$ and $f_K(x; p_2, c_2)$ given by

$$D\left(f_K(x; p_1, c_1) \| f_K(x; p_2, c_2)\right) = \int_{-\infty}^{+\infty} \log\left(\frac{f_K(x; p_1, c_1)}{f_K(x; p_2, c_2)}\right) f_K(x; p_1, c_1) dx.$$

In the above expressions $f_K(\cdot)$ is the Bessel K probability density function introduced in (4.1).

Given two images $\{I_1, I_2\}$ and a bank of filters $\{\mathcal{F}_j, j = 1, 2, \cdots, J\}$, we evaluate a set of filtered images $\{\mathcal{I}_{(n,j)} = I_n * \mathcal{F}_j,\ n = 1, 2;\ j = 1, \cdots, J\}$. After estimating the parameter $p_{(n,j)}$ and $c_{(n,j)}$, each image is mapped to J points in the density space. The distance between two images are calculated by

$$d_I(I_1, I_2) = \sum_{j=1}^{J} d_I(\mathcal{I}_{(1,j)}, \mathcal{I}_{(2,j)}), \quad (4.5)$$

and

$$d_{KL}(I_1, I_2) = \sum_{j=1}^{J} d_{KL}(\mathcal{I}_{(1,j)}, \mathcal{I}_{(2,j)}), \quad (4.6)$$

where $d_I(\mathcal{I}_{(1,j)}, \mathcal{I}_{(2,j)})$ and $d_{KL}(\mathcal{I}_{(1,j)}, \mathcal{I}_{(2,j)})$ are defined in (4.3) and (4.4).

The purpose of using the two distance measures is to balance accuracy and computational efficiency. K-measure is an accurate measure of similarity of two probability density functions. However, it cannot be obtained in closed form for Bessel K forms. Numerical evaluation of K-measure is computationally expensive. The pseudo-metric (4.3) has closed form for Bessel K forms, which means that the computation cost is relatively low. The major drawback of the pseudo-metric is its lower precision compared with the K-measure. To measure the difference between two histograms fast and with relatively high precision, we combine these two distance measures. First, we use the fast method, the pseudo-metric, to evaluate the distance between the input image and all templates in the database. If the pseudo-metric has multiple minima close in their values, there will be a potential misclassification. The precise metric, the K-measure, is then used to re-calculate the distances and make the final decision. By setting threshold properly, we obtain relatively fast and reliable result. The comparison is shown in Section 4.3.

4.2 Recognition based on Two Frames

Consider a scenario where a set of UAVs perform an area search. UAVs monitor the ground continuously at a slow rate (for instance, 2-5 frames per second). We assume that an UVA while passing a target is capable of acquiring only a few (1-4 frames) containing this target. Now, if an UAV detects a potential target within a frame, it may appeal to its neighbors to perform additional monitoring of the area. Thus, this scenario may result in collecting a relatively large number of optical frames containing information about a target. In this section we describe a multivariate Bessel K form for improved recognition.

The multivariate Bessel K forms can be formed as a mixture of Gaussian variables, where the mixing variable is a scaled Gamma distributed random variable with parameters p and c. Multivariate Bessel K forms are a special case of a larger family, namely, the generalized hyperbolic distributed family (see Barndorff-Nielson et al [3] for details). Denote by $\mathbf{v}$ a d-dimensional random vector following Guassian distribution with zero mean and identity covariant matrix. Let z be a random variable following Gamma distribution with parameters p and c, and Γ be a positive definite matrix. Form

$$\mathbf{x} = \sqrt{z}\Gamma^{\frac{1}{2}}\mathbf{v}.$$

Then $\mathbf{x}$ is a d-dimensional random vector following Bessel K distribution with parameters p, c and Γ. The probability density function of $\mathbf{x}$ is given by

$$f_K(\mathbf{x}; p, c, \Gamma) = \frac{2}{Z_M(p,c)} \left(\sqrt{\mathbf{q(x)}}\right)^{p-\frac{d}{2}} K_{(p-0.5)} \left(\sqrt{\frac{2}{c}\mathbf{q(x)}}\right), \tag{4.7}$$

where $\mathbf{q(x)} = \mathbf{x}^T\Gamma^{-1}\mathbf{x}$ and $Z_M(p,c)$ is a normalization given by

$$Z_M(p,c) = \pi^{\frac{d}{2}}\Gamma(p)(2c)^{0.5p+0.25d}.$$

When $d = 1$, (4.7) reduces to (4.1).

As illustrated in Fig. 4.3, the pair of images $I(\alpha_1)$ and $I(\alpha_2)$ are taken from the same object but at different poses. They can be jointly represented by $3J$ sets of parameters.

To estimate the parameters p, c and Γ, we first find the mean and covariance matrix as

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \quad \text{and} \quad \hat{\Gamma} = \frac{\hat{C}}{\left(\det\hat{C}\right)^{\frac{1}{d}}},$$

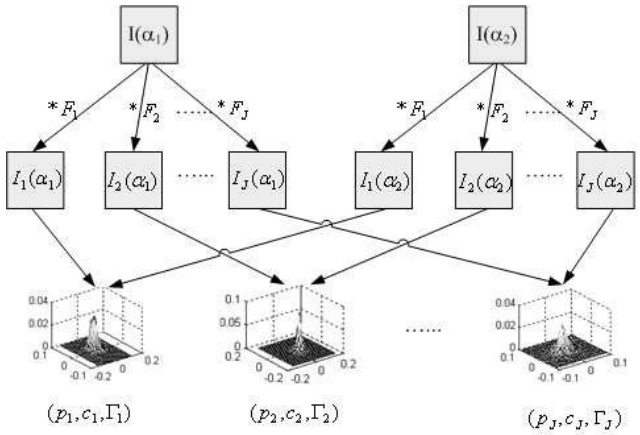


Figure 4.3. Representation of a pair of images $I(\alpha_1)$ and $I(\alpha_2)$ by $3J$ Bessel parameters.

where N is the sample size and $\hat{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T$. Then we generate a new random vector $\mathbf{y}_i = \hat{\Gamma}^{-\frac{1}{2}}(\mathbf{x}_i - \hat{\mu})$, which follows d-dimensional Bessel K distribution with zero mean, identity covariance matrix, the shape parameter p and the scale parameter c. The marginal distribution of $y_{ki},\ k = 1, \cdots, d$ follows univariate Bessel K form. So we can estimate p and c as $\hat{p} = \frac{1}{d}\sum_{k=1}^{d}\hat{p}_k$ and $\hat{c} = \frac{1}{d}\sum_{k=1}^{d}\hat{c}_k$, where $\hat{p}_k$ and $\hat{c}_k$ are the estimates from the k^{th} projection.

We use the K measure to qualify the distance between two pairs of images. Fig. 4.4 shows the bivariate fitting results of two tank images with relative angle 10 degrees.

4.3 Numerical Results

In this section, we use the two datasets described in Chapter 3 to evaluate the recognition performance given different samples per target. Compared with the traditional Principle Component Analysis (PCA) method, the recognition method based on Bessel K forms performs substantially better. The influence of environmental and camera effects on recognition performance is also discussed. We further analyze recognition results from a single frame and multiple frames.

4.3.1 Recognition Performance and Computational Cost

In this section, we will test the capability of recognition methods in dealing with pose changes of targets. The recognition algorithms operate in two modes: training and testing. All the clear images in the datasets introduced in Chapter 3 are divided into nonoverlapping training and testing sets. Some of the images are used for training and

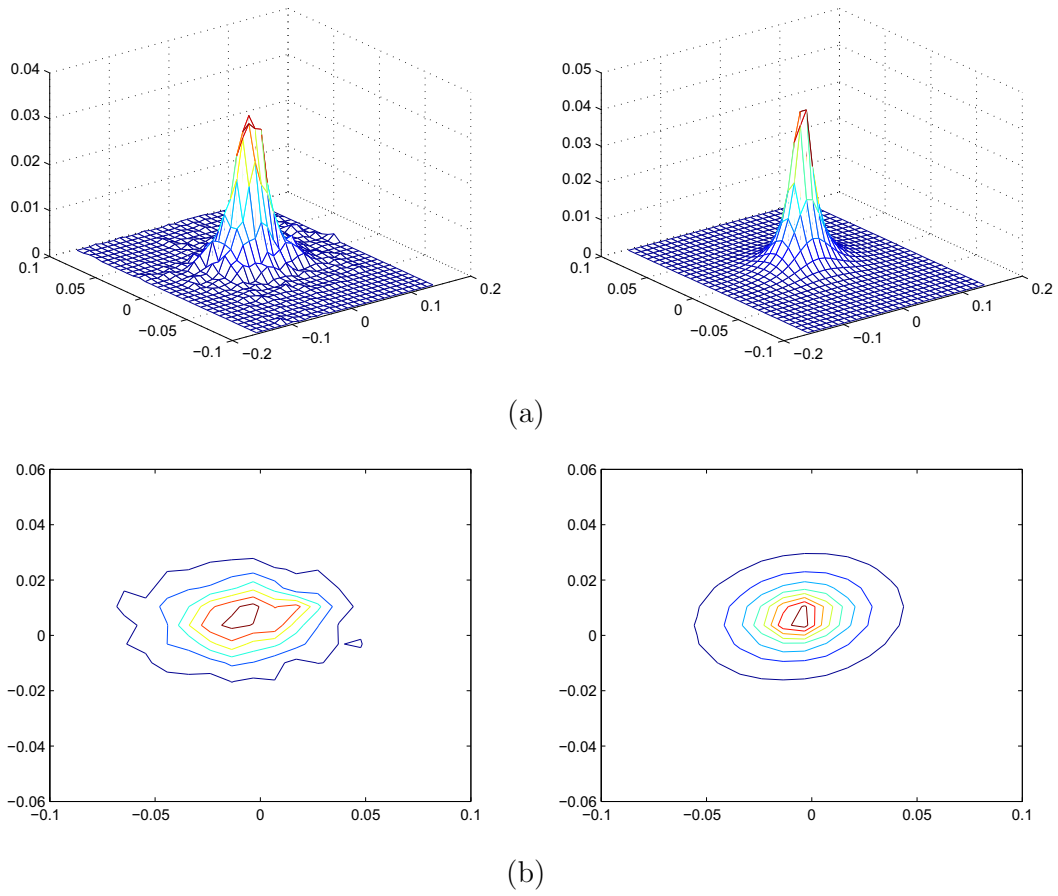


Figure 4.4. The observed (left) and estimated (right) bivariate densities of two tank images with relative angle 10 degree plotted as meshes (a) and contours (b).

the remaining for testing. To process data, we used a bank of 38 filters including Gaussian filters, Laplacian of Gaussian filters and Gabor filters.

Since the recognition performance of our classifier depends on the number of samples per target used for training the system, we present recognition results for a number of values that the ratio of training and testing samples can form. For COIL-100 dataset, since there is only rotation changes, the training samples per target are selected with equal intervals of rotation angles. For ATR dataset, there are 864 (3 distance, 4 elevation angles and 72 rotation angles) clear images per target. We only use part of the images at elevation 15 and distance 10 as training samples and the remaining for testing. Table 4.1 summarizes the results. Note that Bessel K forms generally outperform PCA method.

Here the results of Bessel K forms are obtained using the combined metric described in Section 4.1. Table 4.2 concludes that compared with the method using Pseudo-metric

Table 4.1. Correct Recognition Rates for COIL-100 and ATR datasets Using PCA and Bessel K Forms.

COIL-100 dataset			ATR dataset		
Train/Test per target	PCA	Bessel K	Train/Test per target	PCA	Bessel K
36/36	98.23%	98.48%	36/828	95.77%	96.86%
24/48	97.54%	96.78%	24/840	94.80%	96.83%
18/54	92.26%	93.43%	18/846	89.16%	96.89%
8/64	70.31%	69.18%	8/856	75.35%	91.16%
4/68	46.79%	47.19%	4/860	63.02%	91.32%

only and using K-measure only, the combination algorithm balances the accuracy and speed. The experiments are performed using Matlab 7.0 and Pentium 4 CPU 3.20GHz. No optimization is applied. With proper optimization and using C/C++ coding, the speed of the recognition method based on Bessel K forms may be improved up to 15 times and can meet the requirement to perform online.

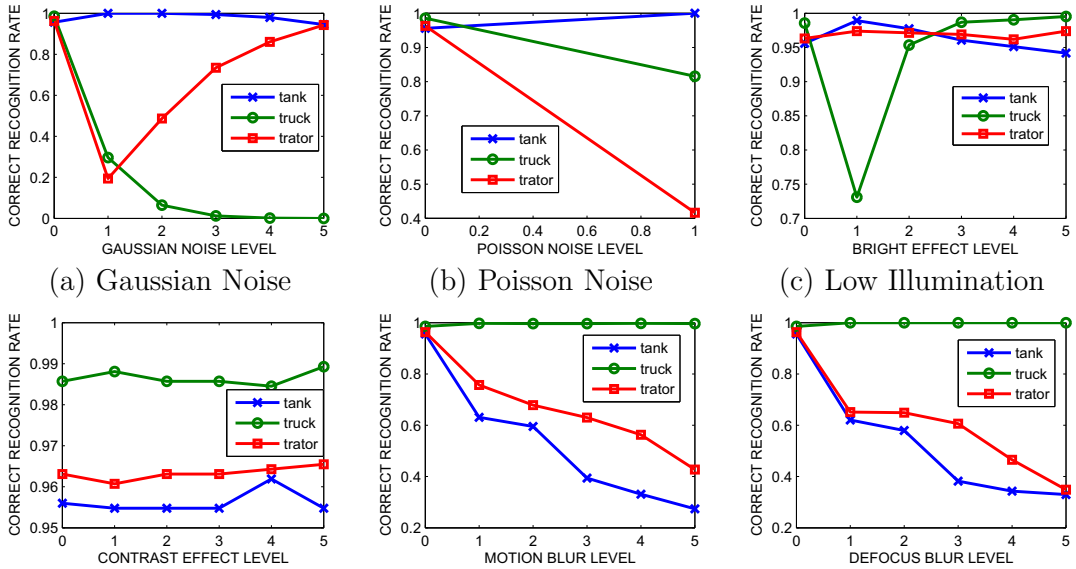
Table 4.2. Speed and Performance Using different metrics

Train Samples: 24 per target	Pseudo-metric Only	K-measure Only	Combination Algorithm
Average Test time per sample (second)	0.013	0.1	0.02
Correct Recognition Rate (ATR dataset)	95.75%	98.81%	96.83%
Correct Recognition Rate (COIL-100 dataset)	92.99%	98.30%	96.78%

4.3.2 Influence of Environmental and Camera Effects on Recognition Performance

To test the influence of environmental and camera effects on recognition performance, we fix the training samples to be 24 per target. The undistorted images of all targets at orientations $0, 15, 30, \cdots, 345$, elevation 15 and distance 10 form the training set. All distorted images at all positions are used in the testing mode. Each effect is evaluated individually. We generate a number of distorted images parameterized by a varying distortion level: from lowest level to highest level. The correct recognition rate of different

objects under six types of distortions for ATR dataset are shown in Fig. 4.5. We observe that the performance changes for all three objects are not consistent. In most cases, the performances of two objects decrease while the performance of the third one increases. The average correct recognition rate as a function of distortion level parameterized by various effects for two datasets are shown in Fig. 4.6. Level 0 corresponds to the case when no distortion is added.



(a) Gaussian Noise (b) Poisson Noise (c) Low Illumination

(d) Low Contrast (e) Motion Blur (f) Defocused Blur

Figure 4.5. Recognition performance as functions of various environmental and camera effects for ATR dataset.

From Fig. 4.6 we conclude that the average recognition performance decreases when distortion level increases, except the illumination and contrast effects. This is due to a light correction procedure performed prior to classification. To be more specific, prior to recognition all test images are resized to 64×64 and then normalized to minimize the effect of different lighting conditions. The procedure of normalization is as follows:

$$I^{-}(x,y) = \frac{I(x,y) - \mu}{c\sigma}, \; c \, \epsilon R^{+}, \tag{4.8}$$

where $I(x,y)$ is the pixel value within the sub-window during detection scanning. μ and σ are the mean and the standard deviation of $I(x,y)$.

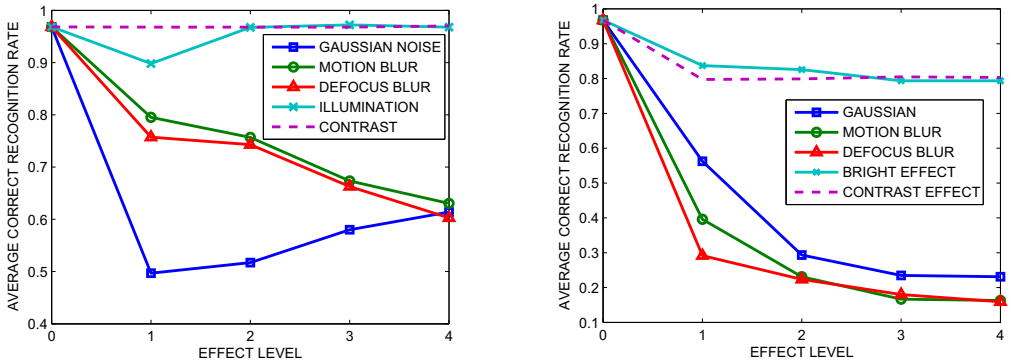


Figure 4.6. Average recognition rate under different effects for (a) ATR dataset and (b) COIL-100 dataset.

4.3.3 Recognition Performance: Single and Multi-frame Cases

In this experiment, we involve only the clear images in the databases. We assume that the relative rotation angle between the test image pair is known. Two different relative angles, 5- and 10-degrees, are tested. For each relative angle, there are 24 sets of multivariate Bessel K parameters describing each target.

For ATR dataset, we only use images generated at 4 different elevation angles, $0, 15, 30$ and 45 degree while keeping the distance at 10. Then the total number of testing pairs is $4 \times 4 \times 72$ per target. Table 4.3 summarizes the results of testing of the multivariate Bessel K recognition algorithm. The correct recognition and error rates are presented in the form of a confusion matrix for a single and two-frame cases (with 5 and 10 degree relative orientation). Note that multivariate Bessel K forms result in considerably improved performance when the relative orientation between two images is 10 degrees, that is, when data are less correlated compared to the case of the relative orientation of 5 degrees.

4.4 Summary

In this chapter, we introduce a recognition method based on Bessel K forms. We concluded that use of the low order statistic properties of the filtered images is sufficient for analysis of recognition. The distribution of the histogram of filtered images is modeled using Bessel K distribution with two parameters. This parametric method is beneficial since it guarantees the storage and the recognition speed by providing a close form metric. The combination algorithm using both Pseudo-metric and K-measure balances the accuracy and speed. Bessel K form based recognition method outperforms PCA method

Table 4.3. Recognition Performance Using Single and Two images for ATR dataset.

Confusion Matrix	Simulated ATR dataset
Single	$\begin{bmatrix} 0.9545 & 0 & 0.0114 \\ 0 & 0.9811 & 0 \\ 0.0455 & 0.0189 & 0.9886 \end{bmatrix}$
Two (5 degree)	$\begin{bmatrix} 0.9931 & 0.0642 & 0 \\ 0 & 0.9358 & 0 \\ 0.0069 & 0 & 1 \end{bmatrix}$
Two (10 degree)	$\begin{bmatrix} 0.9991 & 0.0069 & 0 \\ 0.0009 & 0.9931 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

at different training to testing data ratios. The influence of camera and environmental effects on recognition performance is also evaluated. As the distorted level increases, the recognition performance degrades except the illumination and contrast effects because of the lighting normalization performed prior to classification. The recognition method, given two frames based on Bessel K forms, is also generated and tested using images with relative angle 5 and 10 degrees. The performance given two images is generally better than the single image case.

CHAPTER 5

MODIFIED ATR SYSTEM

As shown in Fig. 1.1, an ATR system includes a detector as a first processing block, which scans an input image with the purpose to locate potential objects. Each window is then sent for further processing into a classifier with a proper size to identify the objects. Ideally, each window should have an object in the center with a minimum background clutter included. In practice, however, the output windows from a detector may only cover parts of objects, or the objects may not be located in the center of the window and/or only occupy small parts of windows. To recognize objects, the size of windows need to be adjusted properly. As a result of adjusted windows, the detector may have a large number of false alarms. This further may result in a large number of misclassifications, since typical classifiers are trained to recognize only a certain number objects, that is, an output of the detector (a window) that enters the classifier will be recognized as one of the objects in an object library. Thus, there is a potential that clutter will be recognized as one of the objects. One approach to remove detected windows containing clutter only is to perform a postprocessing after the classification step. The modified structure of the ATR system is shown in Fig. 5.1. In this chapter, we will first introduce the detection method based on Haar-like features, then we will perform perform recognition using Bessel K forms, and finally we will approach the challenge of practical issues described above by introducing a clutter rejection block based on alignment distances between the testing image and the index object and the recognition scores. The system performance will also be discussed.

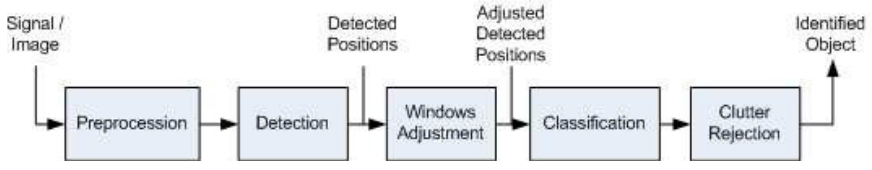


Figure 5.1. Modified ATR structure.

5.1 Target Detection based on Haar-like Features

Detection of a possible object of interest is one of the most critical steps in object recognition problems, since the results of postprocessing depend on this step. In this work, we will use a local Haar-like filter based detection method which is very popular in the field of face recognition. This rapid target detection scheme is based on the idea of a boosted classifier cascade [62].

The classifier cascade is trained on a set of positive images (targets) and a set of negative images (non-targets). For each training image, an over-complete set of Haar-like feature pool is calculated and AdaBoost algorithm of Schapire and Singer [49] is used to build a stage classifier. After the classifier cascade is trained, the detection algorithm is applied to a query image. A search window is sled over the query image. At each window location and scale the content of the window is classified as target or non-target.

In each round of boosting, a weak learning algorithm is applied to select a single rectangle feature which best separates the positive and negative samples. For each feature, the weak learner determines the optimal threshold classification function, such that the minimum number of examples are misclassified. Thus, a weak classifier $h_j(\cdot)$ is a binary valued function obtained by comparing the j-th feature value $f_j(\cdot)$ with a threshold θ_j:

$$h_j(x) = \begin{cases} \alpha_j & if \quad f_j(x) > \theta_j \\ \beta_j & otherwise. \end{cases} \tag{5.1}$$

Here x is a sub-window of an image. The value of the feature is equal to weighted differences of integrals over rectangular subregions. α_j and β_j are positive or negative votes of each feature set by AdaBoost during the learning process. θ_j is the optimal threshold obtained by the weak learner.

The form of the final stage classifier returned by AdaBoost is a thresholded linear combination of weak classifiers (see Fig. 5.2). The stage classifier is given by:

$$C(x) = \begin{cases} 1, & if \ \sum_j h_j(x) > T, \\ 0, & otherwise, \end{cases} \tag{5.2}$$

where T is the stage threshold set by AdaBoost during the learning process.

In order to improve computational efficiency and also reduce the false positive rate, a sequence of increasingly more complex classifiers called cascade is used. A cascade of classifiers is a degenerated decision tree where at each stage almost all objects of interest are detected while only a certain fraction of the non-object patterns are rejected. The more an input window looks like an object, the larger the number of classifiers are

evaluated on it and the longer it takes to classify the window. Since most windows of an image do not look like objects, they are quickly discarded as non-objects. Fig. 5.3 illustrates a cascade.

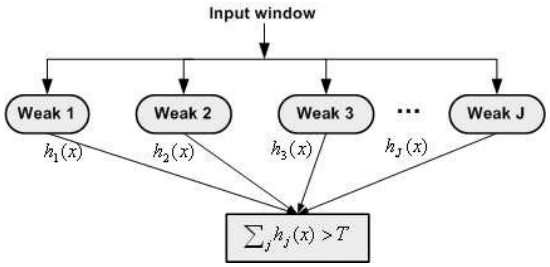


Figure 5.2. Stage classifier.

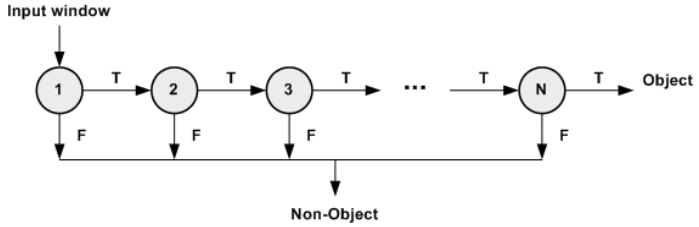


Figure 5.3. Cascade of classifiers.

5.2 Window Adjustment using B-spline based Segmentation

Once the images pass the detector, a set of windows which contain potential targets are produced as the output. Ideally, the targets should be centered in the window with minimum background included. However, in practice, the targets may not be fully covered by windows. Since we use a global recognition method, to avoid the effect of different background and the window positions, the targets should be separated from the background and the windows should be adjusted properly such that the targets are centered and occupy the most of the windows.

Since the targets are roughly at the center of the windows, we are able to build a fully automatic target extraction system. We use B-splines as deformable templates to describe the target boundary. This results in separation of the whole window into two nonoverlapping regions: the target and the background. To extract the target, we involve a region-based approach [18]. This approach implies that target and background regions

are described by two distinct stochastic models. Potential models may include Gaussian model, Gaussian mixture or Poisson models. Since both the target boundary and the parameters of models are unknown, they have to be estimated using available data. This problem can be solved iteratively, i.e., fixing the parameters of models, we perform estimation of the boundary between regions; then fixing the boundary parameters, we re-estimate the models.

For a sake of clarity, we first present a brief review of B-splines; for a more detailed account, see [15] and [16]. We will then introduce our optimization algorithms.

5.2.1 B-splines

Given $m+1$ non-decreasing real numbers $\{t_0 \leq t_1 \leq \cdots \leq t_m\}$, which are also called knots, a B-spline of degree n is a parametric curve $\mathbf{S} : [t_0, t_m] \rightarrow \mathcal{R}^2$ composed of basis B-splines of degree n

$$\mathbf{S}(t) = [x(t),\ y(t)] = \sum_{i=0}^{m-n-1} \mathbf{c}_i B_i^n(t),\ \ t \in [t_n, t_{m-n}].$$

The $\mathbf{c}_i$ are called control points. The $m-n$ basis B-splines of degree n can be defined using the recursion formula

$$B_i^0(t) = \begin{cases} 1 & \text{if} \quad t_i \leq t_{i+1} \\ 0 & \text{otherwise,} \end{cases}$$

$$B_i^n(t) = \frac{t - t_i}{t_{i+n} - t_i} B_i^{n-1}(t) + \frac{t_{i+n+1} - t}{t_{i+n+1} - t_i} B_{i+1}^{n-1}(t).$$

Since the basis functions are based on knot differences, the shape of basis functions is only dependent on the knot spacing and not specific knot values. Here we use the uniform B-splines which means that the knots are equidistant.

To describe closed curves, the periodic extension of the knot sequence, $\{\tilde{t}_j, j \in \mathcal{Z}\}$ with $\tilde{t}_j = t_{j \ mod \ k}$ is defined. The basis functions are also be expanded periodically

$$\tilde{B}_i^n(t) = \sum_{j=-\infty}^{+\infty} B_{i+j(t_k - t_0)}^n(t).$$

Now an m-knot closed curve is represented as a linear combination of m periodic basis functions

$$\mathbf{S}(t) = \sum_{i=0}^{m-1} \mathbf{c}_i \tilde{B}_i^n(t),\ \ t \in \mathcal{R},$$

which can be further written in a matrix form, $\mathbf{S} = \mathbf{Bc}$, where $\mathbf{B}$ is a matrix and $\mathbf{c}$ is a vector.

5.2.2 Region-based Approach and Optimization Algorithms

We use the probability model to define the coherence of different image regions to group the pixels. Given the contour, the image pixels are assumed to be independently distributed. All pixels inside or outside of the contour have a common distribution characterized by a parameter vector θ_{in} or θ_{out} respectively. Denote $\mathcal{A}_{in}(\mathbf{S})$ and $\mathcal{A}_{out}(\mathbf{S})$ as the inside and outside region of the contour $\mathbf{S}$. The likelihood function of the image, given the contour and the model parameters, is

$$p(\mathbf{I}|\mathbf{c}, w_{in}, w_{out}, \theta_{in}, \theta_{out}) = \prod_{(i,j)\in\mathcal{A}_{in}(\mathbf{S})} w_{in} p(I_{(i,j)}|\theta_{in}) \prod_{(i,j)\in\mathcal{A}_{out}(\mathbf{S})} w_{out} p(I_{(i,j)}|\theta_{out}),$$

with $\mathbf{S} = \mathbf{Bc}$ and $I_{(i,j)}$ denoting pixel value at the location (i, j). $p(I_{(i,j)}|\theta_{in})$ and $p(I_{(i,j)}|\theta_{out})$ are the probability functions of the inner and outer regions. $w_{in} = p((i, j) \in \mathcal{A}_{in})$ and $w_{out} = p((i, j) \in \mathcal{A}_{out})$ are the priors such that $w_{in} + w_{out} = 1$.

Now the unsupervised segmentation problem has to estimate, from the observed image $\mathbf{I}$, not only the position of control points $\mathbf{c}$, but also the parameters of probability model $w_{\mathcal{M}}$ and $\theta_{\mathcal{M}}$, for $\mathcal{M} \in \{in,\ out\}$. Estimation of the boundary of a smooth object, a continuous function, from a finite amount of data is an ill-posed problem, which means that an infinite continuous solutions may result in the same observed data. To regularize the contour estimate, that is to restrict the estimate to a certain class of functions providing a unique solution, we use a penalty term that favors smooth contours [67]. Then the segmentation problem solves the following optimization equation,

$$\hat{\mathbf{c}}, \hat{w}_{\mathcal{M}}, \hat{\theta}_{\mathcal{M}} = \arg \min_{\mathbf{c}, w_{\mathcal{M}}, \theta_{\mathcal{M}}} \Big[-\log(p(\mathbf{I}|\mathbf{c}, w_{\mathcal{M}}, \theta_{\mathcal{M}})) + \lambda\kappa \Big], \tag{5.3}$$

where λ is a regularization parameter and κ is the curvature of the boundary given by

$$\kappa(x, y) = \frac{x'y'' - x''y'}{\left(x'^2 + y'^2\right)^{3/2}},$$

among which x', y', x'' and y'' are the first and second order derivatives of $x(t)$ and $y(t)$.

In general, the number of control points m is also a parameter that needs to be optimized. In [18], the Minimum Description Length (MDL) is used as a criterion to find the optimal m. In our case, since we need to segment image very efficiently, we fix

the number of control points during the experiments. The number of control points is selected such that both the accuracy of the segmentation and the computation cost are satisfied.

The minimization of (5.3) is performed iteratively. Each iteration consists of two phases: First, we solve $\mathbf{c}$ with $w_{\mathcal{M}}$ and $\theta_{\mathcal{M}}$ fixed. Second, we fix the contour and re-estimate the parameters of probability model.

Phase 1: Boundary Optimization by B-spline

To find the shape of an unknown object we use the form of a gradient projection method described in [18]. Given $w_{\mathcal{M}}$ and $\theta_{\mathcal{M}}$, for $\mathcal{M} \in \{in,\ out\}$, and the control points $\hat{\mathbf{c}}^{(k)}$, the contour is estimated by minimizing the log likelihood function augmented with a regularization term.

1. Set $p = 0$, $\hat{\mathbf{c}}^{(p)} = \hat{\mathbf{c}}^{(k)}$. Build B and compute $B^{\perp} = (B^T B)^{-1} B^T$

2. Calculate the gradient with respect to the contour

$$\partial\vec{\mathbf{S}} = \nabla\left[-\log(p(\mathbf{I}|\hat{\mathbf{c}}^{(p)}, w_{\mathcal{M}}, \theta_{\mathcal{M}})) + \lambda\kappa(\hat{\mathbf{c}}^{(p)})\right],$$

   where ∇ is the gradient operation.

3. Update the contour estimate according to

$$\hat{\mathbf{S}}^{(p+1)} = \mathbf{S}^{(p)} + \epsilon\partial\vec{\mathbf{S}},$$

   where ϵ controls the step size. The control points are updated by $\hat{\mathbf{c}}^{(p+1)} = B^{\perp}\hat{\mathbf{S}}^{(p+1)}$.

4. If a stopping criterion is met, stop and $\hat{\mathbf{c}}^{(k+1)} = \hat{\mathbf{c}}^{(p^*)}$, where $\hat{\mathbf{c}}^{(p^*)}$ is a stationary point; if not, set $p = p + 1$, go back to step 2.

Phase 2: Image Model Estimation

At the *Phase 2*, we fix the contour and minimize (5.3) with respect to $w_{\mathcal{M}}$ and $\theta_{\mathcal{M}}$. In our work, we use single Gaussian distribution to model both the object and the background regions. Then the model parameters are the mean and variance of the pixel values inside and outside of the contour, i.e., $\theta_{\mathcal{M}} = [\mu_{\mathcal{M}}, \sigma^2_{\mathcal{M}}]$, for $\mathcal{M} = \{in,\ out\}$. By taking the derivatives of the log likelihood with regard to all parameters and setting them to zero, we are able to obtain the optimal parameters. $\mu_{\mathcal{M}}$ and $\sigma^2_{\mathcal{M}}$ are sample mean and variance of image intensity in the inner and outer regions. $w_{\mathcal{M}}$ is the ratios of the number of pixels in $\mathcal{A}_{\mathcal{M}}$ and the total number of pixels in the image.

5.3 Clutter Rejection Mechanism

The potential candidates detected by the detector may include true targets or may not include targets (False Alarms). An important capability of an ATR system is to reject input patterns that cannot be classified in any of the classes in an object library with a sufficiently high degree of confidence. In [9], a linear classifier, based on absolute and relative scores, ranks and their dispersion is used to accept/reject the final result. In [38], the statistic properties of the scores are used to define a threshold or the bounds of threshold to reject non-target.

In our work, after a window on the output of the detector is classified as an object in the object library, the classified window will be registered with the indexed image in the library. The registration is the process of establishing point-by-point correspondence between two images. After registration, the test window will be aligned with the indexed image by rotation, translation and other geometric transforms. If the test window is classified correctly, then both the recognition score and the distance between the registered test window and the indexed image in the library should be small. Because we estimate the object and its position jointly, the registration method that we used here is a gradient based method [34, 39].

Given two images $F(x, y)$ and $G(x, y)$ containing similar regions, pixel locations in $G(x, y)$ are related to those in the reference $F(x, y)$ as:

$$G(x, y) = F(x\cos\theta - y\sin\theta + t_x, y\cos\theta + x\sin\theta + t_y),$$

where t_x and t_y are the horizontal and vertical translations and θ is the rotation, which takes $F(x, y)$ to the image of $G(x, y)$. By applying the first-order Taylor series expansion, we have

$$G(x, y) = F(x, y) + (t_x - y\theta - x\frac{\theta^2}{2})\frac{\partial F(x, y)}{\partial x} + (t_y + x\theta - y\frac{\theta^2}{2})\frac{\partial F(x, y)}{\partial y}.$$

This allows for an error expression between $G(x, y)$ and the transformed $F(x, y)$ parameterized by the registration terms t_x, t_y and θ,

$$E(t_x, t_y, \theta) = \sum \Bigg[F(m, n) + (t_x - n\theta - m\frac{\theta^2}{2})\frac{\partial F(m, n)}{\partial m}$$

$$+(t_y + m\theta - n\frac{\theta^2}{2})\frac{\partial F(m, n)}{\partial n} - G(m, n)\Bigg]^2,$$

where the sum is taken over the overlapping portions of $G(m,n)$ and $F(m,n)$ and variables of summation have been changed to m and n to represent the discrete horizontal and vertical pixel locations.

To find the estimates of (t_x, t_y, θ) parameters, $E(t_x, t_y, \theta)$ can be minimized with respect to each of parameters, yielding the system

$$\begin{bmatrix} \sum F_m^2 & \sum F_m F_n & \sum A F_m \\ \sum F_m F_n & \sum F_n^2 & \sum A F_n \\ \sum A F_m & \sum A F_n & \sum A^2 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ \theta \end{bmatrix} = \begin{bmatrix} \sum F_m G_t \\ \sum F_n G_t \\ \sum A G_t \end{bmatrix},$$

where $F_m = \frac{\partial F(m,n)}{\partial m}$, $F_n = \frac{\partial F(m,n)}{\partial n}$, $G_t = G(m,n) - F(m,n)$, and $A = mF_n - nF_m$. The sum is again taken over overlapping pixels of $F(m,n)$ and $G(m,n)$.

Since the approximations above are based on Taylor expansions about the estimated parameters, t_x, t_y and θ must be relatively small for the estimation to be accurate. To allow for larger registration parameters, a multi-resolution iterative technique is employed. First, both images $F(m,n)$ and $G(m,n)$ are artificially reduced in resolution to a minimum size via a Gaussian pyramid [39]. The base of the pyramid is the image at its original-resolution, each upper tier is derived from the lower one by convolving the image with a 2-D Gaussian kernel and down sampling by a factor of 2 in each direction. Performing the registration at lower resolutions allows large translations to reduce to small values which can be accurately estimated. Once registration has converged at a specific resolution level, the algorithm then moves up one resolution level and proceeds to refine the registration parameters.

Once the two images are aligned, the intensity distance between the overlapping object areas can be obtained. Combined with the scores from the recognition, a new classifier can be trained to reject the non-target while keeping the true recognition results. Here we use the regression tree to remove non-targets. A set of targets and non-targets are classified using PCA and Bessel K method. Each image is aligned with the identified images in the library to calculate the intensity distance. A regression tree is trained using the distances and the recognition scores. To avoid removing the targets, we adjust the weights when training the regression tree.

5.4 Experiments and System Performance

To train Haar-like feature based detector, we first generate a set of positive images with targets in the center and a set of negative images, which do not include any targets. We use the 3D tool described in Chapter 3 to generate the positive images and negative

images. Tank, truck and tractor are targets. Each target is projected at 36 rotation angles and 5 elevation angles. These 180 images per target are further shifted up, down, left and right by 3 pixels. Thus, there are total 2700 positive images. All the other objects included in the 3D tool, such as trees, airplanes, balls, etc., are combined and projected randomly to form the negative set. Some real images are also used as negative images. The total number of negative images generated is 3342. During the training process, all the positive images are normalized to the size 24×24. In each training stage, a set of negative samples are selected from the negative images using windows with all possible sizes. A 17 stage cascade classifier with 349 weak classifiers in it is formed.

To evaluate the performance of the detector, we generated 287 images with total 1116 objects, including 784 tanks, 192 trucks and 140 tractors. The color imagery is converted into grey-scale imagery. The performance of detector depends on 4 parameters, the scale factor, the minimum number of neighbors, the maximum size difference and the maximum position difference between the true windows and the detected windows. The scale factor controls the strength of the window scanning. Small value indicates that a large number of windows in the test images are scanned. The minimum number of neighbors is used to group retrieved windows in order to filter out noise. The larger the value is, the larger is the number of windows to be grouped. The last parameters are used to determine the coincidence of the true and detected windows. Here the scale factor is 1.2, the minimum number of neighbors is 1, the maximum size difference is 1.5 and the maximum position difference is set to be 0.3. The test results are shown in Table 5.1. The missed targets are mostly occluded or located along the boundary of images.

To keep the hitting rate at a high value, the number of false alarms is allowed to be very large. In practice, we find that some targets are included in more than one windows and some detected regions cover two or more targets. We use a heuristic method to filter out large windows which cover more than two small windows and combine overlapped windows. If the window size is less than a threshold, it will be removed. After combination and removal, the number of false alarms decreases dramatically while the number of hits remains the same. The processing time per image also reduces. The ROC curves of the detected results before and after window combination and removal are shown in Fig. 5.4. We can see that the number of false alarms decreases a lot after window combination.

Table 5.1. Detection Results on 287 Images with 1116 Targets

	Hits	Miss	False Alarms
After detector	1004	112	437
After combination and removal	1003	113	187

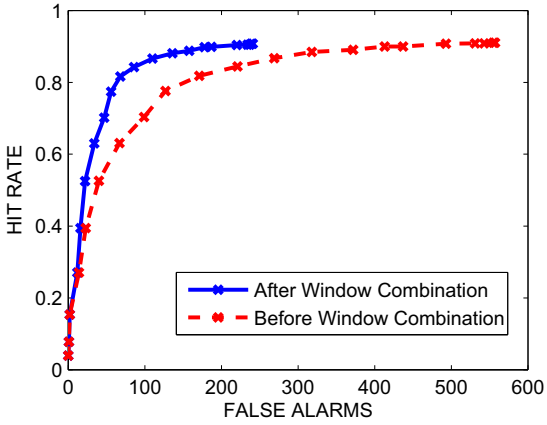


Figure 5.4. The ROC curves of detected results before window combination (dashed line) and after window combination (solid line).

We further apply the proposed segmentation method to each detected region to have the targets located in the center of windows. Since the targets may have some parts outside of the detected windows, we first enlarge the window by 50% and then retrieve the targets. To speed up the segmentation, the windows are normalized to the size 64×64 and the number of control points is set to 10. The detection results are shown in Fig. 5.5. In the top image, two overlapped red windows which contain the same tank are combined using one green window and further adjusted based on segmentation results to fit the yellow window. In the middle image, there are 4 targets covered by two red windows. The larger windows are removed after the window combination procedure. In the bottom image, a red window of small size is removed as non-target. Overall the proposed processing after detection reduces false alarms and allows improved location of targets. After combination and removal, we segment targets in each adjusted window. Although the segmentation increases the processing time, it results in a considerable improvement in the recognition performance. We apply PCA and Bessel K based recognition method introduced in Chapter 4 to the detected regions with and without segmentation. 24 images per target are used as the training samples. Table 5.2 summarizes recognition results. Note that the recognition performance after segmentation is considerably better than the performance without segmentation. We can also observe that in both scenarios Bessel K based method outperforms PCA.

We trained a regression tree for each recognition method to reject the clutters. The regression trees are then applied to the detection results after window adjustment using combination and segmentation. Table 5.3 shows the results after rejection for both PCA

Table 5.2. Correct Recognition Rate with and without Segmentation using PCA and Bessel K methods

	PCA	Bessel K
Without Segmentation	68.29%	85.05%
With Segmentation	86.83%	95.47%

and Bessel K methods. Compared with Table 5.1, although several targets are missed by false rejection, a lot of false alarms are removed.

Table 5.3. Detection Results After Rejection for PCA and Bessel K Methods

Process	Hits	Miss	False Alarms
PCA: After rejection	988	128	46
Bessel K: After rejection	991	125	77

5.5 Summary

We proposed a complete detection-recognition system with imposed capabilities to reject the clutter in detected regions of interest. This addresses a practical approach to designing an ATR system. A rapid object detection method using a boosted cascade of Haar-like features is adopted. The detected regions of each image are further combined or removed using a heuristic method based on the relative locations of the windows to reduce the number of false alarms. B-spline based segmentation method is then utilized to retrieve the separate targets from the clutter within detected regions. This extra step can help to locate the targets in the center of windows and to classify targets imposed on different backgrounds. The experimental results indicate that the proposed postprocessing steps result in considerable performance improvements both for PCA and Bessel K based recognition methods. The Bessel K-based recognition method is more robust compared to PCA method in the presence of clutters and occlusions.

The method for the rejection of an unknown object is introduced. The final reject/accept rule is trained using the regression tree. System performance is evaluated.

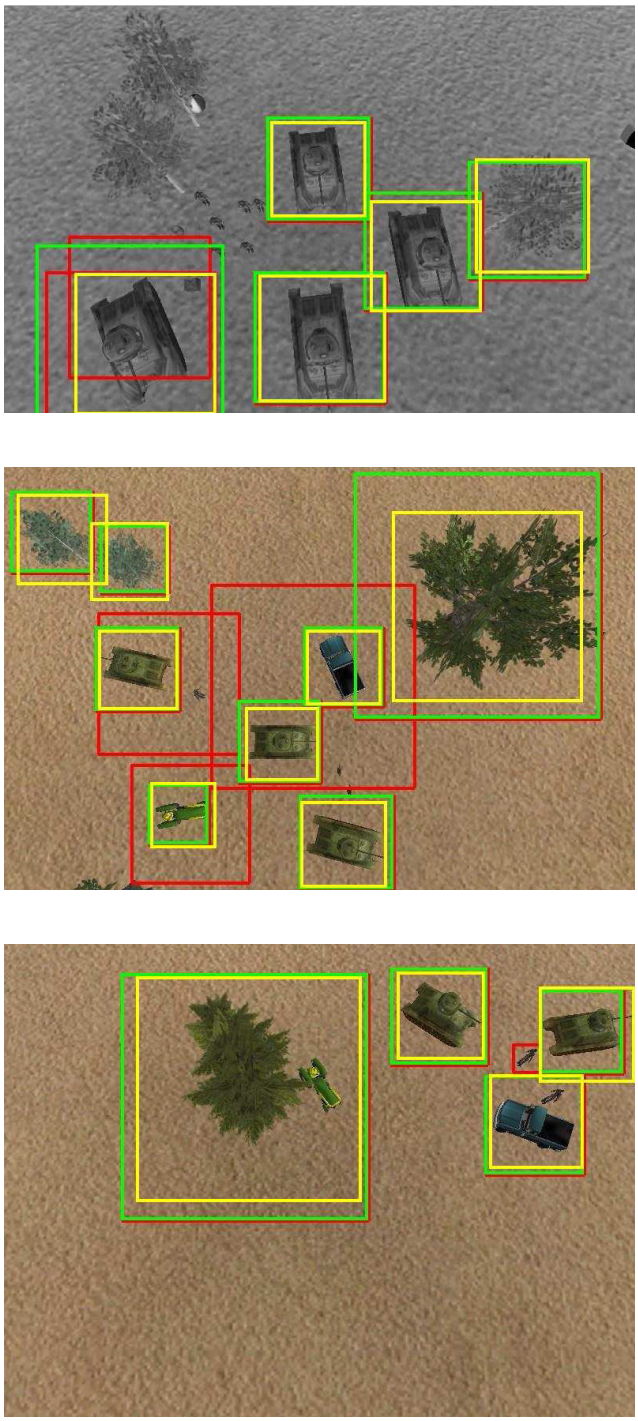


Figure 5.5. Sample images including detection regions. Detector outputs are marked in red, results after combination and removal are marked in green and results after segmentation are marked in yellow.

CHAPTER 6

SYSTEM PERFORMANCE USING DIE CAST DATABASE

In the previous chapters, we designed an ATR system using Bessel K forms applied to imagery acquired by optical sensors and evaluated the recognition capacity of a system implementing PCA coding applied to both single and multiple frames. In Chapter 3, a simulated database was described, and effects of CCD camera and weather effects were simulated. Some effects, such as shadowing and occlusions are challenging for modeling and hard to mimic using the simulation tool. Thus, there is a strong need in generating realistic effects and collecting real data. In this Chapter, we will introduce a real image database and evaluate the detection and recognition performance using the methods described in Chapters 4 and 5.

6.1 Die Cast Database

As described in Chapter 3, to evaluate performance of ATR systems, we involve a database generated using the 3D simulation tool provided by Augusta System, Inc. Although we can conveniently obtain images of objects from arbitrary view angle, we observed a number of drawbacks in the images obtained using the tool: (1) Synthetic images do not look realistic; (2) The tool can not model camera or weather effects with high precision, for example, shadows; and (3) The number of backgrounds is limited. The backgrounds are relatively simple and can not imitate the real world. Since simulated data do not possess features of real images, to understand capabilities of designed ATR system, we need a real database.

Besides the COIL-100 database, there are several other public databases available online, which are traditionally used for object detection and recognition, (for example, UIUC database [24] and UBC database [26].) Some datasets contain imagery that is not applicable for detection. The other datasets do not consider targets at different elevation angles. None of existing databases contains images that would allow us to comprehensively evaluate camera and environmental effects.

In the past we purchased 56 die cast vehicles, including tanks, trucks, trailers, tractors and airplanes, with the purpose to build an extended dataset useful for a variety applications. In this work, we select 6 die cast cars or 1/72 scale copies of military objects as targets. Images are recorded by Nikon CCD color camera (D80). The resolution of the images is set to be 1936×1296. The camera was equipped with Tamron AF 70-300mm f/4.0-5.6 LD Macro zoom lens. Aperture was close to f4.2. Zoom was fixed at 70mm for the clear images. We set the camera shutter speed (intensity integration time) as 1/60 second. Build-in flash was used. For each die cast model, a set of clear images are captured from different views as a bench mark. Fig. 6.1 illustrates these 6 objects in top view. Note that object 1 is M1A1HA with mine plough; object 2 is M1A2; object 3 is UK Challenger; object 4 is a Wurfamen 40 tank; object 5 is a HMMWV M998 "gun truck" and object 6 is Hummer. Images are taken using Nikon D80. The resolution of the images is 1936×1296. The pixel count of objects in the images shown in Fig. 6.1 is about 800×800 pixels. Each object is projected at 3 elevation angles, 0, 20.3 and 35.6 degrees. At each elevation angle, a calibrated turntable is utilized to control the rotation angle of the objects. The selected rotation step angle is 5 degree. Thus there are total 216 (72×3) poses for each object.

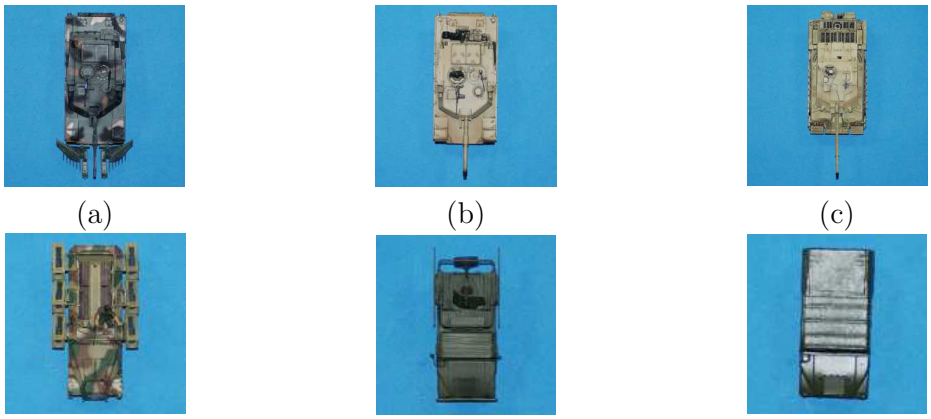


Figure 6.1. Top views of clear images from (a) to (f) are Object 1 to 6.

By adjusting the lighting conditions and the camera settings, we can easily obtain images with shadows, low contrast, and defocus blur. Each effect is measured at 4 levels of degradation. The clear and distorted sample images of object 2 in Fig. 6.1 are listed in Fig. 6.2. The detailed lighting and camera settings are listed in Table 6.1. Thus total 16848 images taken indoor including clear and distorted image form the real image database for classification.

Table 6.1. Lighting Condition and Camera Settings for Real Image Database.

Mode		Focus Length	Shutter Speed	Flash	Lighting
Clear		Auto	1/60 sec	ON	2 lamps, side
Dark	Level 1	Same as Clear mode	1/10 sec	OFF	1 lamp, top
	Level 2		1/20 sec		
	Level 3		1/40 sec		
	Level 4		1/60 sec		
Shadow	Level 1	Same as Clear mode	1/15 sec	OFF	1 lamp, 2 inch from side, 55 feet high
	Level 2		1/25 sec		1 lamp, 2 inch from side, 20 feet high
	Level 3		1/10 sec		1 lamp, 4 inch from side, 55 feet high
	Level 4		1/5 sec		1 lamp, 6 inch from side, 55 feet high
Blur	Level 1	8 inch	1/60 sec	ON	2 lamps, side
	Level 2	10 inch			
	Level 3	12 inch			
	Level 4	14 inch			

6.2 Preprocessing and Recognition Using Bessel K Forms

The raw images captured by the camera are first cropped in a square window with the object in the center. Since in practice, the objects are projected in different backgrounds. To recognize objects under the real backgrounds, we further generate masks for each object in the images. Two recognition methods are performed, PCA-based method and Bessel K forms based method, which are introduced in Chapter 4. For Bessel K forms based method, only the pixels belong to objects are used to estimate the parameters. For PCA method, all segmented objects are placed in the black background to perform classification.

First, we test the capability of recognition methods in dealing with pose changes of objects. All images in the clear mode are divided into two nonoverlapping sets, training and testing. A bank of 38 filters including Gaussian filters, Laplacian of Gaussian filters and Gabor filters are utilized to process data. The recognition performance is evaluated as the number of training samples change. The training images are selected with equal intervals of rotation angles and at elevation angle 20.3. All images are further resized to 64×64 and are normalized using equation (4.8). Table 6.2 summarizes the recognition

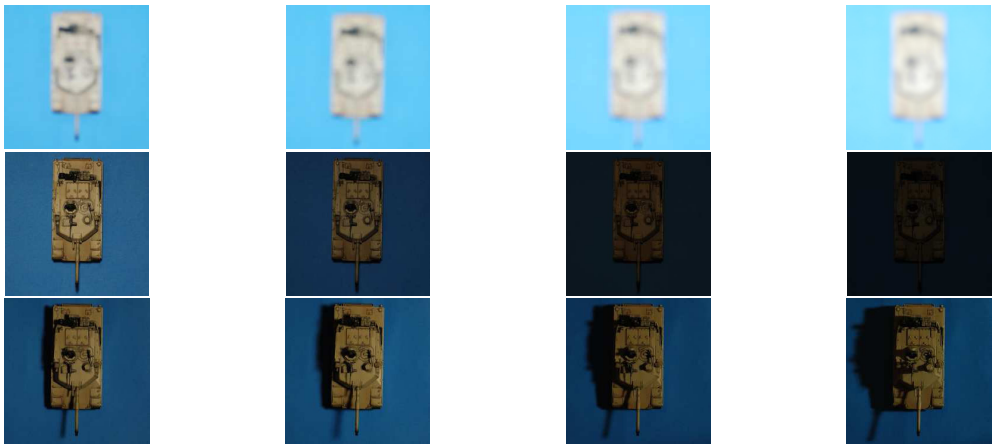


Figure 6.2. Sample distorted images of object 2. From the first row to the third row are images with defocus blur, low illumination and shadow. From the left column to the right column are distorted images from Level 1 to Level 4.

results using PCA and Bessel K forms. Note that in general, Bessel K forms outperform PCA method.

Table 6.2. Correct Recognition Rates for Die Cast dataset Using PCA and Bessel K Forms.

Die Cast dataset		
Train/Test per object	PCA	Bessel K
36/180	87.41%	88.70%
24/192	84.38%	89.58%
18/198	79.21%	87.46%
8/208	60.82%	71.39%
4/212	49.84%	49.69%

The results in Table 6.2 are obtained using the combined metric described in Section 4.1. Given 24 training images per object, the correct recognition rates are 86.98% using Pseudo-metric only, 92.79% using K-measure only and 89.58% using the combined metric. Similar to the results in Table 4.2, we can conclude that the combination algorithm balances the accuracy and speed.

Then we fix the training samples to be 24 clear images per target, and test the influence of camera and lighting effects on recognition performance. The average correct recognition rates as a function of distortion level are shown in Fig. 6.3. Level 0 corresponds to the case when clear images are tested. From the results, we observe that the

recognition performance drops dramatically even for the lowest distortion level for all blur, low illumination and shadow conditions.

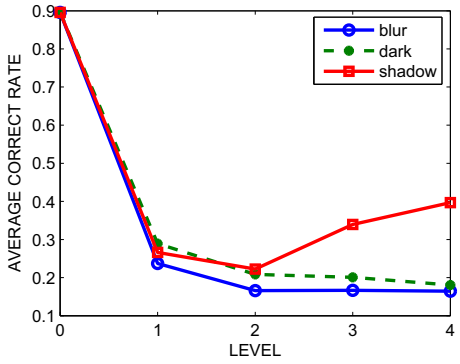


Figure 6.3. Recognition performance under different effects.

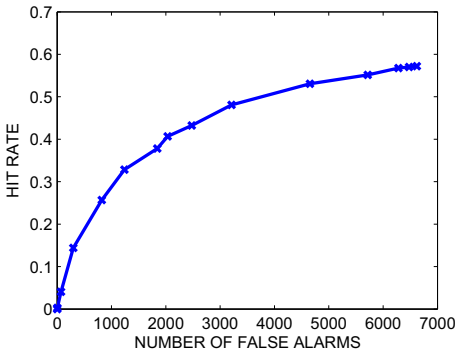


Figure 6.4. The ROC curve of detected results after adjustment.

6.3 Experiments and Results in Detection and Recognition Systems

The 6 objects combined with 3 non-target vehicles are selected for taking outdoor images. To generate real images for detection, all vehicles are arbitrarily placed on the ground, grass, sand and other real world backgrounds. The camera is located 4 to 5 meters apart from the objects. Different zoom scales are applied. Images are taken at different time with different lighting conditions. Thus a set of images are captured given different camera poses. We manually departed the images used for detection into 3 categories, simple background, complex background and complex background with occlusion. Sample images of each categories are shown in Fig. 6.5. All images for detection are resized to 800×536. The number of images and objects included for each category are listed in Table 6.3.

Table 6.3. Information of Detection Images.

Background	Number of Images	Number of Objects
Simple	299	997
Complex	100	422
Complex with occlusion	50	128
Total	499	1547

To train the detector for all objects, we use 3888 positive images and 172 real negative images. All the positive images are modified from the images in the Die Cast dataset. Chroma keying technique, which removes the blue background, is used to extract the objects. The objects after keying are then resized properly and put in a window from a set of background images. The windows from the background images are randomly selected to avoid the detector to learn features from the backgrounds. Sample positive images used for training the detector are shown in Fig. 6.6. All positive images are further resized to 24×24. A 20 stage cascade classifier with 910 weak classifiers in it is formed.

All images listed in Table 6.3 are used to evaluate the performance of the detector. The detection results are shown in Table 6.4. Since only the segmented objects can be processed in recognition, we further use the segmentation results to adjust the windows location. We can see that after adjustment, the number of hits increases and the number of miss and false alarm decrease. The ROC curve after adjustment is drawn in Fig. 6.4. The detection results are further distributed into 3 categories in Table 6.4. We can conclude that the complexity of background and the occlusion constitute more challenges for detection.

Table 6.4. Detection Results with and without Adjustment.

	Hits	Miss	False Alarms
Without Adjustment	764	783	3620
With Adjustment	788	759	3504
Simple	598	399	2066
Complex	148	274	658
Complex with occlusion	42	86	780

We further apply PCA and Bessel K based recognition methods to the detected regions with segmentation. 24 images per object are used as the training samples. Only 15.42% detected objects can be recognized correctly using PCA method. The performance of Bessel K based method is 31.98%. By analyzing the detected results in Fig. 6.7, we observe that some detected regions only occupy part of objects and the real lighting and backgrounds make the segmentation difficult. All these factors make the final system performance very low.

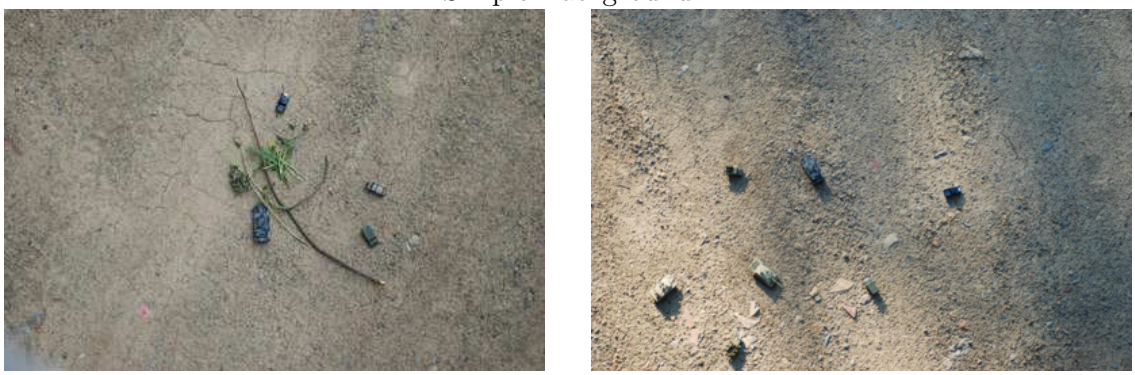
We manually select 35 background textures from 7 negative images. These background textures are further filtered using the designed filter bank. Pairwise distances are calculated among the background textures and each detected window by using the

Pseudo-metric of two distribution functions. K-mean cluster is then applied to separate all the background windows and each detected window into two parts. If the detected window is clustered with any background texture, then it is rejected as clutter. The number of false alarm after clutter rejection is 1392 and the number of hits is 750. This method removes most false alarms while keep the hit rate.

6.4 Summary

In this Chapter, a real image dataset is formed using 6 die cast military models. The objects are projected from different view points. At each view point, the clear and distorted images are taken for each object. These distortions include defocus blur, low illumination and shadow. Bessel K based method outperform PCA based method at different training to testing data ratios. The influence of camera and lighting effects on recognition performance is also evaluated. The detection performance using Haar-like features is tested based on 499 images. The recognition is further performed on the detected regions after segmentation. The complexity of backgrounds and the occlusion decrease both the detection and recognition performance.

Simple Background



Complex Background

Complex Background with Occlusion

Figure 6.5. Sample real images on real background.

Figure 6.6. Sample positive images used to train a detector.

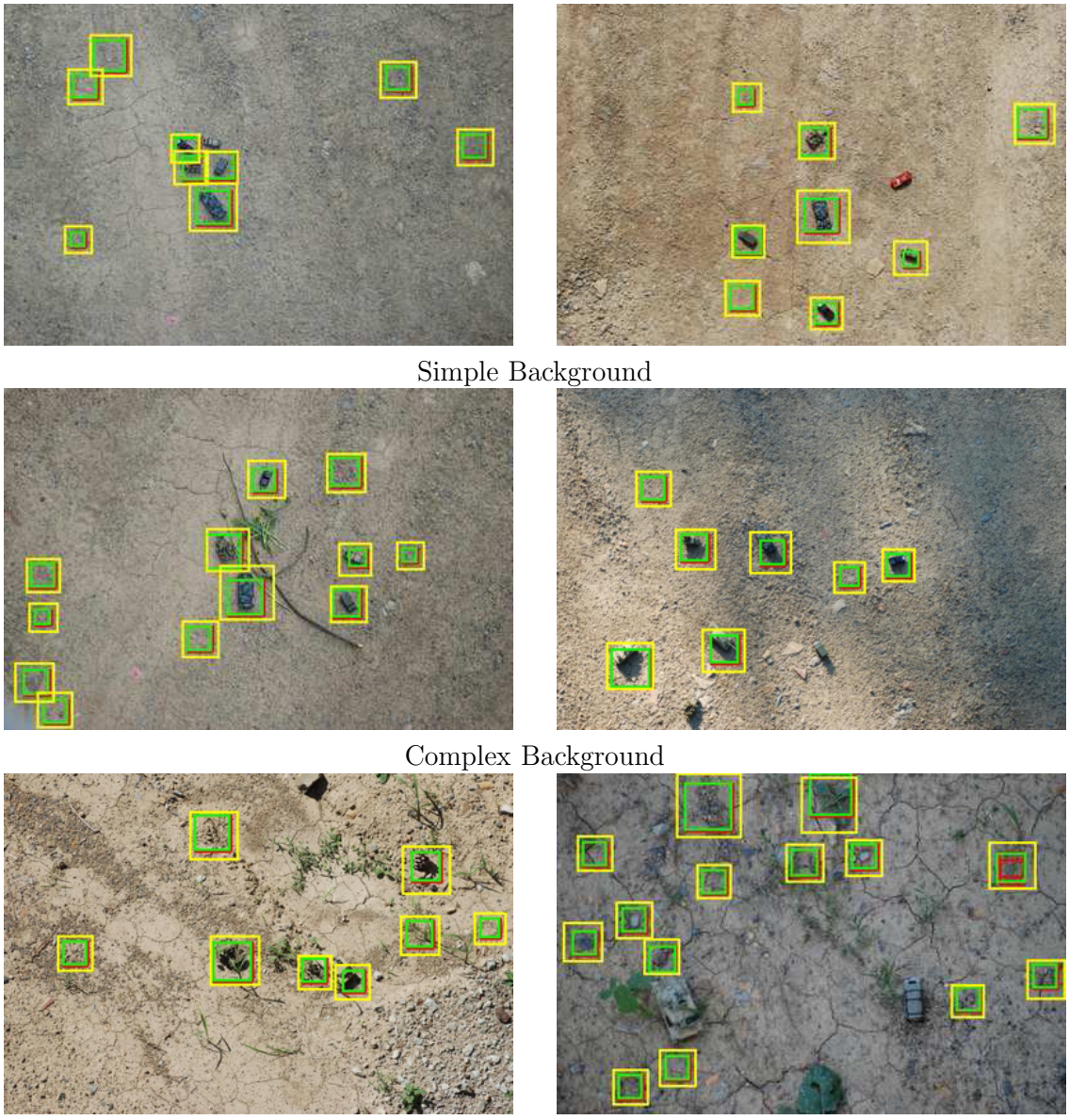


Figure 6.7. Sample images including detection regions. Detector outputs are marked in red, results after combination are marked in green and results after segmentation are marked in yellow.

CHAPTER 7

RECOGNITION CAPACITY OF ATR SYSTEMS

In many large scale recognition applications knowledge of the limiting capabilities of designed recognition systems is crucial. These limits, however, are determined by a variety of factors including a source coding technique used to process data, quality, complexity, and variability of the collected data. Given an encoding technique, the remaining factors can be attributed to recognition channel introduced and characterized by Schmid and O'Sullivan [50] and further analyzed by Westover and O'Sullivan [63, 64]. Similar to a communication channel, a recognition channel is characterized by its capacity, with the difference being recognition capacity. In an object recognition problem, recognition capacity is the exponent in an exponential approximation to the maximum number of objects/targets that can be successfully recognized with probability of error close to zero when the number of informative samples gets large. In this chapter, we briefly summarize the results by Schmid and O'Sullivan on recognition capacity and then evaluate the capacity of Principal Component Analysis (PCA)-based Object Recognition systems. We consider both the case of a single frame and the case of multiple frames of the same object.

7.1 Recognition Capacity

Consider an object recognition system. A majority of practical recognition systems are designed to operate in two modes: enrollment / training mode and recognition / testing mode. A block diagram of a typical recognition system is displayed in Fig. 7.1. During the training mode, the signals or images are encoded to obtain a set of attributes or feature vectors, which are further stored in a library or database. During the recognition mode, an encoded query image / signal, that is, image / signal submitted for recognition and thus presumably containing information about an object to be recognized, is compared against each entry in the library of objects. The recognition system then outputs the identity of the object. It was shown in [50] that a traditional object recognition system, given encoded data stored in an object library and models for encoded data, can be

viewed as a recognition channel, an analog of a communication channel. Fig. 7.2 presents a block diagram of a recognition channel. Given encoded data and an appropriate model, the problem of object recognition can be restated as a maximum likelihood (ML) decoding problem [12, 20, 61]. Similar to a communication channel, a recognition channel is characterized by its capacity, with the difference being recognition capacity. In an object recognition problem, recognition capacity can be thought as being the maximum number of objects/targets that can be successfully recognized with probability of error close to zero when the number of informative samples gets large.

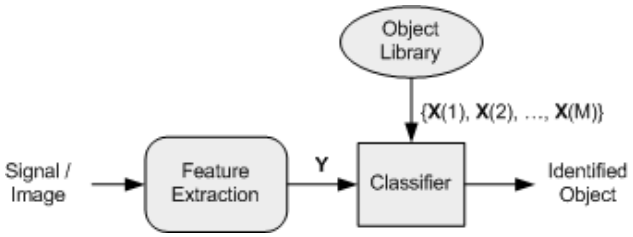


Figure 7.1. Structure of a recognition system. $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(M)$ are images or encoded data characterizing M object classes. The vector $\mathbf{Y}$ is a query image or encoded data.

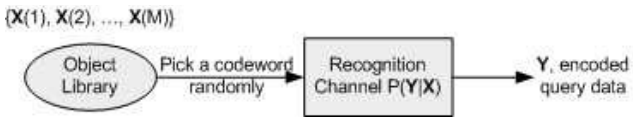


Figure 7.2. Structure of a recognition channel. $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(M)$ are independent codewords (images or encoded data). $\mathbf{Y}$ is a distorted noisy version of one of the codewords in the object library.

Suppose that an object library is composed of templates (processed and encoded images) $x^n(1), \ldots, x^n(M)$ of M distinct objects. Here n is the length of a template (codeword). Denote by y^n a template submitted for recognition. We assume that y^n contains information about one of the objects stored in the object library. From [50] the templates in the library can be modeled as realizations of M independent and identically distributed (i.i.d.) random vectors $X^n(1), \ldots, X^n(M)$ and the template submitted for recognition as a realization of a random vector Y^n. The random vector Y^n is assumed to be a distorted noisy version of one of the M random vectors characterizing M encoded objects in the object library. During recognition/testing mode, the query template y^n is compared against each template in the object library. Since y^n is a distorted version of a

library template, then one of the templates in the database and the template submitted for identification will have some information in common and thus can be described by a joint probability density function $p_{X^n,Y^n}(\cdot,\cdot)$. The remaining $M-1$ templates in the object library and the template submitted for identification do not have information in common and thus can be described by the product of probability density functions $p_{X^n}(\cdot) \times p_{Y^n}(\cdot)$ with $p_{X^n}(\cdot)$ and $p_{Y^n}(\cdot)$ being the marginals of $p_{X^n,Y^n}(\cdot,\cdot)$. Under this setting, the problem of object recognition can be stated as an M-ary hypothesis testing problem. The vector of test statistics is an M-dimensional vector of information densities, which are in this case are i.i.d. components. A more detailed description and analysis to this problem can be found in [50].

Given mathematical models for an encoded image of an object, and for noise and distortions that the encoded image contains, one may derive an expression for the *capacity* (constrained capacity) of an object recognition system. Again, the object library combined with a query object template can be viewed as a recognition channel, where the query object template, a distorted, noisy version of a template in the library, is observed on the output of the channel. Theoretical value for the constrained capacity of a recognition system, given a source encoding technique, can be obtained by following Information Theoretical framework. Given templates, their probability distribution can be empirically evaluated using classical parametric and modern nonparametric estimation techniques. The evaluated joint and marginal probability distributions for a template of an object to be recognized and for a template from the object library can then be used to form the information density:

$$i_n(\cdot,\cdot) = \frac{1}{n}\log\frac{p_{X^n,Y^n}(\cdot,\cdot)}{p_{X^n}(\cdot)p_{Y^n}(\cdot)}, \tag{7.1}$$

where we assume that the ratio of the joint density to the product of marginal densities is well defined. When the template distributions are known, the *constrained recognition capacity* is the mutual information rate defined as

$$\bar{I}(X,Y) = \lim_{n\to\infty} E[i_n], \tag{7.2}$$

where the expected value is with respect to the joint distribution.

In practical cases, given encoded data (templates), their probability distribution are empirically evaluated using classical parametric and modern nonparametric estimation techniques from a set of training data. Then the expression under the expected value

in (7.2) will contain estimated parameters and will not present a deterministic sequence any more. Thus, in practice, we deal with random sequences.

7.1.1 Empirical Mutual Information Rate

Let $p(x^n : \Theta)$ be a probability model of a template from an object library. The model $p(x^n : \Theta)$ is a parametric probability density function (p.d.f.) parameterized by a vector of K parameters $\Theta = [\theta_1, \ldots, \theta_K]^T$. Let $p(y^n : \Theta, \Psi)$ be a probability model of noisy transformed templates from an object library. The vector of L parameters $\Psi = [\psi_1, \ldots, \psi_L]^T$ represents the parameters of the distortions introduced by a recognition channel. The parameters Θ and Ψ are unknown and therefore have to be estimated from a set of training data (assume labeled data). All other assumptions about the probability models for templates are similar to the assumptions in Sec. 7.1.

Denote by $\hat{\Theta}$ and $\hat{\Psi}$ the vectors of estimated parameters. Then the information density with unknown parameters of the p.d.f.s replaced by the estimated parameters becomes:

$$i_n(\hat{\Theta}, \hat{\Psi}) = \frac{1}{n} \log \frac{p_{X^n,Y^n}(\cdot, \cdot : \hat{\Theta}, \hat{\Psi})}{p_{X^n}(\cdot : \hat{\Theta}) p_{Y^n}(\cdot : \hat{\Theta}, \hat{\Psi})}. \tag{7.3}$$

Taking the expected value with respect to the joint p.d.f. and substituting the estimated parameters in the final expression, we obtain a plug-in estimate of the empirical mutual information rate, denote it by $I_n(\hat{\Theta}, \hat{\Phi})$:

$$I_n(\hat{\Theta}, \hat{\Psi}) = E_{p_{X^n,Y^n}(\cdot,\cdot:\hat{\Theta},\hat{\Psi})}[i_n(\hat{\Theta}, \hat{\Psi})]. \tag{7.4}$$

Due to the estimated parameters $\hat{\Theta}$ and $\hat{\Psi}$ the sequence of $I_n(\hat{\Theta}, \hat{\Psi})$ is a sequence of random variables. In the following we will interchangeably use $I_n(\hat{\Theta}, \hat{\Psi})$ and $I_n(M)$ to denote the empirical mutual information rate. The latter notation indicates implicit dependence of estimated parameters on the number of classes, M, to recognize.

Definition 1 *An estimate is a plug-in estimate of a function if unknown parameters of the function are replaced by their estimates.*

It is potentially possible to state and prove convergence of the sequence in (7.4) in probability or with probability one, provided we are dealing with a "nice" family of p.d.f.s (for instance, the exponential family) and maximum likelihood (ML) estimates of unknown parameters. In general, to state conditions for convergence of the sequence in (7.4) and prove convergence with probability one appears to be hard.

In this work, we are interested in deriving the expression (7.4) under the condition that images of objects are encoded using a PCA-based approach, an empirical version

of Karhunen-Loeve expansion [61]. We are further interested in observing and analyzing a trend of the empirical mutual information rate as the number of objects and template length grow in a certain proportion. Ultimately we would like to find the point of empirical recognition capacity that we define as follows.

From (7.4) the empirical mutual information rate is parameterized by the length of templates, n, and by the number of classes to recognize, M. Therefore, we can form a sequence of empirical mutual information rates. Following the definition of the operational capacity in communication theory (see for example, [12, Ch.8] for details), we define the recognition rate as $R = \log(M)/n$. If we had a sequence of PCA codes $(n, 2^{nR})$ with the recognition rate R, we would be able to evaluate empirically the trend of the sequence of $I_n(M)$ as a function of the rate R, since both $I_n(M)$ and R are functions of M and n. Then *empirical recognition capacity* can be defined as a point of intersection of the empirical mutual information rate curve plotted as a function of the recognition rate and the diagonal line bisecting the first quadrant.

7.2 Recognition Capacity under the Constraint of PCA Encoded Data

In this section we will derive an expression for the constrained capacity of noisy recognition channel under the constraint of Principle Component Analysis (PCA)-encoded data. PCA([30, 32]) is a technique which has been widely used in data analysis and compression. This is a linear transform method, which projects the data into a lower dimensional space (eigenspace) chosen to maximally capture variability in the data. A typical PCA-based recognition system operates in two modes: training and testing. During the training mode, the PCA space is empirically evaluated using labeled training data. During the testing mode, a set of signals/images is projected onto the estimated PCA space, resulting in templates that are further stored in a dataset or library. Then the performance of the object recognition system is evaluated by projecting query images onto the PCA space and comparing query templates against all templates in the library. In this work, we apply the global PCA method, which uses a single training image per object class.

The images of M objects from the object library when projected onto the eigenspace form templates $x^n(1), \ldots, x^n(M)$. To recognize an object based on an acquired image, the image is projected onto the same eigenspace and is represented by a set of components that can be arranged in a n-dimensional vector, y^n.

7.2.1 Model for PCA Encoded Data: Single Image Case

Suppose that a set of templates associated with a set of objects is available. We model templates $x^n(1), \ldots, x^n(M)$, n-dimensional vectors of components in PCA representation of the objects $1, \ldots, M$, as realizations of i.i.d. Gaussian random vectors X^n with mean zero and unknown diagonal covariance matrix Λ with entries equal to the eigenvalues of the empirical covariance matrix, that is, $X^n \sim \mathcal{N}(0, \Lambda)$. We assume that distortions and noise in encoded images can be modeled as a realization of white Gaussian vector with mean zero and unknown diagonal covariance matrix Λ_N, that is, $W^n \sim \mathcal{N}(0, \Lambda_N)$. Then a noisy template y^n presented for identification is modeled as a realization of a random vector Y^n, $Y^n = X^n + W^n$. Thus Y^n is also Gaussian distributed with mean zero and diagonal covariance matrix $\Lambda + \Lambda_N$. The unknown matrices are estimated using labeled training data. The details of the procedure are described in later sections. The information density, for this case is given by:

$$i_n(\hat{\Lambda}, \hat{\Lambda}_N) = -\frac{1}{2n} \sum_{i=1}^{n} \left(\frac{X_i^2}{\hat{\sigma}_i^2} - 2\frac{X_i Y_i}{\hat{\sigma}_i^2} + \frac{\hat{\lambda}_i Y_i^2}{\hat{\sigma}_i^2(\hat{\lambda}_i + \hat{\sigma}_i^2)} - \log\left(1 + \frac{\hat{\lambda}_i}{\hat{\sigma}_i^2}\right)\right), \tag{7.5}$$

where $\hat{\lambda}_i$ and $\hat{\sigma}_i^2$ are the i^{th} entry along the diagonal in the estimated matrices $\hat{\Lambda}$ and $\hat{\Lambda}_N$, respectively. Note since the expression for $i_n(M)$ contains estimated parameters, it implicitly depends on the number of classes M used in training. The empirical mutual information rate, given the length of templates, n, is the average of the information density in (7.5) with respect to the joint distribution of X^n and Y^n :

$$I_n(\hat{\Lambda}, \hat{\Lambda}_N) = \frac{1}{n} E_{\mathbf{X},\mathbf{Y}}\{i_n(M)\} = \frac{1}{2n} \sum_{k=1}^{n} \log\left(1 + \frac{\hat{\lambda}_k(M)}{\hat{\sigma}_k^2(M)}\right), \tag{7.6}$$

where $\hat{\sigma}_k^2(M)$ is the estimated variance of the k-th component of the noise and $\hat{\lambda}_k(M)$ are estimated eigenvalues of the empirical covariance matrix $\hat{\Lambda}$.

7.2.2 Model for PCA Encoded Data: Multiple Image Case

To derive an expression for the empirical mutual information rate of the PCA-based Object Recognition system when more than a single image from the same object is available we assume that the relative pose of the same object in two different images is known or can be estimated. For two image case, let X_1^n and X_2^n be two PCA-based templates characterizing an object at the unknown poses α_1 and α_2. We model the

combined vector $[X_1^n, X_2^n]^T$ as being Gaussian distributed with zero mean and block diagonal covariance matrix

$$\Lambda_1 = \begin{bmatrix} \Lambda & \Lambda\rho_{1,2} \\ \Lambda\rho_{1,2} & \Lambda \end{bmatrix} \tag{7.7}$$

where $\Lambda\rho_{1,2}$ is the matrix with components $\lambda_i\rho_i(\delta\alpha)$, $i = 1, \ldots, n$, and $\delta\alpha = |\alpha_1 - \alpha_2|$ is a known parameter or can be estimated.

The combined noisy templates $[Y_1^n, Y_2^n]^T$ submitted for recognition are PCA components from a randomly selected pair of templates from one of M hypothesis. The noise in Y_1^n and the noise in Y_2^n are independent Gaussian with the diagonal covariance matrix Λ_N.. Then the combined vector $[Y_1^n, Y_2^n]^T$ follows Gaussian distribution with zero mean and block diagonal covariance matrix:

$$\Lambda_2 = \begin{bmatrix} \Lambda + \Lambda_N & \Lambda\rho_{1,2} \\ \Lambda\rho_{1,2} & \Lambda + \Lambda_N \end{bmatrix}. \tag{7.8}$$

To further evaluate the empirical mutual information rate, we find the joint distribution of $[X_1^n, X_2^n]^T$ and $[Y_1^n, Y_2^n]^T$. If the pair of templates and a pair of candidates have a signal in common, the joint distribution is assumed to follow Gaussian distribution with mean zero and covariance matrix R_1 composed of four block matrices given in (7.7) and (7.8):

$$R_1 = \begin{bmatrix} \Lambda_1 & \Lambda_1 \\ \Lambda_1 & \Lambda_2 \end{bmatrix}. \tag{7.9}$$

If the pairs do not have a signal in common, their joint distribution follows Gaussian distribution with mean zero and covariance matrix R_0, which is equal to R_1 with off-diagonal block matrices set to zero.

The joint empirical mutual information rate in this case is calculated by replacing X^n and Y^n in (7.6) with the vectors $[X_1^n, X_2^n]^T$ and $[Y_1^n, Y_2^n]^T$, respectively.

If $\{\rho_j(\delta\alpha) \neq \pm 1\}$, for all $j = 1, \cdots, n$, then both R_0 and R_1 are symmetric, positive definite matrices. Assume that the matrices R_0 and R_1 are replaced by their estimates $\hat{R}_0$ and $\hat{R}_1$. The joint information density for the pairs of templates is given by

$$i_n(\hat{R}_0, \hat{R}_1) = -\frac{1}{2n}\left\{[X_1^n, X_2^n, Y_1^n, Y_2^n](\hat{R}_1^{-1} - \hat{R}_0^{-1})[X_1^n, X_2^n, Y_1^n, Y_2^n]^T + \log\det(\hat{R}_1\hat{R}_0^{-1})\right\}. \tag{7.10}$$

Replacing $\hat{R}_1$ and $\hat{R}_0$ with their expressions in terms of block matrices results in the expression for the PCA-based joint empirical mutual information rate:

$$\bar{I}_n^{(2)}(\hat{R}_0, \hat{R}_1) = \frac{1}{2n} \sum_{i=1}^{n} \log \left[\left(1 + \frac{\hat{\lambda}_i}{\hat{\sigma}_i^2} \right)^2 - \left(\frac{\hat{\lambda}_i \hat{\rho}_i(\delta\alpha)}{\hat{\sigma}_i^2} \right)^2 \right], \tag{7.11}$$

where the superscript in $\bar{I}_n^{(2)}$ indicates that the empirical mutual information rate is evaluated for the case of two frames per object. To evaluate the empirical mutual information rate, we would need to form a sequence of $\bar{I}_n^{(2)}(\hat{R}_0, \hat{R}_1)$ parameterized by the increasing parameters n and M in a certain proportion as discussed in the following sections.

Two special cases of (7.11) are of interest. When $\{\rho_j(\delta\alpha) = 0\}$, for all $j = 1, \ldots, n$, which indicates that $X_1^n(\alpha_1)$ and $X_2^n(\alpha_2)$ are uncorrelated, the joint empirical mutual information rate from (7.11) is equal to twice the empirical mutual information rate calculated using a single image per object. When $\{\rho_j(\delta\alpha) = \pm 1\}$, for any $j = 1, \ldots, n$, both R_1 and R_0 are not full rank matrices, thus the inverse matrices do not exist. In general, assume $\rho_j(\delta\alpha) = 1$ or -1, $j = 1, \ldots, p$, the combined vector $[X_1^n, X_2^n]^T$ can be reordered as $\left[X_1^p, X_2^p, X_1^{n-p}, X_2^{n-p}\right]^T$, where X_i^p and X_i^{n-p}, $i = 1, 2$ are the first p entries and the last $n - p$ entries of X_i^n. Since X_2^p is fully dependent on X_1^p, the probability of the vector $[X_1^n, X_2^n]^T$ is equal to the probability of the reduced vector $\left[X_1^p, X_1^{n-p}, X_2^{n-p}\right]^T$. The corresponding noisy vector $[Y_1^n, Y_2^n]^T$ is reduced to $\left[Y_1^p, Y_1^{n-p}, Y_2^{n-p}\right]^T$. Now, the empirical mutual information rate of the reduced pair of templates and the corresponding pair of candidates is the summation of two parts. The first part is the empirical mutual information rate of the recognition channel formed by the first p components of a single frame. The other part is the empirical mutual information rate of the recognition channel formed by the $n - p$ components from two frames separated by the relative orientation $\delta\alpha$. Thus the joint empirical mutual information rate in this case is given by:

$$\bar{I}_n^{(2)}(\hat{R}_0, \hat{R}_1) = \frac{1}{2n} \sum_{i=p+1}^{n} \log \left[\left(1 + \frac{\hat{\lambda}_i}{\hat{\sigma}_i^2} \right)^2 - \left(\frac{\hat{\lambda}_i \hat{\rho}_i(\delta\alpha)}{\hat{\sigma}_i^2} \right)^2 \right] + \frac{1}{2n} \sum_{i=1}^{p} \log \left(1 + \frac{\hat{\lambda}_i}{\hat{\sigma}_i^2} \right), \tag{7.12}$$

where $\hat{\lambda}_i$ and $\hat{\sigma}_i^2$ are the estimated parameters. When all correlation coefficients take values 1 or -1, the expression for the empirical mutual information rate (7.12) degenerates into the empirical mutual information rate of recognition channel based on a single template as in (7.6).

The results above can be extended to obtain the empirical mutual information rate using multiple images of the same object. Suppose that S distinct images are available from the same object. Let $X_1^n, \ldots, X_S^n$ be S PCA-encoded images acquired from the same object at unknown poses $\alpha_1, \ldots, \alpha_S$. The noisy candidates $Y_1^n, \cdots, Y_S^n$ are S PCA-encoded images of an object submitted for recognition. Similar to the two image case,

we assume that the combined vector $[X_1^n, \cdots, X_S^n]^T$ and $[Y_1^n, \cdots, Y_S^n]^T$ follow Gaussian distribution with zero mean and estimated covariance matrix Λ_1 and Λ_2 given by:

$$\Lambda_1 = \begin{bmatrix} \Lambda & \Lambda\rho_{1,2} & \cdots & \Lambda\rho_{1,S} \\ \Lambda\rho_{1,2} & \Lambda & \cdots & \Lambda\rho_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda\rho_{1,S} & \Lambda\rho_{2,S} & \cdots & \Lambda \end{bmatrix},$$

$$\Lambda_2 = \begin{bmatrix} \Lambda+\Lambda_N & \Lambda\rho_{1,2} & \cdots & \Lambda\rho_{1,S} \\ \Lambda\rho_{1,2} & \Lambda+\Lambda_N & \cdots & \Lambda\rho_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda\rho_{1,S} & \Lambda\rho_{2,S} & \cdots & \Lambda+\Lambda_N \end{bmatrix},$$

where $\Lambda\rho_{s,t}$, $s,t = 1,\ldots,S$, is the diagonal covariance matrix of X_s^n and X_t^n, with diagonal element $\lambda_i\rho_i(\delta\alpha_{s,t})$ and $\delta\alpha_{s,t} = |\alpha_s - \alpha_t|$. The matrices are estimated using training data characterizing M distinct objects. Again we assume that the absolute orientation of an object is unknown. However, the test statistic, information density in this case, and thus the empirical mutual information rate, is a function of the relative orientation. For stationary objects or for slowly moving objects, the relative orientation is either known or can be estimated with high fidelity.

The joint distribution of $[X_1^n, \ldots, X_S^n]^T$ and $[Y_1^n, \ldots, Y_S^n]^T$ in multiple images case is similar to the distribution of two encoded images. If the set of templates and the set of candidates have a signal in common, then the joint distribution follows Gaussian distribution with zero mean and covariance matrix R_1 defined in 7.9. Otherwise, their joint distribution follows Gaussian distribution with mean zero and covariance matrix R_0, which is equal to R_1 with off-diagonal block matrices set to zero. If no correlation coefficients equals to 1 or -1, R_0 and R_1 are symmetric, positive definite matrices. By substituting R_1 and R_0 into the expression for the joint empirical mutual information rate, we can find the empirical mutual information rate of the recognition channel for multiple images case. For example, when $S = 3$, the joint empirical mutual information rate is

$$I_n^{(3)}(\hat{R}_0, \hat{R}_1) = \frac{1}{2n}\sum_{i=1}^{n} \log\left[\left(1+\frac{\hat{\lambda}_i}{\hat{\sigma}_i^2}\right)^3 - \left(1+\frac{\hat{\lambda}_i}{\hat{\sigma}_i^2}\right)\left(\frac{\hat{\lambda}_i}{\hat{\sigma}_i^2}\right)^2 A_1 + 2\left(\frac{\hat{\lambda}_i}{\hat{\sigma}_i^2}\right)^3 A_2\right], \quad (7.13)$$

where $A_1 = \hat{\rho}_i^2(\delta\alpha_{1,2}) + \hat{\rho}_i^2(\delta\alpha_{1,3}) + \hat{\rho}_i^2(\delta\alpha_{2,3})$ and $A_2 = \hat{\rho}_i(\delta\alpha_{1,2})\hat{\rho}_i(\delta\alpha_{1,3})\hat{\rho}_i(\delta\alpha_{2,3})$.

If any correlation coefficient equals to 1 or -1, the redundant bits from the input and corresponding output will be removed. The joint empirical mutual information rate will be reduced to the expression similar to the expression in (7.12).

7.3 Model Verification

In this section, we validate the model for PCA-encoded data described above. We assume that a number of images of the same object can be acquired and processed. Two special cases are considered: (1) an object is represented by a single class in the database; (2) an object is represented by a set of classes in the database. This is the case when multiple images of the same object are acquired from different orientation and elevation angles. We present one example of empirical capacity evaluation for the first special case. We involve data from COIL-100. The object classes are formed using images acquired at zero degree orientation (frontal views) of 100 objects (subset I). To evaluate the recognition performance, we use synthetically distorted images at Level 3.

To illustrate the second special case, we involve a dataset generated using a 3D simulation tool and COIL-100 dataset. We present two examples. In the first example, we use data generated from the ATR Training Tool. The generated clean and distorted data (images) are assumed to mimic data acquired by optical cameras mounted on board of a network of Unmanned Aerial Vehicles. In the second example, we involve images of 11 toy cars from COIL-100 dataset (subset II). In each example, the data is then subdivided into two nonoverlapping subsets: training and testing. The clean images at orientation $0, 15, 30, \cdots, 345$ form the training set with each angle representing a single hypothesis. For the generated ATR dataset, the training images are projected at elevation 15 degree. These 24 images per object are used to generate PCA-based templates of object as well as to estimate covariance matrix Λ. The images with orientations $\alpha - 5, \alpha + 5$ are treated as transformed realizations of the image at the orientation α and represent the same hypothesis. The total number of classes formed using the partition above is equal to 72 (24×3) for ATR dataset and 264 (24×11) for the COIL-100 dataset.

7.3.1 Parameter Estimation

Prior to recognition, an object is located and placed in a canonical (or object-centered) reference frame suitable for recognition. All training and testing images from ATR dataset and COIL-100 dataset are further normalized to image size 64×64.

PCA is a global encoding algorithm [30]. Consider an object library with M classes. Assume that M preprocessed images (one per class) $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_M$ are available for train-

ing of a recognition system that uses PCA encoded data for recognition. Training of a PCA recognition system is reduced to estimation of a set of parameters including scatter matrix and its eigenvalues and eigenvectors. Once a PCA system is trained, a set of images, one per class, are projected onto the estimated PCA space, that is, encoded. The encoded images are stored in the database in the form of templates.

In this work we assume that resolution, r, of an image is fixed and consider the following two cases of training a PCA recognition system. In the case when $M >> r$, the estimate of the scatter matrix, Σ, is given by:

$$\Sigma = \frac{1}{(M-1)} \sum_{m=1}^{M} (\mathbf{I}_m - \overline{\mathbf{I}})(\mathbf{I}_m - \overline{\mathbf{I}})^T = \frac{1}{(M-1)} \mathbf{A}\mathbf{A}^T,$$

where $\overline{\mathbf{I}}$ is the sample mean and $\mathbf{A} = [\mathbf{I}_1 - \overline{\mathbf{I}}, \mathbf{I}_2 - \overline{\mathbf{I}}, \ldots, \mathbf{I}_M - \overline{\mathbf{I}}]$. In the case when $M << r$, nonzero eigenvalues of the scatter matrix Σ are the same as the eigenvalues of $\frac{1}{(M-1)}\mathbf{A}^T\mathbf{A}$. The eigenvectors are obtained by multiplying the eigenvectors of $\frac{1}{(M-1)}\mathbf{A}^T\mathbf{A}$ by $\mathbf{A}$ from the left. The estimated matrix Σ is decomposed using eigenvalue decomposition $\Sigma = \mathbf{Q}\Lambda\mathbf{Q}^T$, where Λ is the matrix of eigenvalues and $\mathbf{Q}$ is the orthogonal matrix with columns composed of eigenvectors of Σ. In practice, only a small number n of largest eigenvalues is selected from the total number of eigenvalues equal to $\min\{r, M\}$. Then a new matrix $\widetilde{\mathbf{Q}}$ with vector columns corresponding to the essential eigenvalues is formed (see for details [30]).

The template $X^n(m)$ of the mth class, an n-dimensional vector, is obtained by projecting image $(\mathbf{I}'_m - \overline{\mathbf{I}})$ onto the space formed by the columns of matrix $\widetilde{\mathbf{Q}}$

$$X^n(m) = \widetilde{\mathbf{Q}}^T(\mathbf{I}'_m - \overline{\mathbf{I}}),$$

where "prime" indicates that the data used to form templates do not overlap with training data. A query image is also projected onto the space defined by the eigenvectors of Σ. The query image is now represented by a set of weights that can be arranged in an n-dimensional vector, y^n.

Suppose that L additional training images per class are available to estimate unknown noise variance. Denote by $y^n(m; l)$, $m = 1, \ldots, M$, $l = 1, \ldots, L$, the lth template of length n representing the class m. To estimate the variance of the additive noise, we form a set of vectors $\{y^n(m; l) - x^n(m)\}$. To find the variance of the noise in the kth component of a noisy template, we appeal to sample estimates

$$\hat{\sigma}_k^2 = \frac{1}{ML-1}\sum_{m=1}^{M}\sum_{l=1}^{L}\left(y_k^n(m;l) - x_k^n(m) - \bar{z}_k^n\right)^2,\ k = 1, 2, \cdots, n,$$

where $\bar{z}_k$ is the kth sample mean formed as

$$\bar{z}_k^n = \frac{1}{ML}\sum_{m=1}^{M}\sum_{l=1}^{L}\left(y_k^n(m;l) - x_k^n(m)\right),\ k = 1, 2, \cdots, n.$$

To evaluate the correlation coefficient $\rho(\delta\alpha)$, between the PCA templates of the same object acquired $\delta\alpha$ degrees apart, we use a sample correlation coefficient. The kth correlation coefficient is

$$\hat{\rho}_k(\delta\alpha) = \frac{\sum_{m=1}^{M}\sum_{l=1}^{L}(x_{1,k}^n(m;l) - \bar{x}_{1,k}^n)(x_{2,k}^n(m;l) - \bar{x}_{2,k}^n)}{(ML-1)s_{1,k}^n s_{2,k}^n},$$

where $\bar{x}_1^n$ and $\bar{x}_2^n$ are the sample means of X_1 and X_2, s_1^n and s_2^n are the sample standard deviations of X_1 and X_2 and L is the number of samples per class.

7.3.2 Model Verification: Single Image Case

To verify the stochastic model described in Sec. 7.2.1, we use two non-parametric statistical methods: Kolmogorov-Smirnov test [11] and Shapiro-Wilk test [52] for normality. The Kolmogorov-Smirnov test (D statistic) is to find the greatest discrepancy between the observed and expected cumulative relative frequencies given by,

$$D = \max_x |F_N(x) - F(x)|,$$

where $F(x)$ is the hypothesized distribution (here is a normal distribution) and the empirical distribution function $F_N(x)$ for N observations $x_i,\ i = 1, \cdots, N$ is defined as

$$F_N(x) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases} 1 & if\ \ x_i \le x \\ 0 & otherwise. \end{cases}$$

A p-value p is the smallest significance level at which the null hypothesis would be rejected for the given observation [11]. Let d_{obs} represent the observed value of the D test statistic. Then the p-value is $2 \times \min\{P(D \ge d_{obs}), P(D \le d_{obs})\}$ using the null distribution of D. Small p-value indicates departures of the observations from normality. By setting an acceptable significance value p_{crit} (the critical p-value), the null hypothesis is rejected if $p < p_{crit}$.

The Shapiro-Wilk test uses the weighted order statistics to calculate the test statistic W defined as

$$W = \frac{\left(\sum_{i=1}^{N} a_i x_{(i)}\right)^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2},$$

where x_i, $x_{(i)}$ and $\bar{x}$ are the original data, the ordered data and the sample mean of the observations. The constants a_i are derived from the order statistics of a sample from the standard normal distribution. A p-value is then compared with p_{crit} to make decisions.

In our test, the null hypothesis is that PCA components X^n follows Gaussian distribution with zero mean and estimated variances along the diagonal of Λ. The critical p-value is set to 0.05. The results applying Kolmogorov-Smirnov test and Shapiro-Wilk test to each dataset are reported in Table 7.1, where n is the number of random variables and M is the sample size. Note that normality hypothesis is not rejected for majority of individually treated random components of the vector X^n based on the test results. These results confirm that in spite of using estimated parameters in place of these unknown parameters of the model, it provides a reasonable fit.

Table 7.1. Results of Kolmogorov-Smirnov test and Shapiro-Wilk test for ATR dataset and COIL dataset

Dataset	Subset I of COIL-100	Simulated ATR set	Subset II of COIL-100
n (length of PCA components)	98	71	105
M (number of hypothesis)	100	72	264
Rejected by Kolmogorov -Smirnov test	11	0	2
Rejected by Shapiro-Wilk test	45	17	26

7.3.3 Model Verification: Multiple Image Case

To validate the model for the case of two images, we form the matrix of correlation coefficients between X_1^n and X_2^n and a matrix of p-values, $P_{n\times n}$, for testing "no correlation" hypothesis. The p-value is computed by transforming the correlation to form a t-statistic with $N-2$ degrees of freedom [17], where N is the number of observations. Each entry in the matrix P is evaluated on its significance. If an entry takes value smaller than 0.05, then the corresponding correlation element is significant. Fig. 7.3 displays three matrices of p-values for the cases of $\alpha = 0, 5$ and 10 given two datasets. The entries with $P(i,j) < 0.05$ are marked in black while entries with $P(i,j) > 0.05$ are marked in white.

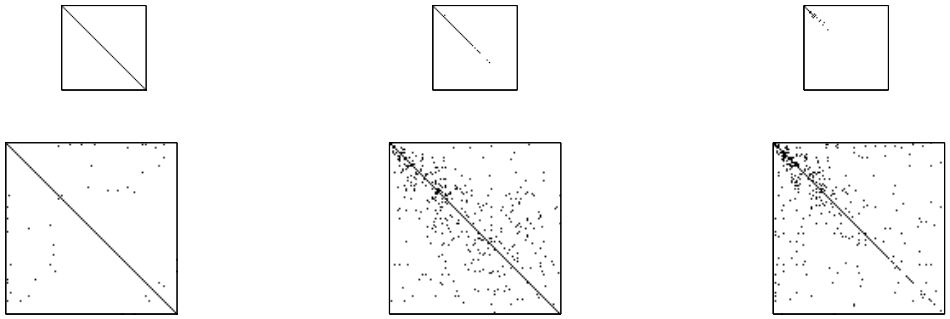


Figure 7.3. The matrix of p-values for three cases of the relative angle $\delta\alpha$: (a) 0, (b) 5 and (c) 10. The first row shows 71×71 matrix from ATR dataset and the second row shows 105×105 matrix from COIL-100 dataset. Black points correspond to $P(i,j) < 0.05$. White points correspond to $P(i,j) > 0.05$.

From the results shown in Fig. 7.3, we can see that only a small portion of elements which are not along the main diagonal are significantly correlated. This is in good agreement with the model introduced in Sec. 7.2.2. The dependence of $\rho_k(\delta\alpha), k = 1, \cdots, n$ on k for two values of $\delta\alpha$ is shown in Fig. 7.4 for both ATR dataset and COIL-100 dataset. As $\delta\alpha$ increases, correlation coefficients decrease. Thus we conjecture that the correlation coefficients are related with the number of hypotheses and $\delta\alpha$.

7.4 Evaluation of The Empirical Capacity

In this section, we evaluate the empirical capacity of recognition systems using both synthetic and real images. We consider two cases: (1) high resolution images and a relatively small number of classes and (2) low resolution images and a large number of classes. To be more specific, let r be the number of pixels that represent an object within an image and M be the number of considered classes. Case I assumes that r is much larger than M. Case II assumes that M is $5 \sim 10$ times larger than r. In our experiments, the parameters are estimated following procedures in Sec. 7.3.1.

7.4.1 Case I: High Pixel Count

In this case, we will fix the image resolution at 64×64, which corresponds to $r = 4096$ describing objects in ATR and COIL-100 datasets.

The total number of classes formed to evaluate the capabilities of the recognition system for this case is 100 for subset I of COIL-100, 72 (24×3) for ATR set and 264 ($24 \times$

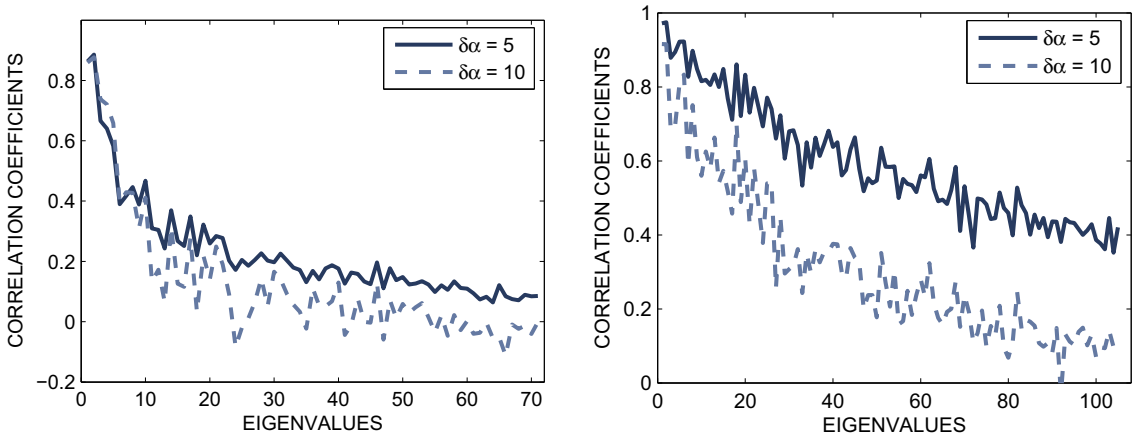


Figure 7.4. The value of diagonal elements in the matrix of correlation coefficients as a function of their number for two cases of $\delta\alpha$. (a) ATR dataset (b) COIL-100 dataset.

11) for subset II of COIL-100. For details on how these hypotheses are formed see Sec. 7.3. In our experiments, we follow the traditional PCA method that generates estimates of n eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ and n corresponding eigenvectors by forming the empirical covariance matrix from M distinct images of M distinct objects. We retain only eigenvectors corresponding to the eigenvalues with the value above 0.5% of the largest eigenvalue $\hat{\lambda}_1$. The values of eigenvalues decrease as the number of hypotheses M increases. When $r >> M$ the estimation problem is ill-posed, since the amount of observed data (available data) is insufficient to estimate unknown parameters, a scatter matrix Σ of size $r \times r$. As described in Sec. 7.3.1, in this case PCA estimates only a small number of parameters out of a large set of unknown parameters. The remaining parameters are assumed to be zero. The empirical information measures that we evaluate in this section contain estimated parameters, the eigenvalues and eigenvectors of the scatter matrix and noise variances. Since the problem of finding the estimates of the eigenvalues and eigenvectors of the scatter matrix is regularized by estimating only a small number of parameters and setting the remaining parameters to zero (see [42] for details on regularization and ill-posed problems), the plug-in estimates of the mutual information rate, empirical capacity, and random coding exponent are also regularized. That is, the estimates are sought in reduced spaces. In this particular case, the reduced space is due to a small number of estimated eigenvalues and eigenvectors with the remaining values set to zero. Therefore, the empirical capacity point may not be the true solution compared

to the case when the estimation of the scatter matrix, Σ, is a well posed problem. The case of a well posed problem is considered in the next subsection.

The values of eigenvalues decrease as the number of hypotheses M increases. Fig. 7.5 and Fig. 7.6 demonstrate the behavior of estimated eigenvalues $\{\hat{\lambda}_i(M)\}$ and noise variance $\{\hat{\sigma}_i^2(M)\}$ as M increases. Note that the eigenvalue $\hat{\lambda}_i$ is evaluated only if $M > i$. It is interesting to observe that the ith eigenvalue and ith noise variance have very similar trends as M and n increase, and the ratio $\hat{\lambda}_i(M)/\hat{\sigma}_i^2(M)$ is approximately constant for all i.

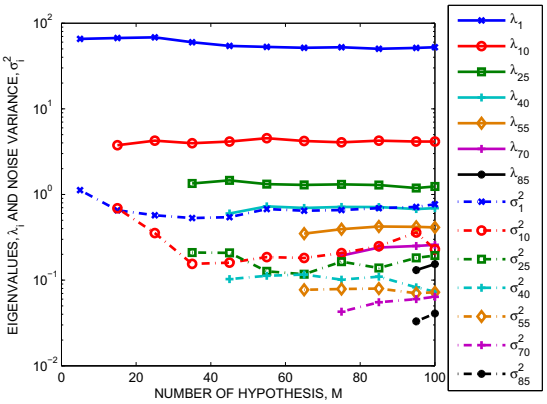


Figure 7.5. Values of eigenvalues and noise variance as a function of the number of hypothesis for subset I of COIL-100.

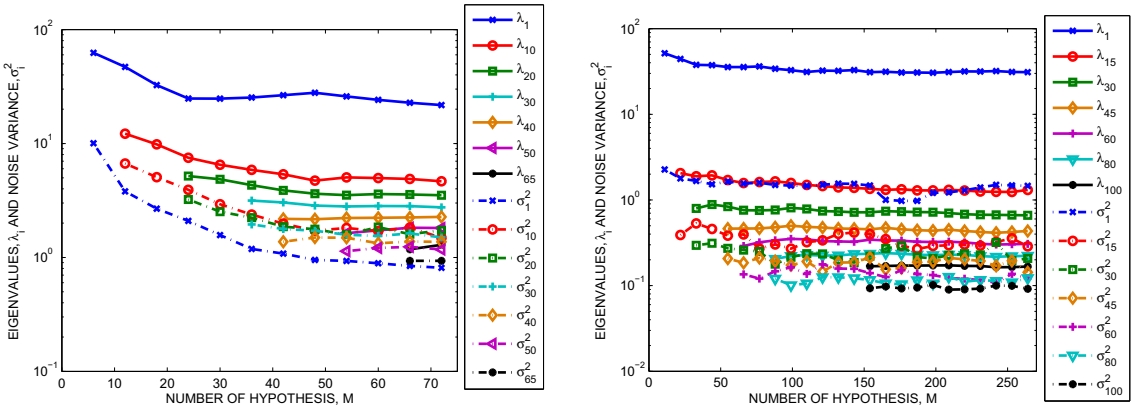


Figure 7.6. Values of eigenvalues and noise variance as a function of the number of hypothesis for ATR dataset (left panel) and subset II of COIL-100 (right panel).

For the multiple image case, we estimate the correlation coefficients between two images. Similar to the single image case, limiting values of ratios of parameters in (7.11)-

(7.13) are evaluated empirically. In our experiments, $\rho_i \neq 1, -1$, for all $i = 1, \cdots, n$. Given a single or multiple image case, the PCA-based empirical mutual information rate can be rewritten as

$$I_n(M) = \frac{1}{2n} \sum_{i=1}^{n} \log[f\left(\hat{\lambda}_i(M), \hat{\sigma}_i^2(M), \hat{\rho}_i(M; \delta\alpha)\right)],$$

where

$$f(\cdot) = 1 + \frac{\hat{\lambda}_i(M)}{\hat{\sigma}_i^2(M)}$$

for single image case and

$$f(\cdot) = \left(1 + \frac{\hat{\lambda}_i(M)}{\hat{\sigma}_i^2(M)}\right)^2 - \left(\frac{\hat{\lambda}_i(M)\hat{\rho}_i(M; \delta\alpha)}{\hat{\sigma}_i^2(M)}\right)^2$$

for two image case and

$$f(\cdot) = \left(1 + \frac{\hat{\lambda}_i(M)}{\hat{\sigma}_i^2(M)}\right)^3 - \left(1 + \frac{\hat{\lambda}_i(M)}{\hat{\sigma}_i^2(M)}\right)\left(\frac{\hat{\lambda}_i(M)}{\hat{\sigma}_i^2(M)}\right)^2 A_1 + 2\left(\frac{\hat{\lambda}_i(M)}{\hat{\sigma}_i^2(M)}\right)^3 A_2$$

for a three image case, where A_1 and A_2 are defined in (7.13). The trends of $f(\cdot)$ as M increases given images from ATR dataset are shown in Fig. 7.7 (a) for a single image case, in (b) for two image case with $\delta\alpha = 5$, in (c) for two image case with $\delta\alpha = 10$ and in (d) for three image case. From Fig. 7.7, it appears that the random sequence $f(\cdot)$ converges to a specific value as M and i increase.

To find the value of PCA-based empirical capacity, given a finite amount of data, we numerically analyze the behavior of the sequence of the empirical mutual information rates,

$$I_n(M) = \frac{1}{2n} \sum_{i=1}^{n} \log\left[f\left(\hat{\lambda}_i(M), \hat{\sigma}_i^2(M), \hat{\rho}_i(M; \delta\alpha)\right)\right].$$

We first select a set of recognition rates, R, and form sequences of $(n, 2^{nR})$ codes for each selected R. For each combination of n and $M = 2^{nR}$ parameterized by a given R, we find the value of the empirical mutual information rate. For each combination of n and $M = 2^{nR}$ (given R) we also find the empirical value of the random coding exponent. If the sequence $(n, 2^{nR})$ parameterized by R produces positive values of the empirical random coding exponent, we say that R is achievable. With available data, it is easy to form and analyze sequences $(n, 2^{nR})$ for small values of the rate, R. However, large values

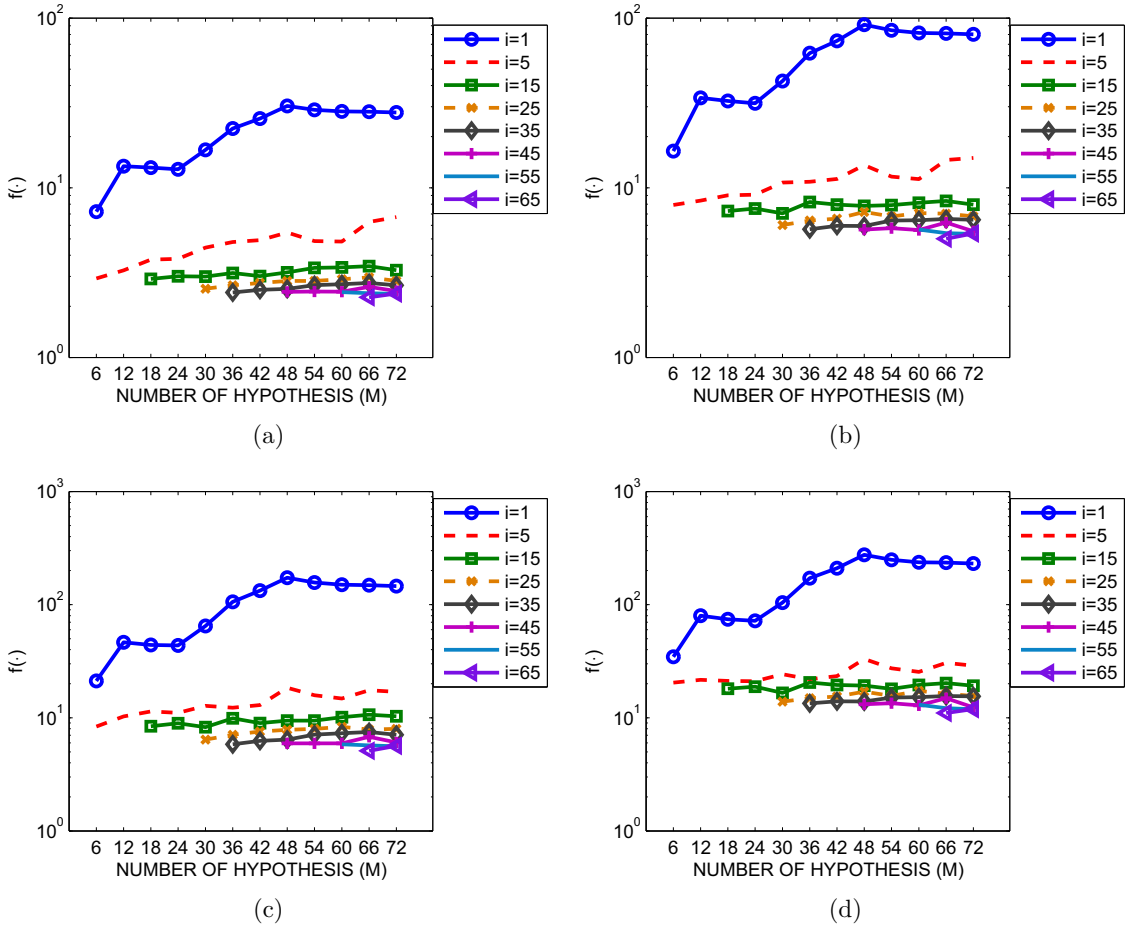


Figure 7.7. Values of $f(\cdot)$ as a function of the number of hypothesis M for ATR dataset and (a) single image case, (b) two image case with $\delta\alpha = 5$, (c) two image case with $\delta\alpha = 10$ and (d) three image case.

of R may have very sparse representation, two or three entries in a sequence. Therefore, the results obtained for large values of R are much less reliable.

Since the channel coding theorem [12, Ch.8], [6] states that transmission is impossible at the rates R exceeding capacity, we use this principle in our numerical evaluation of the empirical capacity. The point of empirical capacity is the point of intersection between the empirical mutual information rate curve plotted as a function of the recognition rate and the straight line bisecting the first quadrant.

The left panel in Fig. 7.8 demonstrates a set of the plots of the empirical mutual information rate as a function of the number of classes, M. Each curve is parameterized by a fixed value R, with M and n growing in proportion. The right panel in Fig. 7.8

shows the plot of the empirical mutual information rate versus the rate for a fixed value of the number of classes, $M = 100$. The empirical capacity is evaluated at the point where the empirical mutual information rate is equal to the recognition rate. The results in Fig. 7.8 are provided for the subset I of COIL-100 dataset. Fig. 7.9 illustrate similar results for ATR dataset (test images at distortion Level 1) and subset II of COIL-100 (test images at distortion Level 3).

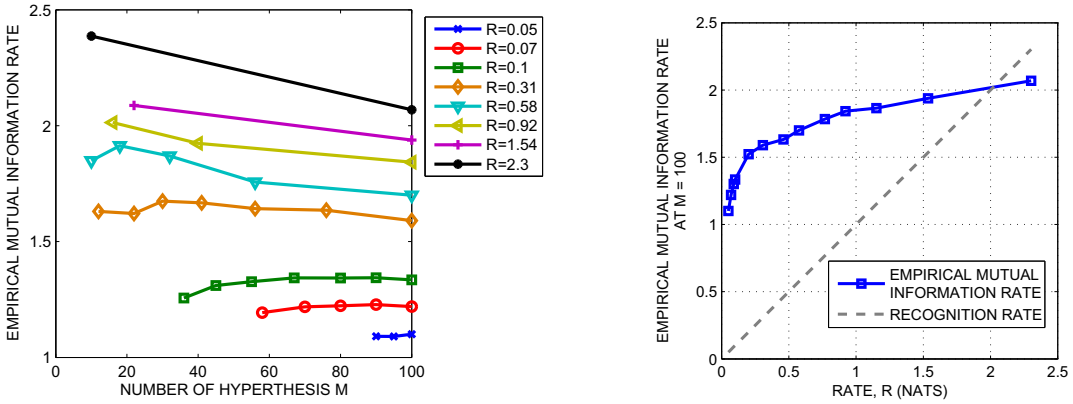


Figure 7.8. The left panel shows the PCA-based empirical mutual information rate as a function of the number of classes parameterized by a set of recognition rates, R. The right panel displays the points of the empirical mutual information rate at $M = 100$ as a function of the recognition rate R. The results are provided for the subset I of COIL-100 with test images at distortion Level 3.

Similar to the case of a single image, we use the empirical mutual information rate to calculate the numerical value of capacity for multiple image case using ATR dataset and subset II of COIL-100. We also change the distortion levels of test images and the results are shown in Fig. 7.10-7.12. Table 7.2 summarizes the empirical value of the recognition capacity for the case $M << r$. Note that the case of high pixel count does not allow us to find the limiting value of the PCA-based recognition capacity.

7.4.2 Case II: Low Pixel Count

This case is motivated by the fact that high resolution images (order of 1024×1024) acquired by UAVs at the elevation of $1000 \sim 1500$ feet may contain objects of interest described by relatively small windows of size 32×32, 24×24 or even smaller. For a window of size 32×32, $r = 1024$. For a window of size 24×24, $r = 576$. To generate the empirical mutual information rate in (7.6), the number of classes has to be much larger than the number of pixels representing objects. To accumulate a large number

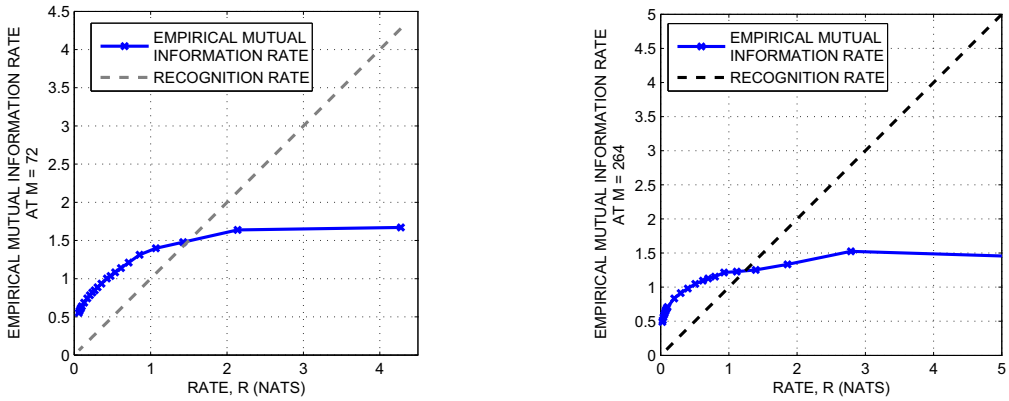


Figure 7.9. The left panel shows the points of the empirical mutual information rate at $M = 72$ as a function of the recognition rate R. The results are provided for the ATR dataset with test images at distortion Level 1. The right panel displays the points of the empirical mutual information rate at $M = 264$ as a function of the recognition rate R. The results are provided for the subset II of COIL-100 with test images at distortion Level 3.

of classes, we treat the images of the same object but acquired at different orientations and elevations as different classes. For ATR database, we use images of 15 objects at 72 orientation angles and 6 elevation angles as training data. The total number of generated hypothesis is $15 \times 72 \times 6$. For COIL-100 database, we use images of 100 objects at 72 orientation angles as training data. The maximum number of hypothesis is 100×72. The testing data per object class consist of 6 distorted images with Level 3 distortion parameters. All images are resized to 32×32 and 24×24. The sample low resolution images are shown in Fig. 7.13.

Similar to the case of high pixel count, we study the empirical mutual information rate as a function of the number of classes, M, parameterized by a set of recognition rates R. The empirical capacity is evaluated at the point where the empirical mutual information rate is equal to the recognition rate. Fig. 7.14 and Fig. 7.15 demonstrate the results for ATR dataset when $r = 576$ and 1024, respectively. Similar results for COIL-100 dataset are shown in Fig. 7.16 and 7.17. Note that the empirical random coding exponent at $r = 576$ and $r = 1024$ are almost identical. Table 7.3 summarizes the empirical capacity at different image resolution levels for both datasets.

We can conclude that for a given encoding technique, the value of empirical recognition capacity is determined not only by distortions due to the environmental conditions and an acquisition device, but also depends on complexity of the data, complexity of the original objects, resolution of the data, and definition of classes.

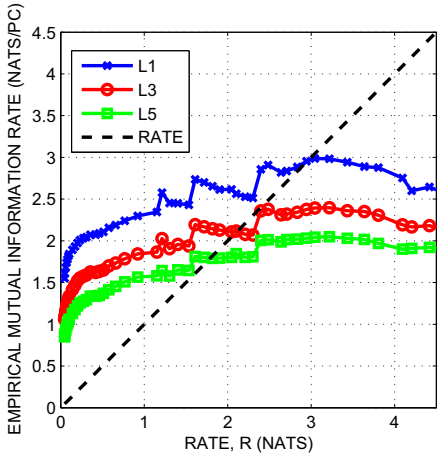


Figure 7.10. The points of the empirical mutual information rate at $M = 100$ as a function of the recognition rate R. The results are provided for the subset I of COIL-100 with test images at distortion Level 1, 3 and 5.

7.5 Summary

This chapter suggests steps towards evaluation of empirical recognition capacity of a recognition system under the constraint of PCA-based encoding. We model encoded single and multiple frames containing the same object as realizations of Gaussian vectors with zero mean and unknown structured covariance matrices with the covariance being a function of the relative orientation angle. The unknown parameters are estimated using training data. The proposed probabilistic models are verified using various goodness of fit tests.

The empirical capacity is evaluated using both simulated data and real images. Two cases of the relationship between the number of pixels representing an object in an image and the number of recognition classes are analyzed: (1) the case of high resolution images and a relatively small number of classes and (2) the case of low resolution images and a large number of object classes.

To obtain the empirical capacity point, we plot the empirical mutual information rate as a function of the recognition rate, R, and find the point of intersection between this curve and the diagonal line bisecting the first quadrant. Given a value of empirical capacity and a length of templates (assume long), we can evaluate the maximum number of object classes in an object library that can be recognized with probability of error close to zero. With respect to single and multiframe cases, we demonstrated that two encoded templates of the same object separated by the orientation angle of 10 degrees,

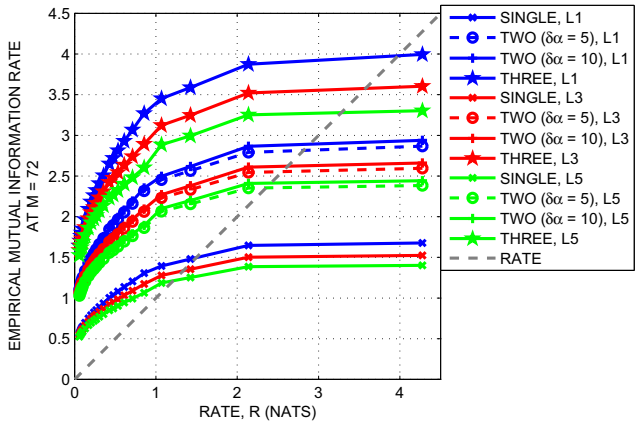


Figure 7.11. The points of the empirical mutual information rate at $M = 72$, obtained using single image and multiple images, as a function of the recognition rate R. The results are provided for the ATR dataset with test images at distortion Level 1, 3 and 5.

when probabilistically fused using the proposed Gaussian model, result in considerable increase in the value of the empirical capacity.

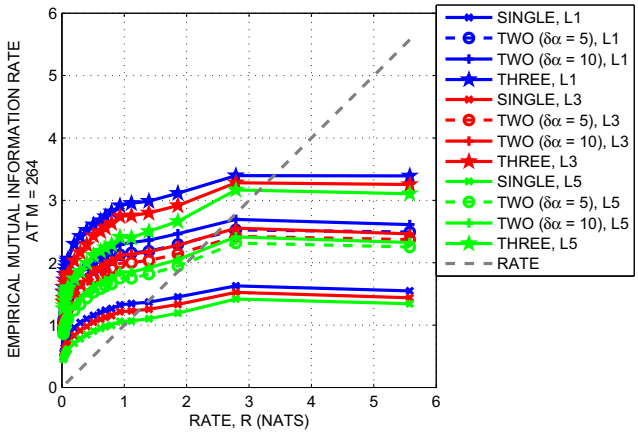


Figure 7.12. The points of the empirical mutual information rate at $M = 264$, obtained using single image and multiple images, as a function of the recognition rate R. The results are provided for the subset II of COIL-100 with test images at distortion Level 1, 3 and 5.

(a) (b)

Figure 7.13. Sample images of size 64×64, 32×32 and 24×24 from (a) ATR dataset (b) COIL-100 dataset.

Table 7.2. Empirical Recognition Capacity under the constraint of PCA encoding in the case of high pixel count (nats/pc)

Dataset	Noise Level	Single	Two ($\delta\alpha = 5$)	Two ($\delta\alpha = 10$)	Three
Subset I of COIL-100	Level 1	2.5194	N/A	N/A	N/A
	Level 3	2.0206	N/A	N/A	N/A
	Level 5	1.6754	N/A	N/A	N/A
Simulated ATR set	Level 1	1.4965	2.8158	2.8913	3.9793
	Level 3	1.3343	2.5539	2.6208	3.5768
	Level 5	1.2123	2.3551	2.4136	3.2796
Subset II of COIL-100	Level 1	1.3652	2.4394	2.6637	3.3961
	Level 3	1.2392	2.2633	2.4568	3.2767
	Level 5	1.0639	2.0037	2.1747	3.1590

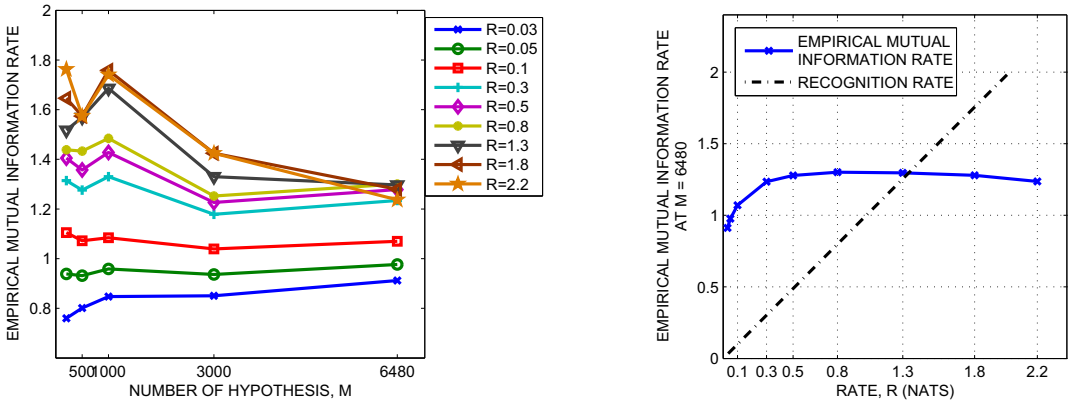


Figure 7.14. The left panel shows the PCA-based empirical mutual information rate as a function of the number of classes parameterized by a set of recognition rates, R. The right panel displays the points of the empirical mutual information rate at $M = 100$ as a function of the recognition rate R. The results are provided for the ATR dataset with the image resolution 24×24.

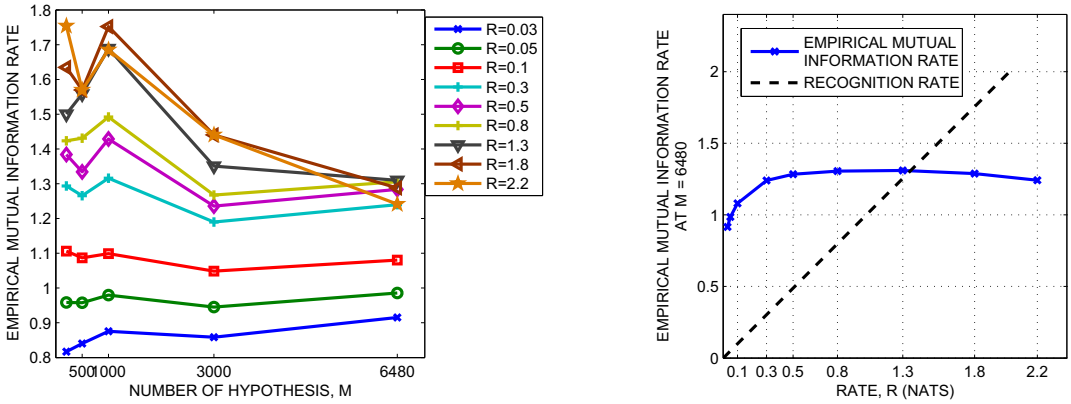


Figure 7.15. The left panel shows the PCA-based empirical mutual information rate as a function of the number of classes parameterized by a set of recognition rates, R. The right panel displays the points of the empirical mutual information rate at $M = 100$ as a function of the recognition rate R. The results are provided for the ATR dataset with the image resolution 32×32.

Table 7.3. Empirical Recognition Capacity under the constraint of PCA encoding in the case of low pixel count (nats/pc)

Dataset	Resolution, $r = 576$	Resolution, $r = 1024$
ATR dataset	1.2952	1.3074
COIL dataset	1.7262	1.7444

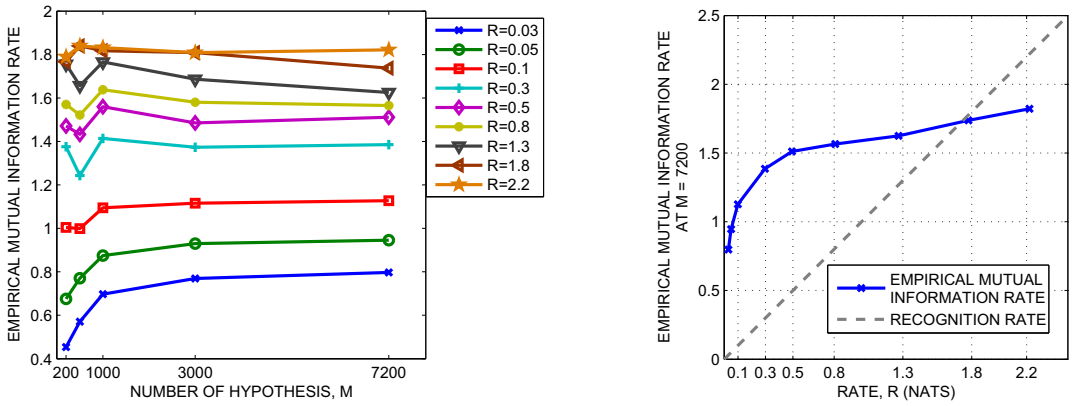


Figure 7.16. The left panel shows the PCA-based empirical mutual information rate as a function of the number of classes parameterized by a set of recognition rates, R. The right panel displays the points of the empirical mutual information rate at $M = 100$ as a function of the recognition rate R. The results are provided for the COIL dataset with the image resolution 24×24.

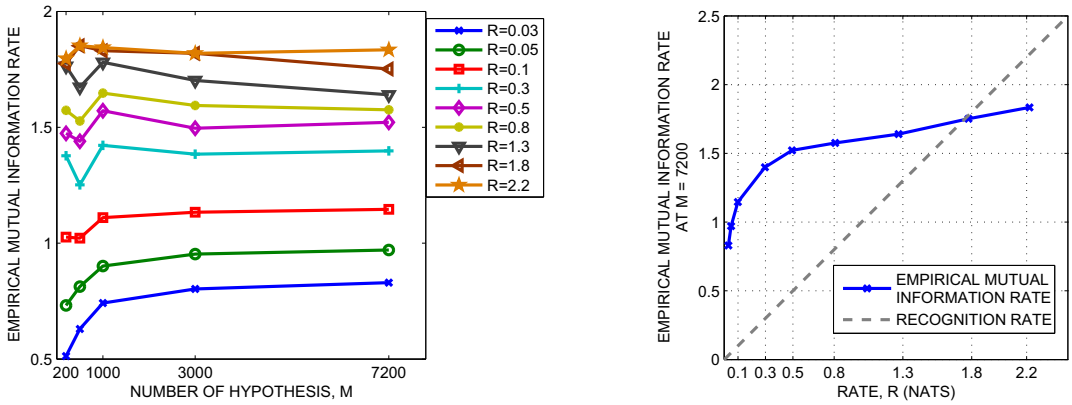


Figure 7.17. The left panel shows the PCA-based empirical mutual information rate as a function of the number of classes parameterized by a set of recognition rates, R. The right panel displays the points of the empirical mutual information rate at $M = 100$ as a function of the recognition rate R. The results are provided for the COIL dataset with the image resolution 32×32.

CHAPTER 8

RECOGNITION ERROR EXPONENT

From information theory, the channel capacity is the maximum rate that can be used to communicate information over a channel reliably, that is, with vanishing probability of errors. Information Theory states that codes with arbitrarily small probability of decoding error exist at any rate smaller than the channel capacity C as the code word length becomes large. In practice, however, the code word length is finite. Therefore the problem can be restated as finding the code word length for a given value of probability of error [51]. In this chapter, we study the relationship between code word length and probability of decoding error for a recognition channel. We involve the bounds on the recognition reliability function, which is defined by analogy with the channel reliability function in communication problems. The limiting expressions and the results of numerical evaluation are applied to PCA encoded images.

8.1 Reliability Function

Consider the average probability of error

$$P(error) = \frac{1}{M} \sum_{m=1}^{M} P(error|X^n(m)),$$

where $X^n(m)$ is the mth template (a codeword), which is defined in Section 7.1, n is the length of the codewords, M is the number of hypothesis and $P(error|X^n(m))$ is the conditional probability of error, given that a codeword is a realization of the random vector generated by the m-th hypothesis. The minimum probability of error decision rule is the same as the maximum likelihood decision rule: select the most likely observation given $X^n(m)$.

Then the reliability function is defined as

$$E(R) = -\liminf_{n\to\infty} \frac{1}{n} \log P(error),$$

where the infimum is taken over all decoding rules. That is, in channel coding problems, there is an optimization over the codewords and the decision rules. The variable R in the left hand side is the code rate defined as $R = \log(M)/n$. This additional complication makes the determination of the channel reliability function difficult. In the recognition problem posed here, the distribution of the codewords is determined by the statistics of object signatures and a measurement process. Provided that the encoding algorithm is specified, the recognition reliability function is the random coding bound for the channel reliability function [6]. When the code rate R is relatively small, the reliability function is unknown. Only the bounds can be found [6].

8.2 Random Coding Lower Bound

In this section we will derive random coding exponent for recognition channel under the constraint of PCA encoding technique. Previously, Shannon [51] has derived the lower bounds on the reliability function of a real Gaussian channel. The bounds were derived using random constructions under a geometrical approach. In [51], the signals have the same power P along each dimension and the noise has the same variance N for each component. All M code words are assumed to lie on the surface of a sphere of radius $\sqrt{nP}$. However, in the PCA-based recognition channel, both the signals and the noise have different variances in different dimensions. Here we choose a more general approach from [6]. The random coding bound is defined as

$$E_r(R) = \lim_{n\to\infty} \max_{s\in[0,1]} \max_{p} \left[-sR - \frac{1}{n} \log \left(\int_Y \left[\int_X p(X) p(Y|X)^{\frac{1}{1+s}} dX \right]^{1+s} dY \right) \right].$$

The derivation of the random coding bound is presented in Appendix A.

In a single-image case, the templates (codes), X, stored in the library follow Gaussian distribution with zero mean and diagonal covariance matrix Λ. The encoded test image can be modeled as $Y = X + W$, where X and W are independent. W is a Gaussian noise with mean zero and diagonal covariance matrix Λ_N. Then the random coding bound becomes

$$E_r(R) = \lim_{n\to\infty} \max_{s\in[0,1]} \left[-sR - \frac{1}{n} \log \left(\int_Y \left[\int_X p(X)p(Y|X)^{\frac{1}{1+s}} dX \right]^{1+s} dY \right) \right], \quad (8.1)$$

where

$$p(X) = \frac{1}{(2\pi)^{\frac{2}{n}} |\Lambda|^{\frac{1}{2}}} \exp\left[-\frac{X^T \Lambda^{-1} X}{2} \right], \quad (8.2)$$

and

$$p(Y|X) = \frac{1}{(2\pi)^{\frac{2}{n}} |\Lambda_N|^{\frac{1}{2}}} \exp\left[-\frac{(X-Y)^T \Lambda_N^{-1} (X-Y)}{2} \right]. \quad (8.3)$$

After submitting (8.2) and (8.3) into (8.1), we obtain expression for the random coding bound under the constraint of PCA encoding (see Chapter 7 on PCA-based model),

$$E_r(R) = \lim_{n\to\infty} \max_{s\in[0,1]} \left\{ -Rs + \frac{s}{2n} \sum_{k=1}^{n} \log[1 + \frac{\lambda_k}{\sigma_k^2(1+s)}] \right\}. \quad (8.4)$$

For practical cases, since codewords have a finite length, n, and since the unknown parameters are estimated using training data, we introduce empirical random coding bound:

$$\hat{E}_r(R) = -\hat{R}_n(s^*)s^* + \frac{s^*}{2n} \sum_{k=1}^{n} \log[1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s^*)}], \quad (8.5)$$

where $s^* \in [0,1]$ is the parameter that maximizes the expression in the parentheses in (8.4). $\hat{\lambda}_k$ and $\hat{\sigma}_k^2$, $k = 1, \cdots, n$, are the estimated parameters,

$$\hat{R}_n(s^*) = \frac{1}{2n} \sum_{k=1}^{n} \log[1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s^*)}] - \frac{s^*}{1+s^*} \frac{1}{2n} \sum_{k=1}^{n} [1 - \frac{1}{1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s^*)}}].$$

Note when $s^* = 0$, the empirical random coding exponent is $\hat{E}_n(R) = 0$ and $\hat{R}_n(s^*) = \frac{1}{2n} \sum_{k=1}^{n} \log[1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}]$, which is the value of the constrained empirical mutual information rate under PCA encoding.

In classical information theory, the true function $E(R)$ is known only for rates greater than the critical rate R_{crit}, at which the random coding bound $E_r(R)$ deviates from a straight line of slope -1. For lower rates, only bounds on $E(R)$ are known.

Similar to the case of a single image, the empirical random coding bound based on two images with the relative orientation $\delta\alpha$ is

$$\hat{E}_r^{(2)}(R) = \max_{s\in[0,1]} \left\{ -\hat{R}_n s + \frac{s}{2n} \sum_{k=1}^{n} \log \left[\left(1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)} \right)^2 - \left(\frac{\hat{\lambda}_k \hat{\rho}_k}{\hat{\sigma}_k^2(1+s)} \right)^2 \right] \right\}, \quad (8.6)$$

where $\hat{\rho}_k$ is the kth estimated correlation coefficient between the templates of two images with relative angle $\delta\alpha$. Here we assume that no correlation coefficients take values 1 or -1. When $s = 0$ achieves the maximum value, $\hat{E}_r^{(2)}(R) = 0$, the rate $\hat{R}_n$ equals to the joint empirical mutual information rate defined in (7.11).

The empirical random coding bound based on three images with the relative orientation $\delta\alpha_{1,2}$, $\delta\alpha_{1,3}$ and $\delta\alpha_{2,3}$ is given by

$$\hat{E}_r^{(3)}(R) = \max_{s\in[0,1]} \left\{ -\hat{R}_n s + \frac{s}{2n} \sum_{k=1}^{n} \log \left[\left(1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)} \right)^3 \right. \right. \tag{8.7}$$

$$\left. \left. - \left(1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)} \right) \left(\frac{\hat{\lambda}_k \hat{\rho}_k}{\hat{\sigma}_k^2(1+s)} \right)^2 A_1 + 2 \left(\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)} \right)^3 A_2 \right] \right\},$$

where A_1 and A_2 are defined in (7.13).

This expression can be generalized to an arbitrary finite number of frames of the same object/target.

8.3 Space Partitioning Upper Bound

The random coding bound is known as one of two lower bounds on the channel reliability function. A popular upper bound on the channel reliability function is the bound by partitioning [6]. Given the PCA encoded data, the empirical upper bound by partitioning is given

$$\hat{E}_p(R) = \max_{s\geq 0} \left\{ -\hat{R}_n s + \frac{s}{2n} \sum_{k=1}^{n} \log[1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)}] \right\}, \tag{8.8}$$

where s is the parameter of optimization. The derivation of the expression 8.8 can be found in Appendix B.

For the case of two PCA encoded images of the same object with a relative angle $\delta\alpha$, the empirical upper bound by partitioning is

$$\hat{E}_p^{(2)}(R) = \max_{s\geq 0} \left\{ -\hat{R}_n s + \frac{s}{2n} \sum_{k=1}^{n} \log \left[\left(1 + \frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)} \right)^2 - \left(\frac{\hat{\lambda}_k \hat{\rho}_k}{\hat{\sigma}_k^2(1+s)} \right)^2 \right] \right\}, \tag{8.9}$$

where we assume that the correlation coefficients $\hat{\rho}_k,\ k = 1, \cdots, n$ do not take values 1 or -1.

The empirical upper bound for the case of three images is

$$\hat{E}_p^{(3)}(R) = \max_{s\geq 0}\left\{-\hat{R}_n s + \frac{s}{2n}\sum_{k=1}^{n}\log\left[\left(1+\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)}\right)^3\right.\right. \tag{8.10}$$

$$\left.\left. -\left(1+\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)}\right)\left(\frac{\hat{\lambda}_k\hat{\rho}_k}{\hat{\sigma}_k^2(1+s)}\right)^2 A_1 + 2\left(\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2(1+s)}\right)^3 A_2\right]\right\},$$

where A_1 and A_2 are defined in (7.13).

From the above expressions we can see that if the optimization of $E_p(R)$ is limited to $s \in [0, 1]$, then the upper bound $E_p(R)$ is equal to the lower bound $E_r(R)$. Therefore, in this range of s we obtain the empirical reliability function.

8.4 Experiments and Results

In this section, the empirical exponents, random coding and by partitioning, under the assumption of PCA encoded data are evaluated using ATR dataset, subset I and subset II of COIL-100 dataset. Following the descriptions in Sec. 7.4, the case with high resolution images and a relatively small number of classes is considered.

To empirically evaluate the random coding exponent, we numerically optimize (8.5-8.10) with respect to the parameter s. The empirical bounds as a function of the recognition rate, R, are evaluated. By changing the number of classes M and the code word length n, we can obtain different values of R. Since the same value of the rate can be obtained by invoking different combinations of M and n, we evaluate the empirical bounds for each combination and select the value with maximum M and n characterizing the same value of the rate R. Since the lower rate are easy to form, we obtain reliable results at lower rates and less reliable results at higher rates. Using the terminology of Information Theory, for each given R, we form a sequence of codes $(n, 2^{nR})$ and empirically evaluate achievability of the rate, R.

The random coding and partitioning exponents are first evaluated using the subset I of COIL-100 dataset. In this set, an object is represented by a single class. The total number of classes formed is 100. The quary images are obtained by distorting clean images at Level 3. The empirical bounds as a function of R are shown on the left panel in Fig. 8.1. The rate at which the bound attains zero is approximately 2.2 nats. By comparing the value of the empirically evaluated capacity 2.1 nats/PC on the right panel in Fig. 8.1 (the point of intersection of the diagonal line and the empirical mutual

information curve), we can see that numerically evaluated recognition capacity points are approximately equal. Apart from the point of capacity, the empirical bounds provide an approximate value of the exponent in the exponential approximation to probability of recognition error.

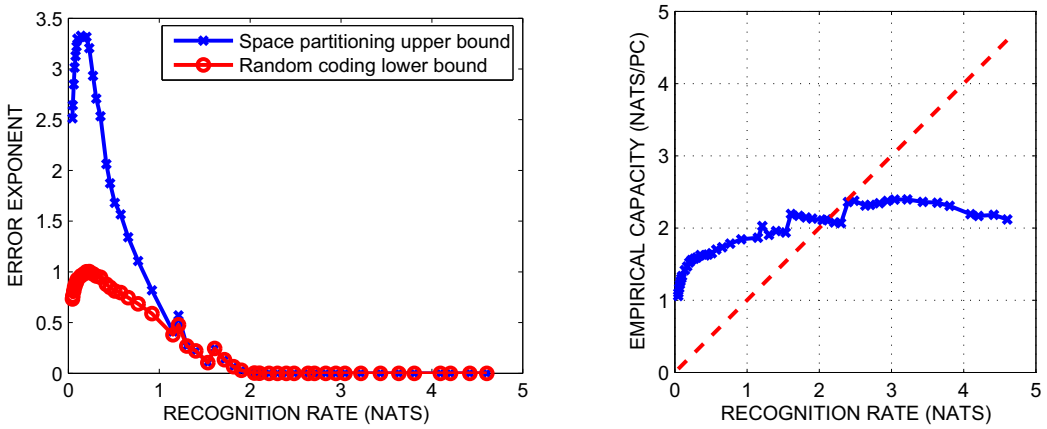


Figure 8.1. The left panel shows the empirical Error Exponents as a function of the recognition rate R. The right panel displays the empirical mutual information rate as a function of the recognition rate R. The results are provided for the subset I of COIL-100 dataset.

The case when an object is represented by a set of classes in the database is illustrated by two examples. One is the ATR dataset generated using a 3D simulated tool and the other is the subset II of COIL-100 dataset. The empirical bounds as a function of R given single and multiple images for ATR dataset are demonstrated in Fig. 8.2 (a) - (d). The corresponding empirical mutual information rates as a function of R at the maximum value of M and n are shown in Fig. 8.3. Note that the empirical capacity values evaluated from the plot of the empirical mutual information rate as a function of R and the empirical random coding and partitioning exponents are consistent.

Similar results for the subset II of COIL-100 dataset are illustrated in Fig. 8.4 and Fig. 8.5.

The empirical error exponents are also evaluated using query images distorted at different distortion levels. These results are shown on the left panel in Fig. 8.6-8.8. Again, the values of empirical capacity obtained using plots of the empirical mutual information rate in Fig. 8.6-8.8 and using empirical error exponents are in a relatively good agreement.

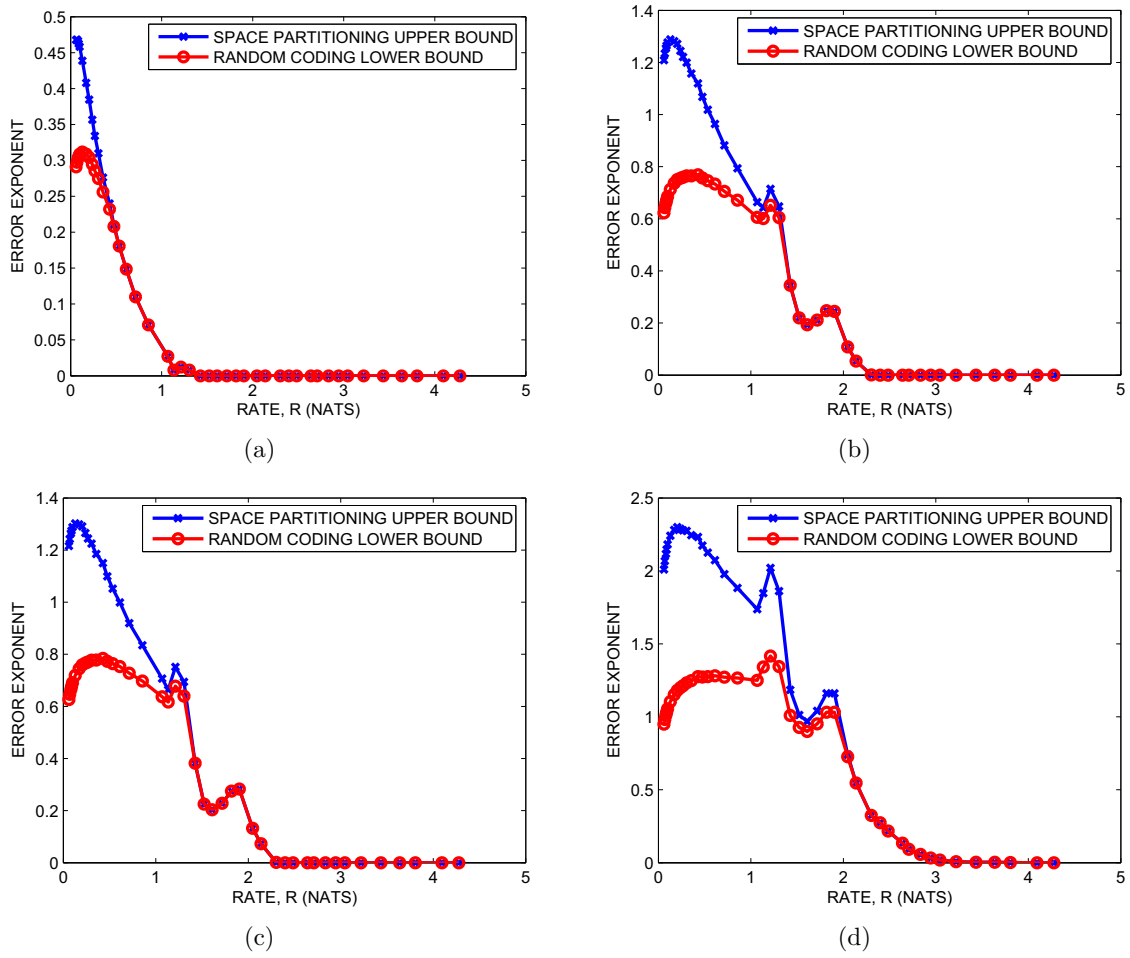


Figure 8.2. The empirical Error Exponents for ATR dataset given (a) Single-image case, (b) Two-image case with relative angle of 5 degrese, (c) Two-image case with relative angle of 10 degrees and (d) Three-image case.

8.5 Summary

In this Chapter, we derived two bounds on the reliability function of a recognition system under the constraint of PCA-based encoding. The empirical random coding exponent and the space partitioning exponent, two approximations to the reliability function, are defined for both single and multiple image cases. The empirical bounds are evaluated using the subset I and subset II of COIL-100 dataset and ATR dataset. The points where the exponents touch the x-axis (become zero) are the values of empirical capacity. These values and the values of the empirical capacity obtained using the empirical mutual information rate curve as a function of R are in a good agreement. For the subset

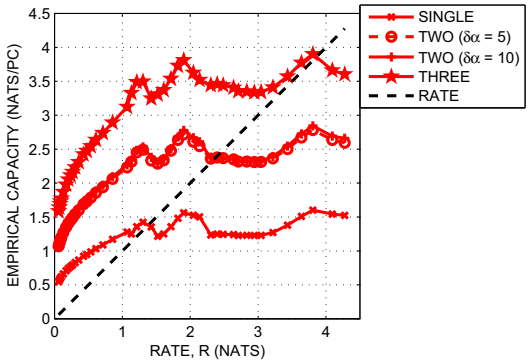


Figure 8.3. The empirical mutual information rate as a function of R for ATR dataset for a single-image case, two-image case with the relative angle 5 degree, 10 degree and a three-image case.

II of COIL-100 dataset and ATR dataset, the upper and lower exponents from multiple images are compared with the corresponding exponents from a single image. The results confirm that the empirical recognition capacity and error rates from multiple images are higher compared to the capacity and error rates from a single image.

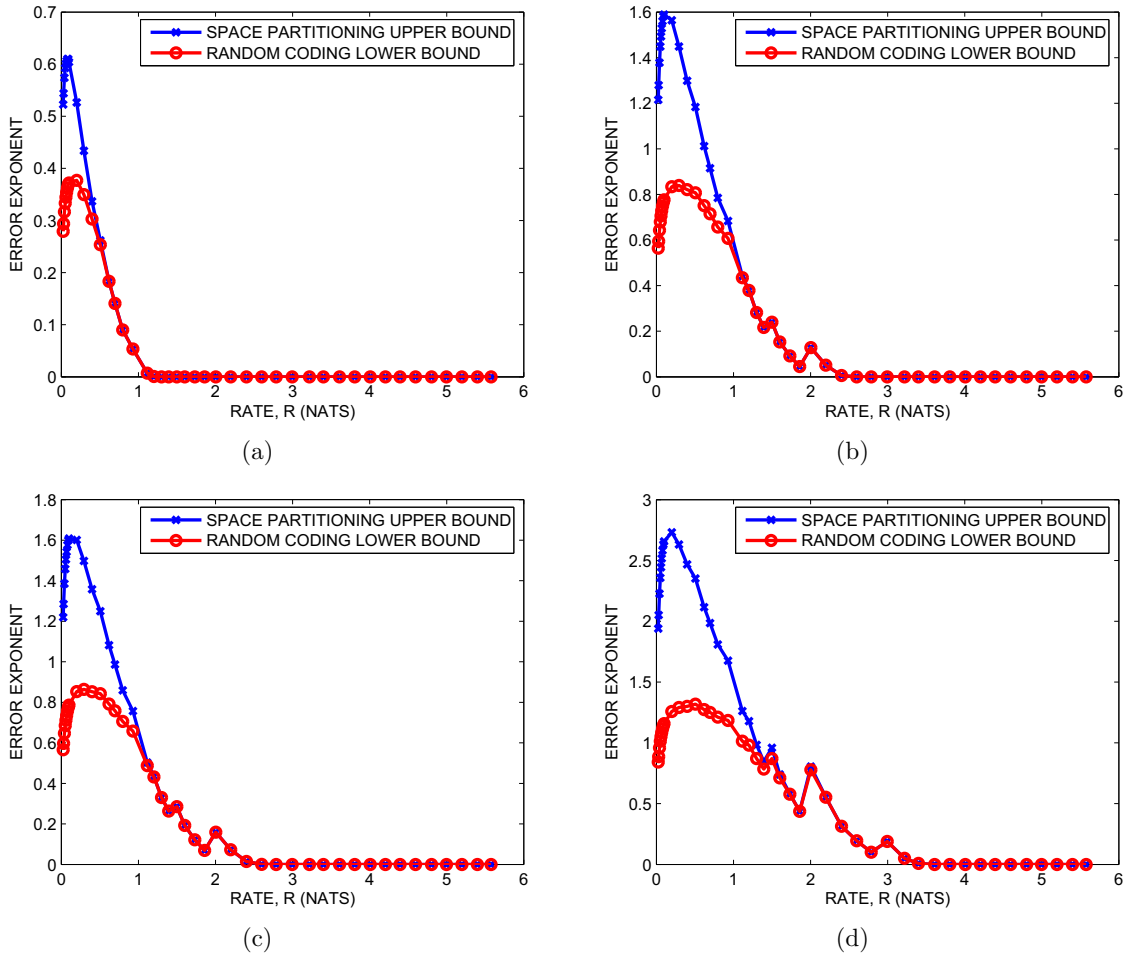


Figure 8.4. The empirical Error Exponents for the subset II of COIL-100 dataset for (a) Single-image case, (b) Two-image case with relative angle 5 degree, (c) Two-image case with relative angel 10 degree and (d) Three-image case.

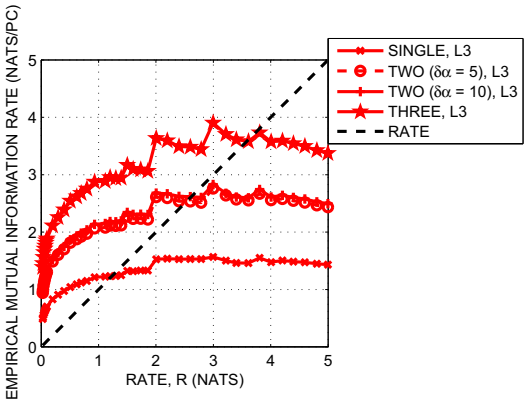
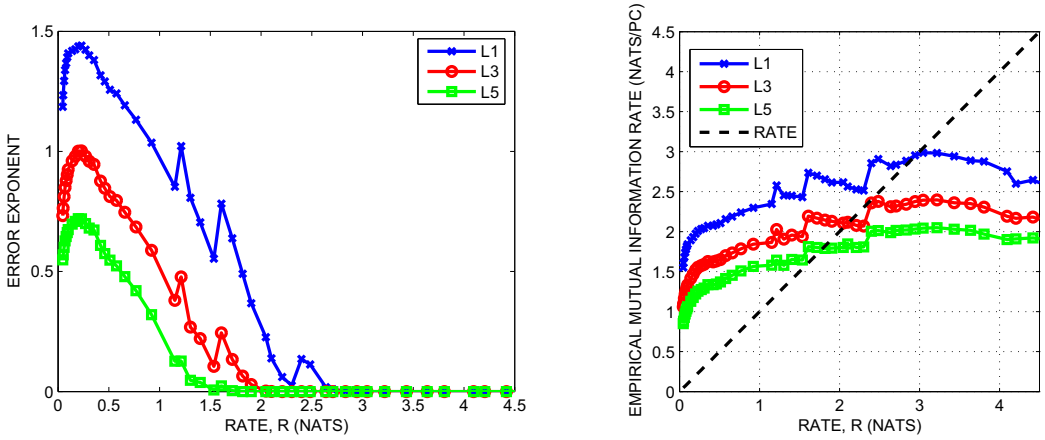
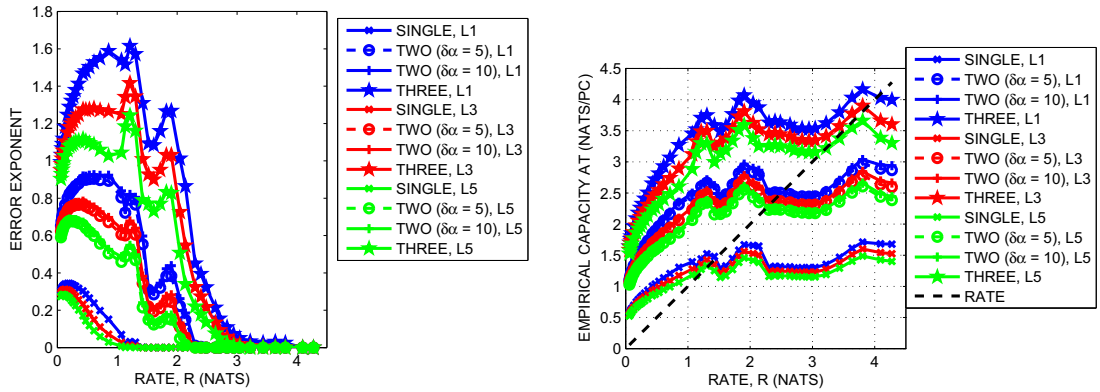


Figure 8.5. The empirical mutual information rate as a function of R for the subset II of COIL-100 dataset for a single-image case, two-image case with the relative angle 5 degree, 10 degree and a three-image case.



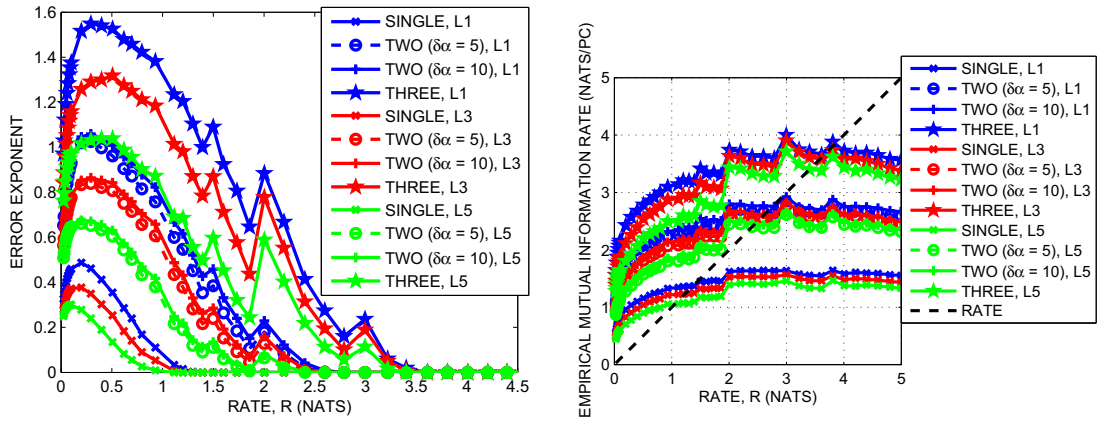
(a) (b)

Figure 8.6. The left panel shows the empirical random coding exponents. The right panel displays the empirical mutual information rates as functions of the recognition rate R. The results are provided for the subset I of COIL-100 dataset with test images at distortion Level 1, 3 and 5.



(a) (b)

Figure 8.7. The left panel shows the empirical random coding exponents. The right panel displays the empirical mutual information rates as functions of the recognition rate R. The results are provided for the ATR dataset with test images at distortion Level 1, 3 and 5, for a single image, two images with the relative angle 5 and 10 degree and three image cases.



(a) (b)

Figure 8.8. The left panel shows the empirical random coding exponents. The right panel displays the empirical mutual information rates as functions of the recognition rate R. The results are provided for the subset II of COIL-100 dataset with test images at distortion Level 1, 3 and 5, for a single image, two images with the relative angle 5 and 10 degree, and three image cases.

CHAPTER 9

RECOGNITION PROBABILITY OF OUTAGE

In Chapter 7, we defined the constrained recognition capacity given an encoding method and evaluated the PCA-based recognition capacity using large datasets of images of objects. The capacity is related with the maximum number of classes which can be correctly recognized with arbitrarily small probability of error when the length of code words becomes large. However, since the capacity does not specify the value of the probability of error, in Chapter 8 we appealed to bounds on the channel reliability function. These bounds characterize the asymptotic exponential rate of the probability of recognition error. However, evaluation of the error exponent require both the number of users, M, and the length of codewords n to become very large. This assumption is rarely satisfied in practice. To address practical cases of relatively large but finite M and n, we invoke the performance measure used in communication theory called Probability of information Outage. It evaluates the probability that the empirical capacity estimated using codewords of M distinct classes, each of length n does not drop below the recognition rate $R = \log M/n$. In this Chapter, we discuss results related to the outage probability of recognition channel.

9.1 Probability of Outage

In fading communication channel, the outage probability is defined as the percentage of time in which information transfer with a predetermined reliability is not possible. The performance parameters which characterize the reliable transfer can be signal-to-inference ratio (SINR) [13], the instantaneous capacity [31] and bit error probability [60]. In a cellular system, the outage probability is related with the probability of not being able to successfully place a call, whereas the average probability of error relates to the average quality of such calls [31]. In the recognition channel, we define two types of the probability of outage.

1. The Type I outage probability is the probability that the empirical capacity is lower than the recognition rate, R,

$$PO_r^I = Pr[\hat{C} < R],$$

where $\hat{C}$ is the empirical capacity and the subscript $_r$ stands for recognition. For a single-image case

$$I_n(M) = \frac{1}{2n}\sum_{k=1}^{n}\log\left[1+\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}\right],$$

for a two-image case

$$I_n(M)^{(2)} = \frac{1}{2n}\sum_{k=1}^{n}\log\left[\left(1+\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}\right)^2-\left(\frac{\hat{\lambda}_k\hat{\rho}_k}{\hat{\sigma}_k^2}\right)^2\right],$$

and for a three-image case

$$I_n(M)^{(3)} = \frac{1}{2n}\sum_{k=1}^{n}\log\left[\left(1+\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}\right)^3-\left(1+\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}\right)\left(\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}\right)^2 A_1+2\left(\frac{\hat{\lambda}_k}{\hat{\sigma}_k^2}A_2\right)\right],$$

where A_1 and A_2 are defined in (7.13). Here $\hat{\lambda}_k$, $\hat{\sigma}_k^2$ and $\hat{\rho}_k$ are estimated using all training and testing images from all classes. Given a set of images from M classes, we can calculate the Type I outage probability by iteratively selecting one image per class as a training image and the other images as testing images.

2. The Type II outage probability is the probability that the instantaneous capacity (defined below) is lower than the recognition rate, R,

$$PO_r^{II} = Pr[C_{inst} < R],$$

where C_{inst} is the instantaneous capacity. C_{inst} for a single and multiple images have the same expressions as the corresponding $\hat{C}$ defined above, with $\hat{\sigma}_k^2$ replaced by $\tilde{\sigma}_k^2$. Here $\tilde{\sigma}_k^2$ is estimated using only testing images of one class, while $\hat{\sigma}_k^2$ is estimated using all testing images from all classes. The Type II outage probability is calculated by comparing C_{inst} of each class with the recognition rate. It is related to the percentage of times when submitted for identification codewords are not recognized correctly.

9.2 Experiments and Results

Following the experimental setting in Chapter 8, we evaluate the outage probability using the subsets I and II of COIL-100 dataset and ATR dataset. Instead of fixing the training images, we at random select one image per class to be included in a training set. The remaining images per class are used for testing and estimating the noise variances. This procedure is repeated 20 times resulting in 20 data realizations. Testing images at different distortion levels are involved.

To estimate the Type I probability of outage, for each realization, we find the empirical capacity first and then compare it with the recognition rate. The number of events when the value of the empirical capacity drops below a fixed value of the rate is counted and normalized by the number of data realizations. The plots of the Type I probability of outage as a function of the recognition rate, R, are shown in Fig. 9.1 (a) for the subset I of COIL-100 dataset, Fig. 9.2 (a) for ATR dataset and Fig. 9.3 (a) for the subset II of COIL-100 dataset.

To find the Type II probability of outage, the instantaneous capacity is estimated using the testing images from each class. The probability of outage is evaluated as the relative frequency of the event that the instantaneous capacity drops below a given value of the rate. The total number of events is M. This procedure is repeated 20 times. The average value of Type II outage probability as a function of the recognition rate is shown in Fig. 9.1 (b) for the subset I of COIL-100 dataset, Fig. 9.2 (b) for ATR dataset and Fig. 9.3 (b) for the subset II of COIL-100 dataset.

In Fig. 9.1-9.3, the blue, red and green curves are obtained using the distorted images at Levels 1, 3 and 5, respectively. The curves marked by 'x', 'o', '+' and '$\star$' correspond to the case of a single, two image with the relative angle $\delta\alpha = 5$, two image with the relative angle $\delta\alpha = 10$, and three images, respectively. From the results we can see that for the subset I of COIL-100 dataset and ATR dataset, the larger the distortion is, the higher is the outage probability for the same recognition rate. For the subset II of COIL-100 dataset, the increasing of the distortion level does not degrade the performance significantly. We can also conclude on improvements due to use of multiple images separated by a small rotation angle compared to a single image case.

The empirical mutual information rate and the empirical random coding lower bound averaged over 20 data realizations as a function of the recognition rate, R, are also displayed in Fig. 9.1 (c) and (d) for the subset I of COIL-100 dataset, in Fig. 9.2 (c) and (d) for ATR dataset and in Fig. 9.3 (c) and (d) for the subset II of COIL-100 dataset. By comparing the values of the empirically evaluated capacity estimated using

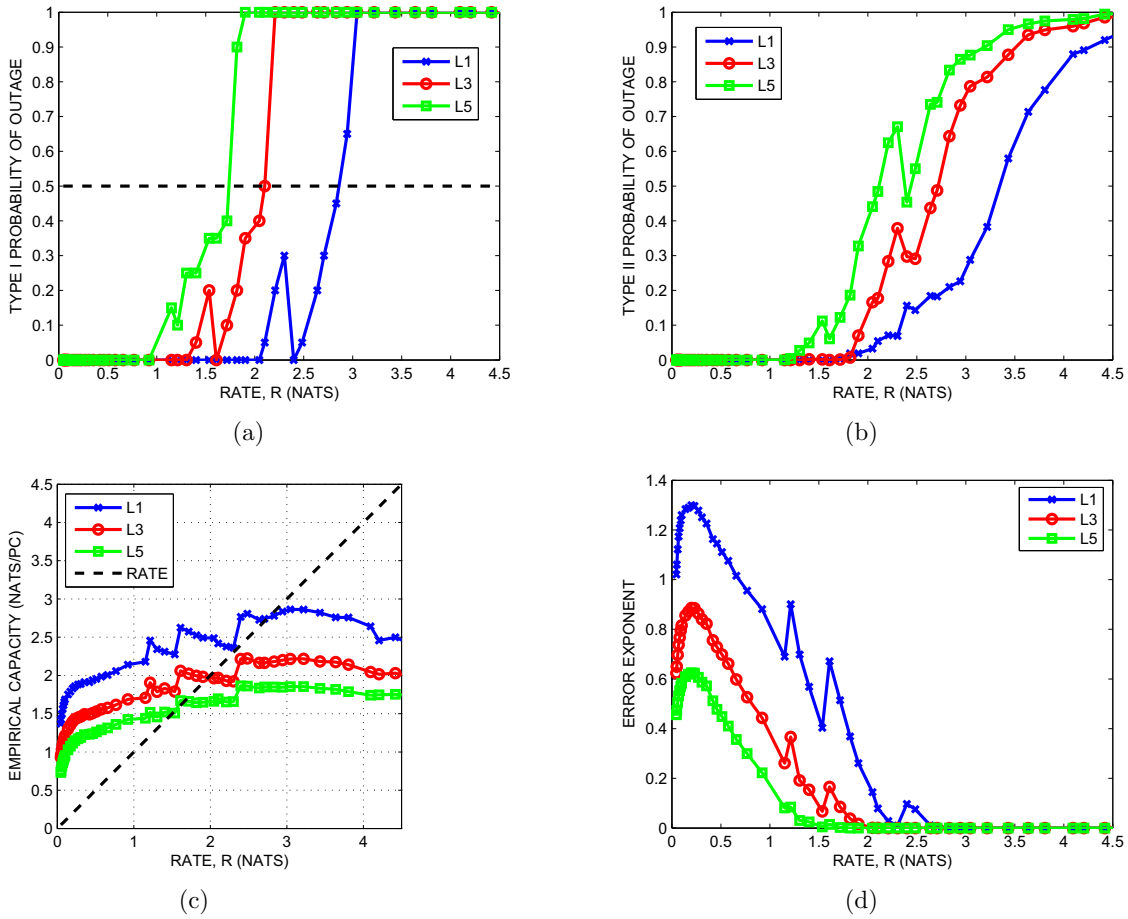


Figure 9.1. (a) The Type I probability of outage, (b) the average Type II probability of outage, (c) the average empirical mutual information rate and (d) the average empirical random coding bounds. The results are provided for the subset I of COIL-100 dataset with test images at distortion Level 1, 3 and 5.

the sequence of empirical mutual information (the point of intersection of the diagonal line and the empirical mutual information rate curve) and the values evaluated using the random coding bounds (the rate at which the bound attains zero), we can see that numerically evaluated recognition capacity points are approximately equal. The point where the Type I probability of outage equals to 0.5 is clearly related to the empirical capacity of the PCA-based recognition systems. Furthermore, by analyzing plots of the Probability of Outage as a function of the rate, R, we can identify these regions of recognition rates: (1) the region of perfect recognition, where the probability of outage is zero; (2) transition region, where the probability of outage takes values between zero and one, and (3) the region of complete outage, where recognition is impossible. Within the

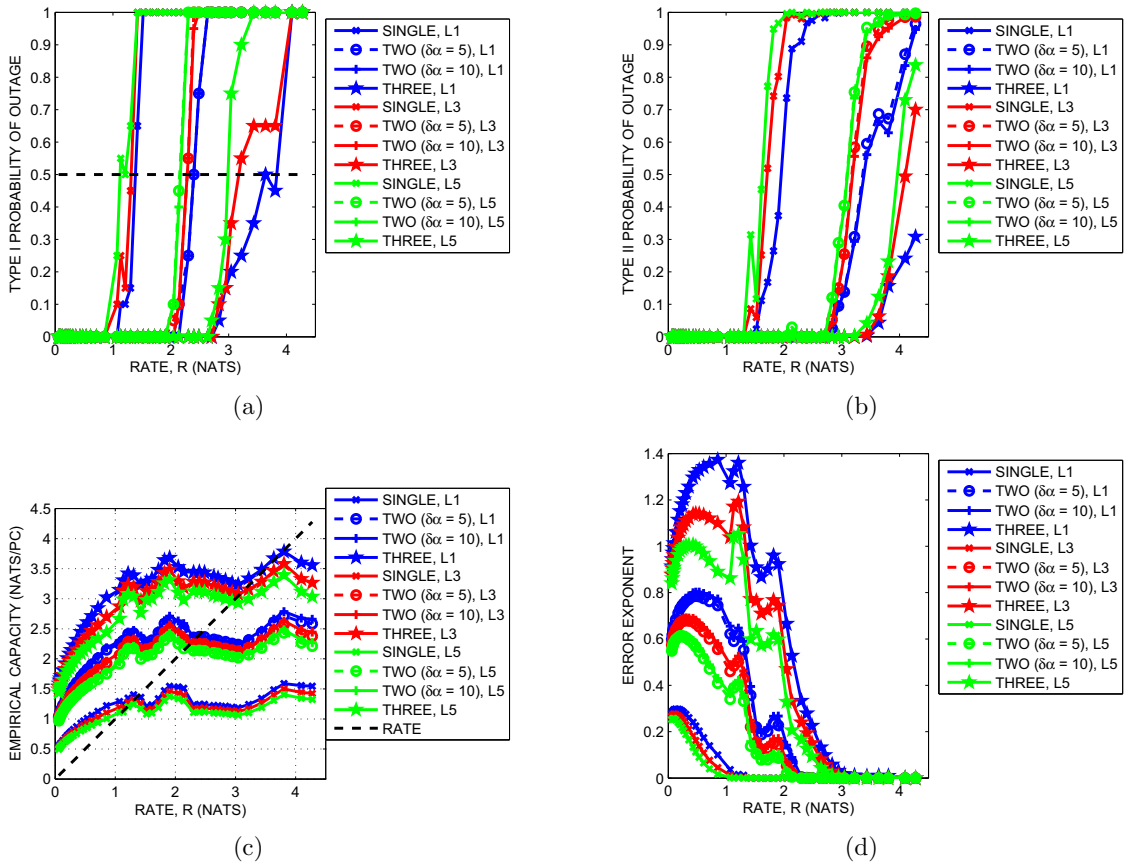


Figure 9.2. (a) The Type I probability of outage, (b) the average Type II probability of outage, (c) the average empirical mutual information rate and (d) the average empirical random coding bounds. The results are provided for the ATR dataset with test images at distortion Level 1, 3 and 5, given single image, two images with relative angle 5 and 10 degree and three images.

transition region, a value of probability of outage can be interpreted as the percentage of times when an object is not recognized correctly provided that a number of attempts is made to recognize an object. Note that the estimated capacity is within the transition region, which indicates that the probability of recognition error is not zero for the largest n, resulting in the code $(n, 2^{nR})$, where $R = \hat{C}$. In summary, for a dataset of a finite size, recognition with empirically evaluated zero probability of error is possible only at the rates significantly smaller $\hat{C}$, that is, at the rates within the region of perfect recognition. These conclusions hold for Type I probability of outage. However, it is harder to establish the relationship between the point of the empirical recognition capacity and the Type II probability of outage. Since the Type I and Type II probabilities of outage are two distinct estimates of the probability of outage, they exhibit distinct behavior as the rate, R, grows.

The actual values of empirical capacity evaluated from the sequence of the empirical mutual information rate, the empirical bounds and the Type I probability of outage are compared in Fig. 9.4.

9.3 Summary

Similar to the communication channel, the recognition channel is characterized by the outage probability. We define two types of outage probability. The Type I outage probability involves the empirical capacity, while the Type II outage probability is based on the instantaneous capacity. The subset I and II of COIL-100 and ATR dataset are used to evaluate the empirical outage probability. The rate at which the Type I outage probability equals to 0.5 is in agreement with the rate at which the average empirical capacity is equal to the rate. We also analyze the outage probability involving images at different distortion levels. Generally, the heavy distortions result in high outage probability at a given rates. For the subset II of COIL-100 dataset and the ATR dataset, we compared the outage probability for the two image case with the images separated by the relative angles $\delta\alpha = 5$ and 10 and three image case against the result for the single image case. The results show that recognition based on multiple images provides more reliable recognition compared to recognition based on a single image case.

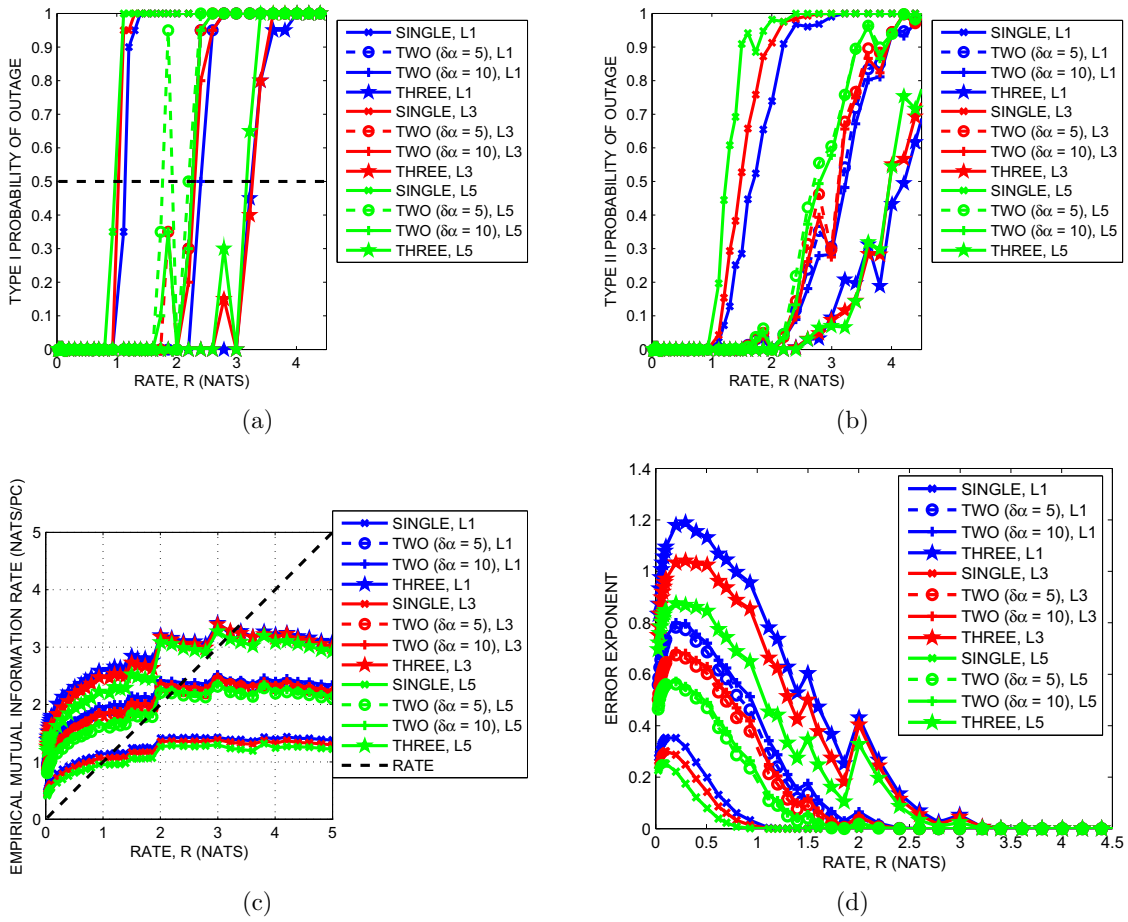


Figure 9.3. (a) The Type I probability of outage, (b) the average Type II probability of outage, (c) the average empirical mutual information rate and (d) the average empirical random coding bounds. The results are provided for the subset II of COIL-100 dataset with test images at distortion Level 1, 3 and 5, given single image, two images with relative angle 5 and 10 degree and three images.

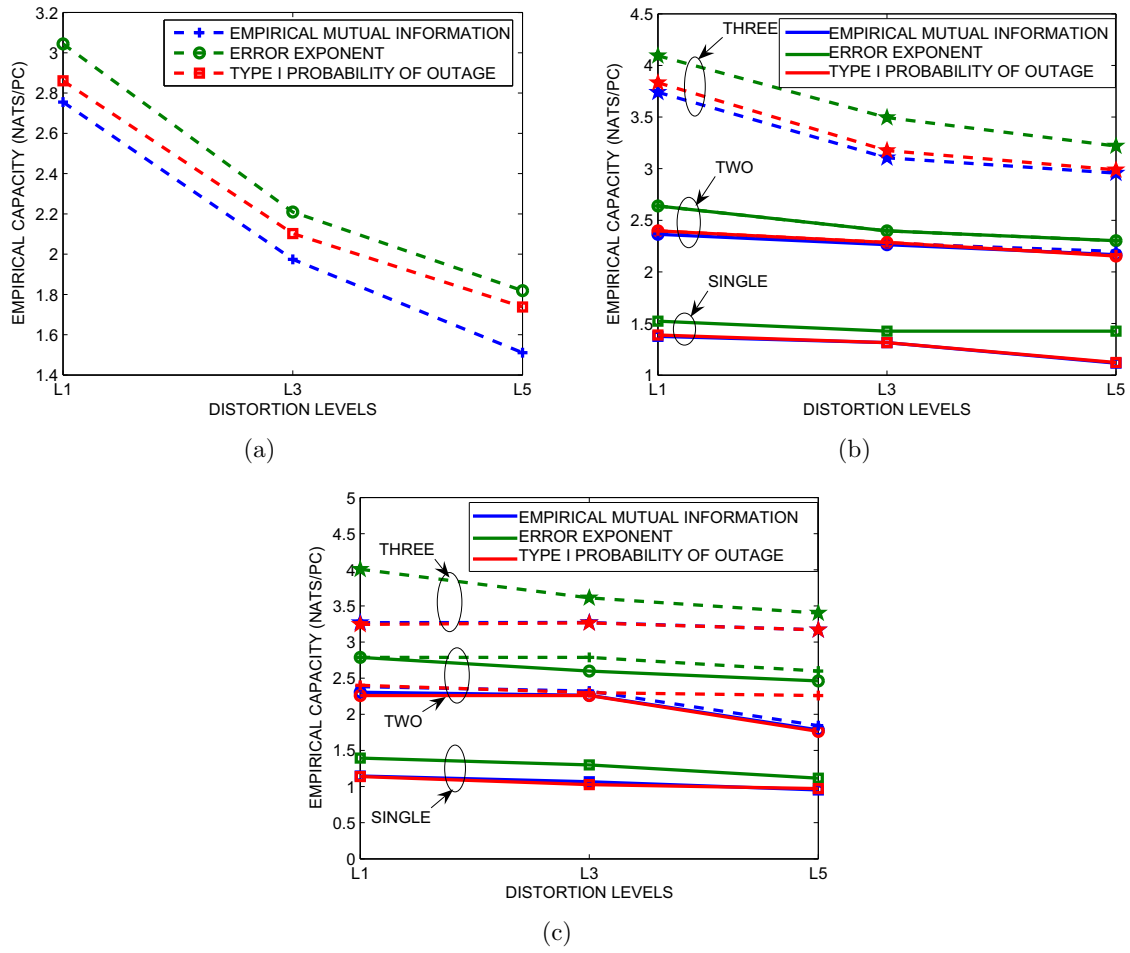


Figure 9.4. The consistence of empirical capacity estimated from the sequence of the empirical mutual information rate (blue), the empirical random coding bounds (green) and the Type I probability of outage (red), given test images at distortion Level 1, 3 and 5. The results are provided for (a) the subset I of COIL-100 dataset, (b) the ATR dataset and (c) the subset II of COIL-100 dataset.

CHAPTER 10

A TWO NODE RECOGNITION SENSOR NETWORK

Theoretically, the recognition systems benefit from using more sample images of the same target. However, in a sensor network, the limited communication between nodes imposes additional constraints. This can decrease the overall system performance. It is important to optimize the ATR system performance under the constraint on communication channel or input codewords. In this chapter, a two node recognition network is considered, and various scenarios of its operation are analyzed. The performance for each scenario will be evaluated numerically.

10.1 Scenarios of Operation of a Two Node Network

Consider a two node recognition network. Each node is autonomous unit equipped with a sensor (for instance, camera), radio transmitter and receiver, a processing unit (a computer), power block, and a storage unit. The node is autonomous since it can independently acquire, process, store, transmit, and receive data. Nodes can be stationary ground nodes or flying UAVs. In this chapter we are not concerned with control problem. The main focus of this work is to analyze large scale recognition performance at the network. We assume that the recognition network can operate in three modes or follow three scenarios. In Scenario I, the images obtained by Node 1 are encoded on board using the PCA method. The PCA codes are further converted to bitstream and the bitstream is transmitted to the ground station through a wireless communication channel. Then the ground station performs classification based on the transmitted data. The flow chart of Scenario I is shown in Fig. 10.1 (a). In Scenario II and III, the classification is performed based on two images. In Scenario II, after converting the PCA codes of images to bitstream on board, Node 1 transmit the information to Node 2. Then Node 2 classifies objects according to the transmitted data from Node 1 and the encoded information from Node 2. Thus, the recognition is performed by Node 2. Compared with Scenario II, in Scenario III the final decision will be made by the ground station using

the transmitted codes from both Node 1 and Node 2. The diagrams in Fig. 10.1 (b) and (c) show the two scenarios.

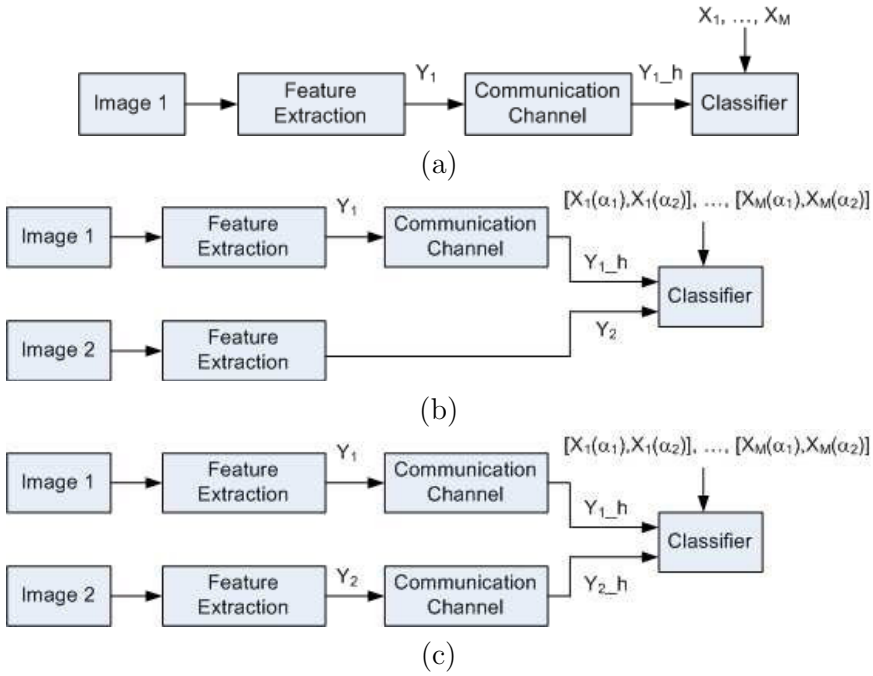


Figure 10.1. Block Diagrams: (a)Scenario I, (b)Scenario II and (c)Scenario III.

10.2 Performance Measures

In this section, the performance measures used to characterize recognition channel, communication channel and the overall system are described.

To characterize a recognition channel, the empirical mutual information rate defined in Chapter 7, the empirical random coding exponent, the space partitioning exponent defined in Chapter 8 and the outage probability defined in Chapter 9 are utilized. For the purpose of comparison, the probability of recognition error, one of traditional performance measures, is involved.

In a communication channel, transmission bit error rate (BER) is a common performance measure, which is defined as the number of errorneous bits received divided by the total number of bits transmitted. In the case when a wireless channel experiences Rayleigh fading, another appropriate measure of performance is the outage probability

of the communication channel. The probability of outage is defined as the probability that the instantaneous capacity is less than the transmission rate, R_c,

$$PO_c = Pr[C(SNR) < R_c],$$

For a block fading channel, the unconstrained instantaneous capacity is given by

$$C(SNR) = W \log(1 + SNR), \tag{10.1}$$

where W is the bandwidth and the signal to noise ratio, SNR, is a random variable for each fading status. SNR is often modeled as being Gamma distributed. Note that this capacity is for an unconstrained input. Therefore, (10.1) assumes arbitrary modulation. This expression also does not depend on the type of channel coding. It would be more appropriate to use the capacity based on a particular modulation and coding method, however, the difference could only be a couple of dBs. So it is still a good approximation.

Given L blocks per frame, the instantaneous capacity of one frame is

$$C(SNR) = \frac{1}{L} \sum_{i=1}^{L} W \log(1 + SNR_i),$$

where SNR_i corresponds to the signal-to-noise ratio of each block.

The overall system performance will be evaluated using probability of recognition outage, which is an important metric that quantifies the leakage of information at a given rate. The system probability of outage is evaluated using the transmitted PCA codewords, i.e., $\hat{Y}_1$ for Scenario I ,$\{\hat{Y}_1, Y_2\}$ for Scenario II and $\{\hat{Y}_1, \hat{Y}_2\}$ for Scenario III.

10.3 Experiments and Results

In this section, the performance of each scenario will be evaluated numerically by varying the parameters of the recognition channel and the communication channel. We use ATR dataset to analyze the performance of the recognition network. The dataset is composed of 72 classes, 12 images per class.

In our simulations the wireless communication channel will assume Rayleigh block fading. We select frequency shift keying (FSK) modulation. FSK is a modulation scheme in which digital information is transmitted through discrete frequency changes of a carrier wave. Encoding and decoding employ cdma2000. The detailed information on cdma2000 can be found in [59].

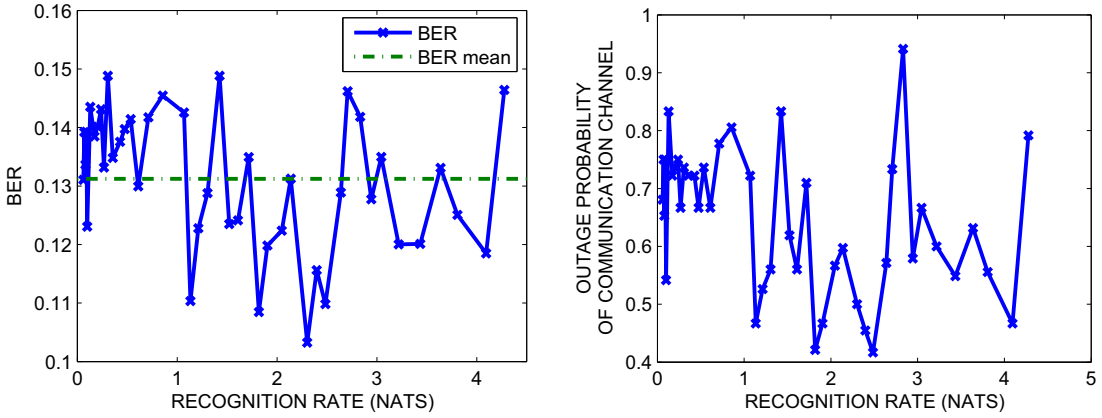
The PCA encoding and parameter estimation are described in Chapter 7.3.1. Since transmission of data requires convertion of the fractional PCA codes of one image into a bitstream, we chose a fixed-point conversion. The range and precision of the fixed-point conversion depends on the dataset. Here, we use $ws = 18$ bits per PCA component. The precision is $p = 0.0001$ and the bias is $b = 0$. Then the range is $[p \times (-2^{ws-1}) + B, p \times (2^{ws-1} - 1) + B]$, that is $[-13.1072, 13.1071]$ for our case.

Given the length n of a codeword from a single encoded image, the total number of bits transmitted over a wireless channel is $18 \times n$ bits per encoded image. Assume that all PCA encoded components converted into a bit stream from a single image are transmitted over the communication channel within one frame. In this case the number of bits per frame, K, should be greater than $18 \times n$. For the ATR dataset, the maximum length of a PCA encoded image, n, is 72. Thus the number of required bits per frame is at least 1296. According to the standard of cdma2000, we select $K = 1530$ bits. In each frame, the first $18 \times n$ bits contain PCA codewords converted into bit streams. The remaining bits are set to be zeros or ones at random.

The recognition performance is evaluated as a function of the recognition rate, R. By changing the number of classes, M, and the code word length, n, we can obtain different values of R. For each given R the total of $M \times 12$ images are converted to bitstreams and further transmitted over the communication channel. The performance of communication channel characterized by BER and PO_c are evaluated based on all K bits per frame and averaged over $M \times 12$ frames (assume one frame per image). The bitstreams after the communication channel are converted back to fractional numbers and compared with the templates in the database to make classification. The system probability of outage (Type II) is obtained by comparing the instantaneous recognition capacity against the recognition rate.

First, we fix the parameters of communication channel, i.e., 4-ary non-coherent FSK modulation, 20 blocks per frame and cdma2000 encoding rate 1/2. The E_b/N_o of communication channel is set to be $6dB$. The BER of the communication channel is shown in Fig. 10.2 (a). The average BER is 0.13. The outage probability of the communication channel is shown in Fig. 10.2 (b).

The recognition performance for all three scenarios is compared with the performance without transmission using a single image or two images. Fig. 10.3 (a)-(d) show the empirical mutual information rate, the Type II outage probability of the recognition channel, the error exponents and the probability of recognition error as a function of recognition rate, respectively. We can see that after transmission of the PCA encoded images over the communication channel, the recognition performance drops for both a



(a) (b)

Figure 10.2. (a) BER and (b) Outage probability of communication channel as a function of the recognition rate. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.

single image and two image cases. The errors introduced during data transmission result in highly unstable recognition performance in the region of small recognition rates. This result is due to definition of recognition rate. Small values of R are achieved in practice using long codewords. This requires a large number of bits to be transmitted. The transmitted codewords with larger n contain more errors than the codes with smaller n. Another observation is that the performance of the network set to operate according to results Scenario 3 (with both sensor nodes communicating their PCA codewords to the central station, marked in green) are worse than the results under Scenario 2 (with a PCA codeword communicated by one of sensor nodes, marked in red). However, under both scenarios, the results when two frames per object with relative angles $\delta\alpha = 5$ (marked by 'o') and $\delta\alpha = 10$ (marked by '+') do not differ significantly.

We also test the impact of performances of the communication channel on the system performance. Here the performanceis analyzed only under Scenario 1. We first vary E_b/N_o of the communication channel. We set it to $6dB$, $8dB$, $10dB$ and $20dB$. The performance of communication channel is displayed in Fig. 10.4, and the performance of the recognition channel is displayed in Fig. 10.6 (a) to (d). We can see that as the E_b/N_o of the communication channel increases, the recognition performance under Scenario 1 improves.

We further vary the transmission rate of the communication channel under fixed $E_b/N_o = 6dB$. We investigate three transmission rates supported by cdma2000, specif-

ically $R_c = 1/2, 1/3$, and $1/4$. Fig. 10.5 shows the communication BER and PO_c as a function of the recognition rate. The recognition performance as a function of the recognition rate, R, is displayed in Fig. 10.7 (a)-(d). It is hard to interpret the performance difference under 3 different transmission rates. We further plot the system outage probability (Type II) as a function of code rate R_c and recognition rate R_r. Fig. 10.8 illustrates the results. We can see that as the code rate of the communication channel increases, the performance under Scenario 1 drops.

We further change the order of FSK modulation to 8-ary FSK and 16-ary FSK modulation while keeping the other parameters fixed. As the size of constellation increases, the ratio of energy per symbol to the noise variance, E_s/N_o increases, and therefore the overall communication performance inproves. Fig. 10.9 illustrates the results. From the the recognition performance plotted in Fig. 10.10 (a)-(d), we can also conclude that the recognition and system performance increase as the order of FSK modulation increases.

The impact on the performance of varying number of blocks per frame in the communication channel is also investigated. The number of blocks per frame varies from 2, 20 to 100. The communication and recognition performance results are shown in Fig. 10.11 and Fig. 10.12. We can see that as the number of blocks per frame increases, the channel is more ergodic. Although the values of outage probability of communication channel for different number of blocks per frame are similar, the actual communication bit error rates are different. The BER with 20 blocks per frame is higher than the BER with 2 or 100 blocks per frame. This higher communication error rate of the channel with 20 blocks per frame degrades recognition performance compared with the recognition results of the channel with 2 or 100 blocks per frame.

10.4 Summary

In this Chapter, the topic of joint recognition and communication was discussed. We considered a two node sensor network and assumed 3 scenarios of its operation. The system probability of outage was defined as probability that encoded images after their transmission over a wireless communication channel are not recognized correctly at the recognition rate, R. The bit error rate (BER) and the communication probability of outage were used to analyze the performance of communication channel. The empirical mutual information rate, the Type II outage probability, the error exponents of recognition channel and the probability of recognition error were applied to measure the performance of recognition channel. ATR dataset was utilized to obtain numerical results. The wireless communication channel was the Rayleigh block fading channel using

non-coherent FSK modulation and $K = 1530$ bit cdma2000 turbo code. Given the communication channel with $E_b/N_o = 6dB$, using 4-ary NFSK modulation and code rate 1/2, we found that the recognition performance of all scenarios decrease when compared with the performance without transmission. The system probability of outage under Scenario 3 is higher than the system outage probability under Scenario 2. However, the system performance based on combined information from two sensor nodes (Scenario 2 or 3) is higher than the performance based on the performance based on a single encoded image transmitted over communication channel (Scenario 1).

We also tested the impact of various parameters on system performance under Scenario 1 by varying the parameters of the communication channel. As E_b/N_o of the communication channel increases, the recognition performance improves and the system probability of outage decreases. For a high signal-to-noise ratio in communication channel, the performance of the entire system based on transmitted PCA codewords exhibits similar performance to the performance of the network without data transmission. We further investigated the influence of varying transmission rate on the performance. Three code rates of communication channel supported by cdma2000 were tested. The system probability of outage as a function of the transmission rate and the recognition rate was evaluated. The order of FSK modulation was also varied. The use of the higher order of FSK modulation results in higher system performance. Finally, the impact of varying number of blocks per frame on the performance was evaluated.

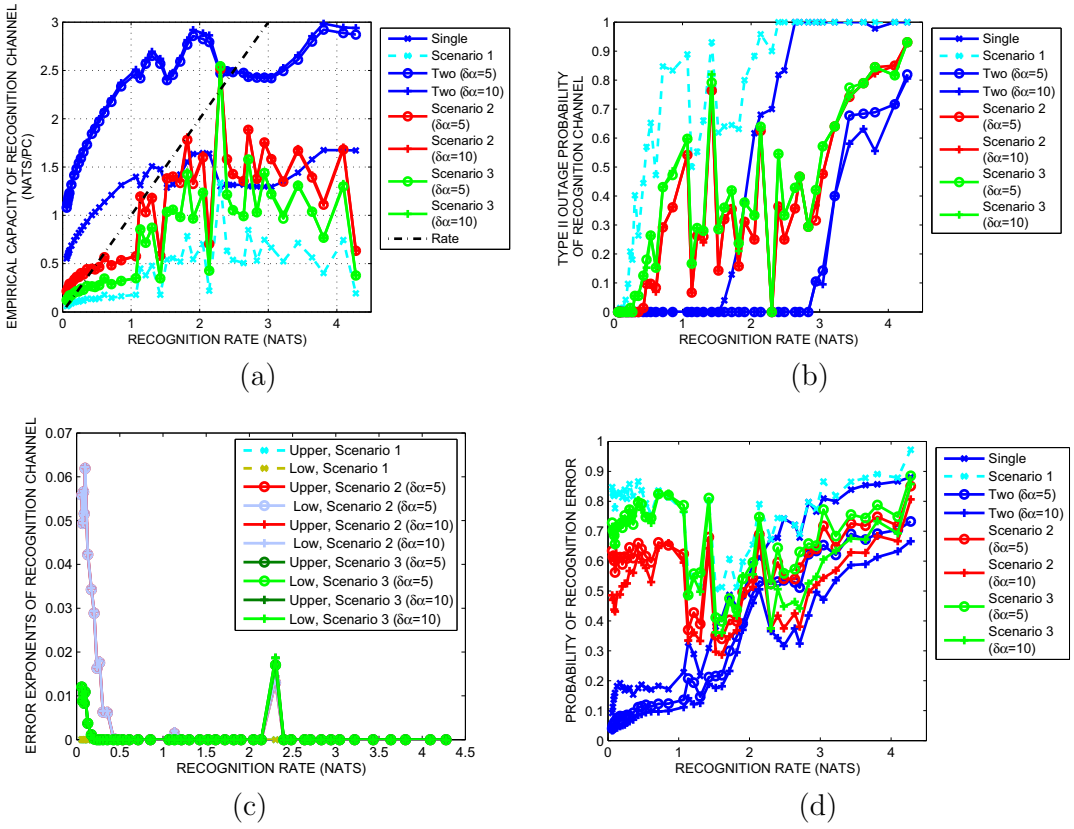
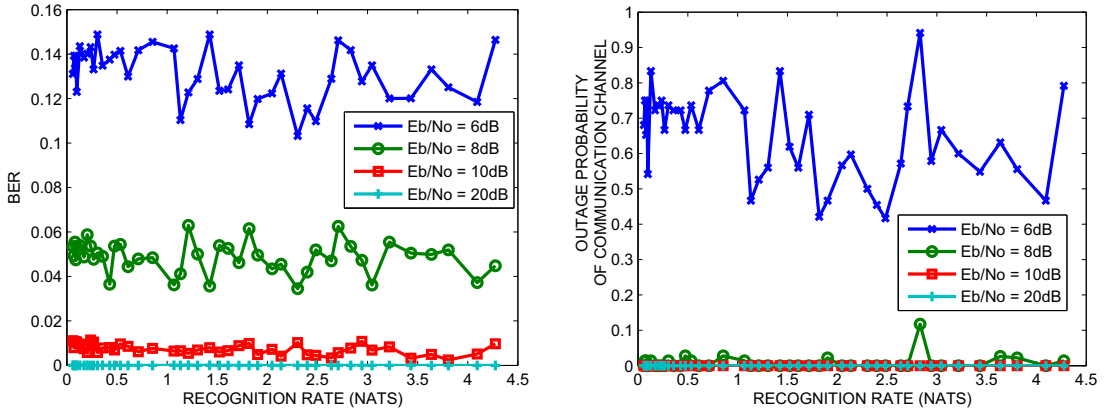
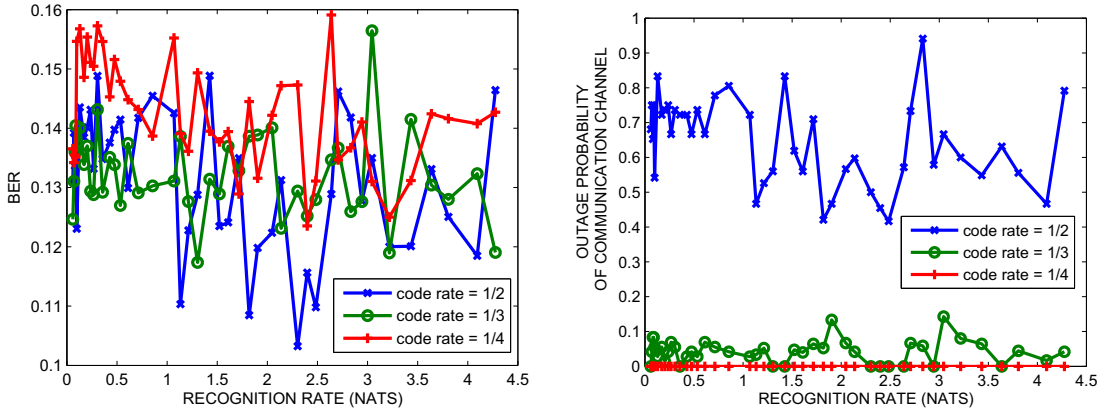


Figure 10.3. (a) Empirical mutual information rate, (b) Type II outage probability, (c) Random coding error exponents and (d) Probability of recognition error as a function of recognition rate. The results of Scenario 1 (cyan lines), Scenario 2 (red lines) and Scenario 3 (green lines) are compared with the results without data transmission (blue lines). The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.



(a) (b)

Figure 10.4. (a) BER and (b) Outage probability of communication channel for each recognition rate when E_b/N_o of the communication channel is $6dB$, $8dB$, $10dB$ and $20dB$. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.



(a) (b)

Figure 10.5. (a) BER and (b) Outage probability of communication channel parameterized by the transmission rate R_c set to $1/2$, $1/3$ and $1/4$. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code are used.

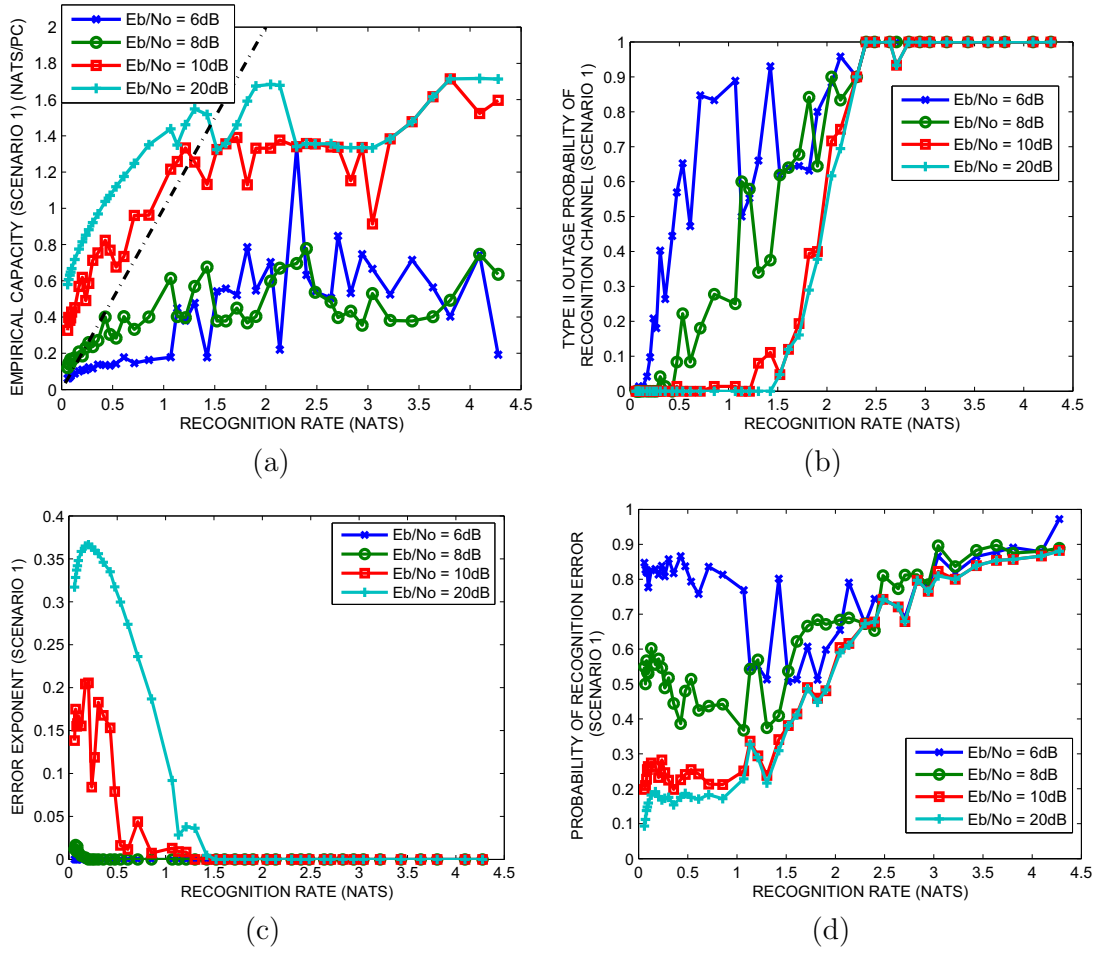


Figure 10.6. Scenario 1: (a) Empirical mutual information rate, (b) Type II outage probability, (c) Error exponents of the recognition channel and (d) Probability of recognition error as a function of recognition rate. E_b/N_o of the communication channel is $6dB$, $8dB$, $10dB$ and $20dB$. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.

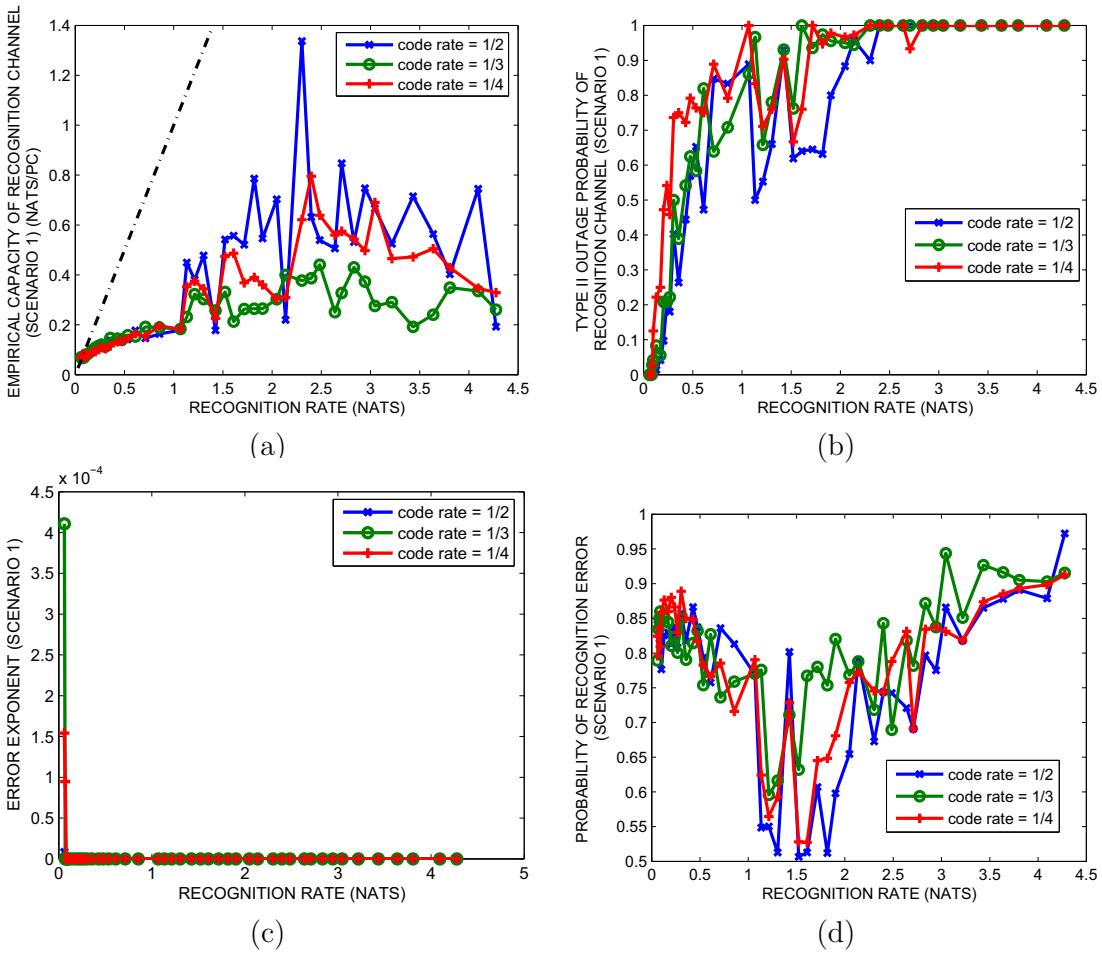


Figure 10.7. Scenario 1: (a) Empirical mutual information rate, (b) Type II outage probability of the recognition channel, (c) Random coding lower bound and (d) Probability of recognition error as a function of recognition rate. Code rate R_c is 1/2, 1/3 and 1/4. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code are used.

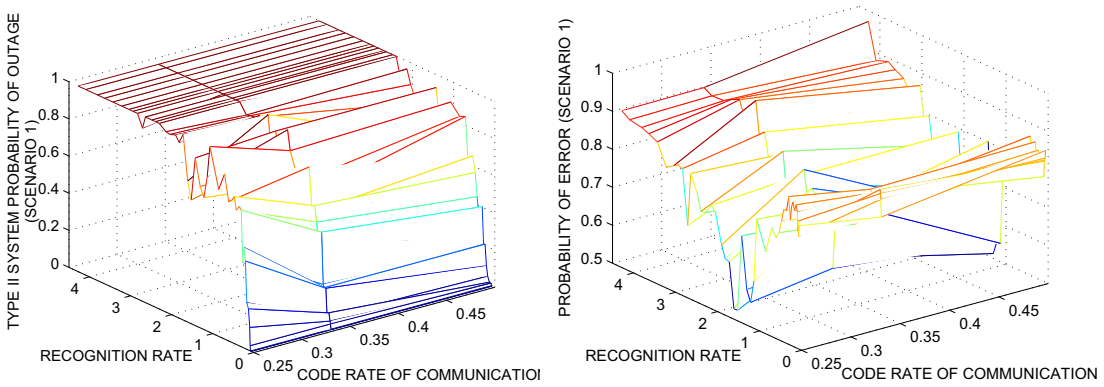
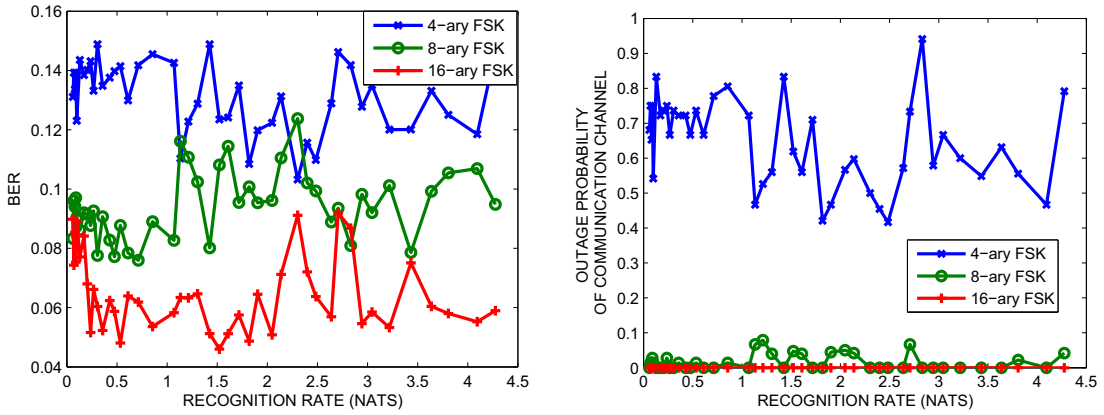


Figure 10.8. Type II system probability of outage (left panel) and probability of error (right panel) for Scenario 1 as a function of transmission rate and recognition rate. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code are used.



(a) (b)

Figure 10.9. (a) BER and (b) Outage probability of communication channel as a function of the recognition rate for three FSK modulations: 4-ary, 8-ary and 16-ary. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.

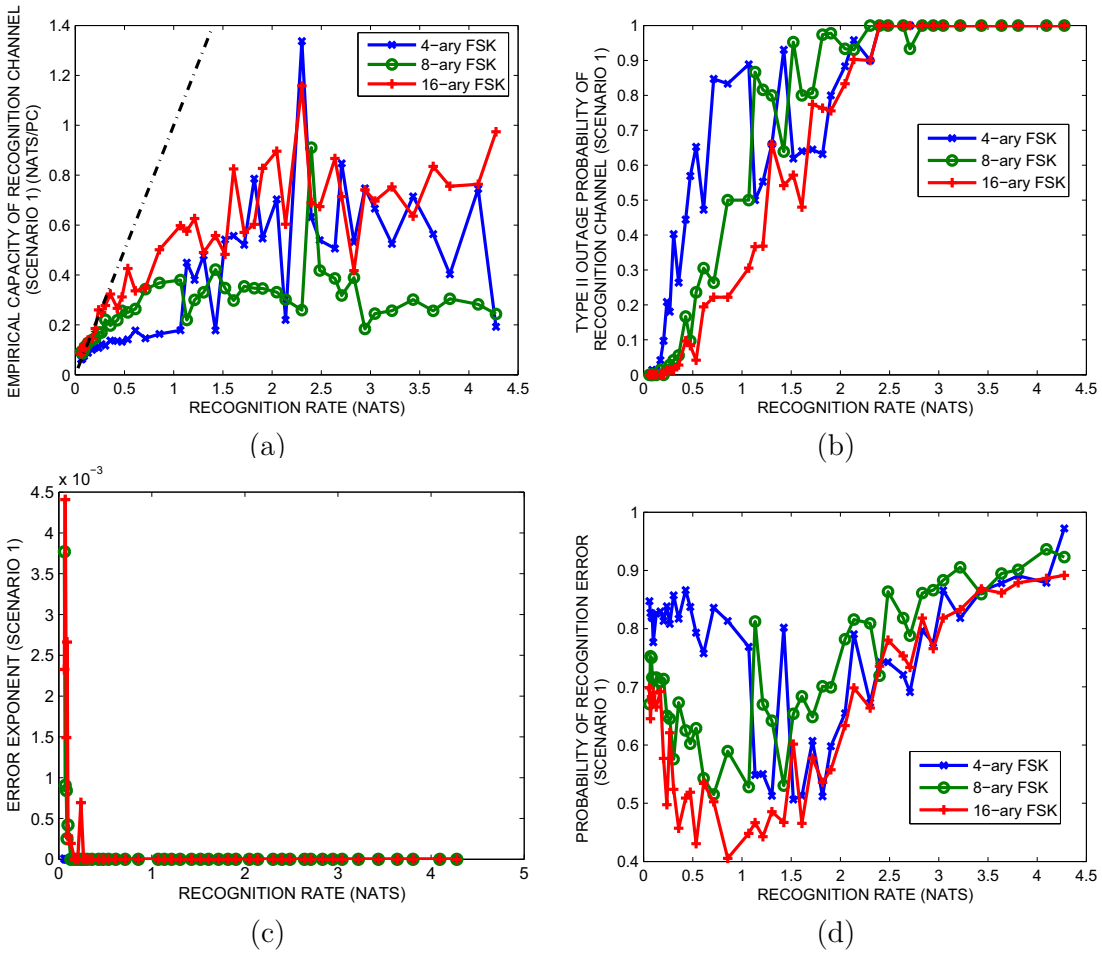
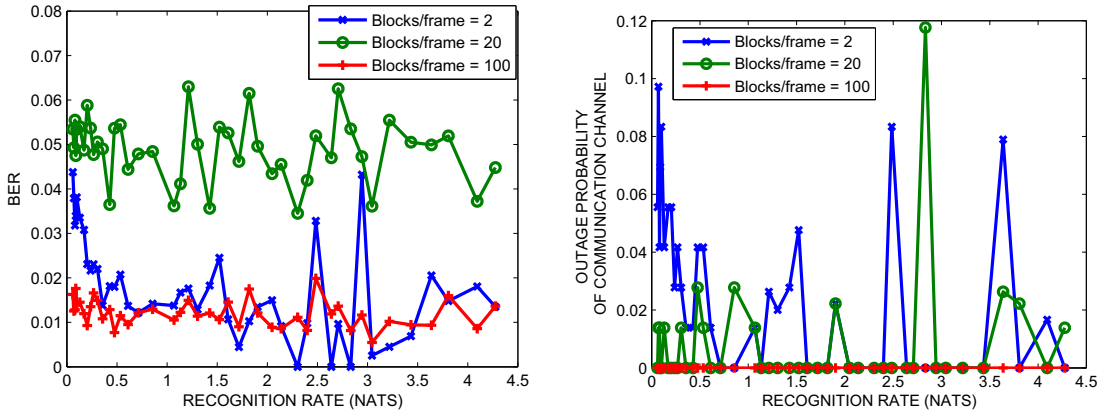


Figure 10.10. Scenario 1: (a) Empirical mutual information rate, (b) Type II outage probability of the recognition channel, (c) Random coding exponent and (d) Probability of recognition error as a function of the recognition rate. The order of FSK modulation is 4-ary, 8-ary and 16-ary. The performance is evaluated over a Rayleigh block fading channel with 20 blocks per frame and $E_b/N_o = 6dB$. NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.



(a) (b)

Figure 10.11. (a) BER and (b) Outage probability of communication channel as a function of the recognition rate when the number of blocks per frame is set to 2, 20 and 100. The performance is evaluated over a Rayleigh block fading channel with $E_b/N_o = 8dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.

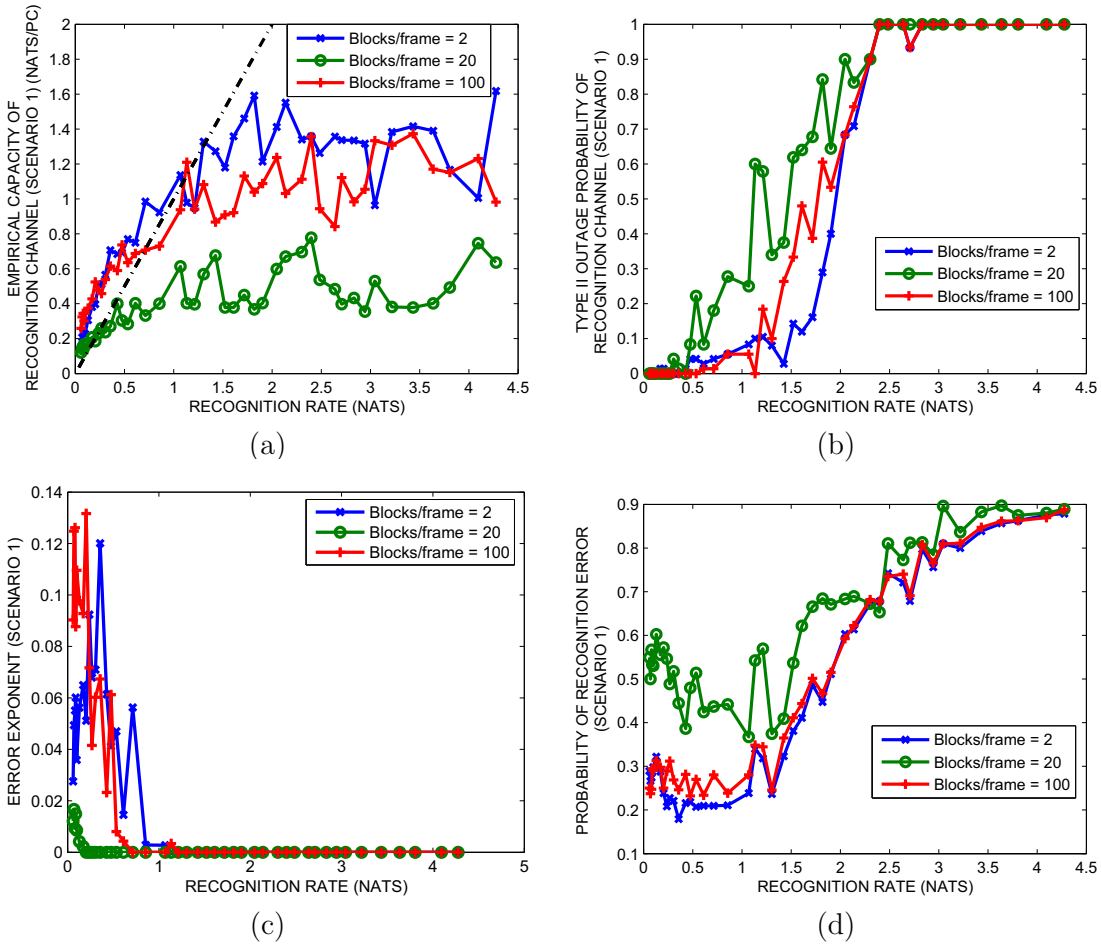


Figure 10.12. Scenario 1: (a) Empirical mutual information rate, (b) Type II outage probability of the recognition channel, (c) Random coding lower bound and (d) Probability of recognition error as a function of the recognition rate. The number of blocks per frame is 2, 20 and 100. The performance is evaluated over a Rayleigh block fading channel with $E_b/N_o = 8dB$. 4-ary NFSK modulation and the $K = 1530$ bit cdma2000 turbo code with code rate $R_c = 1/2$ are used.

CHAPTER 11

CONCLUSION AND FUTURE WORK

11.1 Conclusion

Researchers have invested a substantial amount of effort in studying Automatic Target Recognition (ATR) or generally, object recognition for decades of years. New sensors and technologies in control and systems provide new platforms and challenges for ATR. Over the past several years, there is an increasing trend towards operating swarms of Unattended Aerial Vehicles (UAVs) as large-scale sensor networks in the air. This trend offers opportunities of integration ATR systems with a UAV-based sensor network to improve the recognition performance. To build an ATR system using UAVs successfully, researches are needed in lots of fields, such as sensors modeling, signal/image processing, detection, tracking, etc. In this thesis, we focus on three problems: designing a complete automatic recognition system, evaluating capabilities of this system to perform recognition in complex environments and evaluating performance limits of a two node recognition network.

We first introduce a model-based recognition method. Instead of modeling the image itself, we decompose images into their spectral components using a bank of filters. Bessel K forms are then applied to model the low dimensional statistics of these filtered components. In our thesis, we choose the histogram of the filtered images. These parametric forms provide a close form of the distance metric, pseudo-metric, between two images. This metric is then combined with the metric based on K-measure to balance accuracy and speed. This model-based method is further extended to recognition using two correlated images, by modeling the joint histogram of a pair of filtered images using multivariate Bessel K forms. The recognition performance is tested using both simulated dataset and the real dataset.

A complete detection-recognition system is then built. Images with potential targets are first detected using a Haar-like feature based method. The detected regions of each image are then combined or removed using a heuristic method based on the relative locations of the windows. After detection, a B-spline based segmentation method is applied

to separate the images of potential targets from the clutter within each detected region. The recognition method is then applied to each segmented regions. These procedures allow to detect and recognize targets regardless of the type of backgrounds. Two methods of rejecting an unknown object are introduced. This first one is based on classification index. The final reject or accept rule is based on a regression tree. System performance is evaluated on the simulated 3D dataset. The second one uses cluster method to separate sample clutter images and the detected windows into two groups. This method is applied in the die cast dataset collected using the real images. The performance of the Bessel K based recognition method is compared with the performance of PCA method. The Bessel K based method is proved to be more robust in the presence of clutters and occlusions.

To analyze the performance of recognition system, a general methodology from channel capacity view point is used. We derive the expression for capacity of a recognition system under constraint of the PCA-based encoding. Encoded single and multiple frames containing the same target are modeled as realizations of Gaussian vectors. These models are verified using statistical goodness of fit tests. The relationship of the empirical mutual information rate and the recognition rate is established. The empirical capacity is then evaluated from the sequence of empirical mutual information rates using two datasets for both high resolution images and low resolution images. The empirical capacities for the distorted images at different levels are also obtained.

The recognition channel capacity is related with the maximum number of classes which can be correctly recognized with arbitrarily small probability of error when the length of code words becomes large. However, the constrained recognition capacity does not specify the value of the probability of error, we further study the relationship between code length and the probability of decoding error for a recognition channel. The recognition reliability function is involved and bounded using the random coding lower bound and the space partitioning upper bound. The empirical bounds are evaluated using two datasets for both single and two frames. The rates where the bounds become zero are consistent with the crossing point where the empirical mutual information rate is equal to the rate.

Similar to the communication channel, the recognition channel is also characterized by the outage probability. It evaluates the confidence measure that the empirical capacity estimated using codewords of M classes, each of length n does not drop below the recognition channel rate $R = \log M/n$. Two types of outage probability are defined. The empirical outage probability is evaluated using two datasets. The relationship between the outage probability and the recognition rate is studied. We further analyze the

outage probability involving images at different distortion levels. Generally, the heavy distortions result in high outage probability at given rates.

Finally, we go back to the problem of recognition sensor networks. The joint recognition and communication problem is discussed. Since in the swarmed UAV system, the limited communication between UAVs imposes additional constraints, the benefit of the recognition system obtained from using more sample images of the same target will decrease. We consider a small sensor network composed of two autonomous nodes. Three scenarios of its operation are considered. The system probability of outage is defined to evaluate the entire system performance. The simulated 3D dataset is used to obtain numerical results. Here the wireless communication channel is the Rayleigh block fading channel using non-coherent FSK modulation and $K = 1530$ bit cdma2000 turbo code. The system performance of different scenarios are compared. We also test the impact of performance by varying the parameters of communication channels.

11.2 Future Work

The model-based recognition method based on Bessel K forms as was shown is more robust compared to the PCA method especially in the presence of clutters and occlusion. However, the filter bank used to filter each image requires adjustment with each new dataset. Additional studies about the number of filters, the type of filters and the parameters of each type of filter are needed to find an optimal set up. Although the combination method using both pseudo-metric and k-measure can speed up the classification, more efforts may be involved in algorithm optimization to achieve the requirement of real time recognition. One limitation of this model-based recognition method is that it is a texture based method. Therefor, it may not work well for targets without rich textures. One way is to involve other features, such as shape as overall structure, to improve the recognition performance.

Although the detection-recognition system performance is relatively good when tested on the simulated 3D dataset, the performance evaluated using the die cast dataset indicates that the designed system needs adjustments when dealing with complex environments. The real situation is very complex, including uneven lighting conditions, shadows, camouflage of targets, etc. To improve the system performance, a proper image processing, efficient detection method and a robust segmentation are needed. A proper image processing may include adaptive illumination adjustment and shadow detection. When the number of targets and the complexity of target increase, the Haar-like feature based detector trained using all targets at an arbitrary pose becomes inefficient. The number

of false alarms increases dramatically. One possible way to solve the problem is to train each target at a range of poses separately and then combine all detectors together. A more robust segmentation method is needed for targets displaying a lot of textures on a background of clutter. One way to deal with the problem is to use the prior information about the clutter. For camouflaged targets, optical cameras are not an appropriate imaging modality. Infrared sensors can be mounted on board of UAVs to obtain additional useful information about objects of interest.

In the part with performance evaluation, we analyze the recognition limits of recognition channel under the constraint of PCA encoding. Other encoding techniques can be used to encode images. The distributions of data under matching and nonmatching hypotheses can be modeled using parametric and nonparametric techniques. The capacity and other information theoretic limits under constraints on encoding techniques can be evaluated.

In this thesis, targets at different pose are treated as distinct classes. The problem can be viewed as a joined pose estimation-target recognition problem. However, it is not clear how to evaluate the object recognition capacity of the channel. Therefore, we would like to address this problem in our future work.

We discussed a joint recognition and communication system in the last part of the thesis. Three scenarios of operation of a two node sensor network were proposed and a comprehensive performance evaluation was performed. Further optimizations of the two node sensor network need to be addressed. One possible solution is tandem structure with resource control, that is optimal allocation of recognition and communication bits given the overall rate constraints. Since there are more than one sensor collecting data in the networks, the multi-terminal optimization should also be considered.

APPENDIX A

RANDOM CODING LOWER BOUND

The random coding lower bound can be obtained from the hypothesis test. Given $R = \log(M)/n$, we want to prove that there exists a data code of size M and blocklength n for the channel with conditional probability $Q(Y|X)$ whose probability of decoding error p_e satisfies

$$p_e \leq \min_{s\in[0,1]} \min_{p(X)} \left\{ e^{nsR} \int_Y \left[\int_X p(X) Q(Y|X)^{\frac{1}{1+s}} dX \right]^{1+s} dY \right\}.$$

Detailed proof can be found in [6], here we only give the brief descriptions.

Let $\{\mathbf{C}_0, \cdots, \mathbf{C}_{M-1}\}$ denote the set of codewords, and let the decoding rule be that the mth codeword is decoded when Y is received if $p(Y|\mathbf{C}_m) > p(Y|\mathbf{C}_{m'})$ for all $m' \neq m$. All possible Y that decode into $\mathbf{C}_m$ form a set $\mathcal{U}_m$,

$$\mathcal{U}_m = \{Y : p(Y|\mathbf{C}_m) > p(Y|\mathbf{C}_{m'}) \ \ for\ all\ m' \neq m\}.$$

The characteristic function of this set $\phi_m(Y)$ is defined as

$$\phi_m(Y) = \begin{cases} 0 & if\ Y \notin \mathcal{U}_m \\ 1 & if\ Y \in \mathcal{U}_m. \end{cases}$$

If $\phi_m(Y) = 0$, then $p(Y|\mathbf{C}_m) \leq p(Y|\mathbf{C}_{m'})$ is true for at least one m'. Thus for $s \geq 0$, we have

$$1 \leq \left[\frac{p(Y|\mathbf{C}'_m)}{p(Y|\mathbf{C}_m)} \right]^{\frac{1}{1+s}}.$$

Since the ratio $\frac{p(Y|\mathbf{C}'_m)}{p(Y|\mathbf{C}_m)}$ is nonnegtive for any m', we have

$$1 - \phi_m(Y) \leq \left\{ \sum_{m'\neq m} \left[\frac{p(Y|\mathbf{C}'_m)}{p(Y|\mathbf{C}_m)} \right]^{\frac{1}{1+s}} \right\}^s. \tag{A.1}$$

This inequality is also true when $\phi_m(Y) = 1$ because the right side is nonnegative.

Given that the mth codeword is transmitted, the probability of decoding error is given by

$$\begin{aligned} p_{e|m} &= \int_{Y \notin \mathcal{U}_m} p(Y|\mathbf{C}_m) dY \\ &= \int_Y p(Y|\mathbf{C}_m)[1 - \phi_m(Y)] dY \\ &\leq \int_Y p(Y|\mathbf{C}_m) \left\{ \sum_{m' \neq m} \left[\frac{p(Y|\mathbf{C}'_m)}{p(Y|\mathbf{C}_m)} \right]^{\frac{1}{1+s}} \right\}^s dY \\ &= \int_Y p(Y|\mathbf{C}_m)^{\frac{1}{1+s}} \left[\sum_{m' \neq m} p(Y|\mathbf{C}'_m)^{\frac{1}{1+s}} \right]^s dY. \end{aligned} \tag{A.2}$$

Assume that the codewords are selected independently, with probability of selecting X as a codeword equal to $p(X)$. Then the expected value $p_{e|m}$ is

$$\begin{aligned} E[p_{e|m}] &\leq E\left\{ \int_Y p(Y|\mathbf{C}_m)^{\frac{1}{1+s}} \left[\sum_{m' \neq m} p(Y|\mathbf{C}'_m)^{\frac{1}{1+s}} \right]^s dY \right\} \\ &= \int_Y E\left\{ p(Y|\mathbf{C}_m)^{\frac{1}{1+s}} \left[\sum_{m' \neq m} p(Y|\mathbf{C}'_m)^{\frac{1}{1+s}} \right]^s \right\} dY \\ &= \int_Y E\left[p(Y|\mathbf{C}_m)^{\frac{1}{1+s}} \right] E\left[\sum_{m' \neq m} p(Y|\mathbf{C}'_m)^{\frac{1}{1+s}} \right]^s dY, \end{aligned}$$

since the codewords are selected independently.

Now suppose $0 \leq s \leq 1$, then $E[t^s] \leq [E(t)]^s$ holds according to Jensen's inequality. Therefore,

$$\begin{aligned} E[p_{e|m}] &\leq \int_Y E\left[p(Y|\mathbf{C}_m)^{\frac{1}{1+s}} \right] \left\{ \sum_{m' \neq m} E\left[p(Y|\mathbf{C}'_m)^{\frac{1}{1+s}} \right] \right\}^s dY \\ &= \int_Y E\left[Q(Y|X)^{\frac{1}{1+s}} \right] \left\{ (M-1) E\left[Q(Y|X)^{\frac{1}{1+s}} \right] \right\}^s dY \\ &= (M-1)^s \int_Y \left\{ E\left[Q(Y|X)^{\frac{1}{1+s}} \right] \right\}^{1+s} dY \\ &\leq M^s \int_Y \left[\int_X p(X) Q(Y|X)^{\frac{1}{1+s}} dX \right]^{1+s} dY. \end{aligned} \tag{A.3}$$

Then the average probability of error has the expected value

$$\begin{aligned} E[p_e] &= E\left[\sum_m Pr[\mathbf{C}_m] p_{e|m}\right] \\ &= \sum_m Pr[\mathbf{C}_m] E[p_{e|m}] \\ &\leq M^s \int_Y \left[\int_X p(X) Q(Y|X)^{\frac{1}{1+s}} dX\right]^{1+s} dY. \end{aligned}$$

Because the expected value over the ensemble of codes satisfies this bound, there must be at least one code that itself satisfies the bound. Hence there exists a code whose probability of decoding error satisfies

$$p_e \leq \min_{s\in[0,1]} \min_{p(X)} \left\{ e^{nsR} \int_Y \left[\int_X p(X) Q(Y|X)^{\frac{1}{1+s}} dX\right]^{1+s} dY \right\}.$$

According to the definition of reliability function, the random coding bound is

$$E_r(R) = \lim_{n\to\infty} \max_{s\in[0,1]} \max_p \left[-sR - \frac{1}{n} \log\left(\int_Y \left[\int_X p(X) Q(Y|X)^{\frac{1}{1+s}} dX\right]^{1+s} dY \right) \right].$$

APPENDIX B

SPACE PARTITIONING UPPER BOUND

Let $\{\mathbf{C}_0, \cdots, \mathbf{C}_{M-1}\}$ denote the set of codewords, all possible Y that decode into $\mathbf{C}_m$ form a set $\mathcal{U}_m$. We will start from a binary hypothesis testing problem:

$$H_0 : \mathbf{C}_m = codewordtransmitted$$
$$H_1 : \mathbf{C}_m \neq codewordtransmitted$$

The hypotheses are characterized by $q_0(Y|X)$ for H_0 and $q_1(Y|X)$ for H_1. By defining the error exponent function $e(r)$ as

$$e(r) = \min_{\hat{q} \in \mathcal{P}_r} \int_X p(X) \int_Y \hat{q}(Y|X) \log \frac{\hat{q}(Y|X)}{q_1(Y|X)} dY dX,$$

where

$$\mathcal{P}_r = \{\hat{q} : \int_X p(X) \int_Y \hat{q}(Y|X) \log \frac{\hat{q}(Y|X)}{q_0(Y|X)} dY dX \leq r\}.$$

Then according Theorem 4.5.3. in [6], we have

$$\alpha \geq e^{-nq_\lambda(Y|X) \log \frac{q_\lambda(Y|X)}{q_0(Y|X)} - o(n)}$$
$$\beta \geq e^{-nq_\lambda(Y|X) \log \frac{q_\lambda(Y|X)}{q_1(Y|X)} - o(n)}$$

where α, β are type I and type II errors of a hypothesis-testing problem.

$$q_\lambda(Y|X) = \frac{q_0^{1-\lambda}(Y|X) q_1^\lambda(Y|X)}{\int_Y q_0^{1-\lambda}(Y|X) q_1^\lambda(Y|X) dY},$$

and λ is chosen such that the threshold T satisfies,

$$T = \int_Y q_\lambda(Y|X) \log \frac{q_0(Y|X)}{q_1(Y|X)} dY.$$

Then we defined the channel reliability exponent $E(R)$ as

$$E(R) = \max_{p} \min_{\hat{Q} \in \mathcal{L}_R(p)} \int_x p(X) \int_Y \hat{Q}(Y|X) \log \frac{\hat{Q}(Y|X)}{Q(Y|X)} dX dY,$$

where

$$\mathcal{L}_R(p) = \{\hat{Q} : \int_X p(X) \int_Y \hat{Q}(Y|X) \log \frac{\hat{Q}(Y|X)}{\int_X p(X)\hat{Q}(Y|X)dX} dX dY \leq R\}.$$

By Theorem 10.1.4 and 10.1.5 in [6], we have

1. Suppose that p^* and Q^* achieve $E(R)$, and $q^*(Y) = \int_X p^*(X)Q^*(Y|X)dX$. Then

$$E(R) = \max_{p} \min_{\hat{Q} \in \mathcal{L'}_R(p)} \int_X p(X) \int_Y \hat{Q}(Y|X) \log \frac{\hat{Q}(Y|X)}{Q(Y|X)} dX dY,$$

   where

$$\mathcal{L'}_R(p) = \{\hat{Q} : \int_X p(X) \int_Y \hat{Q} \log \frac{\hat{Q}(Y|X)}{q^*(Y)} dX dY \leq R\}.$$

2.
$$E(R) = \max_{p} \max_{s \geq 0} \left[-nsR - \log \int_Y \left(\int_X p(X) Q(Y|X)^{\frac{1}{1+s}} dX \right)^{1+s} dY \right]$$

Now let us consider the channel coding problem. Let $q^*(Y)$ be the probability distribution on the channel output that achieves $E(R^*)$ for any R^*. Select m so that $\int_{Y \in \mathcal{U}_m} q^*(Y)dY \leq \frac{1}{M}$. The hypotheses are now characterized by $Q(Y|X)$ and $q^*(Y)$, which replace $q_0(Y|X)$ and $q_1(Y|X)$, respectively. The probability of error $p_{e|m}$ replaces the type I error α and $\frac{1}{M}$ replaces the type II error β. By the Theorem 4.5.3 in [6], we have the lower bound on $p_{e|m}$:

$$p_{e|m} = \int_{Y \in \mathcal{U}_m^c} Q(Y|\mathbf{C}_m) \geq exp\{-n \int_X p(X) \int_Y Q^*(Y|X) \log \frac{Q^*(Y|X)}{Q(Y|X)} dY dX - o''(n)\}.$$

Referring to Theorem 10.1.4, these becomes

$$p_{e|m} \geq e^{-nE(R^*)-o''(n)} \geq e^{-nE(R-o'(n))-o''(n)}, \tag{B.1}$$

since $R \leq R^* + o'(n)$ and $E(R)$ is a decreasing function.

Of the M codewords, remove those $M/2$ codewords whose probability of error is largest. This results in a code having $M/2 = e^{nR-\log 2}$ codewords that must satisfy B.1. Let $\mathcal{C}'$ denote the set of codewords in the purged code. Let $(p'_e)_{\max}$ denote the maximum error of any codeword in this purged code. Then,

$$(p'_e)_{\max} \geq e^{-nE(R-\log 2/n-o'(n))-o''(n)}.$$

The average error of the original code is bounded as follows:

$$\begin{aligned} p_e &= \frac{1}{M}\sum_m p_{e|m} \geq \frac{1}{M}\sum_{m\notin\mathcal{C}'} p_{e|m} \\ &\geq \frac{1}{M}\sum_{m\notin\mathcal{C}'} (p'_e)_{\max} = \frac{1}{2}(p'_e)_{\max} \\ &\geq \frac{1}{2}e^{-nE(R-\log 2/n-o'(n))-o''(n)} \\ &= e^{nE(R)-o(n)}. \end{aligned}$$

BIBLIOGRAPHY

[1] Aggarwal, J. K. *Multisensor Fusion for Computer Vision.* Springer-Verlag, New York, 1991.

[2] Augustyn, Kenneth. A new approach to automatic target recognition. *IEEE Trans. on Aerospace and Electronic Systems 28*, 1 (January 1992), 105–114.

[3] Barndorff-Nielsen, O., Kent, J., and Sorensen, M. Normal variance-mean mixtures and z distributions. *Int'l Statistical Rev. 50* (1982), 145–159.

[4] Belongie, S., Malik, J., and Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24*, 4 (April 2002), 509–522.

[5] Bhanu, B. Automatic target recognition: State of the art survey. *IEEE Trans. Aerospace Electron. Syst. AES-22*, 4 (July 1986), 364–379.

[6] Blahut, R. E. *Principles and Practice of Information Theory.* Addison-Wesley, New York, 1987.

[7] Bonabeau, E., Dorigo, M., and Theraulaz, G. *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press, 1999.

[8] Bone, E., and Bolkcom, C. Unmanned aerial vehicles: Background and issues for congress. pp. 1–52.

[9] Brunelli, R., and Falavigna, D. Person identification using multiple cues. *IEEE Trans. Pattern Analysis and Machine Intelligence 17*, 10 (October 1995), 955–966.

[10] Chapple, P. B., Bertilone, D. C., Caprari, R. S., and Newsam, G. N. Stochastic model-based processing for detection of small targets in non-gaussian natural imagery. *IEEE Trans. on Image Processing 10*, 4 (April 2001), 554–564.

[11] Conover, W. J. *Practical Nonparametric Statistics, Third Edition.* John Wiley and Sons, New York, 1999.

[12] Cover, T. M., and Thomas, J. A. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[13] Cui, J., and Sheikh, A.U.H. Outage probability of radio cellular systems using maximal ratio combining in the presence of multiple interferers. *Vehicular Technology Conference 2* (May 1998), 731–735.

[14] Dasgupta, P. Distributed automatic target recognition using multi-agent UAV swarms. In *Proc. of AAMAS* (Hakodate, Hokkaido, Japan, May 2006).

[15] deBoor, C. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

[16] Dierchx, P. *Curve and Surface Fitting with Splines*. Oxford Univ. Press, Oxford, U.K., 1993.

[17] Dunn, O. J., and Clark, V. A. *Applied Statistics: Analysis of Variance and Regression*. John Wiley and Sons, New York, 1974.

[18] Figueiredo, M.A.T., Leitao, J.M.N., and Jain, A.K. Unsupervised contour representation and estimation using b-splines and a minimum description length criterion. *IEEE Trans. on Image Processing 9*, 3 (May 2000), 1171–1182.

[19] Freund, J., and Perles, B. A new look at quartiles of undergrouped data. *American Stat. 41* (1987), 200–203.

[20] Gallager, R. G. *Information Theory and Reliable Communications*. Wiley & Sons, New York, 1968.

[21] Gaudiano, F., and Bonabeau, E. Control of UAV swarms: What the bugs can teach us. In *Proc. of 2nd AIAA Unmanned Unlimited* (2003), pp. 582–589.

[22] Grenander, U., Miller, M. I., and Srivastava, A. Hilbert-schmidt lower bounds for estimators on matrix lie groups for ATR. *IEEE Trans. on Pattern Analysis and Machine Intelligence 20*, 8 (August 1998), 790–802.

[23] Gyer, M. Automatic target recognition panel report. In *Report ET-TAR-002 Science Applications International Corp* (San Diego, CA, December 1988).

[24] http://www cvr.ai.uiuc.edu/ponce_grp/data/.

[25] http://www1.cs.columbia.edu/CAVE/software/softlib/coil 100.php.

[26] http://www.cs.ubc.ca/ murphyk/Vision/objectRecognitionDatabases.html.

[27] Jahne, B., Haubecker, H., and Geibler, P. vol. 2. Academic Press, San Diego, CA, 1999.

[28] Jain, A., Moulin, P., Miller, M. I., and Ramchandran, K. Information-theoretic bounds on target recognition performance based on degraded image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24*, 9 (September 2002), 1153–1166.

[29] Jain, A. K. *Fundamentals of digital image processing.* Prentice Hall, New Jersey, 1989.

[30] Jolliffe, I.T. *Principal Component Analysis.* Springer-Verlag, 1986.

[31] Kaplan, G., and Shamai, S. Error exponents and outage probabilities for the block-fading gaussian channel. *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (September 1991), 329–334.

[32] Kendall, M. *Multivariate Analysis.* Charles Griffin & Company Limited, 1975.

[33] Kerekes, R., Narayanaswamy, B., Beattie, M., Kumar, B. V. K. Vijaya, and Savvides, M. Multiframe distortion-tolerant correlation filtering for video sequences. *Proceeding of the SPIE 6245* (June 2006).

[34] Keren, D., Peleg, S., and Brada, R. Image sequence enhancement using sub-pixel displacements. In *IEEE Conference on Computer Vision and Pattern Recognition* (June 1988), pp. 742–746.

[35] Kottke, D., and Fiore, P. D. Systolic array for acceleration of template-based atr. *Proceedings of the International Conference on Image Processing 1* (1997), 869–872.

[36] Kullback, S. *Information Theory and Statistics.* Dover Publications, 1997.

[37] Lanterman, A. D., Grenander, U., and Miller, M. I. Bayesian segmentation via asymptotic partition functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence 22*, 4 (April 2000), 337–347.

[38] Li, F., and Wechsler, H. Open set face recognition using transduction. *Proceedings of the International Conference on Image Processing 27*, 11 (November 2005), 1686–1697.

[39] Lucas, B., and Kanade, T. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artifical Intelligence* (Auguest 1981), pp. 674–679.

[40] Miller, M. I., Grenander, U., O'Sullivan, J. A., and Snyder, D. L. Automatic target recognition organized via jump-diffusion algorithms. *IEEE Trans. on Image Processing 6*, 1 (January 1997), 157–174.

[41] Murphy, Robin R., and Taylor, Brent. A survey of machine learning techniques for automatic target recognition. *submitted to IEEE Trans. on Pattern Analysis and Machine Intelligence* (2002).

[42] O'Sullivan, J. A., Blahut, R. E., and Snyder, D. L. Information-theoretic image formation. *IEEE Trans. on Information Theory 44*, 6 (Oct 1998), 2094–2013.

[43] Patnaik, R., and Casasent, D. Mstar object classification and confuser and clutter rejection using minace filters. *Proceeding of the SPIE 6234* (June 2006).

[44] Ratches, James A., Walters, C. P., Buser, Rudolf G., and Guenther, B. D. Aided and automatic target recognition based upon sensory inputs from image forming systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence 19*, 9 (September 1997), 1004–1019.

[45] Roth, Michael W. Survey of neural network technology for automatic target recognition. *IEEE Trans. on Neural Networks 1*, 1 (March 1990), 28–43.

[46] Sadeque, A. Z., and Alam, M. S. Target detection in flir imagery using independent component analysis. *Proceeding of the SPIE 6234* (June 2006).

[47] Salembier, P., Brigger, P., Casas, J. R., and Pardas, M. Morphological operators for image and video compression. *IEEE Trans. on Image Processing 5*, 6 (June 1996), 881–898.

[48] Savakis, A. E., and Trussell, H. J. Blur identification by residual spectral matching. *IEEE Transactions on Image Processing 2* (April 1993), 141–151.

[49] Schapire, R., and Singer, Y. Improving boosting algorithms using confidence-rated predictions.

[50] Schmid, Natalia A., and O'Sullivan, J. A. Performance prediction methodology for biometric systems using a large deviations approach. *IEEE Trans. on Signal Processing, Supplement on Secure Media 52*, 10 (October 2004), 3036–3045.

[51] Shannon, C. E. Probability of error for optimal codes in a Gaussian channel. *Bell Syst. Tech. J. 38*, 3 (1959), 611–656.

[52] Shapiro, S., and Wilk, M. An analysis of variance test for normality. *Biometrika 52*, 3-4 (December 1965), 591–611.

[53] Shusterman, E., Miller, M. I., and Rimoldi, B. Rate-distortion theory applied to automatic object recognition systems. *IEEE Trans. on Information Theory 45*, 5 (August 2000), 1921–1927.

[54] Smith, S. W. *The Scientist and Engineer's Guide to Digital Signal Processing.* California Technical Publishing, 1997.

[55] Srivastava, A. Stochastic models for capturing image variability. *IEEE Signal Processing Magazine 19*, 5 (September 2002), 63–76.

[56] Srivastava, A., Grenander, U., Jensen, G. R., and Miller, M. I. Jump-diffusion markov processes on orthogonal groups for object pose estimation. *Journal of Statistical Planning and Inference 103* (2002), 15–37.

[57] Srivastava, A., Liu, X., and Grenander, U. Universal analytical forms for modeling image probabilities. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 9 (September 2002), 1200–1214.

[58] Srivastava, A., Miller, M. I., and Grenander, U. *Bayesian Automated Target Recognition.* Academic Press, New York, 2000.

[59] Third Generation Partnership Project 2 (3GPP2). Physical layer standard for cdma2000 spread spectrum systems, release C. pp. 115–122.

[60] Tralli, V., and Verdone, K. Performance characterization of digital transmission systems with cochannel interference. *IEEE Trans. on Vehicular Technology 48*, 3 (May 1999), 733–745.

[61] Trees, H. L. Van. *Detection, Estimation, and Modulation Theory, Part I.* Wiley & Sons, New York, 2001.

[62] Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. *Proceeding of Computer Vision and Pattern Recognition 2001 1* (2001), 511–518.

[63] Westover, M. B., and O'Sullivan, J. A. Towards an information theoretic framework for object recognition. In *International Symp. on Information Theory (ISIT)* (Chicago, Illinois, 2004), p. 219.

[64] Westover, M. B., and O'Sullivan, J. A. Achievable rates for pattern recognition: Binary and gaussian cases. In *International Symp. on Information Theory (ISIT)* (Adelaide, Australia, 2005), pp. 28–32.

[65] Xiao, J., and Shah, M. Automatic target recognition using multi-view morphing. *Proceeding of the SPIE 5426* (April 2004), 391–399.

[66] Young, S. S., Kwon, H., Der, S. Z., and Nasarabadi, N. M. Adaptive target detection in forward-looking infrared imagery using the eigenspace separation transform and principal component analysis. *Optical Engineering 43*, 8 (August 2004), 1767–1776.

[67] Zhu, S.Ch., and Yuille, A. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *IEEE Trans. On Pattern Analysis and Machine Intelligence 18*, 9 (1996), 884–900.