

**MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY**

<http://committees.comsoc.org/mmc>

R-LETTER

Vol. 5, No. 2, April 2014



IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from the Review Board	2
Prediction of Video Popularity Based on Cross-Domain Knowledge Transfer	3
A short review for “Towards Cross-Domain Learning for Social Video Popularity Prediction”	3
Edited by Karine Pires and Gwendal Simon.....	3
Lighting the design of replication algorithms for P2P VoD system in practice	5
A short review for “On Replication Algorithm in P2P VoD”	5
Edited by Lifeng Sun	5
The Design of Next Generation Multimedia Synchronization Algorithms	7
A short review for “Evolution of temporal multimedia synchronization principles:A historical viewpoint”	7
Edited by Irene Cheng	7
Converting 2D to 3D by Learning from Examples	9
A short review for “Learning-based, automatic 2D-to-3D image and video conversion”	9
Edited by Jun Zhou	9
Sparse Representation Assists Video Tagging	11
A short review for “Video-to-Shot Tag Propagation by Graph Sparse Group Lasso”. ..	11
Edited by Vladan Velisavljević	11
Using Near-Infrared Image for Dehazing.....	13
A short review for “Near-Infrared Guided Color Image Dehazing”	13
Edited by Gene Cheung	13
Exploring frequency-domain oversampling for multicarrier transmissions in Underwater Acoustic Communications	15
A short review for “Frequency-Domain Oversampling for Zero-Padded OFDM in Underwater Acoustic Communications”	15
Edited by Weiyi Zhang	15
Paper Nomination Policy.....	17
MMTC R-Letter Editorial Board.....	18
Multimedia Communications Technical Committee (MMTC) Officers	18

Message from the Review Board

Introduction

While the innovation of social media, mobile devices and network capability facilitates the delivery of multimedia content, these technological advances, together with the increasing consumer demands, also introduce new challenges and necessitate more sophisticated strategies in handling data streaming and synchronization. In this issue, you will find interesting papers addressing related topics. In particular, you will find a valuable review of “temporal multimedia synchronization” from a historical perspective covering the evolution in the last 30 years, and a discussion of future development.

Distinguished Category

Video streaming has become the most prominent network application in recent years. How to predict and then provide better video to users is one of the most important questions to answer. Along with big data analysis using social media and Internet of Things (machine-to-machine) applications, video streaming provisioning is one of the most practical and significant problem to study.

The **first paper**, published in *IEEE Transactions on Multimedia*, presents a novel approach to extract knowledge from a “social stream” with the ambition of a better prediction of popularity bursts.

The **second paper**, published in *IEEE/ACM Transactions on Networking*, provides replication algorithms for P2P VoD system.

Regular Category

The topics of the regular category include multimedia synchronization, image/video conversion, video tagging, and underwater acoustic communication.

The **first paper**, published in the *ACM Transactions on Multimedia Computing, Communications and Applications* and edited

by *Irene Cheng*, reviews the history of multimedia synchronization principles and proposes a multidimensional synchronization framework to address challenges of current and future applications.

The **second paper**, published in the *IEEE Transactions on Image Processing* and edited by *Jun Zhou*, presents means for the conversion of 2D-to-3D image and video content based on machine learning.

The **third paper** is edited by *Vladan Velisavljević* and has been published within the *IEEE Transactions on Multimedia*. It proposes a method to enable the tag propagation from the whole video to individual shots.

The **forth paper**, published in the *IEEE International Conference on Image Processing* and edited by *Gene Cheung*, targets dehazing of images using information from a near-infrared version of the same image.

Finally, the **fifth paper**, published in *IEEE Journal of Oceanic Engineering* and edited by *Weiyi Zhang*, explores multicarrier transmission in underwater acoustic communication using frequency-domain oversampling.

We would like to thank all the authors, nominators, reviewers, editors, and others who contribute to the release of this issue.

IEEE ComSoc MMTc R-Letter

Director:

Irene Cheng, University of Alberta, Canada

Email: locheng@ualberta.ca

Co-Director: Distinguished Category

Weiyi Zhang, AT&T Research, USA

Email: maxzhang@research.att.com

Co-Director: Regular Category

Christian Timmerer

Alpen-Adria-Universität Klagenfurt, Austria

Email: christian.timmerer@itec.aau.at

Prediction of Video Popularity Based on Cross-Domain Knowledge Transfer

A short review for "Towards Cross-Domain Learning for Social Video Popularity Prediction"

Edited by Karine Pires and Gwendal Simon

Suman Deb Roy, Tao Mei, Wenjun Zeng, and Shipeng Li, "Towards Cross-Domain Learning for Social Video Popularity Prediction", IEEE Transactions on Multimedia, vol. 15, no. 6, pp. 1255-1267, October 2013.

The prediction of videos popularity has become a critical topic, which matters from both technical and business regards. For the technical aspect, popularity prediction is a key parameter of infrastructure optimization process, which allows a proper management of the needed delivering resources and the improvement of Quality of Experience (QoE). For the business aspect, popularity prediction can increase the accuracy of advertising campaign and attract investment. Unsurprisingly, many studies have been made in this direction over the past years, including time series, video characterization, popularity growth patterns, and geographical analysis.

However, fewer studies have explored the (intuitive) connection between videos and the "conversation" in social networks.

In this paper authors presents a novel approach to extract knowledge from a "social stream" with the ambition of a better prediction of popularity bursts. The proposed algorithm is expected to learn topics from social streams, and then to derive the social prominence of a video. The presented framework is able to scale regardless the bursts on the inputs.

Before entering into the details, we would like to emphasize that, from this intuitively interesting ideas, authors succeeded in developing a quite complete study, which includes a great deal of theoretical contributions as well as an accurate evaluation based on real-traces. Such a study opens perspectives for scientists having a domain study that overlaps social networks and multimedia systems.

The functionality of the framework can be divided into two main tasks: detecting the topics, and measuring the popularity of the video topic. For the former task, by using the Online Stream LDA (OSLDA) algorithm, the framework collects topics from social stream (Twitter) in real-time and creates a topic space to be used by the other parts of the framework. On the latter

task, the so-called SocialTransfer algorithm uses the previously created topic space to classify videos by social media topics. This classification corresponds to a labeling a video with a learned social domain topic. It allows the framework to continuously adapt to the new topic entries that are found on the input stream. In summary, the system keeps updating the topic space and connecting the new topics to the videos. It is done through the creation of a transfer graph, which is the association of topics to videos. A calculation of the social prominence of each of the topics is made, which allows the system to distinguish those special videos having bursty popularity rises. Authors conclude that this behavior shows the evidence that popularity signals traverses across different domains and affects video popularity.

As previously said, the optimization framework occupies a large part of the paper, with many interesting developments, including a Power Iteration that is used to retrieve Laplacian eigenvalues needed to create the transfer graph.

Experiments, using 10.2 million tweets and 3.5 million YouTube videos, show that the developed solution outperforms traditional learners for more than 60%. However, we have to admit that the evaluation of these real traces is a bit below our expectations. It should be possible to provide a better evaluation of the framework, and to go deeper into the analysis of both video popularity and social network analysis.

A very important and well-conducted discussion is made over the scalability of the framework, which is a fundamental requirement for such multimedia services. The fact that experimentations were conducted on a commodity machine somehow demonstrates that the framework can indeed scale with increasing number of inputs, contrarily to other evaluated solutions. The framework itself can be tuned in terms of number of topics and number of

IEEE COMSOC MMTc R-Letter

iterations of the generative process, which gives some additional margin to the system with respect to the load.

Overall, the paper deals with a topic that has key industrial implications. It starts from an intuitive idea and it provides nice theoretical advances. The paper proves again that using social interconnection between different services can improve the overall knowledge of a third-party service.



Gwendal Simon is Associate Professor at Telecom Bretagne. He received his Master Degree in Computer Science in 2000 and his PhD degree in Computer Science in December 2004 from University of Rennes 1 (France). From 2001 to 2006 he was a researcher at Orange Labs, where he

worked on peer-to-peer networks and social

media innovations. Since 2006, he has been Associate Professor at Telecom Bretagne, a graduate engineering school within the Institut Mines-Telecom. He has been a visiting researcher at University of Waterloo from September 2011 to September 2012. His research interests include large-scale networks, distributed systems, optimization problems and video delivery systems.

Karine Pires is a PhD candidate in Telecom Bretagne and Université Pierre et Marie Curie. Advised by Dr. Gwendal Simon and Dr. Sébastien Monnet her studies are on Massive Live Streaming on Over-The-Top (OTT) Structures. She received her Master Degree in Computer Science in 2011 from Universidade Federal do Paraná (Brasil).



Lighting the design of replication algorithms for P2P VoD system in practice

A short review for "On Replication Algorithm in P2P VoD"

Edited by Lifeng Sun

Yipeng Zhou, Tom Z. J. Fu, and Dah Ming Chiu, "On Replication Algorithm in P2P VoD", IEEE/ACM Transactions on Networking, vol. 21, no. 1, February 2013.

Traditional video-on-demand (VoD) is based on the client-server approach. It is expensive and not scalable. The effectiveness of using the P2P approach for content distribution has been proven by many deployed systems [1, 2, 3]. In recent years, the peer-to-peer (P2P) approach was first demonstrated to work for live content streaming [2], and later for VoD streaming as well [4]. Various efforts are working on building a P2P-based VoD platform, for example using set-top boxes [5] where the peers are assumed to be under the control of the content provider.

P2P VoD streaming is definitely harder to accomplish (than live content streaming) since peers are less likely to have the same content to share with each other. To compensate, the new genre of P2P systems requires each user to contribute a small amount of storage (usually 1GB) instead of only the playback buffer in memory as in the P2P streaming systems. This additional resource opens up vast new opportunities for arranging suitable patterns of content replication to meet diverse user demands. Essentially, the new system is a highly dynamic P2P replication system, plus a sophisticated distributed scheduling mechanism for directing peers to help each other in real time.

The goal of the replication strategy is to make the chunks as available to the user population as possible to meet users' viewing demand while without incurring excessive additional overheads. This is probably the most critical part of the P2P-VoD system design. A fundamental question is what the relationship between the storage capacity (at each peer), the number of videos, the number of peers, and the resultant off-loading of video server bandwidth.

Movie popularities can impact server's workload, the conventional wisdom of using the proportional replication strategy is "sub-optimal". [6] expanded the design space to both "passive

replacement policy" and "active push policy" to achieve the optimal replication ratios. In paper [7], a tractable analytical framework was constructed using control theory and dynamic programming, and the optimal strategy was derived with the knowledge of segment popularity. The paper [8] assumes that each peer stores two movies (one of the two movies is the current one being viewed) and each peer will provide upload service with higher priority to peers watching the same movie. In [9], the authors proposed an architecture for P2P VoD system, including an ARIMA module to predict the popularity of each movie. Based on the prediction, they proposed a heuristic algorithm to do replication.

In this paper, the authors used a statistical model to formulate how does the server load scale with the critical system parameters, such as the number of peers in the system, the number of movies, and the storage provided by peers. They propose and analyze a generic replication algorithm Random with Load Balancing (RLB) that balances the service to all movies for both deterministic and random (but stationary) demand models and both homogeneous and heterogeneous peers (in upload bandwidth). Specific contributions of this paper were as follows:

1. A Perfect Fair-Sharing service model and a stationary statistical demand model were proposed for P2P VoD, so that a class of near-optimal movie replication algorithms can be analyzed. In those models, user demand follows a stationary popularity model for movies, and the server is assumed to be offloaded by a processor-sharing-style peer service scheduling.
2. A generic replication algorithm Random with Load Balancing (RLB) was proposed and analyzed, which balances the service to

IEEE COMSOC MMTc R-Letter

all movies for both deterministic and random (but stationary) demand models and both homogeneous and heterogeneous peers (in upload bandwidth). This generic replication algorithm allows to derive (both a closed-form lower bound and upper bound of) the server bandwidth utilization as a function of the key system parameters. The result not only gives the scaling properties of what P2P replication can achieve; it also serves as a rule-of-thumb for needed peer storage for given number of movies.

3. A distributed adaptive RLB (ARLB) algorithm was further proposed, which changes movie replication using viewed movies and can converge to a balanced state given movie popularity.

The breakthrough result of this paper is a general scaling law that P2P VoD has to satisfy in terms of peer population, movie number, and replication capacity at each peer, which leads to several fundamental insights for P2P VoD system design in practice. Experimental results show that the adaptive RLB can outperform many other well-known algorithms.

References:

- [1] “Emule”, <http://www.emule.com/>.
- [2] “PPLive”, <http://www.pplive.com/>.
- [3] “Joost”, <http://www.joost.com/>.
- [4] Y. Huang, T. Z. J. Fu, D.M. Chiu, J. C. S. Lui, and C. Huang, “Challenges, design and analysis of a large-scale P2P VoD system,” in Proc. ACM SIGCOMM, 2008, pp. 375–388.
- [5] N. Laoutaris, P. Rodriguez, and L. Massoulie, “ECHOS: Edge capacity hosting overlays of nano data centers,”

Comput. Commun. Rev., vol. 38, no. 1, pp. 51–54, 2008.

- [6] J. Wu and B. Li, “Keep cache replacement simple in peer-assisted VoD systems,” in Proc. IEEE INFOCOM, 2009, pp. 2591–2595.
- [7] B. R. Tan and L. Massoulie, “Optimal content placement for peer-to-peer video-on-demand systems,” in Proc. IEEE INFOCOM, 2011, pp. 694–702.
- [8] W. Wu and J. C. Lui, “Exploring the optimal replication strategy in P2P-VoD systems: Characterization and evaluation,” in Proc. IEEE INFOCOM, 2011, pp. 1206–1214.
- [9] C. Loeser, G. Schomaker, A. Brinkmann, M. Vodisek, and M. Heidebuer, “Content distribution in heterogeneous video-on-demand P2P networks with ARIMA forecasts,” Lecture Notes Comput. Sci., vol. 3421, pp. 800–809, 2005.



Lifeng Sun received the B.S. and Ph.D. degrees in System Engineering from National University, of Defense Technology China, in 1995 and 2000, respectively. He joined the Department of Computer Science and Technology, Tsinghua University (THU), Beijing, China, in 2001. Currently, he is a Professor in CST of THU. His research interests include video streaming, 3D video processing and Social Media. He is a Member of IEEE.

The Design of Next Generation Multimedia Synchronization Algorithms

A short review for "Evolution of temporal multimedia synchronization principles: A historical viewpoint"

Edited by Irene Cheng

Zixia Huang and Klara Nahrstedt, "Evolution of Temporal Multimedia Synchronization Principles: A Historical Viewpoint", ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 9, No. 1s, Article 34, October 2013.

New media and communication technologies have introduced new challenges to the design of multimedia synchronization algorithms. The landscape of multimedia content is no longer restricted to image and video, but includes more dynamic and multi-modal streams such as motion capture and sensor-based interaction data. In order to understand the past and flourish future development, this article "gives an excellent overview of Synchronization *through the ages* until today," as described by the nominator. As well, it proposes a synchronization framework to tackle new challenges. This article is informative for readers who are interested in the development of temporal multimedia synchronization techniques from a historical perspective, which includes a discussion on "The picture of the future" back in 1969 [1] archived by the Bell Lab. On the other hand, this article is also inspiring for readers who wonder where the technology will likely to go.

The transition from analog modulation to single playback media, to multi-modal multi-channel content, and so on has necessitated new strategies. Traditional approaches in communications aim at delivering optimal Quality of Services (QoS), which is measured largely based on quantitative and probabilistic analysis for the best resources distribution, e.g. bandwidth and time. With increasing attention on Quality of Experience (QoE), the measuring criteria tend to put weight on the satisfaction of consumers instead of simply relying on the numerical comparison, e.g. PSNR. Thus, human perceptual factors become a main contributor in the evaluation of multimedia synchronization techniques; in particular for temporal multimedia that involves visualization on heterogeneous displays. "Synchronization Perception" is discussed in various subsections in the article. Nevertheless, it is interesting to point out that the start of synchronization perception studies can be traced back to 1978 [2] which offers references to demonstrate the impact of jitter on intra-media synchronization of digital audio. As the result of this work, the maximum tolerable intra-stream skew for 16-bit high quality audio is 200 ns in one sampling period. Related work can be found in 1972 [3], which

suggests a maximum allowable intra-stream skew of no more than 5-10 ns.

Before the literature review, the authors explain the synchronization formulation and explain the hierarchical structure in a continuous multimedia data model, which is composed of Session, Bundle, Media Modality, Sensory Stream and Media Frame. Because of the hierarchical multisite multisensory nature of the multimedia data, four synchronization layers, namely Intro-stream, Intra-media, Intro-bundle and Intra-session, are depicted. The authors give a definition of synchronization skew in continuous multimedia as "the delay difference of two time correlated media objects (media frame, sensory stream, media modality, or participating site), traveling from the media sources to the current location." It is pointed out that due to the multilayer synchronization hierarchy, a media object can be represented in multiple forms, leading to the possibility of multiple skews instead of a single skew to describe the entire multimedia session.

The authors review the synchronization studies for continuous multimedia developed over the last 30 years. The evolution is categorized into three generations: Years of Birth (On and before 1988), Years of Understanding (1989-1994), Years of Blossoms (1995-2001) and Years of Leaps (On and after 2002).

- Years of Birth – The rapid development of digital computing and communication technology with unreliable characteristics aroused researchers' attention to the digital multimedia synchronization problem.
- Years of Understanding – Extensive research was done to investigate the synchronization problem, in order to catch up with the technological advances of the Internet protocols (IP).
- Years of Blossoms – Multimedia synchronization stayed as a hot and important topic as a result of the evolutionary change in Internet quality.

IEEE COMSOC MMTC R-Letter

- Years of Leaps – This is also the generation of increasing challenges due to high accessibility of computation and communication resources, as well as advances in hardware and software.

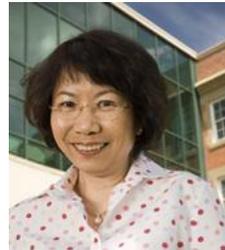
The historical review presents the readers an understanding of how synchronization formulations evolve leading to new challenges. The authors then conclude with several observations:

1. There is no classification model that captures all synchronization requirements.
2. New multimedia systems are far broader than traditional conferencing and on-demand applications, and thus the synchronization reference has to be chosen dynamically and adaptably.
3. Insufficient study has been performed to investigate the heterogeneity of computation demand, and its impact on synchronization.

Motivated by these observations, a new multidimensional synchronization framework is introduced in this article to address the next generation challenges encountered in emerging applications, e.g. interactive 3D Tele-immersion. These next generation multimedia systems come with the demands of scale and device heterogeneity, multi-location synchronization controls, and diverse applications on a single multimedia platform. In order to validate the proposed framework, comparison and evaluation results are presented in this article. Detailed technical discussion can be found in the original paper.

References:

- [1] Bell Lab., “The picture of the future,” Bell Labs Record 47, 134-186
- [2] Blesser, B., “Digitization of audio: A comprehensive examination of theory, implementation, and current practice,” J. Audio Engineering Society, 1978, 26, 739-771.
- [3] Stockham, T., “A/D and D/A converters: Their effect on digital audio fidelity,” In Digital Signal Processing, IEEE Press, 1972, 55-66.



Irene Cheng, SMIEEE is the Scientific Director of the Multimedia Research Centre, and an Adjunct Professor in the Faculty of Science, as well as the Faculty of Medicine & Dentistry, University of Alberta, Canada. She is also a Research Affiliate with the Glenrose Rehabilitation Hospital in Alberta, Canada. She

was a visiting profession in INSA Lyon, France and a Co-Chair of the IEEE SMC Society, Human Perception in Vision, Graphics and Multimedia Technical Committee; was the Chair of the IEEE Northern Canada Section, Engineering in Medicine and Biological Science (EMBS) Chapter, and the Chair of the IEEE Communication Society, Multimedia Technical Committee 3D Processing, Render and Communication (MMTC) Interest Group. She is now the Director of the Review-Letter Editorial Board of MMTC. Over the last ten years, she has more than 130 international peer-reviewed publications including 2 books and more than 30 journals. Her research interests include multimedia communication techniques, Quality of Experience (QoE), Levels-of-detail, 3D Graphics Visualization and Perceptual Quality Evaluation. In particular, she introduced applying human perception – Just-Noticeable-Difference – following psychophysical methodology to generate multi-scale 3D models.

Converting 2D to 3D by Learning from Examples

A short review for "Learning-based, automatic 2D-to-3D image and video conversion"

Edited by Jun Zhou

Janusz Konrad, Meng Wang, Prakash Ishwar, Chen Wu, and Debargha Mukherjee. "Learning-based, automatic 2D-to-3D image and video conversion", IEEE Transactions on Image Processing, Vol. 22, No. 9, pages 3485-3496, 2013.

3D content production has become increasingly needed with the wide spread of consumer 3D display hardware and rapid development of 3D movie industries. This has made 2D-to-3D image or video conversion an urgent research topic, in which two most important tasks are depth estimation for a given 2D image and depth-based rendering.

Depth estimation approaches can be classified as automatic or semi-automatic. Semi-automatic approach has been widely accepted by research communities and industries because of its high accuracy. It requires a human operator to assign depth to pixels or regions in images or several frames of videos. Such knowledge can be learned by computational methods and then be used to guide the depth estimation for other images or frames [1,2]. Automatic depth estimation is much more challenging. Recent efforts include shape-from-shading [3], graph cut [4], and learning based methods [5, 6].

In this paper, the authors tackle the depth estimation problem from two aspect of view: how to generate depth map from a single image, and how to use a collection of images or video frames to estimate a global depth map. The novel idea is using machine learning approach to explore big data in large image or video datasets. The goal is to perform 2D-to-3D conversion on arbitrary scenes without human intervention.

The depth estimation on single image aims at assigning a depth value to each pixel in the scene, which is therefore considered as a local level estimation task. This method is based on the assumption that image or video attributes, such as color, spatial location, and local motion, have some sort of dependency with the depth information. Such dependency can be learned from large number of training data so as to generate statistical attribute-depth relations. For example, what is the probability that blue color, which normally correspond to sky, is associated with large depth values; or when motion is

concerned, the probability that moving objects is close to the viewer than the background. The final output of the learning, which is based on a linear regression model, is a set of weights for the attribute-depth transformations. The linear combination of these transformations forms the final depth field.

What's more interesting in this paper is the 2D-to-3D conversion by learning from large image sets or videos which is considered as a global method. An assumption here is that a 3D dataset is available in advance. Given a query image, which is treated as the left image in the stereo pair, this method tries to find the optimal images or frames to provide reliable estimation of the depth map and the right image for the stereo pair. Such strategy is implemented in four steps. The first step is to perform K-nearest neighbour search for the 2D query by comparing the histograms of oriented gradients between query and target images in the dataset. This step selects images that are mostly or partly consistent with the query. Then a smooth depth field is generated by applying a median operator on the depth map of all retrieved images. To allow more edges be recovered to better match the query image, a cross-bilateral filtering method [7] is applied. Finally, a disparity map is generated to estimate the right image.

The proposed approach was tested on two benchmark datasets. The results show that both local and global methods are quite effective in reconstructing the depth map. The global method has outperformed two alternative methods in most cases, and has demonstrated excellent efficiency by three orders of magnitude faster than two alternative methods.

In summary, this paper has addressed a very interesting problem in multimedia content generation. The proposed learning from examples approach has shown promising performance in both efficiency and consistency between estimated depth and the ground truth. It will

IEEE COMSOC MMTC R-Letter

enlighten more learning based 2D-to-3D conversion research, and will form the foundation of developing more practical approaches for the multimedia industry.

References:

- [1] R. Phan, and D. Androutsos. "Robust semi-automatic depth map generation in unconstrained images and video sequences for 2D to stereoscopic 3D conversion". *IEEE Transactions on Multimedia*, Vol. 16, No. 1, pages 122-136, 2013.
- [2] M. Guttman, L. Wolf, and D. Cohen-Or, "Semi-automatic stereo extraction from video footage," in *Proceedings of the IEEE International Conference on Computer Vision*, pages 136-142, 2009.
- [3] R. Zhang, P. S. Tsai, J. Cryer, and M. Shah, "Shape-from-shading: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 8, pages 690-706, 1999.
- [4] R. Phan, R. Rzeszutec, and D. Androutsos, "Semi-automatic 2D to 3D image conversion using scale-space random walks and a graph cuts based depth prior," in *Proceedings of the IEEE International Conference on Image Processing*, pages 865-868, 2011.
- [5] B. Liu, S. Gould, and D. Koller, "Single image depth estimation from predicted semantic labels," in *Proceedings of the IEEE Conference on Computer Vision and*

Pattern Recognition, pages 1253-1260, 2010.

- [6] A. Saxena, M. Sun, and A. Ng, "Make3D: Learning 3D scene structure from a single still image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 5, pages 824-840, 2009.
- [7] L. Angot, W.. Huang, and K. Liu, "A 2D to 3D video and image conversion technique based on a bilateral filter," *Proceedings of the SPIE*, Vol. 7526, pages 75260D, 2010.



Jun Zhou received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Canada, in 2002, and the

Ph.D. degree in computing science from University of Alberta, Canada, in 2006.

He joined the School of Information and Communication Technology in Griffith University as a lecturer in June 2012. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA. His research interests are in statistical pattern recognition, interactive computer vision, and their applications to hyperspectral imaging and environmental informatics.

Sparse Representation Assists Video Tagging

A short review for "Video-to-Shot Tag Propagation by Graph Sparse Group Lasso"

Edited by Vladan Velisavljević

X. Zhu, Z. Huang, J. Cui, H. T. Shen, "Video-to-Shot Tag Propagation by Graph Sparse Group Lasso," IEEE Transactions on Multimedia, Vol. 15, No. 3, pp. 633-646, April 2013.

Video tagging is an important aspect of automatic video content annotation, where textual labelling is associated to various semantic concepts (e.g. objects, persons, locations, etc.) [1]. Within general video tagging, the problem of tag propagation from the entire video to video segments of the same scene called shots plays a significant role in refining video tagging and providing the users with better structured label information. Such an approach has not been properly addressed in the existing solutions, such as [2], where tagging has been implemented only at the level of the entire video.

The authors of this paper exploit the correlation between visual features in a test shot and training videos using a reconstruction model to enable propagation of the tags from the videos to the shots. The model is based on a novel method for so-called, shot reconstruction, the process of linear reconstruction of visual features of a given shot from visual features of the known videos. In this process, the shots in video sequences are considered as the bases, whereas each video sample is represented as a group of shots.

The proposed shot reconstruction algorithm uses the least absolute shrinkage and selection operator (lasso) [3]. The lasso operator searches for the best basis representation given data samples by minimizing the Euclidean distance weighted by the representation sparsity metric. While the l_0 -norm would serve as the ideal sparsity metric, the associated best basis search process becomes NP-hard and, thus, inconvenient for practical implementation. Instead, the l_1 -norm is used as an approximation, which leads to an efficient process and satisfying results under practical conditions [4]. Such an approximated search based on the l_1 -norm is also referred to as basis pursuit [3].

The proposed method in this paper tailors the lasso operator for application in video shot reconstruction. The authors have identified two major issues in the classical lasso operator. First, the regularization property inherited from using

the l_1 -norm leads to a suboptimal number of training videos used in the shot reconstruction. The main reason for this issue lies in the fact that the training videos are not considered as groups of video shots. Second, orthogonalization implemented in basis representation makes it difficult to update the shot reconstruction in case any video content in the training set is changed.

To improve sparsity of the lasso operator, the authors exploit two modifications of the classical operator and also relax the orthogonality condition. First, to enforce sparsity across video groups, the regularization factor is modified to include a prior knowledge of the training videos. By such a modification, a test shot is represented by the training video instead of a collection of video shots and, when the factor is chosen to measure the l_2 -norm, it ensures the maximally sparse representation of the shot across the video groups. This operator is called group lasso [5]. Second, to improve sparsity within the group, another l_1 -norm regularization factor is added to the optimization basis search to penalize the entire reconstruction error. Such an operator is called sparse group lasso and it is a generalization of the previous two lasso operators [6].

Armed with the group and sparse group lasso operators, the authors propose a novel graph-based sparse group lasso (GSGL). The GSGL assigns temporal and spatial weights to each shot to capture the information within individual video shots and across shots within each video, respectively. In addition, similarity weights are computed as a metric to measure the similarity between the test shot and the training videos. These three types of weights are used jointly in a new optimization objective function that combines the regularization factors for each weight. Finally, to simplify the optimization basis search, the authors prove that the novel complex objective function can be transformed to a similar form as used by the sparse group lasso to benefit from the existing lasso

IEEE COMSOC MMTC R-Letter

optimization techniques.

In the presented experiments, the authors demonstrate the efficiency and superior performance of the proposed method. The obtained metrics are compared to their previous achievements and to several related methods, where the novel method indeed outperforms the others. Furthermore, the authors exhaustively analyze the influence of the parameter selection on the final performance, which gives their paper an additional credibility.

In summary, the proposed method exploits sparsity of video representation for shots and video sequences to address two scenarios. First, tags from annotated videos are localized to each individual shot. Second, semantic tags are automatically assigned to selected video test shots. For both scenarios, the method is based on modified sparse group lasso operator that enforces sparse representation among video shots and also within individual video shots and preserves similarity across video shots and sequences. The method exhibits an improved performance as compared to the underlying methods without the modifications. The work presents a solid mathematical expansion on top of the well-developed basis pursuit optimization problem. It also demonstrates that sparsity in video representation can boost the performance of even such a high-level semantic video processing, such as annotation and tagging.

References:

- [1] Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S. Chua, and X.-S. Hua, "Visual query suggestion: towards capturing user intent in internet image search," *ACM Trans. Multimedia Comput. Commun. Appl.*, Vol. 6, No. 3, pp. 1-19, 2010.
- [2] Y. Li, Y. Tian, L.-Y. Duan, J. Yang, T. Huang, and W. Gao, "Sequence multi-labeling: a unified video annotation scheme with spatial and temporal context," *IEEE Trans. Multimedia*, Vol. 12, No. 8, pp. 814-828, 2010.
- [3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, Vol. 20, pp. 33-61, 1998
- [4] B. Efron, T. Hastie, L. Johnstone, and R.

Tibshirani, "Least angle regression," *Ann. Statist.*, Vol. 32, pp. 407-499, 2004.

- [5] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc., Series B*, Vol. 68, pp. 49-67, 2006.
- [6] J. Peng, J. Zhu, A. Bergamaschi, W. Han, D.-Y. Noh, J. R. Pollack, and P. Wang, "Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer," *Ann. Appl. Statist.*, Vol. 4, No. 1, pp. 53-77, 2010.



Vladan Velisavljević received the B.Sc. and M.Sc. (Magister) degree from the University of Belgrade, Serbia, in 1998 and 2000, respectively, and the Master and Ph.D. degree from EPFL, Lausanne, Switzerland, in 2001 and 2005.

From 1999 to 2000, he was a member of academic staff at the University of Belgrade. In 2000, he joined the Audiovisual Communications Laboratory (LCAV) at EPFL as teaching and research assistant, where he was working on his Ph.D. degree in the field of image processing. In 2003, he was a visiting student at Imperial College London. From 2006 to 2011, Dr. Velisavljević was a Senior Research Scientist at Deutsche Telekom Laboratories, Berlin, Germany. Since October 2011, he is Senior Lecturer (Associate Professor) at Bedfordshire University, Luton, UK, where he also serves as a Deputy Head of the Centre for Wireless Research.

He has co-authored more than 50 research papers published in peer-reviewed journals and conference proceedings and he has been awarded or filed 4 patents in the area of image and video processing. Vladan is a Lead Guest Editor for a special issue on visual signal processing for wireless networks in the *IEEE JSTSP* and he also co-organized a special session at *IEEE ICIP-2011* on compression of high-dimensional media data for interactive navigation. He has been a co-chair of the *Multimedia Computing and Communications Symposium (MCC)* at *ICNC-2013* and *2014*. His research interests include image, video and multiview video compression and processing, wavelet theory, multi-resolution signal processing and distributed image/video processing.

Using Near-Infrared Image for Dehazing

A short review for "Near-Infrared Guided Color Image Dehazing"

Edited by Gene Cheung

Chen Feng, Shaojie Zhuo, Xiaopeng Zhang, Liang Shen, Sabine Susstrunk, "Near-Infrared Guided Color Image Dehazing," *IEEE International Conference on Image Processing*, September, 2013.

When taking a landscape photo using a consumer-grade camera, the presence of haze or mist can degrade the quality of the acquired image [1]. In particular, haze effect is due to the presence of atmospheric particles of size comparable to the wavelength in the visual band (around 0.1 μ m) that absorb and scatter light. The net result is that the reflected light from distant objects is attenuated and diffused. It is thus desirable to remove the haze effect—called *dehazing*—after a color image has been captured via image processing techniques. The key challenge in image dehazing is twofold: i) to estimate and remove the airlight color, and ii) to recover lost details in the color image. The review paper addresses this image dehazing problem.

Previous attempts at the dehazing problem fall into two categories. Methods in the first category [2, 3, 4] tried to remove haze given a single RGB image, using certain image assumptions such as dark channel [2]. However, when scene details are lost in the lone image, it is very difficult to recover them, and thus the quality in dense haze regions in the recovered image tends to be poor. Methods in the second category [1] assumed multiple images of the same scene are available for dehazing. For example, [1] assumed that two images are captured with different mediums, while [5] assumed that images are taken with different degrees of polarization, for example, by rotating a polarizing filter attached to the camera between shots. Neither of these approaches are realistic or practical for outdoor scenes where objects like trees and clouds move quickly.

Leveraging on the authors' previous work [6], the novel setup in the review paper is to assume the availability of an additional near-infrared (NIR) image of the same scene for dehazing. NIR has a longer wavelength (about 1 μ m), which has the advantage of deeper penetration and thus is capable of unveiling scene details not observable in the corresponding RGB image. It is argued [6]

that an off-the-shelf camera can be modified slightly to acquire NIR spectrum.

Given the available NIR image, the authors address the previously described two challenges as follows. In the first stage, by exploiting dissimilarity between NIR and other color bands, the airlight color is estimated. Second, by enforcing a NIR gradient constraint through an optimization framework, details in the RGB image are recovered in an image dehazing stage. We briefly describe the details of the two stages below.

The image model used is from [2]:

$$\mathbf{I}(x) = t(x) \mathbf{J}(x) + (1-t(x)) \mathbf{A} \quad (1)$$

where for each pixel x , $\mathbf{I}(x)$ is the observed pixel, $\mathbf{J}(x)$ is the haze-free image pixel, \mathbf{A} is the global airlight color, and $t(x)$ is the medium transmission describing the portion of light that reaches the camera correctly. In the first stage, the goal is to estimate \mathbf{A} . The authors consider that transmission t depends on the scene depth and the density of the haze, while color \mathbf{J} depends on the illumination of the scene and surface reflectance. Thus in a small patch, one can assume that t and \mathbf{J} are uncorrelated. The idea is then to find a local patch Ω with pixels having large similarities, followed by searching an airlight color \mathbf{A} that leads to the smallest possible correlation between t and \mathbf{J} .

In the second stage of image dehazing, the authors formulate an optimization problem. Basically, we are required to recover \mathbf{J} given \mathbf{I}^{RGB} and \mathbf{I}^{NIR} . Statistically, the problem can be reformulated as finding the largest joint probability of (\mathbf{J}, t) given \mathbf{I}^{RGB} and \mathbf{I}^{NIR} . Based on Bayes' theorem, one can write:

IEEE COMSOC MMTC R-Letter

$$P(\mathbf{J}, t | \mathbf{I}^{RGB}, \mathbf{I}^{NIR}) = \frac{P(\mathbf{I}^{RGB}, \mathbf{I}^{NIR} | \mathbf{J}, t)P(\mathbf{J}, t)}{P(\mathbf{I}^{RGB}, \mathbf{I}^{NIR})} \quad (2)$$

$$\propto P(\mathbf{I}^{RGB} | \mathbf{J}, t)P(\mathbf{I}^{NIR} | \mathbf{J}, t)P(\mathbf{J})P(t)$$

where $P(\mathbf{I}^{RGB} | \mathbf{J}, t)$ corresponds to the haze model in (1), $P(\mathbf{I}^{NIR} | \mathbf{J}, t)$ represents the relationship between the color and NIR images (called the *NIR constraint*), and $P(\mathbf{J})$ and $P(t)$ are the color image prior and transmission prior, respectively.

Based on the statistical analysis in (2), the authors propose an objective function for reconstructed image \mathbf{J} as follows:

$$\arg \min_{(\mathbf{J}, t)} \|t\mathbf{J} + (1-t)\mathbf{A} - \mathbf{I}^{RGB}\|^2$$

$$+ \lambda_1 w |\nabla \mathbf{J} - \nabla \mathbf{I}^{NIR}|^\alpha + \lambda_2 |\nabla \mathbf{J}|^\beta + \lambda_3 \|\nabla t\|^2 \quad (3)$$

where α, β are in (0,1). The first term originates from the chosen image model in (1). The second term means the details in the NIR image can be transferred to the reconstructed image, given the longer wavelength of NIR penetrates better than visible band. The third and fourth terms are smoothness priors for the reconstructed image \mathbf{J} and transmission map t .

The optimization problem (3) is solved using *Iteratively Reweighted Least Squares* (IRLS). Experimental results show that the proposal can reproduce more details in the reconstructed image, particularly for distant objects.

References:

[1] S. K. Nayar, S. G. Narasimhan, "Vision in Bad Weather," *IEEE International Conference on Computer Vision (ICCV)*, 1999.

[2] K. M. He, J. Sun, X. O. Tang, "Single Image Haze Removal using Dark Channel Prior," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[3] R. Fattal, "Single Image Dehazing," *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 2008.

[4] C. O. Ancuti, C. Ancuti, C. Hermans, P. Bekaert, "A Fast Semi-inverse Approach to Detect and Remove the Haze from a Single Image," *10th Asian Conference on Computer Vision (ACCV)*, 2008.

[5] S. Shwartz, E. Namer, Y. Schechner, "Blind Haze Separation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[6] N. Salamati, A. Germain, S. Susstrunk, "Removing Shadows from Images using Color and Near-Infrared," *IEEE Conference on Image Processing (ICIP)*, 2011.



Gene Cheung received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively. He was a senior researcher in Hewlett-Packard

Laboratories Japan, Tokyo, from 2000 till 2009. He is now an associate professor in National Institute of Informatics in Tokyo, Japan. His research interests include image & video representation, immersive visual communication, and graph signal processing. He has served as associate editor for *IEEE Transactions on Multimedia* from 2007 to 2011. He currently serves as associate editor for *DSP Applications Column* in *IEEE Signal Processing Magazine* and *APSIPA journal* on signal and information processing, and as area editor in *EURASIP Signal Processing: Image Communication*.

Exploring frequency-domain oversampling for multicarrier transmissions in Underwater Acoustic Communications

A short review for "Frequency-Domain Oversampling for Zero-Padded OFDM in Underwater Acoustic Communications"

Edited by Weiyi Zhang

Zhaohui Wang, Shengli Zhou, Georgios B. Giannakis (Fellow of IEEE), Christian R. Berger, and Jie Huang, "Frequency-Domain Oversampling for Zero-Padded OFDM in Underwater Acoustic Communications", IEEE Journal of Oceanic Engineering, pp. 14-24, VOL. 37, NO. 1, January 2012.

Zero-padded (ZP) orthogonal frequency division multiplexing (OFDM) has been extensively investigated for high data rate underwater acoustic communications [1,2,3].

Following Doppler shift compensation and an overlap-add operation, fast Fourier transform (FFT) is performed on the received block to obtain frequency-domain samples, that are used for subsequent channel estimation and data detection. In spite of its known suboptimality, the overlap-add operation is used in most ZP-OFDM receivers. This is because on channels that are linear, time-invariant, or can be approximated as such after proper processing [1], [2], the overlap-add operation preserves the orthogonality among subcarriers, which enables low-complexity equalization and demodulation. This is no longer the case on strongly time-varying channels [3], where intercarrier interference (ICI) impairs subcarrier orthogonality, thus requiring adjacent subcarriers to be jointly demodulated.

This paper proposes a frequency-domain oversampling method to improve the system performance of zero-padded OFDM underwater acoustic communication system with large Doppler spread. Their smart design allows the receiver with frequency-domain oversampling outperforms the conventional one significantly.

In detail, the authors consider the same ZP-OFDM signal design as in [4], [5] that separate data subcarriers from pilot subcarriers using interspersed null subcarriers. This way, channel estimation and data detection can be carried out separately at the receiver, even in channels with large Doppler spread. The authors further develop a frequency-domain oversampling receiver, which relies on compressed sensing techniques for sparse channel estimation and minimum mean-square error (MMSE) equalization for data detection. The receiver complexity is only increased marginally by the frequency-domain oversampling: the FFT size increases proportionally and the equalizers process more inputs – but the

equalizer complexity is dominated by the matrix inversion which scales with the number of data symbols – not the observations. Besides the consideration of the rectangular pulse-shaping window, this work also considers raised-cosine windows in the signal design to further alleviate the ICI. This paper evaluates the performance of the proposed receiver using both simulated and real data collected from the SPACE08 experiment, conducted off the coast of Martha's Vineyard, Massachusetts, October 2008, and the WHOI09 experiment, conducted in the Buzzards Bay, Massachusetts, December 2009. Simulation results demonstrate that frequency-domain oversampling improves the system performance considerably, where the performance gain increases as the channel Doppler spread increases. Experimental results verify the benefits of frequency-domain oversampling in achieving similar performance with fewer phones than the receiver without oversampling. Interestingly, although a raised-cosine pulse-shaping window improves performance relative to a rectangular window, the performance gain is less pronounced when using frequency-domain oversampling. Overall, numerical and experimental results demonstrated that the proposed frequency-domain oversampling improves the system performance considerably, and the gain becomes larger as the channel Doppler spread increases.

References:

- [1] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, "Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts," IEEE J. Ocean. Eng., vol. 33, no. 2, Apr. 2008.
- [2] M. Stojanovic, "Low complexity OFDM detector for underwater channels," in Proc. of MTS/IEEE OCEANS Conf., Boston, MA, Sept. 18-21, 2006.
- [3] C. R. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," IEEE Trans. Signal Processing, vol. 58, no. 3, pp. 1708–1721, Mar. 2010..

IEEE COMSOC MMTC R-Letter

- [4] S. Mason, C. R. Berger, S. Zhou, K. Ball, L. Freitag, and P. Willett, "An OFDM design for underwater acoustic channels with Doppler spread," in Proc. of the 2009 DSP & SPE Workshop, Marco Island, FL, Jan. 2009.
- [5] —, "Receiver comparisons on an OFDM design for Doppler spread channels," in Proc. of MTS/IEEE OCEANS Conf., Bremen, Germany, May 2009.

Weiyi Zhang is currently a Senior Research Staff Member of the Network Evolution Research Department at AT&T Labs Research, Middletown, NJ. Before join AT&T Labs Research, he was an Assistant Professor at the Computer Science Department, North Dakota State University, Fargo, North Dakota, from 2007 to 2010. His research interests include routing, scheduling, and cross-layer design in wireless networks, localization and coverage issues in wireless sensor networks, survivable design and quality-of-service provisioning of communication networks. He has published more than 80 refereed papers in his research areas, including papers in prestigious conferences and journals



such as IEEE INFOCOM, ACM MobiHoc, ICDCS, IEEE/ACM Transactions on Networking, ACM Wireless Networks, IEEE Transactions on Vehicular Technology and IEEE Journal on Selected Areas in Communications. He received AT&T Labs Research Excellence Award in 2013, Best Paper Award in 2007 from IEEE Global Communications Conference (GLOBECOM'2007). He has been serving on the technical or executive committee of many internationally reputable conferences, such as IEEE INFOCOM. He was the Finance Chair of IEEE IWQoS'2009, and serves the Student Travel Grant Chair of IEEE INFOCOM'2011.

Paper Nomination Policy

Following the direction of MMTC, the R-Letter platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia include, but are not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication.

Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings or other distinguished journals/conferences, within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to R-Letter Editorial Board Directors:

Irene Cheng (locheng@ualberta.ca),
Weiyi Zhang (maxzhang@research.att.com), and
Christian Timmerer
(christian.timmerer@itec.aau.at)

The nomination should include the complete reference of the paper, author information, a

brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper when possible.

Review Process

Each nominated paper will be reviewed by members of the IEEE MMTC Review Board. To avoid potential conflict of interest, nominated papers co-authored by a Review Board member will be reviewed by guest editors external to the Board. The reviewers' names will be kept confidential. If two reviewers agree that the paper is of R-letter quality, a board editor will be assigned to complete the review letter (partially based on the nomination supporting document) for publication. The review result will be final (no multiple nomination of the same paper). Nominators external to the board will be acknowledged in the review letter.

R-Letter Best Paper Award

Accepted papers in the R-Letter are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board).

For more details, please refer to <http://committees.comsoc.org/mmc/rletters.asp>

MMTC R-Letter Editorial Board

DIRECTOR

Irene Cheng
University of Alberta
Canada

CO-DIRECTOR

Weiyi Zhang
AT&T Research
USA

CO-DIRECTOR

Christian Timmerer
Alpen-Adria-Universität Klagenfurt
Austria

EDITORS

Koichi Adachi
Institute of Infocom Research, Singapore

Pradeep K. Atrey
University of Winnipeg, Canada

Gene Cheung
National Institute of Informatics (NII), Tokyo, Japan

Xiaoli Chu
University of Sheffield, UK

Ing. Carl James Debono
University of Malta, Malta

Guillaume Lavoue
LIRIS, INSA Lyon, France

Joonki Paik
Chung-Ang University, Seoul, Korea

Lifeng Sun
Tsinghua University, China

Alexis Michael Tourapis
Apple Inc. USA

Vladan Velisavljevic
University of Bedfordshire, Luton, UK

Jun Zhou
Griffith University, Australia

Jiang Zhu
Cisco Systems Inc. USA

Pavel Korshunov
EPFL, Switzerland

Marek Domański
Poznań University of Technology, Poland

Hao Hu
Cisco Systems Inc., USA

Cyril Concolocato
Telecom ParisTech, France

Carsten Griwodz
Simula and University of Oslo, Norway

Frank Hartung
FH Aachen University of Applied Sciences, Germany

Gwendal Simon
Telecom Bretagne (Institut Mines Telecom), France

Roger Zimmermann
National University of Singapore, Singapore

Michael Zink
University of Massachusetts Amherst, USA

Multimedia Communications Technical Committee (MMTC) Officers

Chair Jianwei Huang

Steering Committee Chair Pascal Frossard

Vice Chair – North America Chonggang Wang

Vice Chair – Asia Yonggang Wen

Vice Chair – Europe Luigi Atzori

Vice Chair – Letters & Member Communications Kai Yang

Secretary Liang Zhou

MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.