MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
*http://mmc.committees.comsoc.org/*

# IEEE COMSOC MMTC Communications – Review

Vol. 9, No. 1, February 2018

IEEE COMMUNICATIONS SOCIETY

## TABLE OF CONTENTS

# Message from the Review Board Directors

Welcome to the February 2018 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises a number of reviews that cover multiple facets of multimedia communication research including social event analysis, wireless networks, cloud security, and quality of experience. These reviews are briefly introduced below.

The first paper, published in IEEE Communications Surveys and edited by Wei Wang, provides a comprehensive survey of wireless multimedia communication techniques over cognitive radio network technologies.

The second and third papers are related to multimedia vision area. The second paper is published in IEEE Transactions on Multimedia and edited by Chidansh Bhatt. It looks into indoor localization issue using magnetic and visual sensors. The third paper, published in IEEE Transactions on Circuits and Systems for Video Technology Volume and edited by Chidansh Bhatt, presents a context-aware viewpoint recommendation system for capturing high-quality photographs.

The fourth paper, published in IEEE Transactions on Multimedia and edited by Jun Zhou, resolves friend recommendation issues in social networks. It investigates a two-stage probabilistic topic model to analyze the relationship between data from different domains.

The fifth paper, published in IEEE International Conference on Multimedia and Expo and edited by Bruno Macchiavello, studies the projection as an additional reference frame in estimation for structure-based frame video coding prediction.

All the authors, nominators, reviewers, editors, and others who contribute to the release of this issue deserve appreciation with thanks.

IEEE ComSoc MMTC Communications – Review Directors

Pradeep K. Atrey
State University of New York at Albany, USA
Email: patrey@albany.edu

Wei Wang
San Diego State University, USA
Email: wwang@mail.sdsu.edu

Qing Yang
Montana State University, USA
Email: qing.yang@montana.edu

Jun Wu
Tongji University, China
Email: wujun@tongji.edu.cn

# Review of Wireless Multimedia Communication in Cognitive Radio Networks

*A short review for "Wireless Multimedia Cognitive Radio Networks: A Comprehensive Survey"*
Edited by Wei Wang

*Muhammad Amjad, Mubashir Husain Rehmani, and Shiwen Mao, "Wireless Multimedia Cognitive Radio Networks: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, in print 2018. doi: 10.1109/COMST.2018.2794358.*

Multimedia applications are regarded as the time-critical and bandwidth hungry applications. Support of multimedia application in cognitive radio networks (CRNs) has been studied from various dimensions. This paper [1] has encompassed state-of–the-art work related to the transmission of multimedia applications using the CRNs. Wireless multimedia CRNs (WMCRNs) exploit the idle spectrum (white spaces) while remaining within the scarce spectrum resources [2]. This paper surveyed every possible perspective involving WMCRNs.

Diverse range of applications including video-on-demand, video conferencing, medical, and 3D applications can now be supported by CRNs [3]. These delay-sensitive and time-critical applications have been reviewed and compared with their supporting architectures. Different studies considering these studies have also been surveyed with their evaluation metrics. Most of the existing work on WMCRNs considers the simple video streaming applications; however other critical applications such as video caching and video conferencing have not been explored much by researchers.

Routing protocols for WMCRNs have been classified into three classes namely: QoS-aware, hop-count based, and routing protocols for streaming videos. The existing work on QoS-aware routing has further been studied with respect to CR-adhoc networks, Mobile CRNs, CR-mesh networks, CR-smart grid, and CR-cellular networks. It is clear from the review of routing protocols for WMCRNs that mobility and spectrum hand-off has not been considered. This survey paper also highlights that multiple channel support in routing protocols for WMCRNs has not been considered by the existing literature.

Extensive work on medium access control (MAC) protocol for WMCRNs has been done to support the multimedia applications. The work on MAC protocol has been studied by taking into consideration the centralized and distributed nature of MAC protocols. Cross-layer and multiple channel support in conjunction with the work on MAC protocol for WMCNRs has been surveyed with higher level insights [4]. User perspective of the multimedia communications in CRNs has been provided with the name of quality-of-experience (QoE). QoE has been studied with respect to its objective and subjective-based metrics.

Every effort has been made to encompass each potential aspect of WMCRNs with the higher level insights. Cross-layer designs utilized for achieving the multimedia support in CRNs has been reviewed with the notion of QoE and quality-of-service (QoS) support. An-depth insightful review of QoS-aware communications in CR-adhoc networks, CR-mesh networks, CR-sensor networks, CR-smart grid, CR-cellular networks, CR-WLAN, and CR-heterogeneous have been provided with their different evaluation metrics and supported applications.

White space utilization is usually classified into underlay, overlay, interweave, and hybrid CRNs. Therefore, white space utilization in WMCRNs has also been provided with respect to underlay, overlay, interweave, and hybrid WMCRNs. In addition to traditional white space, TV white space utilization for transmitting the time-critical data has also been explored by considering the QoS-support [5]. Cyclostationary and energy-detection based spectrum sensing in WMCRNs have been surveyed with different evaluation metrics and supported applications.

This paper also highlights several concrete future research directions that will open the new vistas in the field of WMCRNs. The concepts of network coding, energy-harvesting, beamforming, polarization, full-duplex, and virtualization have still to be explored in the future [6]. Open issues related to the security, MAC layer, routing layer, spectrum-sensing, green communication, and cross-layer designs with their potential solutions have also been highlighted in detail.

## Acknowledgement

The editors would like to thank the authors for providing the short review summary.

## References:

[1] M. Amjad, M. H. Rehmani and S. Mao, "Wireless Multimedia Cognitive Radio Networks: A Comprehensive Survey,"in *IEEE Communications Surveys & Tutorials*,.in print, 2018.

[2] Y. Saleem, and M.H. Rehmani, Primary radio user activity models for cognitive radio networks: A survey. *Journal of Network and Computer Applications*, *43*, pp.1-16.

[3] C.S. Hyder,, A.A. Al Islam,, L. Xiao, and E. Torng,. Interference aware reliable cooperative cognitive networks for real-time applications. *IEEE Transactions on Cognitive Communications and Networking*, *2*(1), pp.53-67, 2016.

[4] H. Su, and X. Zhang,Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks. IEEE Journal on selected areas in communications, 26(1), 2008.

[5] Fayaz Akhtar, Mubashir Husain Rehmani, and Martin Reisslein, White Space: Definitional Perspectives and their Role in Exploiting Spectrum Opportunities, Telecommunications Policy, Volume 40, Issue 4, Pages 319-331, April 2016.

[6] M. Amjad, F. Akhtar, M. H. Rehmani, M. Reisslein and T. Umer, "Full-Duplex Communication in Cognitive Radio Networks: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2158-2191, 2017..

**Wei Wang** is an Assistant Professor of Computer Science at San Diego State University. He received his Ph.D. degree in Computer Engineering with Significant Computer Science Emphasis from University of Nebraska - Lincoln, NE, USA. He worked as a Software Engineer with Nokia Siemens Networks Ltd (former Siemens Communications Ltd) in Beijing, China, 2005, and as a Co-op with National Broadcasting Research Center in conjunction with Tongshi Data Communications Co., Xian, China, 2001-2005. His research interests include wireless multimedia communications, QoE-QoS issues in network economics, cyber physical system healthcare, and breast cancer imaging. He serves as an Associate Editor of Wiley Security and Communication Networks Journal (SCIE-Index), the Guest Editor of several journals and special issues such as IEEE ACCESS, Springer MONET, IJDSN, etc, the web cochair of IEEE INFOCOM 2016-17, the program chair of ACM RACS 2014-16, the track-chair of ACM SAC 2014-16, workshop co-chair of ICST BodyNets 2013, the chair of IEEE CIT-MMC track 2012, the vicechair of IEEE ICCT-NGN track 2011, and the program chair of the ICST IWMMN 2010, a Technical Program Committee (TPC) member for many international conferences such as IEEE INFOCOM, GLOBECOM and ICC. He also serves as the Co-Director of IEEE MMTC Review Board 2017-2018, and the Co-Director of the IEEE MMTC Publicity Board 2014-2016. He is a Senior Member of IEEE.

# Indoor Localization Utilizing Magnetic and Visual Sensors

*A short review for "Fusion of Magnetic and Visual Sensors for Indoor Localization: Infrastructure-Free and More Effective"*

Edited by Chidansh Bhatt

*Z. Liu, L. Zhang, Q. Liu, Y. Yin, L. Cheng and R. Zimmermann, " Fusion of magnetic and visual sensors for indoor localization: Infrastructure-free and more effective," IEEE Transactions on Multimedia, vol. 19, no. 4, December 2016.*

Accurate and reliable indoor positioning is very useful in a variety of applications [1, 2]. For instance, localizing survivors inside a building in case of any accidental event, guiding robots in a fully automated factory, and navigating a person to a meeting room. However, precise indoor positioning remains an open challenge. In outdoor environments, GPS is commonly used for navigation but it could not provide accurate localization for indoor environments due to blocked satellite signals by walls and/or ceilings. Approaches based on WiFi, infrared and ultrasound have shown great promise in indoor positioning. However, due to the limited coverage of a single signal transmitter/receiver, these approaches heavily rely on specific infrastructure, which is both expensive and difficult to maintain. Also, it might be infeasible to deploy signal transmitters/receivers in certain buildings due to safety or privacy concerns.

In this paper the authors proposed an indoor localization and tracking system by integrating both magnetic and visual sensing of smartphones. Their motivation for utilizing magnetic and visual sensing is twofold. 1) Both visual images and the geomagnetic field are omnipresent across the globe, and do not rely on any pre-deployed infrastructure. They can be conveniently captured by common sensors of a smartphone. 2) Images and the geomagnetic field are complementary in indoor localization because images are usually distinguishable across distant locations while magnetic signals are known to be more locally distinctive [1], [3-6].

The authors have conducted empirical evaluation on strengths of geomagnetic field for indoor localization task. First, drastic geomagnetic field changes across locations are observed in their empirical study. Second, they have found that the field magnitudes at the same location at two different times do not change much. Further, they evaluated the influence from common indoor objects such as turning on/off computers, printers, and refrigerators. The empirical evidence suggests that the influence of turning common equipment on/off is negligible when the distance goes beyond 1.5 m.

Similarly, the advantages of using visual images for indoor localization are studied. Their in-depth empirical evaluation suggests a plethora of strengths of visual images including distinct visual images of different locations, stable visual properties, and limited influence of mobile objects.

Their study revealed the three challenges in fusing magnetic field and visual images for indoor localization. The first challenge concerns the low resolution of magnetic measurements, i.e., the measured magnetic field signals are usually not reliable enough to form a unique location signature due to its low dimensionality. The second issue is noisy sensor readings, while the third issue relates to diverse gait patterns of different users.

Based on these studies, the authors designed a context-aware particle filtering (CPF) framework to track the user. The framework works on a per-step basis. More specifically, after each user step, CPF updates the location estimation of the user. Instead of using only one certain position to estimate the true location, CPF utilizes a probabilistic distribution to depict the potential locations of the user.

Extensive experiments on four different indoor settings including a laboratory, a garage, a canteen, and an office building are conducted to evaluate the proposed method. Experimental results demonstrate the superior performance of the proposed method over the state of the arts. Notably, the fusion of magnetic field and visual images achieves a meter level accuracy on the four different indoor environments. Furthermore, the robustness of the methods is studied and the influence of grid cell size is discussed.

## References:

[1] Y. Shu et al., “Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing,” IEEE J. Sel. Areas Commun., vol. 33, no. 7, pp. 1443–1457, Jul. 2015.

[2] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, “GPS/HPS-and Wi-Fi fingerprint-based location recognition for check-in applications over smartphones in cloud-based LBSs,” IEEE Trans. Multimedia, vol. 15, no. 4, pp. 858–869, Jun. 2013.

[3] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos, “How feasible is the use of magnetic field alone for indoor positioning,” in Proc. Int. Conf. Indoor Position. Indoor Navigat., 2012, pp. 1–9.

[4] J. Haverinen and A. Kemppainen, “Global indoor self-localization based on the ambient magnetic field,” Robot. Auton. Syst., vol. 57, no. 10, pp. 1028–1035, 2009.

[5] M. Liu, “A study of mobile sensing using smartphones,” Int. J. Distrib. Sens. Netw., vol. 9, 2013, Art. no. 272916.

[6] F. Li et al., “A reliable and accurate indoor localization method using phone inertial sensors,” in Proc. ACM Conf. Ubiquitous Comput., 2012, pp. 421–430.

**Chidansh Bhatt** is a Research Scientist at FXPAL. He received a Ph.D. in computer science from National University of Singapore (NUS). His research focuses on context-aware multimedia analytics, recommendation and interactive visualization. Further interests are multimodal data mining, information retrieval, machine learning, natural language processing, big data analytics and IoT. Prior to joining FXPAL, Chidansh was working as an assistant professor at Indian Institute of Technology (IIT), India; as a post-doc researcher at IDIAP Research Institute, Switzerland; as a researcher at Big Data Experimental Laboratory, Hitachi Research & Development Ltd., Singapore and as a software engineer at IBM India. Chidansh actively participates as a technical program committee member / reviewer / organizer of leading international conferences, journals and panel member for national science foundation NSF (TOMM, MTAP, MMSJ, IEEE MM, ACMMM, ICME, ICMR, CSCW, SIGMAP, TOIT, ETRI, IWISC, ISM, VCIP etc.)

# Viewpoint Recommendation for Mobile Photography

*A short review for "ClickSmart: A Context-Aware Viewpoint Recommendation System for Mobile Photography"*

Edited by Chidansh Bhatt

*Y. S. Rawat and M.S. Kankanhalli, "ClickSmart: A Context-Aware Viewpoint Recommendation System for Mobile Photography" IEEE Transactions on Circuits and Systems for Video Technology Volume: 27, Issue: 1, Jan. 2017.*

In photography, viewpoint refers to the geolocation from where an image is captured and is considered as one of the essential factors in the art of photography [1]. It has a large impact on the composition of an image and as a result it also affects the aesthetic quality of a captured image. Advanced digital cameras can provide features like autofocus, face detection, etc., for assisting users in capturing better photos, however, it can be challenging for an amateur user to find a good viewpoint in any tourist location. It is a human tendency to follow others [2] and users generally follow the crowd to find a good viewpoint that can be misleading and therefore an amateur user may end up with bad-quality photos.

In this article the authors proposed a context-aware viewpoint recommendation method which can provide a real-time recommendation to the users for capturing high-quality photographs. This article provided a brief review of state of the art photography assistance research works, focusing on 1) view-based assistance, 2) location-based assistance, 3) offline recommendation, and 4) real-time recommendation. This is followed by a detailed description of the proposed method which includes offline learning and real-time recommendation.

As an essential contribution of this article, the authors proposed a view-point recommendation for photography which is context-aware. It was observed that context (time and weather) plays an important role in the viewpoint selection for landmark photography. For example, it is difficult to get a good-quality image when the camera lens is facing the sun. As the sun moves during the day, the viewpoints for photography will also change with time for a view at a given location. Similarly, weather conditions also affect the viewpoint as factors like visibility, clouds, etc., have impact on lighting conditions, which is known to play an important role in photography

In the past decade, we have seen an increasing trend in people's photo-taking and photo-sharing behavior. There are many social media services such as *Flickr* and *Photo.net*, with a large collection of photos shared by professional and other users. These photos have *Exif* data, which provide context information like time of capture and geolocation of the captured image. Using these details, we can infer the photo-taking behavior of people for popular tourist locations. In addition, the shared photos are augmented with social media cues such as the number of user views, likes, and comments. In this article, the authors integrate the photo-taking behavior with social media cues to develop a recommendation system that can provide real-time viewpoint recommendation to the user for taking better photos.

The authors attempt to bridge the gap between view-based and location-based methods. They propose *ClickSmart*, which provides viewpoint recommendation based on the preview on the user's camera. The proposed method makes use of publicly available photographs via the social media. The proposed recommendation system also considers the presence of people in photographs and also recommends rare but interesting viewpoints for photography.

The framework of *ClickSmart* consists of two phases, offline learning and real-time recommendation. In the offline phase, publicly available images are utilized along with the associated metadata information to train a viewpoint recommendation model. The number of possible views at any tourist location can be numerous and therefore the problem of scene-based viewpoint recommendation is challenging.

Bringing in the time and weather parameters into consideration makes the problem even more difficult. To solve this problem, the authors follow a bottom-up approach and instead of focusing on the complete view they focus on the landmark objects present in the view. The photo-taking behavior of users corresponding to each landmark object is modeled using a generative approach.

The evaluation of the proposed method was done for a dataset collected from *Flickr* for 12 different tourist locations. The authors presented both quantitative and qualitative evaluation. A user study was also conducted to further verify the experimental results. A comparative study with the existing state-of-the-art methods was also conducted. The experimental results show that the proposed method can make effective viewpoint recommendation.

**References:**

[1] M. Freeman, The Photographer’s Eye: Composition and Design for Better Digital Photos, Boston, MA, USA:Focal Press, 2007.

[2] S. R. Musse, D. Thalmann, "A model of human crowd behavior: Group inter-relationship and collision detection analysis" in Computer Animation and Simulation, Vienna, Austria:Springer, pp. 39-51, 1997.

**Chidansh Bhatt** is a Research Scientist at FXPAL. He received a Ph.D. in computer science from National University of Singapore (NUS). His research focuses on context-aware multimedia analytics, recommendation and interactive visualization. Further interests are multimodal data mining, information retrieval, machine learning, natural language processing, big data analytics and IoT. Prior to joining FXPAL, Chidansh was working as an assistant professor at Indian Institute of Technology (IIT), India; as a post-doc researcher at IDIAP Research Institute, Switzerland; as a researcher at Big Data Experimental Laboratory, Hitachi Research & Development Ltd., Singapore and as a software engineer at IBM India. Chidansh actively participates as a technical program committee member / reviewer / organizer of leading international conferences, journals and panel member for national science foundations NSF (TOMM, MTAP, MMSJ, IEEE MM, ACMMM, ICME, ICMR, CSCW, SIGMAP, TOIT, ETRI, IWISC, ISM, VCIP etc.)

# Towards Precise Friend Recommendation in Social Media

*A short review for "Two-Stage Friend Recommendation Based on Network Alignment and Series Expansion of Probabilistic Topic Model"*

Edited by Jun Zhou

*S. Huang J. Zhang, D. Schonfeld, L.Wang, and X. Hua " Two-Stage Friend Recommendation Based on Network Alignment and Series Expansion of Probabilistic Topic Model," IEEE Transactions on Multimedia, vol. 19, no. 6, June 2017.*

Nowadays many people rely on social media to keep contact with friends, access and share information, and expand their connections, making friend recommendation an essential function of the services provided by social media [1]. Similar to real life, finding a good friend is not an easy job. People come from different backgrounds, have diverse interests, and live in a dynamic environment. All these impose great challenges to precise and usable friend recommendation.

The goal of friend recommendation is to provide a ranked candidate friend list to a new user of social network or help to refresh the friend list of existing users. Friend recommendation methods in social media such as Facebook or Twitter are often based on similarity of profiles or common friends between users. Services provided by Flickr or Amazon also use product related information, such as purchasing record, image tags, or comments, to produce recommendations. In some cases, multi-source environment also enables the exploration of cross-domain information from text and images, or cross platform friend lists to boost the precision of recommendation [2,3].

In cross-domain friend recommendation, practical solutions have been provided by modelling the joint distribution of data from multiple sources. Such techniques, however, requires consideration of different factors simultaneously in the model building and the following learning and optimization process. When a large number of factors need to be considered, the solution becomes time consuming and sometimes intractable in identifying the contribution from each factor.

The authors of this paper proposed a novel method to address this problem, and showcase its utility on the Flickr dataset. Instead of integrating all factors into one model, friend recommendation is implemented in two stages. In the first stage, user friend list and relationship between text and users are used to produce and align different social networks. This leads to a list of candidate recommendations. In the second stage, the recommendation result is refined by considering the relationship between images and users via a probabilistic topic model.

The generation of the initial friend list is based on the alignment of contact network and tag similarity network [4]. The contact network is constructed via creating a graph on the existing friendship relationship of users in the dataset. The generation of tag similarity network uses the tags of images upload by users. In both cases, users are treated as the nodes of graphs, and the user to user and user to tag relationships are used to form the edges. During this stage, the importance of tags is determined so redundant information can be removed from further consideration, thus greatly improve the effectiveness of tags.

The novel contribution of this paper comes from the second stage where a probabilistic topic model [5] is used to analyse the relationship between data from different domains. First, the relationship between cross-domain topic models is built with a small number of latent variables related to image property, user interests, and user behavior. This leads to a compact model with advantages in computational complexity. Second, the traditional solution for topic model is either based on Gibbs sampling or variational inference, each has some weakness in handling big data or in determining the accuracy of approximation. This paper proposes to explore a series expansion solution to calculate the coupled integrals as required in the Bayesian inference, which gives more precise recommendations. The series expansion is built on Mellin transform [7] which is a solid tool for deducing the mathematic expression of the coupling of different random variables.

In the experiments, a dataset of 30,000 users, 1,356,294 photos, 628,153 friend links, and 42,739 tagged words are crawled from the Flickr website. Image features are extracted using AlexConvNet [6]. This is a dataset big enough to effectively reduce the bias. Among these data, 4/5 of the users were used as the training set, and the rest were used as the testing set.

Compared with several state-of-the-art methods, the proposed method has achieved higher precision and recall. However, the proposed model is still mathematically complicated, and cannot be easily generated to data of other modalities. Furthermore, the efficiency of the method was not reported in this paper, so its extent of usability is not clear. Nonetheless, this paper opened a new direction of method development for friend recommendation, and may lead to the exploration of more general framework and extension to new applications.

**References:**

[1] S. Wan, Y. Lan, J. Guo, C. Fan, and X. Cheng, “Informational friend recommendation in social media,” in Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1045-1048, 2013.

[2] M. Yan, J. Sang, T. Mei, and C. Xu, “Friend transfer: Cold-start friend recommendation with cross-platform transfer learning of social knowledge,” in Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 1–6, 2013.

[3] C. Guo, X. Tian, and T. Mei, “User specific friend recommendation in social media community,” in Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 1–6, 2014.

[4] S. Huang, J. Zhang, L. Wang, and X. S. Hua, “Social friend recommendation based on multiple network correlation,” IEEE Transactions on Multimedia, vol. 18, pp. 287–299, 2016.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003..

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Red Hook, NY, USA: Curran Associates, Inc., pp. 1097–1105, 2012.

[7] M. D. Springer, The Algebra of Random Variables. New York, NY, USA:Wiley, 1979.

**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 is now a senior lecturer in the School of Information and Communication Technology in Griffith University. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA, Australia. His research interests are in spectral imaging, pattern recognition, computer vision, and their applications to and environmental informatics and remote sensing.

# Frame Prediction for 2D Video Coding using 3D Point Cloud Reconstruction

*A short review for "Point cloud estimation for 3D structure-based frame prediction in video coding"*
Edited by Bruno Macchiavello

*H. Bakhshi Golestani, J. Schneider, M. Wien, Mathias and J.R. Ohm, "Point cloud estimation for 3D structure-based frame prediction in video coding", IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, July, 2017.*

Most video coding standards use a hybrid architecture, based on differential coding modulation (DPCM) and transform coding. DPCM is mostly used to remove the temporal redundancy that exits among contiguous video frames of the same sequence. Basically, the current frame is estimated using past and/or future frames and only the difference between the actual frame and the estimation is encoded. However, in presence of camera motion or scene change, accurate frame estimation can be challenging. In this work the authors propose a frame prediction scheme based on a 3D structure estimated from a subset of decoded frames. The main idea is that both the encoder and decoder can reconstruct the 3D structure of the scene based on some key-frames, and then estimate an intermediate and/or future frame by projecting the 3D model into a specific view point that corresponds to that frame. This type of estimation requires the ability to reconstruct the 3D model of a scene based on only few frames and estimate the exact view point of each frame.

Note that the proposed method is explicitly designed for encoding 2D video sequences, not 3D point clouds. The authors estimate the 3D structure and use it for 2D motion compensation. Therefore, the addressed problem is different from 3D view synthesis [1]. For 3D structure estimation the authors used a Structure from Motion (SfM) method based on uncalibrated images [2]. The input of SfM is a set of images and its output is a 3D point cloud. The structure assume by the authors, is that each frame is considered a view point from a particular camera. Each view point is a projection of points from a real world location. In order to reconstruct the 3D structure, it is required to estimate both the projection matrix of each camera and the real world locations. The projection matrix of each camera can be estimated based on the intrinsic parameters (such as focal length) and the extrinsic parameters (which relates to pose and location of the camera). SfM can provide, along with the 3D point cloud, an estimation of the camera parameters. The authors solved the SfM problem through an algebraic approach in 6 (six) steps.

In the first step a feature extraction is performed. Extracting reliable and sufficient features is a key step in SfM, since the accuracy of camera calibration and number of extracted 3D points are highly dependent on extracted features. SIFT, ORB, Laplacion-of-Gaussian and Harris corners can be used for feature extraction in SfM [3]. Assuming that exits a pair of camera with canonical projection matrices, a fundamental matrix F from two views can be obtained. F encapsulates the relative translation and rotation between two cameras (extrinsic parameters). As long as a sufficient number of correspondences between the 2 images was obtained in the first step, F can be computed as the second step of the reconstruction method. In the third step, F is used to estimate the projection matrices of each of the two cameras involved. With the projection matrices and the projected points (pixels of each frame) the real world locations can be computed in the next step. Note, that these locations are estimations and not necessarily reflect the actual real world structure. The next step is auto-calibration. Auto-Calibration is the problem of recovering the metric reconstruction from the perspective reconstruction. SfM provides a perspective solution. Assuming that a first camera (first frame of the video sequence) is canonical and the other cameras have perspective projections a linear system can be derived which can be used to upgrade the perspective solution into a metric one [4]. The final step is a Bundle Adjustment (BA). The authors used a non-linear optimization that aims to minimize the re-projection error. In other words, the authors use the estimated 3D structure to project the real cameras (frames) and then try to minimize the error between the estimated cameras and their corresponding observation. The authors indicated that they used Newton's method and the Levenberg-Marquardt [5] algorithm for solution of their optimization problem.

Unfortunately, even if several feature points were extracted, normally the derived point cloud is not dense enough to be used directly in mesh reconstruction. Thus, the output of SfM is imported into a dense point cloud estimation algorithm. The initially estimated point cloud is interpolated using a graph representation. The interpolation process used by the authors is based on Delaunay tetrahedralization [7], which is a classical tool in the field of mesh generation and mesh processing. Once the 3D structure is computed, it can be used to project any 2D view point.

In order to verify the quality of the estimated frames through 3D projection, the authors included the projection as an additional reference frame for conventional 2D motion compensation. The video codec used was the HEVC Test Model (HM). The estimation was used both in prediction of future frames, and prediction of intermediate B-frames. The video sequences were of 4K resolution, with moving camera. Results show around 0.8% bit-rate reduction compared to conventional HEVC. This shows that there is some redundancy that can be remove using 3D information. Nevertheless, this initial experiments focus on camera motion, an actual multiview scene where temporal and spatial redundancy exits between several view points can be the next scenario where this technique can be applied.

**References:**

[1] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View Synthesis for Advanced 3D Video Systems", EURASIP Journal on Image and Video Processing, vol. 2008, Article ID 438148, 11 pages, 2008. doi:10.1155/2008/438148.

[2] R. Toldo, R. Gherardi, M. Farenzena and A. Fussielo, "Hierarchical structure-and-motion recovery from uncalibrated images," Computer Vision.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. "ORB: an efficient alternative to SIFT or SURF", in Proceeding of the IEEE International Conference on Computer Vision (ICCV), volume 13, 2011.

[4] R. Hartely and A. Zisserman, "Multiple view geometry in computer vision", Cambridge university press, second edition, 2004.

[5] D. A. Forsyth and J. Ponce, "Computer Vision: A Modern Approach," Prentice Hall, second edition, 2011.

[6] F. Cazals and J. Giesen, "Delaunay Triangulation Based Surface Reconstruction, Mathematics and Visualization," Effective Computational Geometry for Curves and Surfaces, Springer, pp. 231-276, 2006.

**Bruno Macchiavello** is an associate professor at the Department of Computer Science of the University of Brasilia (UnB), Brazil. He received his B. Eng. degree in the Pontifical Catholic University of Peru in 2001, and the M. Sc. and D.Sc. degrees in electrical engineering from the University of Brasilia in 2004 and 2009, respectively. Prior to his current position he helped develop a database system for the Ministry of Transport and Communications in Peru. He is and Are Editor for the Elsevier Journal Signal Processing: Image Communications. His main research interests include video and image coding, image segmentation, distributed video and source coding, multi-view and 3D video processing.

# Paper Nomination Policy

Following the direction of MMTC, the Communications – Review platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia includes, but is 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 Review Board Directors: Pradeep K. Atrey (patrey@albany.edu), Qing Yang (qing.yang@montana.edu), Wei Wang (wwang@mail.sdsu.edu), and Jun Wu (wujun@tongji.edu.cn). 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**

Members of the IEEE MMTC Review Board will review each nominated paper. In order to avoid potential conflict of interest, guest editors external to the Board will review nominated papers co-authored by a Review Board member. The reviewers' names will be kept confidential. If two reviewers agree that the paper is of Review quality, a board editor will be assigned to complete the review (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.

**Best Paper Award**

Accepted papers in the Communications – Review 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://mmc.committees.comsoc.org/.

## MMTC Communications – Review Editorial Board

DIRECTORS

**Pradeep K. Atrey**
State University of New York at Albany, USA
Email: patrey@albany.edu

**Wei Wang**
San Diego State University, USA
Email: wwang@mail.sdsu.edu

**Qing Yang**
Montana State University, USA
Email: qing.yang@montana.edu

**Jun Wu**
Tongji University, China
Email: wujun@tongji.edu.cn

EDITORS

**Chidansh Bhatt**
FXPAL, Palo Alto, USA

**Ing. Carl James Debono**
University of Malta, Malta

**Xiaohu Ge**
Huazhong University of Science and Technology, China

**Carsten Griwodz**
Simula and University of Oslo, Norway

**Frank Hartung**
FH Aachen University of Applied Sciences, Germany

**Bruno Macchiavello**
University of Brasilia (UnB), Brazil

**Mukesh Saini**
Indian Institute of Technology, Ropar, India

**Debashish Sen**
Indian Institute of Technology, Kharagpur, India

**Gwendal Simon**
Telecom Bretagne (Institut Mines Telecom), France

**Cong Shen**
University of Science and Technology of China

**Lifeng Sun**
Tsinghua University, China

**Roger Zimmermann**
National University of Singapore, Singapore

**Michael Zink**
University of Massachusetts Amherst, USA

**Jun Zhou**
Griffith University, Australia

## Multimedia Communications Technical Committee Officers

**Chair:** Shiwen Mao, Auburn University, USA
**Steering Committee Chair:** Zhu Li, University of Missouri, USA
**Vice Chair – America:** Sanjeev Mehrotra, Microsoft, USA
**Vice Chair – Asia:** Fen Hou, University of Macau, China
**Vice Chair – Europe:** Christian Timmerer, Alpen-Adria-Universität Klagenfurt, Austria
**Letters & Member Communications:** Honggang Wang, UMass Dartmouth, USA
**Secretary:** Wanqing Li, University of Wollongong, Australia
**Standard Liaison:** Liang Zhou, Nanjing University of Posts and Telecommunications, China

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.