MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
http://www.comsoc.org/~mmc

MMTC Communications - Frontiers

Vol. 17, No. 4, July 2022

CONTENTS

SPECIAL ISSUE ON Intelligent Communication and Computation Technologies in 6G

Guest Editor: Long Shi, Zhe Wang, and Guanyu Gao
Nanjing University of Science and Technology, Nanjing, China
Email: slong1007@gmail.com, zwang@njust.edu.cn, gygao@njust.edu.cn

Advances of edge intelligence have accelerated the explosion of new research frontiers in 5G and upcoming 6G era (B5G). Using sophisticated learning and big data analytic tools at the edge side, various communication and computation technologies have been developed to solve complicated problems of allocation, optimization, and control across different layers of B5G networks. In return, unique characteristics of B5G networks (e.g., interference, channel conditions, dynamics, bandwidth, etc.) bring a volume of open challenges for the designs of intelligent communication and computation technologies. The aim of this Special Issue is to solicit articles addressing numerous topics within its scope.

The first paper introduced the goal-oriented principle for communication systems redesign to provide efficient communication links at the physical layer for achieving wireless collaborative intelligence. The paper focused on the application of federated learning (FL) and demonstrated how the goal-oriented principle can be leveraged to solve the bottleneck of uplink model aggregation in FL.

The second paper proposed a multi-model deep learning approach to fuse the features of both point clouds and range-Doppler for classifying six activities, i.e., boxing, jumping, squatting, walking, circling and high-knee lifting, based on a millimeter wave radar. The authors adopted a CNN-LSTM model to extract the time-serial features from point clouds and a CNN model to obtain the features from range-Doppler. The proposed method has higher accuracy than utilizing the two kinds of information separately and achieves a recognition accuracy of 97.26%.

The third paper considered the problem of L-channel multiple descriptions and proposed distortion criteria among reconstructions within the same level. The authors showed the new coding strategy strictly improves the VKG scheme. The authors also discussed the special case of zero distortions among reconstructions, in which the new inner bound coincides with the rate-distortion region of SMDC.

Long Shi is currently a Professor at the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing. He received the Ph.D. degree in Electrical Engineering from The University of New South Wales, Sydney, Australia, in 2012. His research interests include wireless communications, mobile edge computing, and blockchain networks.

Zhe Wang is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. She received the Ph.D. degree in electrical engineering from The University of New South Wales, Sydney, Australia, in 2014. Her research interests include applications of optimization, game theory, and machine learning to resource allocation in communications and networking.

Guanyu Gao is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing. He received the Ph.D. degree from Nanyang Technological University, Singapore, in 2017. His research interests include multimedia networking, edge/cloud computing, resource scheduling, and deep reinforcement learning.

Goal-oriented Communication System Redesign for Wireless Collaborative Intelligence

Yulin Shao
Imperial College London
y.shao@imperial.ac.uk

Abstract – The communication requirements of wireless collaborative intelligence (WCI) pose great challenges to existing communication systems. This paper introduces the goal-oriented principle for communication systems redesign to provide efficient communication links at the physical layer for achieving WCI. In particular, we focus on the application of federated learning (FL) and demonstrate how the goal-oriented principle can be leveraged to solve the bottleneck of uplink model aggregation in FL.

Keywords: Wireless collaborative intelligence, 6G, federated learning, over-the-air computation.

1. Introduction

Wireless collaborative intelligence (WCI) aims to address the problem that how a team of wirelessly connected agents can collaborate to achieve joint intelligence based on their sensed data from the environment. The joint intelligence can be a collaborative machine learning (ML) model that transforms the sensed data into other representations for particular tasks [1], [2]. Applications include federated learning (FL) and distributed inference, wherein a group of agents targets to attain a global ML model. The joint intelligence can also be an autonomous protocol among agents [3], [4], whereby they collaborate for a common goal. Applications include vehicular platooning and autonomous channel access, in which the wireless network itself is the intelligence.

A main challenge of WCI is wireless communications, owing to the scarcity, unreliability, and capacity limitation of wireless channels. The collaboration of agents, in contrast, often requires frequent, reliable, low-latency, and large-volume data exchanges [1-2], which are heavy burdens for wireless links. To meet the challenge, a promising approach is goal-oriented communication system redesign at both the physical layer [2], [5-8] and higher layers [9-12] for achieving WCI. In this paper, we consider the application of FL and demonstrate an efficient physical redesign via over-the-air computation (OAC) for gradient aggregation in practical mobile edge networks.

2. System model

FL is a distributed learning algorithm to enable the collaborative learning of a group of M devices without sharing private data. In its basic implementation, orchestrated by a parameter server (PS), each iteration of FL consists of five main steps [1], as illustrated in Fig. 1.

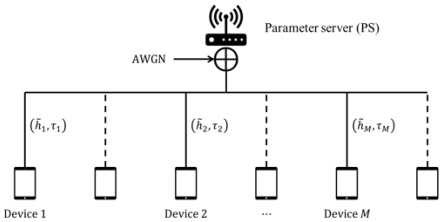


Fig. 1. In FL, edge devices collaboratively train a shared model with the help of a PS.

1) DL broadcast: at the beginning of the iteration, the PS broadcasts the global model $\boldsymbol{\theta} \in \mathcal{R}^d$ to the M edge devices; 2) Local training: each of the M devices trains the global model $\boldsymbol{\theta}$ on its local dataset $\boldsymbol{\mathcal{B}}_m$ of size B_m and obtains a new model $\widetilde{\boldsymbol{\theta}}_{\boldsymbol{m}} \in \mathcal{R}^d$; 3) UL aggregation: each device scales the local model update $\widetilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}$ by B_m and transmits the scaled model update $\boldsymbol{\theta}_{\boldsymbol{m}}^{'} = B_m\left(\widetilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\right)$ back to the PS; 4) Arithmetic-sum estimation: the PS estimates the arithmetic sum of the transmitted model-updates $\boldsymbol{\theta}_{\boldsymbol{m}}^{'}$ from the edge devices: $\boldsymbol{\theta}_+ = \sum_{m=1}^{M} \boldsymbol{\theta}_{\boldsymbol{m}}^{'}$; 5) Model update: the PS updates the global model by $\boldsymbol{\theta}_{\mathbf{new}} = \boldsymbol{\theta} + \frac{1}{\sum_m B_m}\boldsymbol{\theta}_+$. The updated global model $\boldsymbol{\theta}_{\mathbf{new}}$ is then broadcasted in the next iteration and the cycle continues.

Among the above five steps, the uplink model aggregation requires each device to transmit d real numbers to the PS, where d can run into hundreds of millions or more for today's neural networks [13]. This poses the greatest challenge to traditional radio access networks with orthogonal multiple access techniques. In this context, we resort to OAC for a goal-oriented solution.

The essence of OAC is to create and leverage inter-user interferences over the multiple-access channel (MAC) rather than trying to avoid it. When operated with OAC, edge devices transmit their raw model updates $\boldsymbol{\theta}_{\boldsymbol{m}}^{'}$ simultaneously to the PS in a discrete-time analog fashion. The PS, on the other hand, estimates the sum of the model updates $\boldsymbol{\theta}_+$ directly from the received overlapped signal.

The successful operation of OAC hinges on accurate channel-gain precoding and strict synchronization among the participating devices. In practice, however, both requirements may not be perfectly fulfilled. With the residual channel gains and residual asynchronies in the system, which we refer to as the misaligned OAC, an open problem is how to estimate the arithmetic sum of the transmitted symbols from different devices. In this paper, we fill this gap.

The signal flow of the misaligned OAC is detailed as follows. Each device partitions its sequence of scaled model update $\boldsymbol{\theta}_{\boldsymbol{m}}^{'}$ to two subsequences $\boldsymbol{\theta}_{\boldsymbol{m}}^{'} = \left[(\boldsymbol{s}_{\boldsymbol{m}}^{\boldsymbol{r}})^{\top}, (\boldsymbol{s}_{\boldsymbol{m}}^{\boldsymbol{i}})^{\top}\right]^{\top}$, where $\boldsymbol{s}_{\boldsymbol{m}}^{\boldsymbol{r}}, \boldsymbol{s}_{\boldsymbol{m}}^{\boldsymbol{i}} \in \boldsymbol{\mathcal{R}}^{\frac{d}{2}}$, and constructs a complex sequence $\boldsymbol{s}_{\boldsymbol{m}} \in \mathcal{C}^{\frac{d}{2}}$: $\boldsymbol{s}_{\boldsymbol{m}} = \boldsymbol{s}_{\boldsymbol{m}}^{\boldsymbol{r}} + j\boldsymbol{s}_{\boldsymbol{m}}^{\boldsymbol{i}}$. That is, the raw model update information $\boldsymbol{\theta}_{\boldsymbol{m}}^{'}$ is carried on both the real and imaginary parts of $\boldsymbol{s}_{\boldsymbol{m}}$. Time is divided into slots, and each device transmits a packet of L symbols in each slot. Without loss of generality, we focus on signal processing in one slot. The time-domain signal transmitted by the m-th device in one slot is given by

$$x_m(t) = \alpha_m \sum_{\ell=1}^{L} s_m[\ell] p(t - \ell T), \#(1)$$

where $p(t)$ is a rectangular pulse of duration T and α_m is the channel precoding factor. Given an estimated channel coefficient $\bar{h}_m$ at the m\$-th device, α_m is designed to be $\alpha_m = 1/\bar{h}_m$.Each of the M edge devices then calibrates the transmission timing, based on its distance from the PS and its moving speed, so that the signals from different devices arrive at the PS simultaneously.

In practice, however, both the channel-gain precoding and transmission-timing calibration can be imperfect due to the non-ideal hardware and inaccurate estimation of the channel gains and transmission delays. The received signal $r(t)$ at the PS can be written as

$$r(t) = \sum_{m=1}^{M} \tilde{h}_m x_m(t - \tau_m) + z(t), \#(2)$$

where 1) $\tilde{h}_m$ is the time-domain complex channel response and we consider flat and slow fading; 2) without loss of generality, we sort the M devices so that the symbols from the devices with smaller indexes arrive at the receiver earlier. The delay of the first device is set to $\tau_1 = 0$ and the relative delay of the m-th device with respect to the first device is denoted by $\tau_m < T$; 3) $z(t)$ is the zero-mean baseband complex additive white Gaussian noise (AWGN), whose double-sided power spectral density is N_0. Substituting (2) into (1) gives us

$$r(t) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} h_m' s_m[\ell] p(t - \tau_m - \ell T) + z(t), \#(3)$$

where $h_m' = \tilde{h}_m / \bar{h}_m$ is the residual channel-fading coefficient between the m-th device and the PS. The objective of the PS is to estimate the arithmetic sum of the local model updates $\boldsymbol{\theta}_+$. This is equivalent to estimating the arithmetic sum of the transmitted complex symbols $\boldsymbol{s}_+ \triangleq \sum_{m=1}^{M} s_m[i]$. Therefore, we shall focus on the estimation of the complex symbols $\boldsymbol{s}_+$ in the following.

3. Aligned and misaligned OAC

Prior works on OAC considered only the perfectly aligned case, where there is neither channel-gain misalignment nor time misalignment, which we refer to as the aligned OAC. In this case, we have $\alpha_m = 1/\tilde{h}_m$ and $\tau_m = 0$, $\forall m$, and the received signal is given by

$$r(t) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} s_m[\ell] p(t - \ell T) + z(t) \#(4)$$

Matched filtering $r(t)$ by $p(t)$ and sampling at $t = i\mathrm{T}$, $i = 1,2, ..., L$ give us

$$r[i] = \frac{1}{T} \int_{(i-1)T}^{i\mathrm{T}} r(t) dt = \sum_{m=1}^{M} s_m[i] + z[i] = s_+[i] + z[i] \#(5)$$

where the noise sequence $z[i]$ in the samples is independent and identically distributed (i.i.d.), $z[i] \sim \mathcal{CN}\left(0, \frac{N_0}{T}\right)$. As can be seen, the target signal $s_+[i]$ appears explicitly on the right-hand side (RHS) of (5). In this context, the fading MAC degenerates to a Gaussian MAC and the M devices can be abstracted as a single device transmitting the arithmetic sum of the local model updates directly to the PS. In practice, however, the channel-gain precoding and the calibration of transmission timing can be inaccurate. With either channel-gain or time misalignment, clean samples as in (5) with $s_+[i]$ explicitly present are no longer available.

Considering both channel-gain and time misalignments, we first matched filter $r(t)$ by $p(t)$ and then oversample the matched filtered signal at $\{iT + \tau_k : i = 1,2, ..., L, k = 1,2, ..., M\}$ to collect sufficient statistics. In so doing, the samples we get, denoted by $\{r_k[i]\}$ can be written as

$$r_k[i] = \sum_{m=1}^{M} c_{m,k}[i] s_m[i - \mathbb{I}_{m>k}] + \sum_{m=1}^{M} c'_{m,k}[i] s_m[i + \mathbb{I}_{m<k}] + z_k[i] \#(6)$$

where $\mathbb{I}$ is the indicator function and $c_{m,k}[i] = \frac{h'_m}{T}[(1 - \mathbb{I}_{m>k})T + \tau_m - \tau_k]$, $c'_{m,k}[i] = \frac{h'_m}{T}[\mathbb{I}_{m>k}T + \tau_k - \tau_m]$. Given the samples $\{r_k[i]\}$, we now set out to estimate the desired arithmetic sum $\boldsymbol{s}_+$. First, the sequence of samples $y_k[i]$ can be written in a more compact form as

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{z}, \#(7)$$

where the dimensionalities of $\boldsymbol{r}$, $\boldsymbol{s}$, and $\boldsymbol{z}$ are $ML \times 1$, and the dimensionality of matrix $\boldsymbol{A}$ is $ML \times ML$. The desired sequence $\boldsymbol{s}_+$, on the other hand, can be written as a linear transformation of the complex vector $\boldsymbol{s}$: $\boldsymbol{s}_+ = \boldsymbol{V}\boldsymbol{s}$, where the $L \times ML$ matrix $\boldsymbol{V}$ is given by

$$\boldsymbol{V} = \begin{bmatrix} \mathbf{1}_{1\times M} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{1\times M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}_{1\times M} \end{bmatrix},$$

in which $\mathbf{1}_{1\times M}$ represents a $1 \times M$ all-ones matrix.

Multiplying both sides of (7) by $\boldsymbol{V}\boldsymbol{A}^{-1}$ gives us

$$\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{r} = \boldsymbol{s}_+ + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{z}, \#(8)$$

based on which an ML estimator can be devised, as in Definition 1.

Definition 1 (ML estimation for misaligned OAC). Given a sequence of samples $\boldsymbol{r} \in \mathcal{C}^{ML}$, the ML estimate of sequence $\boldsymbol{s}_+ \in \mathcal{C}^{L}$ is $\hat{\boldsymbol{s}}_+^{\boldsymbol{ml}} = \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{r}$.

The ML estimator follows directly from (8) since the likelihood function of $\boldsymbol{s}_+$ is an L-dimensional Gaussian distribution. Specifically, given $\boldsymbol{r}$, the likelihood function of $\boldsymbol{s}_+$ is

$$f(\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{r}|\boldsymbol{s}_+) \propto \mathcal{CN}(\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{r}, \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{z}}\boldsymbol{A}^{-\boldsymbol{H}}\boldsymbol{V}^{\boldsymbol{H}})$$

Differentiating $f(\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{r}|\boldsymbol{s}_+)$ with respect to $\boldsymbol{s}_+$ gives us the ML estimate in Definition 1.

It is worth noting that the ML estimator in Definition 1 has high computational complexity due to matrix inversion. To address this problem, we further put forth in [?] a factor-graph-based ML estimator by exploiting the sparsity of the coefficient matrix.

4. Conclusion

This paper introduces the goal-oriented principle to design an efficient communication physical layer for achieving WCI. In the application of FL, OAC, as a goal-oriented scheme, exploits the property of the MAC that its output is the arithmetic sum of the inputs, and jointly performs computation and communication, thereby significantly speeding up the uplink aggregation in FL. In a broader context, the goal-oriented principle can be used in both the physical and higher layers to refine the legacy communication system design for supporting WCI.

5. Reference

[1] J. Konecny, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon, "Federated learning: strategies for improving communication efficiency," arXiv:1610.05492, 2016.

[2] Y. Shao, D. Gunduz and S. Liew. "Federated edge learning with misaligned over-the-air computation," IEEE Trans. Wireless Commun., vol. 21, no. 6, pp. 3951-3964, 2022.

[3] X. Cheng, C, Chen, W. Zhang, Y. Yang, "5G-enabled cooperative intelligent vehicular framework: When Benz meets Marconi", IEEE Intelligent Systems, vol. 32, no. 3, pp. 53-59, 2017.

[4] Y. Shao, Y. Cai, T. Wang, D. Gunduz. "Learning-based autonomous channel access in the presence of hidden terminals", arXiv:2207.03605, 2022.

[5] Y. Shao, D. Gunduz and S. Liew. "Bayesian over-the-air computation," arXiv:2109.03780, 2022.

[6] M. Kalfa, M. Gok, A. Atalik, B. Tegin, T. Duman, O. Orhan, "Towards goal-oriented semantic signal processing: Applications and future challenges", Digital Signal Processing, 2021.

[7] Y. Shao, S. Liew and D. Gunduz. "Denoising noisy neural networks: A Bayesian approach with compensation," arXiv:2105.10699, 2022.

[8] Y. Shao, S. Liew, and L. Lu. "Asynchronous physical-layer network coding: symbol misalignment estimation and its effect on decoding," IEEE Trans. Wireless Commun., vol. 16, no. 10, pp. 6881–6894, 2017.

[9] E. Strinati, S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications", Computer Networks, 2021

[10] Y. Shao, A. Rezaee, S. Liew, and V. Chan. "Significant sampling for shortest path routing: a deep reinforcement learning solution," IEEE J. Sel. Areas Commun., vol. 38, no. 10, pp.2234–2248, 2020.

[11] Y. Shao, S. Liew, and J. Liang. "Sporadic ultra-time-critical crowd messaging in V2X," IEEE Trans. Commun., vol. 69, no. 2, pp. 817-830, 2020.

[12] Y. Shao, Q, Cao, S. Liew, and H. Chen. "Partially observable minimum-age scheduling: the greedy policy," IEEE Trans. Commun., vol. 70, no. 1, pp. 404-418, 2021.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in IEEE CVPR, 2016, pp. 770–778.

Activity Recognition based on Millimeter Wave Fusing Point Cloud and Range-Doppler Information

*Yuchen Huang, Wei Li, Zhiyang Dou, Wangtong Zou, Anye Zhang, Zan Li**
College of Communication Engineering, Jilin University, Changchun, China
Email:zanli@ jlu.edu.cn

Abstract: Millimeter wave radar has demonstrated its high efficiency in complex environments in recent years, which outperforms LiDAR and computer vision in human activity recognition in the presence of smoke, fog and dust. In previous studies, researchers mostly analyze either 2D (3D) point cloud or range-Doppler information from radar echo to extract activity features. In this paper, we propose a multi-model deep learning approach to fuse the features of both point clouds and range-Doppler for classifying six activities, i.e., boxing, jumping, squatting, walking, circling and high-knee lifting, based on a millimeter wave radar. We adopt a CNN-LSTM model to extract the time-serial features from point clouds and a CNN model to obtain the features from range-Doppler. Then we fuse the two features and input the fused feature into the full con-nected layer for classification. We build a dataset based on a 3D millimeter wave radar from 17 volunteers. The evaluation result based on the dataset shows that this method has higher accu-racy than utilizing the two kinds of information separately and achieves a recognition accuracy of 97.26%, which is about 1% higher than other networks with only one kind of data as input.

Index Terms: Human activity recognition, millimeter wave radar, deep learning, feature fusion, point cloud, range-Doppler.

1. Introduction

With the fast development of communication networks, such as 5G and WiFi net-work, millimeter wave has emerged as a key technology to achieve high data rate trans-mission. With the increase of frequency, the perception ability of wireless signal improves. In the future, wireless signals will not only have communication ability, but also have sensing ability, such as Millimeter wave in 5G and Terahertz in 6G. Compared with the traditional communication technologies, millimeter wave has the capability of sensing. Hence, wireless sensing based on millimeter wave has attracted more research attentions. Human activity recognition is widely investigated in the field of ubiquitous sensing based on inertial sensors, computer vision or LiDAR. Compared with those traditional approaches, human activity recognition based on millimeter wave radar has many unique advantages. Firstly, compared with LiDAR, millimeter wave radar has better penetration ability to floating particles such as smoke and dust. This advantage means that the process of activity recognition of millimeter wave radar has strong environmental adaptability. Secondly, because the millimeter wave radar uses radio frequency signals, it will not return the privacy screen of the target person, which means that the radar monitoring will not reveal personal identity information. This edge enables indoor activity recognition based on millimeter wave radar more widely applicable, such as nursing homes, kindergartens, wards, etc.

In previous studies, researchers generally used to collect data such as point cloud, range-Doppler, echo intensity from millimeter-wave radar echo signals for activity recognition. For instance, RadHAR [1] shows how to voxelize sparse and non-uniform point clouds and feed it into a classifier. Additionally, in [2], since the single-sensor system can only observe the radial component of the micro-Doppler signal, researchers propose a multi-static sensor consisting of two bistatic micro-Doppler sensors to enhance the classification of micro-Doppler signatures to improve activity recognition accuracy. In [3], also using Doppler information, the researchers propose an environmental impact mitigation method by analyzing millimeter wave signals in Doppler-Range domain. In terms of extracting activity features, deep learning is favored by researchers [4], among which Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM) are widely used. Because most human actions have time-series characteristics, these two network models are often combined. For example, some researchers [5] adopt LSTM+CNN network to classify activities.

However, most of these studies only focus on a single type of data which cannot comprehensively depict the activity feature. The approaches based on point clouds merely leverage the distributions of point clouds in 3D space but do not consider the moving speed. In contrast, the methods using range-Doppler does not take the distributions of point clouds into consideration. Our goal is to design a multi-model network by fusing multiple types of input data to increase the recognition accuracy based on mmWave. We are going to extract the features from the two kinds of data, fuse them, and then classify the activities.

In this work, we propose a new method for activity recognition that uses both 3D point cloud data and range-Doppler data as input. We propose a multi-model deep learning model to fuse the point cloud and range-Doppler to achieve activity recognition based on millimeter wave radar. We voxelize and merge the 3D point cloud data in multiple frames into a 4D array. We further merge and convert the range-Doppler data in the same number of frames into a 3D array. Afterwards, we

use two sub-networks to train these two kinds of data separately. The 3D CNN+LSTM network is ap-plied for the point cloud data while the 3D CNN network is used for the range-Doppler data. Next, we fuse the outputs of the two networks, that is, the activity features contained in each kind of data, to obtain a more comprehensive activity feature than utilizing only one kind of data. Finally, the fused features are input into the fully connected layer for classification. Finally, experimental results show that the classification using the fused features of these two kinds of data is about 1% more accurate than merely using point cloud or range-Doppler, which shows that the use of multiple types of information has a positive effect on the recognition accuracy.

2. Activity recognition based on fusion of point cloud and range-doppler

Data Pre-processing

Since the millimeter-wave radar has a wide coverage area, and the distance between the collected person and the radar is relatively fixed, we consider removing some unnecessary data. This is done to prevent the excessive invalid data which reduce the speed and accuracy of recognition. Hence, we filter the data whose range values are between 1m and 2m. In the following part, we explain how to pre-process 3D point cloud data and range-Doppler data respectively.

Considering the irregularity of the number of 3D point cloud data in each frame, we must convert it into a data format with a fixed size and shape, so that it can be effectively input into a convolutional neural network (CNN). Therefore, we first establish a three-dimensional discrete coordinate system. Then we calculate the scale of the x, y, z axes for each frame, in which the two ends of the coordinate axis are the minimum and maximum value of the point cloud data of the axis. Then we create a 10×32×32 array, and fill the point cloud coordinates of each frame into the corresponding position of the array. It is worth mentioning that we do not directly filled in the coordinate value but count the number of points located in each array. This is to avoid having two different coordinates filled in the same position. Thus, we get a three-dimensional array of size 10×32×32 that represents the distribution of the point cloud in this frame. However, when the sampling rate is eight frames per second, a single frame of data cannot reflect the timing characteristics between frames. Therefore, we developed a fourth dimension based on the three-dimensional data, which represents the time series, and merged the eight frames of data to obtain a 4D array with a size of 8×10×32×32. So far, the pre-processing of 3D point cloud data is over.

On the other hand, the pre-processing of range-Doppler data is relatively simple. The range-Doppler data is presented in a picture-like form. Its abscissa is the range, the ordinate is the Doppler speed, and the value in the picture is the thermal value. Hence, we export the range-Doppler data from each frame separately and get a 2D array of size 32×32. Then, to synchronize with 3D point cloud, we merge the range-Doppler data in eight frames, and finally get a 3D array with a size of 8×32×32.

Using only 3D point cloud data as input

Considering that the activities are continuous, there is a timing characteristic between the received frames. Hence, we use a 3D CNN+LSTM network for training and classification. Because the acquisition rate is eight frames per second, and the duration of each activity is about one second, we pre-process the 3D point cloud data into a four-dimensional array of size 8×10×32×32, which includes the merging of 3D point cloud data received in one second. Because the input of LSTM must contain the parameter timesteps, we use the 'TimeDistributed' layer wrapper to encapsulate the convolutional layer, pooling layer and flatten layer, adding a time dimension to the input data. As mentioned above, we set the value on this dimension to 8.

In the convolution part, we use the 'Conv3D-Conv3D-Maxpooling' structure three times. The parameter settings are the same with the first network. Then the extracted features are flattened and input into the bidirectional LSTM layer, whose number of hidden layer units is 16. Finally, a fully connected layer is used for classification after regularization, whose activation function is Softmax.

Using only range-Doppler data as input

Because a piece of range-Doppler data is actually a superposition of multiple images in the time dimension, we try to use a 3D CNN to train the range-Doppler data. The network is similar to the convolutional part of the point cloud network. We used the structure of 'Conv3D-Conv3D-Maxpooling' three times, where the parameter settings are consistent with the network in Part B, without the 'TimeDistributed' layer wrapper encapsulating the layers. Then we input the features to the flatten layer and the fully connected layer for classification successively. The network takes eight frames of range Doppler data as a set of input data, that is, the data with the shape of 8 × 32 × 32 × 1 can be regarded as a three-dimensional single channel stereo image, so the network can be understood as a classification network of three-dimensional images. But different from the ordinary image classification network, the input data has not only spatial characteristics, but also temporal

characteristics. The temporal characteristics can be understood as the series between frames, that is, the characteristics reflected by Doppler velocity. This network can better learn the characteristics of dynamic actions, rather than only extracting static two-dimensional features.

Using both 3D point cloud data & range-Doppler data as input

We know that the more types of information describing an object, the higher the classification accuracy and loss will be. Multimodal fusion is based on this principle [6]. In our study, there are various kinds of information describing an activity, such as point cloud information, echo intensity information, range-Doppler information, and so on. This information describes an activity from different aspects, and the information contains different characteristics of the activity. If we extract these features and fuse them effectively, we can get more comprehensive features than using one kind of information alone. Therefore, we propose to fuse the features contained in the 3D point cloud information and the range-Doppler information. Of course, fusing more kinds of features will get a more comprehensive description of the activity, but the difficulty of their fusion will also increase accordingly, so this paper only fuses the features of point cloud information and range-Doppler information.

Taking the second network of part B and the first network of part C as an example, the network structure is shown in Fig.1. We parallelize the networks of part B and C, and input the 3D point cloud data and range-Doppler data into these two networks respectively. The input of the fully connected layers of the two networks is the features contained in the two kinds of information, so we output the results of the previous layer of the fully connected layer separately, and found that they are all two-dimensional, and the first dimension is the amount of data. Hence, we feed these two features into the concatenate layer, merge them in the second dimension, and get a feature that contains two different features. Finally, we regularize this feature and feed it into a fully connected layer for classification.

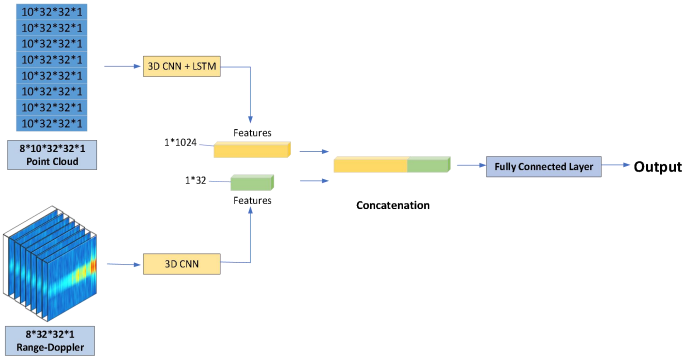


Figure 1. The structure of the parallel network.

3. *Experiment and Results*

Data Collecting and Experiment Setup

We adopt IWR1843 which is produced by Texas Instruments (TI) for data collecting. The IWR1843 device is an integrated single-chip millimeter wave sensor based on FMCW radar technology, which can operate in the 76GHz to 81GHz band. IWR1843 includes three transmitting antennas and four receiving antennas, ADC converters, a DSP subsystem [7].

After configuring, we connect the IWR1843 device to the computer, and place it on a plat-form with a height of 1.2 meters. The side where the antenna array is located faces the person doing the activity. In the process of data collection, the fixation of the millimeter-wave radar must be guaranteed. We design six activities, which are boxing in place, jumping in place, squatting, walking in place, circling in place and high-knee lifting. Each activity is completed by 17 collectors and repeated for 20 seconds per person. During the data collection process, we decode the hexadecimal 3D point cloud data and range-Doppler data contained in the signal received by serial port in real time. After using one-hot encoding to label the data, we set the ratio of the amount of training data to the amount of validation data to be 4:1, that is, the training set accounts for 80% of the total data, while the verification set accounts for 20%.

Evaluation and Analysis

Using only 3D point cloud data as input

As mentioned above in the data preprocessing section, we input the acquired 3D point cloud data into the network and then train them by 100 epochs. The accuracy and loss are shown in Table I. To better evaluate the precision, we visualize the confusion matrix as shown in Fig.2. Each column of the confusion matrix represents the predicted category, and the total number of each column represents the number of the data predicted as that category. While each row represents the true ascribed category of the data, the total number of the data for each row represents the number of data instances of that category. We find that the prediction difficulty of these activities except circling, which is slightly more difficult, is all close.

Using only range-Doppler data as input

Similar to 3D point cloud data, we combine eight frames of range-Doppler data. After training the network with 100 epochs, we make Table II. In order to compare the verification accuracy of each action in 3D CNN network, we draw the confusion matrix, as shown in Fig.3. In this network, we find that high-knee lifting is easy to predict, while circling is relatively difficult to identify. The spectrograms of other activities are similar.

Using both 3D point cloud data & range-Doppler data as input

It can be concluded that, for point cloud data, the 3D CNN + LSTM has the highest accuracy and the lowest loss among the two networks. While for range-Doppler data, the 3D CNN performs best. Therefore, we choose these two networks to be the two parallel branches of the feature fusion network mentioned above. The obtained network structure is shown in Fig.1. After training, we tabulate the specific values as Table III.

TABLE I. ACCURACY AND LOSS OF THE TWO NETWORKS USING POINT CLOUD AS INPUT

Network Structure	***Accuracy***	***Loss***
3D CNN + LSTM	96.59	0.129

TABLE II. ACCURACY AND LOSS OF THE TWO NETWORKS USING POINT CLOUD AS INPUT

Network Structure	***Accuracy***	***Loss***
3D CNN	95.85	0.125

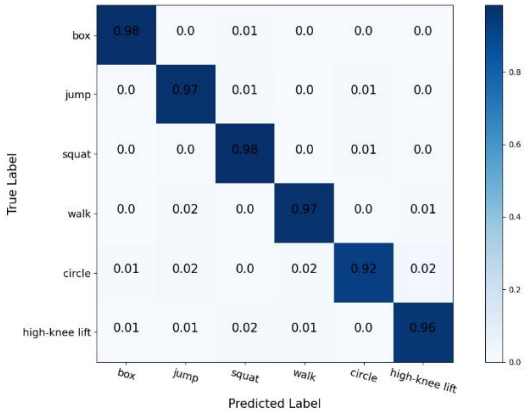


Figure 2. The confusion matrix of 3D CNN + LSTM network.

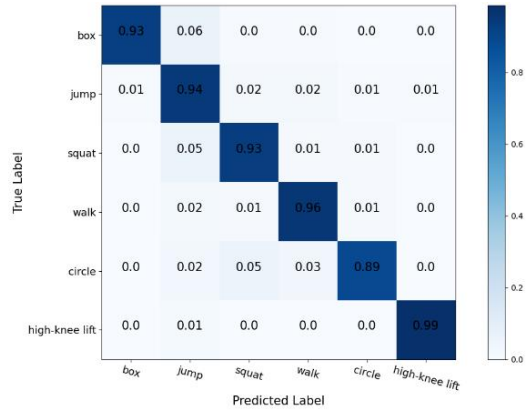


Figure 3. The confusion matrix of 3D CNN + LSTM network.

TABLE III. ACCURACY AND LOSS OF THE TWO NETWORKS USING RANGE-DOPPLER AS INPUT

Input Data Kinds	***Network Structure***	***Accuracy***	***Loss***
Only Range-Doppler	3D CNN	95.85	0.125
Only 3D point cloud	3D CNN + LSTM	96.59	0.129

Both 3D point cloud & Range-Doppler	3D CNN + LSTM & 3D CNN Parallelized	97.26	0.088

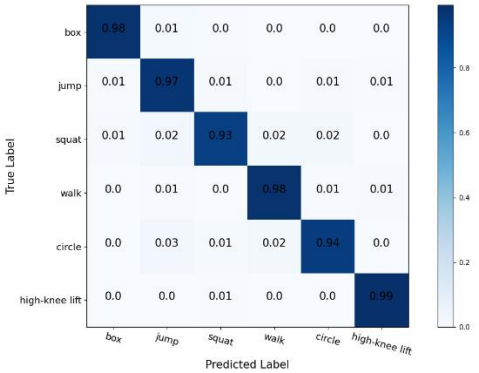


Figure 4. The confusion matrix of 3D CNN + LSTM network.

From the above results, we find that the network using both 3D point cloud data and range-Doppler data as input has better accuracy and loss than using one of them alone. This shows that using more kinds of data can extract more action features, which has a positive factor for the improvement of network performance. We also find that the convergence speed of the fusion network is faster. Like the above two networks, we also draw the confusion matrix, as shown in Fig. 4, so as to analyze the verification accuracy of each action. We find that the prediction difficulty of these activities is all close except squatting and circling, which are slightly more difficult to identify.

4. Conclusion

Nowadays, with the deepening of the interconnection of all things, people pay more attention to millimeter wave radar, which can ensure privacy and performance at the same time. In this paper, depth learning is used to classify people's different actions collected by millimeter wave radar. Different from previous studies, we increase the types of input data, input 3D point cloud data and range-Doppler data into the network for training at the same time, and then fuse the extracted features respectively. Finally, the network preforms better and achieves a recognition accuracy of 97.26%, which is about 1% higher than other networks with only one kind of data as input. This represents the success of our attempt to increase data types to achieve higher accuracy. We also consider that we can continue to expand data types, optimize the process of feature extraction, and better integrate different types of data in order to obtain better network performance, which requires further study.

REFERENCES

[1] Singh, A.D.; Sandha, S.S.; Garcia, L.; Srivastava, M. RadHAR: Human Activity Recognition from Point Clouds Generated through a Millimeter-Wave Radar. In Proceedings of the the 3rd ACM Workshop; Los Cabos, Mexico, October 25 2019.

[2] Fairchild, D.P.; Narayanan, R.M. Determining Human Target Facing Orientation Using Bistatic Radar Micro-Doppler Signals. Proc. SPIE - Int. Soc. Opt. Eng. 2014, 9082, 90–98.

[3] Xie, Y.; Jiang, R.; Guo, X.; Wang, Y.; Cheng, J.; Chen, Y. MmEat: Millimeter Wave-Enabled Environment-Invariant Eating Behavior Monitoring. Smart Health 2022, 23, 100236.

[4] Abdu, F.J.; Zhang, Y.; Fu, M.; Li, Y.; Deng, Z. Application of Deep Learning on Millimeter-Wave Radar Signals: A Review. Sensors 2021, 21, 1951.

[5] Zhao, P.; Lu, C.X.; Wang, J.; Chen, C.; Markham, A. MID: Tracking and Identifying People with Millimeter Wave Radar. In Proceedings of the 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS); Santorini, Greece, May 29 2019.

[6] Han, Z.; Zhang, C.; Fu, H.; Zhou, J.T. Trusted Multi-View Classification. ArXiv Prepr. ArXiv210202051, 2021, 1-16.

[7] IWR1843 Single-Chip 76- to 81-GHz FMCW mmWave Sensor (Rev. A). Available online: https://www.ti.com.cn/document-viewer/cn/IWR1843/datasheet/GUID-4FBB6021-CAC6-45C5-B361-E49F964BFB22#TITLE-SWRS188X2481(accessed on 9 February 2022

An Inner Bound for Multiple Descriptions Problem

Tao Guo
School of Cyber Science and Engineering, Southeast University, Nanjing, China
taoguo@seu.edu.cn

Abstract

We generalize the Venkataramani-Kramer-Goyal (VKG) coding scheme for L-channel multiple descriptions via generating a common description at each layer. The new achievable rate region coincides with the rate region of symmetrical multilevel diversity coding (SMDC) for all $L \geq 2$, in the special case of zero distortions among reconstructions, while the VKG region does only when $L = 2$. Moreover, we prove that the new region strictly contains the VKG region.

Index Terms

multiple descriptions, VKG region, common description, SMDC, superposition coding.

1. Introduction

In the multiple description problem, a memoryless source is encoded by a number of encoders. There are several decoders, each of which can access a certain subset of the encoders and reconstruct the source with a certain distortion criterion. The description by each encoder is either received errorfree or completely lost. The problem is to characterize the rate-distortion region, which is the set of all achievable rate-distortion tuples.

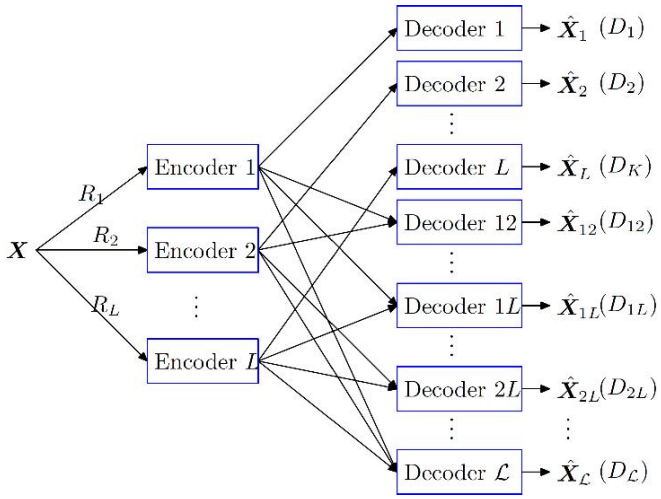
The optimal scheme for generating even two descriptions remains unknown. El Gamal and Cover (EGC) obtained an achievable region for 2-description problem in [1]. Ozarow [2] proved the EGC inner bound is tight for Gaussian source and mean squared error distortion measures. Ahlswede proved the tightness of EGC region for the no excess rate case in [3].

Zhang and Berger [4] obtained a tighter inner bound by generating a common description and refining it with EGC coding scheme. This is the best result till now for the 2-description problem. Venkataramani, Kramer, and Goyal (VKG) generalized Zhang-Berger inner bound into the general L-description case in [5]. Wang et. al. [6] proved that the VKG region, when specialized to the 2-description case, is equivalent to Zhang-Berger region.

In this paper, we focus on the L-channel *multiple description* (MD) problem. A decoder is said to be in level-l, if it can access a set of l encoders. We define a set of additional distortion measures among each level of reconstructions. When these additional distortions are set to be zero and reconstructions within the same level are identical, the multiple description problem is specialized to symmetrical multilevel diversity coding (SMDC), which was introduced by Roche [7] and Yeung [8]. The achievable rate region for SMDC system was determined in [9][10]. In particular, it was shown that a strategy known as superposition coding is optimal, where the sources are encoded separately. In this paper, we confine our discussion to SMDC with distortions, and denote L-channel problem as SMDC-L.

In the special case of zero distortions among reconstructions, we find that Zhang-Berger inner bound and thus VKG-2 coincides with the rate-distortion region of SMDC-2. However, when $L \geq 3$, the corresponding VKG region is exactly contained in the rate region of SMDC, which implies that VKG is not tight in the special case.

We notice that when generalizing Zhang-Berger coding scheme, VKG only involved one common description in the first layer. Here a further generalization is proposed, where we generate a common description at each level, and then refine them using the VKG



coding scheme. We apply the idea of superposition to deal with the common descriptions and the VKG refinement. Similar ideas can also be found in [11], which is called combinatorial message sharing.

The new inner bound, when specialized to the case of zero distortions among reconstructions, coincides with the rate region of SMDC for all $L \geq 2$. Therefore, our new inner bound strictly improves the VKG region. The newly introduced common descriptions compared with VKG scheme play the same role of improvement as Zhang-Berger's common description. The work is partially presented in [12], which involves some detailed proofs.

The rest of the paper is organized as follows. In Section 2, we state our problem and describe some existing results. In Section 3, we discuss the VKG region in the special case where distortions among reconstructions are zero. We propose a new coding strategy and

Fig. 1. The L-channel Multiple Description Model

show its strict improvement against VKG in Section 4. We conclude the paper in Section 5.

2. Problem Formulation and Existing Results

2.1 Problem Statement

Let X^n (denoted by $\boldsymbol{X}$) be the discrete memoryless source sequence with generic random variable X taking value from finite alphabet $\mathcal{X}$. The L-channel multiple description problem is depicted in Fig. 1. The single source is encoded by a set of encoders, indexed by $\mathcal{L} = \{1,2,\cdots,L\}$. A set of decoders are required to reconstruct the source, where each decoder is indexed by the index set $\mathcal{K}$ of its accessible encoders. Decoder $\mathcal{K} \subseteq \mathcal{L}$, $\mathcal{K} \neq \emptyset$ reconstructs the source as $\hat{\boldsymbol{X}}_{\mathcal{K}}$, where $\hat{X}_{\mathcal{K}i}, i = 1,2,\cdots,n$ take values in some finite reproduction alphabet $\hat{\mathcal{X}}_{\mathcal{K}}$. We denote decoders that access l $(l \in \mathcal{L})$ encoders as level-l decoders. As in Fig. 1, we will often simplify notation by dropping braces of sets when no confusion arises.

We define a single-letter distortion measure $d_{\mathcal{K}}: \mathcal{X} \times \hat{\mathcal{X}}_{\mathcal{K}} \to \mathbb{R}^+$ for each reconstruction, where $\mathbb{R}^+$ is the set of nonnegative reals. The distortion measure between the source sequence and respective reconstruction sequences are defined by the average distortion per symbol

$$d_{\mathcal{K}}(\boldsymbol{x}, \hat{\boldsymbol{x}}_{\mathcal{K}}) \triangleq \frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{K}}(x_i, \hat{x}_{\mathcal{K}i}), \quad \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset. \#(1)$$

Additional to the usual set up of multiple description problem, we define distortion measures among the same level reconstructions via

$$d^{(l)}\left(\hat{\boldsymbol{x}}^{(l)}\right) \triangleq \frac{1}{n}\sum_{i=1}^{n} d^{(l)}\left(\hat{x}^{(l)}\right),\ \ l \in \mathcal{L}, \#(2)$$

where $d^{(l)}: \mathcal{X}^{(l)} \to \mathbb{R}^{+}$ is the corresponding single-letter distortion measure, and $\hat{\boldsymbol{x}}^{(l)}$ is a shorthand for $\{\hat{\boldsymbol{x}}_{\mathcal{K}}: |\mathcal{K}| = l\}$ and $\mathcal{X}^{(l)}$ for the Cartesian product of $\mathcal{X}_{\mathcal{K}}, |\mathcal{K}| = l$. We assume that $d_{\mathcal{K}}$ and $d^{(l)}$ are upper-bounded by $d_{\max}$ for all $\mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}$ and $l \in \mathcal{L}$.

Formally, an $\left(n, M_l, \Delta_{\mathcal{K}}, \Delta^{(l)}, l \in \mathcal{L}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}\right)$ code is defined by the encoding functions

$$f_l: \mathcal{X}^n \to \{0,1,2,\cdots, M_l - 1\},\ \ l \in \mathcal{L}, \#(3)$$

and the decoding functions

$$g_{\mathcal{K}}: \prod_{l\in\mathcal{K}} \{0,1,2,\cdots, M_l - 1\} \to \hat{\mathcal{X}}_{\mathcal{K}}^{n},\ \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset \#(4)$$

and the expected distortions

$$\Delta_{\mathcal{K}} = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{K}}\left(X_i, \hat{X}_{\mathcal{K}i}\right)\right],\ \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset \#(5)$$

$$\Delta^{(l)} = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} d^{(l)}\left(\hat{X}^{(l)}\right)\right],\ \ l \in \mathcal{L}. \#(6)$$

A rate-distortion tuple $\left(R_l, D^{(l)}, l \in \mathcal{L}, D_{\mathcal{K}}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}\right)$ is *achievable* if for every $\epsilon > 0$, there exists for sufficiently large n, an $\left(n, M_l, \Delta_{\mathcal{K}}, \Delta^{(l)}, l \in \mathcal{L}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}\right)$ code such that

$$\frac{1}{n}\log M_l \leq R_l + \epsilon,\quad l \in \mathcal{L} \#(7)$$

and

$$\Delta_{\mathcal{K}} \leq D_{\mathcal{K}},\ \mathcal{K} \subseteq \mathcal{L}, \mathcal{K} \neq \emptyset \#(8)$$

$$\Delta^{(l)} \leq D^{(l)},\ \ l \in \mathcal{L}. \#(9)$$

The multiple description rate-distortion region $\mathcal{R}$ is the set of all achievable rate-distortion tuples. Our goal is to find the rate-distortion region $\mathcal{R}$.

2.2 VKG Region

The first general result for the 2-channel case was derived by El Gamal and Cover in [1]. Zhang and Berger [4] proposed a tighter inner bound by generating a common description before applying the EGC coding scheme. Venkataramani, Kramer and Goyal (VKG) generalized the results of EGC and Zhang-Berger to L-description problem in [5]. We now restate their result.

Theorem 1 (VKG [5]): *Let $\hat{X}_{(2^{\mathcal{L}})}$ be any set of 2^L random variables jointly distributed with X, where $\hat{X}_{\emptyset}$ takes values in some finite alphabets $\hat{\mathcal{X}}_{\emptyset}$ and each $\hat{X}_{\mathcal{K}}$ takes values in the reproduction alphabets $\hat{\mathcal{X}}_{\mathcal{K}}, \mathcal{K} \neq \emptyset$. Then the rate-distortion region of the L-channel multiple description problem contains the rate-distortion tuples $(R_l, D_{\mathcal{K}},\ l \in \mathcal{L}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\})$ satisfying*

$$D_{\mathcal{K}} \geq\ \mathbb{E}d_{\mathcal{K}}(X, \hat{X}_{\mathcal{K}}), \#(10)$$

$$R_{\mathcal{K}} \geq (|\mathcal{K}| - 1)I\left(X; \hat{X}_{\emptyset}\right) - H\left(\hat{X}_{(2^{\mathcal{K}})}|X\right) + \sum_{\mathcal{M}\subseteq\mathcal{K}} H\left(\hat{X}_{\mathcal{M}}|\hat{X}_{(2^{\mathcal{M}}-\{\mathcal{M}\})}\right) \#(11)$$

where $R_{\mathcal{K}}$ is a shorthand for $\sum_{l\in\mathcal{K}} R_l$ and $\hat{X}_{(\mathcal{C})}$ for $\{\hat{X}_{\mathcal{M}}: \mathcal{M} \in \mathcal{C}\}$, and we interpret $\hat{X}_{(\{\emptyset\}-\{\emptyset\})} = \hat{X}_{(\emptyset)}$ as a constant.

Remark 1: The VKG region is a convex set, since we can incorporate a time-sharing random variable into $\hat{X}_{(\emptyset)}$.

Remark 2: Since the distortion measures among the reconstructions are first introduced in this paper, it suffices to set these distortion fidelities $D^{(l)} = \infty$ for all $l \in \mathcal{L}$.

2.3 Symmetrical Multilevel Diversity Coding (SMDC)

SMDC was first introduced by Roche [7] and Yeung [8], where the word "symmetrical" refers to the fact that the reconstruction depends on the set of accessible encoders only via its cardinality. The 3-level case was solved in [9], and L-level case in [10].

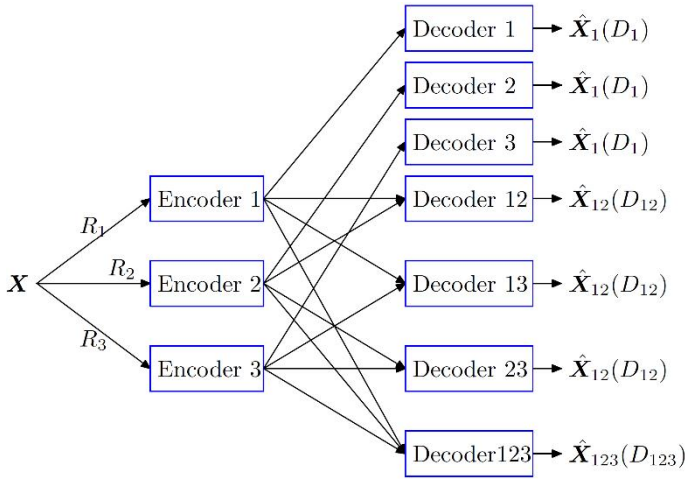


Fig. 2. Symmetrical multilevel diversity coding with distortions (L = 3)

We now confine our discussion to SMDC with distortions, which is illustrated in Fig. 2 as a 3-level example. There is only one source sequence $\boldsymbol{X}$, and the reconstructions within the same level are identical due to the same distortion criteria. The optimality of superposition coding, which is shown in [10], can be easily generalized to the lossy case here.

The superposition rate-distortion region of lossy SMDC is characterized in the following lemma. For simplicity, we write $U_l, l \in \mathcal{L}$ as a shorthand for level-l reproduction random variable $\widehat{X}_{1,2,\cdots,l}$ and D_l for $D_{1,2,\cdots,l}$, when no confusion is caused here.

Lemma 1 (superposition region): A nonnegative rate-distortion tuple $(R_1, \cdots, R_L, D_1, \cdots, D_L)$ is achievable for SMDC if and only if

$$\mathbb{E}d_l(X, U_l) \le D_l, \quad l \in \mathcal{L} \ \#(12)$$

and

$$R_l = \sum_{\alpha=1}^{L} r_l^{\alpha}, \quad l \in \mathcal{L} \ \#(13)$$

for some $r_l^{\alpha} \ge 0, 1 \le \alpha \le L$, satisfying

$$\sum_{l \in \mathcal{K}} r_l^{|\mathcal{K}|} \ge I\left(X; U_{|\mathcal{K}|} | U_1^{|\mathcal{K}|-1}\right), \quad \text{for } \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}, \#(14)$$

where

$$U_1^{|\mathcal{K}|-1} = \left(U_1, U_2, \cdots, U_{|\mathcal{K}|-1}\right). \#(15)$$

3. VKG Region in the Special Case of SMDC

The VKG region, which extends EGC region and Zhang-Berger region, is an inner bound of the L-channel multiple description problem. Wang et. al. [6] proved the equivalence of RV KG and RZB for $L = 2$, which appeared as a consequence of the fact that the final layer in the VKG scheme is dispensable. We restate the result in the following lemma, whose proof can be found in [12].

Lemma 2: $\mathcal{R}_{VKG2} = \mathcal{R}_{ZB}$.

3.1 Equivalence between VKG-2 and SMDC-2

The difference between multiple descriptions and SMDC lies in the fact that reconstructions within the same level are required to be identical due to the same distortion criteria in SMDC. We introduce distortion measures among reconstructions in this paper to unify the two problems.

First we assume that the distortion criteria within the same level are identical, which is

$$d_{\mathcal{K}} = d_{\mathcal{M}} \#(16)$$

and

$$D_{\mathcal{K}} = D_{\mathcal{M}} \#(17)$$

for all $\mathcal{K}, \mathcal{M} \in 2^{\mathcal{L}} - \{\emptyset\}$ such that $|\mathcal{K}| = |\mathcal{M}|$.

Now define the distortion measures among the reconstructions as Hamming distortions

$$d^{(l)}\left(\hat{x}^{(l)}\right) = \begin{cases} 0, \text{ if } \hat{x}_{\mathcal{K}} = \hat{x}_{\mathcal{M}} \text{ for all } |\mathcal{K}| = |\mathcal{M}| = l \\ 1, \text{otherwise} \end{cases} \#(18)$$

for all $l \in \mathcal{L}$. If the corresponding distortion fidelities are set to be zero,

$$D^{(l)} = 0 \#(19)$$

then the multiple description problem reduces to SMDC with distortions. We denote the VKG region under condition (19) as $\mathcal{R}^*_{VKG}$.

Since Hamming distortion fidelities among the reconstructions are set to be zero, the reconstruction random variables in $\mathcal{R}^*_{VKG}$ should satisfy

$$\mathbb{E}d^{(l)}\left(\widehat{X}^{(l)}\right) = 0, \ l \in \mathcal{L}, \#(20)$$

which implies that

$$\widehat{X}_{\mathcal{K}} = \widehat{X}_{\mathcal{M}} \#(21)$$

for all $\mathcal{K}, \mathcal{M} \in 2^{\mathcal{L}} - \{\emptyset\}$ such that $|\mathcal{K}| = |\mathcal{M}|$.

We shall show that the VKG region $\mathcal{R}^*_{VKG}$ and thus the corresponding Zhang-Berger region coincides with the superposition rate-distortion region of SMDC-2. The proof is omitted here and can be found in [12].

Theorem 2: $\mathcal{R}^*_{VKG} = \mathcal{R}_{SMDC2}$.

Since the zero-distortion condition (19) is looser than the requirement of SMDC that reconstructions within the same level are identical, the rate-distortion region of multiple descriptions with condition (19) should contain the rate-distortion region of SMDC. Now, when $L = 2$, the VKG inner bound of multiple description problem coincides with the rate-distortion region of SMDC.

3.2 Non-tightness of VKG-3

Consider the 3-description case with condition (18)(19), which indicates $\widehat{X}_1 = \widehat{X}_2 = \widehat{X}_3$ and $\widehat{X}_{12} = \widehat{X}_{13} = \widehat{X}_{23}$. The VKG inner bound $\mathcal{R}^*_{VKG3}$ is the set of rate-distortion tuples $(R_1, R_2, R_3, D_1, D_{12}, D_{123})$, for which there exists an auxiliary random variable $\widehat{X}_\emptyset$ jointly distributed with X, such that

$$D_1 \geq \mathbb{E}d_1(X, \widehat{X}_1), \#(22)$$

$$D_{12} \geq \mathbb{E}d_{12}(X, \widehat{X}_{12}), \#(23)$$

$$D_{123} \geq \mathbb{E}d_{123}(X, \widehat{X}_{123}), \#(24)$$

$$R_i \geq I\left(X; \widehat{X}_1, \widehat{X}_\emptyset\right), \ i = 1,2,3 \#(25)$$

$$R_i + R_j \geq I\left(X; \widehat{X}_1, \widehat{X}_\emptyset\right) + I\left(X; \widehat{X}_1, \widehat{X}_{12}, \widehat{X}_\emptyset\right) + H(\widehat{X}_1 | X, \widehat{X}_\emptyset), \ 1 \leq i < j \leq 3 \#(26)$$

$$R_1 + R_2 + R_3 \geq I\left(X; \widehat{X}_1, \widehat{X}_{12}, \widehat{X}_{123}, \widehat{X}_\emptyset\right) + 2I(X; \widehat{X}_\emptyset) + 2H(\widehat{X}_1, \widehat{X}_{12} | \widehat{X}_\emptyset) \#(27)$$

where $\widehat{X}_1, \widehat{X}_{12}, \widehat{X}_{123}$ take values in the reproduction alphabets $\widehat{\mathcal{X}}_1, \widehat{\mathcal{X}}_{12}$, and $\widehat{\mathcal{X}}_{123}$, respectively.

Remark 3: Since $D_1 = D_2 = D_3$ and $D_{12} = D_{13} = D_{23}$, we only consider D_1 and D_{12} in the discussion for simplicity, and omit the zero distortion criteria $D^{(1)} = D^{(2)} = 0$.

We now compare regions $\mathcal{R}^*_{VKG3}$ and $\mathcal{R}_{SMDC3}$ in Fig. 3 for some valid $\widehat{X}_1, \widehat{X}_{12}, \widehat{X}_{123}$, where $\mathcal{R}_{SMDC3}$ is specified by blue lines with corner points (Q_i) and $\mathcal{R}^*_{VKG3}$ is characterized by red lines with corner points (P_i).

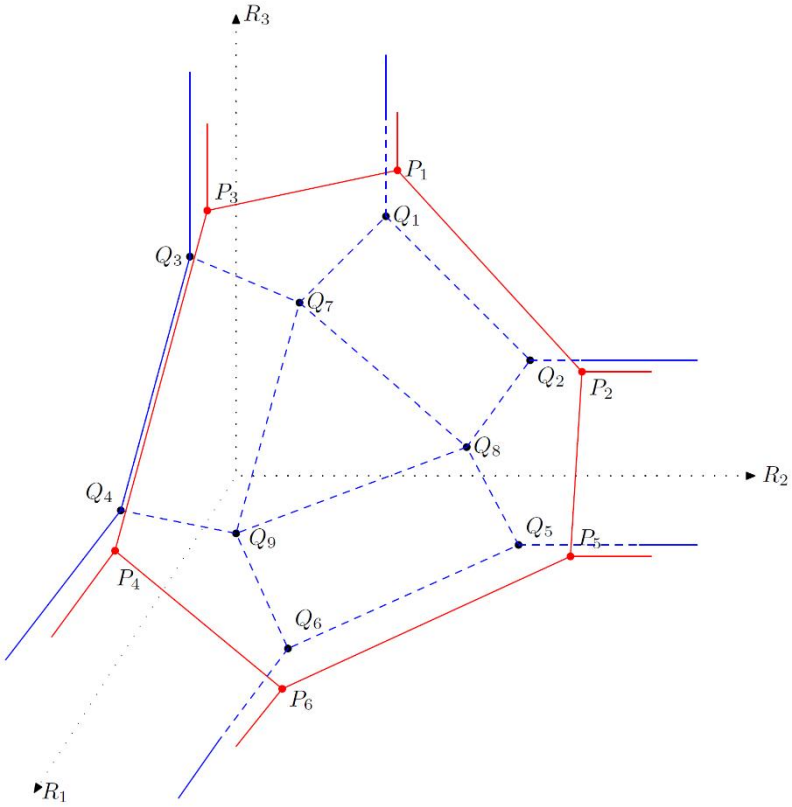


Fig. 3. Comparison between VKG inner bound and the region of SMDC

Focusing on the relation between two points P_1 and Q_1, we find that

$$R_1^{P_1} - R_1^{Q_1} = I(X; \widehat{X}_\emptyset | \widehat{X}_1) \geq 0 \ \#(28)$$

$$R_2^{P_1} - R_2^{Q_1} = I\left(X; \widehat{X}_\emptyset \middle| \widehat{X}_1, \widehat{X}_{12}\right) + H(\widehat{X}_1 | X, \widehat{X}_\emptyset) \geq 0 \#(29)$$

$$R_3^{P_1} - R_3^{Q_1} = I\left(X; \widehat{X}_\emptyset \middle| \widehat{X}_1, \widehat{X}_{12}, \widehat{X}_{123}\right) + I\left(X; \widehat{X}_{12} \middle| \widehat{X}_1, \widehat{X}_\emptyset\right) \#$$
$$+ H\left(\widehat{X}_{12} \middle| X, \widehat{X}_1, \widehat{X}_\emptyset\right) + H\left(\widehat{X}_1, \widehat{X}_{12} \middle| X, \widehat{X}_\emptyset\right) \geq 0 \#(30)$$

which indicates that P_1 is always contained in $\mathcal{R}_{SMDC3}$. Similar results hold for the other corner points $P_2, \cdots, P_6$ of $\mathcal{R}^*_{VKG3}$. Since at least one of the following two conditions happen,

i. Q_7, Q_8, Q_9 are always active,

ii. equalities in (28), (29), and (30) do not hold,

we conclude that $\mathcal{R}_{SMDC3}$ properly contains $\mathcal{R}^*_{VKG3}$. This further implies that $\mathcal{R}^*_{VKG3}$ is not tight, since $\mathcal{R}_{SMDC3}$ is obviously an inner bound of the rate-distortion region of multiple description with condition (19). Similar results hold for $L > 3$.

4. Main Results

Since $\mathcal{R}^*_{VKG}$ is exactly contained in $\mathcal{R}_{SMDC}$ when $L \geq 3$, we want to close the gap. Inspired by the improvement of Zhang-Berger inner bound against EGC region, we propose a new coding scheme via adding a common description at each layer of VKG scheme. The common descriptions perform superposition coding as SMDC, where the description of layer l is recovered due to the same distortion criteria if any l encoders are accessible. We then refine the common description of each layer using the corresponding layer of VKG scheme.

The inner bound derived from the new coding scheme coincides with the SMDC rate-distortion region under the zero-distortion condition (19). Therefore, by introducing the common descriptions, our new coding scheme improves the VKG scheme.

4.1 A New Inner Bound

Our main result is an inner bound to the rate-distortion region for L-channel multiple descriptions. In the following discussion, we define $\widehat{X}_{\emptyset}$ and U_L to be constants. Denote

$$\widehat{X}^{(j)} \triangleq \{\widehat{X}_{\mathcal{K}}: |\mathcal{K}| = j\} \#(31)$$

as the collection of all level-j reconstructions. For example,

$$\widehat{X}^{(1)} = \{\widehat{X}_1, \widehat{X}_2, \cdots, \widehat{X}_L\} \#(32)$$

$$\widehat{X}^{(2)} = \{\widehat{X}_{12}, \widehat{X}_{13}, \cdots, \widehat{X}_{L-1,\ L}\} \#(33)$$

$$\widehat{X}^{(L)} = \{\widehat{X}_{1,2,\cdots,L}\} \#(34)$$

For any $\mathcal{Q} \subseteq \mathcal{L}$ with cardinality $|\mathcal{Q}| = q$, denote

$$\widehat{X}_{\mathcal{Q}}^{(j)} \triangleq \{\widehat{X}_{\mathcal{K}}: \mathcal{K} \subseteq \mathcal{Q}, |\mathcal{K}| = j\} \#(35)$$

as the collection of all level-j reconstructions when only the encoders indexed by $\mathcal{Q}$ are available. We write $\widehat{X}_{(\mathcal{C})}$ as a shorthand for $\{\widehat{X}_{\mathcal{M}}: \mathcal{M} \in \mathcal{C}\}$.

Let auxiliary random variables $U_1, \cdots, U_{L-1}$ take values in some finite alphabets $\mathcal{U}_1, \cdots, \mathcal{U}_{L-1}$, and $\widehat{X}_{\mathcal{K}}$ take values in reproduction alphabets $\widehat{\mathcal{X}}_{\mathcal{K}}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset, \mathcal{L}\}$. Now we state our main theorem as follows.

Theorem 3: A rate-distortion tuple $\left(R_l, D^{(l)}, D_{\mathcal{K}}, l \in \mathcal{L}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\}\right)$ is achievable for L-channel multiple description problem, if

$$\mathbb{E}\mathrm{d}_{\mathcal{K}}\left(X, \widehat{X}_{\mathcal{K}}\right) \leq D_{\mathcal{K}}, \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset\} \#(36)$$

$$\mathbb{E}\mathrm{d}^{(l)}\left(\widehat{X}^{(l)}\right) \leq D^{(l)}, l \in \mathcal{L}. \#(37)$$

and

$$R_l = \sum_{\alpha=0}^{L-1} r_l^{\alpha}, l \in \mathcal{L} \#(38)$$

for some $r_l^{\alpha} \geq 0, 0 \leq \alpha \leq L$-1, satisfying

$$\sum_{l \in \mathcal{K}} r_l^{|\mathcal{K}|} \geq I\left(X; U_{|\mathcal{K}|} | U_1^{|\mathcal{K}|-1}\right), \quad \text{for } \mathcal{K} \in 2^{\mathcal{L}} - \{\emptyset, \mathcal{L}\} \#(39)$$

and

$$r_{\mathcal{Q}}^0 \geq \beta(\mathcal{Q}), \mathcal{Q} \subseteq \mathcal{L} \#(40)$$

where

$$U_1^{|\mathcal{K}|-1} = \left(U_1, U_2, \cdots, U_{|\mathcal{K}|-1}\right), \#(41)$$

and $\beta(\mathcal{Q}), \mathcal{Q} \subseteq \mathcal{L}$ is defined via

$$\beta(\mathcal{Q}) = \sum_{j=1}^{|\mathcal{Q}|} \left[\sum_{\mathcal{K} \subseteq \mathcal{Q}, |\mathcal{K}| = j} H\left(\widehat{X}_{\mathcal{K}} | \widehat{X}_{(2^{\mathcal{K}} - \{\mathcal{K}\})}, U_1, \cdots, U_j\right) \right] - H\left(\widehat{X}_{(2^{\mathcal{Q}})} | X, U_1, \cdots, U_{L-1}\right). \#(42)$$

Proof: The theorem can be proved by a coding scheme that introducing more auxiliary random variables compared to VKG. The details can be found in [12], which are omitted here.

The idea of superposition coding is introduced to obtain Theorem 3, thus we denote the corresponding inner bound as $\mathcal{R}_{sup}$.

4.2 Equivalence between the New Inner Bound and SMDC-L

We now consider the special case of zero distortion fidelities among reconstructions, where conditions (18)(19) are satisfied. Likewise, denote the resulting inner bound as $\mathcal{R}_{sup}^*$. We shall show the new inner bound $\mathcal{R}_{sup}^*$ coincides with the rate-distortion region in Lemma 1. The detailed proof is given in [12] and thus omitted here.

Theorem 4: $\mathcal{R}_{sup}^{*} = \mathcal{R}_{SMDC}$.

Remark 4: Theorem 4 only states that the inner bound $\mathcal{R}_{sup}^{*}$ coincides with $\mathcal{R}_{SMDC}$. But since condition (19) is looser than the symmetrical reconstruction requirement of SMDC, $\mathcal{R}_{SMDC}$ is of course an inner bound for multiple description problem with conditions (18)(19). The tightness of $\mathcal{R}_{sup}^{*}$ has not yet been shown.

4.3 Strict Improvement Against VKG

We now show that our new inner bound $\mathcal{R}_{sup}$ reduces to VKG region $\mathcal{R}_{VKG}$, via setting

$$U_2 = U_3 = \cdots U_{L-1} = 0 \#(43)$$

where the values of $\beta(\mathcal{Q})$ reduce to

$$\beta'(\mathcal{Q}) = \sum_{j=1}^{|\mathcal{Q}|}\left[\sum_{\mathcal{K}\subseteq\mathcal{Q},|\mathcal{K}|=j} H\left(\widehat{X}_{\mathcal{K}}|\widehat{X}_{(2^{\mathcal{K}}-\{\mathcal{K}\})}, U_1\right)\right] - H\left(\widehat{X}_{(2^{\mathcal{Q}})}|X, U_1\right) \#(44)$$

$$= \sum_{\mathcal{K}\subseteq\mathcal{Q}} H\left(\widehat{X}_{\mathcal{K}}|\widehat{X}_{(2^{\mathcal{K}}-\{\mathcal{K}\})}, U_1\right) - H\left(\widehat{X}_{(2^{\mathcal{Q}})}|X, U_1\right) \#(45)$$

and the rates corresponding to common descriptions reduce to

$$r_l^1 \geq I(X; U_1) \#(46)$$

$$r_l^{\alpha} = 0, for\ \alpha = 2,3,\cdots, L-1. \#(47)$$

Thus, the rates of each encoder reduces to

$$R_l = r_l^0 + r_l^1, l \in \mathcal{L} \#(48)$$

which implies

$$R_{\mathcal{Q}} \geq \beta'(\mathcal{Q}) + |\mathcal{Q}| \cdot I(X; U_1) \#(49)$$

$$= (|\mathcal{Q}| - 1)I(X; U_1) - H\left(\widehat{X}_{(2^{\mathcal{Q}})}, U_1|X\right) + H(U_1) + \sum_{\mathcal{K}\subseteq\mathcal{Q}} H\left(\widehat{X}_{\mathcal{K}}|\widehat{X}_{(2^{\mathcal{K}}-\{\mathcal{K}\})}, U_1\right). \#(50)$$

Note that $\widehat{X}_{\emptyset} = 0$ in (50). We can further replace U_1 by $\widehat{X}_{\emptyset}$ to obtain the inner bounds in Theorem 3, which implies

$$\mathcal{R}_{VKG} \subseteq \mathcal{R}_{sup}. \#(51)$$

In the special case where conditions (18)(19) are satisfied, VKG region is strictly contained in $\mathcal{R}_{SMDC}$ (Section 3.2), which coincides with our new inner bound $\mathcal{R}_{sup}^{*}$ (Section 4.2). This indicates that

$$\mathcal{R}_{VKG}^{*} \subseteq \mathcal{R}_{sup}^{*}. \#(52)$$

Therefore, we conclude that our new inner bound strictly improves VKG inner bound.

5. Conclusion and Future Study

This paper considered the problem of L-channel multiple descriptions, where we proposed distortion criteria among reconstructions within the same level. It generalized the work in [5], via generating a common description at each layer and then refining it with the VKG coding scheme. We showed the new coding strategy strictly improves the VKG scheme.

In this paper, we discussed the special case of zero distortions among reconstructions, in which the new inner bound coincides with the rate-distortion region of SMDC. But whether the inner bound is tight under the special case remains unknown and will be explored in future work.

References

[8] A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," IEEE Transactions on Information Theory, vol. IT-28, pp. 851–857, Nov. 1982.

[9] L. Ozarow, "On a source coding problem with two channels and three receivers," Bell Syst. Tech. J., vol. 59, pp. 1909 – 1921, Dec. 1980.

[10] R. Ahlswede, “The rate-distortion region for multiple descriptions without excess rate,” IEEE Transactions on Information Theory, vol. 31, pp. 721–726, 1985.

[11] Z. Zhang and T. Berger, “New results in binary multiple descriptions,” IEEE Transactions on Information Theory, vol. 33, pp. 502–521, Jul. 1987.

[12] R. Venkataramani, G. Kramer, and V. K. Goyal, “Multiple description coding with many channels,” IEEE Transactions on Information Theory, vol. 49, pp. 2106–2114, Sep. 2003.

[13] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, “On the role of the refinement layer in multiple description coding and scalable coding,” IEEE Transactions on Information Theory, vol. 57, pp. 1443–1456, Mar. 2011.

[14] J. R. Roche, distributed information storage. PhD thesis, Stanford University, Stanford, CA, Mar. 1992.

[15] R. W. Yeung, “Multilevel diversity coding with distortion,” IEEE Transactions on Information Theory, vol. 41, pp. 412–422, Mar. 1995.

[16] J. R. Roche, R. W. Yeung, and K. P. Hau, “Symmetrical multilevel diversity coding,” IEEE Transactions on Information Theory, vol. 43, pp. 1059–1064, May 1997.

[17] R. W. Yeung and Z. Zhang, “On symmetrical multilevel diversity coding,” IEEE Transactions on Information Theory, vol. 45, pp. 609–621, Mar. 1999.

[18] K. B. Viswanatha, E. Akyol, and K. Rose, “Combinatorial message sharing and a new achievable region for multiple descriptions,” IEEE Transactions on Information Theory, vol. 62, pp. 769–792, Feb. 2016.

[19] T. Guo and R. W. Yeung, “Extended multiple descriptions with reconstruction consistency constraints,” 2016 IEEE Globecom Workshops, Washington, D.C., Dec. 2016.

MMTC OFFICERS (Term 2020 — 2022)

CHAIR

Jun Wu
Fudan University
China

STEERING COMMITTEE CHAIR

Joel J. P. C. Rodrigues
Federal University of Piauí (UFPI)
Brazil

VICE CHAIRS

Shaoen Wu (North America)
Illinois State University
USA

Liang Zhou (Asia)
Nanjing University of Post and Telecommunications
China

Abderrahim Benslimane (Europe)
University of Avignon
France

Qing Yang (Letters & Member Communications)
University of North Texas
USA

SECRETARY

Han Hu
Beijing Institute of Technology
China

STANDARDS LIAISON

Weiyi Zhang
AT&T Research
USA

MMTC Communication-Frontier BOARD MEMBERS (Term 2016—2018)

Danda Rawat	Director	Howard University	USA
Sudip Misra	Co-Director	IIT Kharagpur	India
Guanyu Gao	Co-Director	Nanjing University of Science and Technology	China
Rui Wang	Co-Director	Tongji University	China
Lei Chen	Editor	Georgia Southern University	USA
Tasos Dagiuklas	Editor	London South Bank University	UK
ShuaiShuai Guo	Editor	King Abdullah University of Science and Technology	Saudi Arabia
Kejie Lu	Editor	University of Puerto Rico at Mayagüez	Puerto Rico
Nathalie Mitton	Editor	Inria Lille-Nord Europe	France
Zheng Chang	Editor	University of Jyväskylä	Finland
Dapeng Wu	Editor	Chongqing University of Posts & Telecommunications	China
Luca Foschini	Editor	University of Bologna	Italy
Mohamed Faten Zhani	Editor	University of Quebec	Canada
Armir Bujari	Editor	University of Padua	Italy
Kuan Zhang	Editor	University of Nebraska-Lincoln	USA
Bin Tan	Editor	Jinggangshan University	China