DISTRIBUTED SOURCE CODING

THEORY AND PRACTICE

SHUANG WANG
YONG FANG
SAMUEL CHENG

with website

WILEY

Distributed Source Coding

Theory and Practice

Shuang Wang, Yong Fang, and Samuel Cheng

WILEY

This edition first published 2017

Registered office John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication data applied for

ISBN: 9780470688991

A catalogue record for this book is available from the British Library.

Cover design by Wiley
Cover image: Top numbers image: a_Taiga/Gettyimages
Bottom molecular image: KTSDESIGN/SCIENCE PHOTO LIBRARY/Gettyimages

Set in 10/12pt WarnockPro by SPi Global, Chennai, India
Printed and bound in Malaysia by Vivar Printing Sdn Bhd

10 9 8 7 6 5 4 3 2 1

Preface

This book intends to provide fundamental knowledge for engineers and computer scientists to get into the topic of distributed source coding, and is targeted at senior undergraduate or first-year graduate students. It should also be accessible for engineers with a basic knowledge of calculus and applied probability. We have included short chapters on information theory, channel coding, and approximate inference in the appendix. The goal is to make the content as self-contained as possible.

Excluding the appendix, the book is divided into three parts. First-time readers who would like to get into applications quickly should work through Chapters 1, 2, 3, 5, and 7, and then go straight to Part III, the application chapters. Chapters 4 and 6 are more advanced and may require more math maturity from readers.

Distributed source coding is an active research area and the authors have not been able to cover every aspect of it. We have included materials that are geared more toward our own research interests. However, we have striven to include the most interesting applications in this book.

Acknowledgment

Book writing is a long endeavor. It is especially so in this case. This book project grew out of several distributed source coding tutorials presented in the US, Taiwan and Switzerland around 2009 by Dr. Vladimir Stankovic, Dr. Lina Stankovic, and the third author. Many materials in this book reflect the immense influence of these tutorials. Drs. Stankovic have been very kind in offering extensive unconditional help and allowed us to include many of their ideas and even their own words in this book. For this we wish to express our most sincere and deepest gratitude.

The third author would also like to express special thanks to his Master and PhD advisor, Dr. Zixiang Xiong, who taught him not just distributed source coding, but also everything relating to research and academia.

Over the years, the authors have received help from many colleagues, friends, and students, including Angelos Liveris, Costas Georghiades, Yang Yang, Zixin Liu, Zongmin Liu, Qian Xu, Shengjie Zhao, Nikoloas Deligiannis, Xiaoqian Jiang, Adrian Munteanu, Peter Schelkens, Andrei Sechelea, Feng Chen, and Lijuan Cui. In particular, we would like to offer special thanks to Liyan Chen and Xiaomin Wang for generating some of the figures. We could not possibly include everyone in the list above, and we would like to ask for forgiveness for any omission. Finally, the authors would like to acknowledge support from the National Science Foundation, the National Institutes of Health (NHGRI K99HG008175), and the National Nature Science Foundation of China.

About the Companion Website

Don't forget to visit the companion website for this book:

www.wiley.com/go/cheng/dsctheoryandpractice

There you will find valuable material designed to enhance your learning, including:

- Example codes

Scan this QR code to visit the companion website

Contents

1

Introduction

Imagine a dense sensor network consisting of many tiny sensors deployed for information gathering. Readings from neighboring sensors will often be highly correlated. This correlation can be exploited to significantly reduce the amount of information that each sensor needs to send to the base station, thus reducing power consumption and prolonging the life of the nodes and the network. The obvious way of exploiting correlation is to enable neighboring sensors to exchange data with one another. However, communication among sensors is usually undesirable as it increases the complexity of the sensors, which in turn leads to additional cost and power consumption. How then is it possible to avoid information exchange among sensors but still be able to exploit the statistical dependency of the readings in different sensor nodes? The solution lies in *distributed source coding*.

Distributed source coding (DSC) enables correlation to be exploited efficiently without the need for communications among sensors. Moreover, in some specific cases it does not incur any loss compared to the case when the sensors communicate.

DSC has been receiving significant attention since the beginning of the 21st century from academics and industrial researchers in different fields of electrical and computer engineering, and mathematical and computer science. Indeed, special sessions and tutorials dedicated to DSC are given at major communications, signal processing, multimedia, and computer engineering conferences. This is no wonder as DSC has potential applications ranging from wireless sensor networks, ad-hoc networks, surveillance networks, to robust low-complexity video coding, stereo/multiview video coding, high-definition television, and hyper-spectral and multi-spectral imaging. This book is intended to act as a guidebook for engineers and researchers to grasp the basic concepts quickly in order to understand and contribute to this exciting field and apply the emerging applications.

Distributed Source Coding: Theory and Practice, First Edition. Shuang Wang, Yong Fang, and Samuel Cheng.

Companion Website: www.wiley.com/go/cheng/dsctheoryandpractice

1.1 What is Distributed Source Coding?

DSC deals with the source coding or compression of correlated sources. The adjective *distributed* stresses that the compression occurs in a distributed or noncentralized fashion. We can, for example, assume that the sources to be compressed are distributed across different nodes in a network. The task is to compress these sources and communicate compressed streams to a decoder for joint decompression. The basis of DSC is that the compressions take place *independently*, that is, the nodes do not exchange their information, whereas decompression is *joint*.

This is illustrated in Figure 1.1 for the simple case of two nodes. Each node has access only to its source, and does not have information about sources present at other nodes. Therefore, "distributed" in this context refers to separate compression at each node. Note that if decompression were also separate for each of the sources, then the problem would boil down to multiple conventional compressions. Throughout the book, distributed compression will always refer to separate encoding and joint decoding, whereas joint compression will refer to joint encoding and joint decoding, if not stated otherwise.

The search of achievable rates of DSC is a major information-theoretical problem and lies in the framework of network information theory, a branch of information theory that tries to find the compression and communication limits of a network of nodes. In the case of discrete sources and perfect reconstruction at the decoder, DSC extends Shannon's Source Coding Theorem, in information theory, from the point-to-point to multipoint scenario. This is referred to as lossless DCS. When we allow for some distortion in reconstruction, the DSC problem becomes a rate-distortion problem and is referred to as lossy DSC.

1.2 Historical Overview and Background

DSC started as an information-theoretical problem in the seminal 1973 paper of Slepian and Wolf [1]. Slepian and Wolf considered lossless separate compression of two discrete sources, and showed that roughly speaking there is no

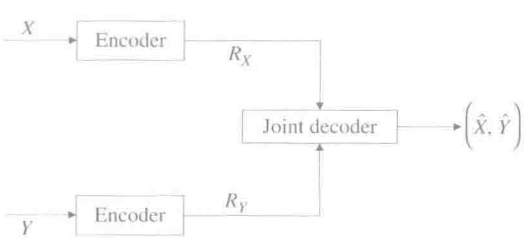

Figure 1.1 DSC concept with two separate encoders who do not talk to each other and one joint decoder. X and Y are discrete, correlated sources; R_X and R_Y are compression rates.

performance loss compared to joint compression as long as joint decompression is performed.

This remarkable result triggered significant information-theoretical research resulting in solutions – in the form of achievable rate regions – for more complicated lossless setups. In 1976, Wyner and Ziv [2] considered a lossy version, with a distortion constraint, of a special case of the Slepian–Wolf (SW) problem, where one source is available at the decoder as side information. Wyner and Ziv showed that for a particular correlation where source and side information are jointly Gaussian, there is no performance loss due to the absence of side information at the encoder. The lossy case of the generalized SW setup, known as multiterminal (MT) source coding, was introduced by Berger and Tung in 1977 [3, 4].

A possible realization of DSC via the use of conventional linear channel codes to approach the SW bound was known as early as 1973, but due to the lack of any potential application of DSC, work on code designs, that is, how to code the sources to approach given bounds, started only at the end of the last century. The first practical design was reported in 1999 [5], followed by many improved solutions. One key insight of these designs is that conventional channel coding can be used for compression. Indeed, correlation between the sources is seen as a virtual communication channel, and as long as this virtual channel can be modeled by some standard communication channel, for example Gaussian, channel codes can be effectively employed. Capacity-approaching designs [6] based on quantization followed by advanced channel coding, for example with turbo codes [7] and low-density parity-check (LDPC) codes [8], come very close to the bounds for two jointly Gaussian sources.

1.3 Potential and Applications

The launch of wireless sensor networks (WSNs) ignited practical DSC considerations in the early years of this century since WSNs naturally call for distributed processing. Closely located sensors are expected to have correlated measurements; thus in theory the DSC setup fulfills the requirement of power-efficient compression for distributed sensor networks. However, many practical problems remain to be solved before DSC is used in mainstream commercial networks. The challenges include the complex correlation structure of real signals, nonGaussian sources, mandatory long codeword lengths, and the complexity of current designs.

Though WSN triggered renewed interest in DSC, another application has emerged: low-complexity video, where the DSC paradigm is used to avoid computationally expensive temporal prediction loop in video encoding. Indeed, loosely speaking, a conventional video encoder needs to find the best matching block to the current one by examining all possible candidates, for