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

R-LETTER

Vol. 5, No. 3, June 2014

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from the Review Board

Introduction

Since November 2012, the Review Board has completed ten R-letters, reviewing various aspects of multimedia communications research which include content streaming, information discovery and exploration, process optimization, surveillance and privacy, smart sensing and mobility, quality assessment and others. The Review Board would like to thank the strong support from the MMTC Chair, the Vice-Chair (Letter & Member Communications) and the community in the last two years. We hope MMTC will continue support the new Review Board under the new MMTC leadership team to be elected in ICC 2014.

Distinguished Category

As multimedia communications have become more immersive in our daily activities, an increasing number of innovative ideas are also inspired and emerged, but not without challenges: the balance between privacy and transparency; and the choice of evaluation metrics for proposed methods. Enjoy the related discussions in the two distinguished articles.

The **first paper**, published in *IEEE Transactions on Image Processing* and *edited by Koichi Adachi*, describe a framework for comprehensive sensing based on secure watermark detection and privacy preserving storage.

The **second paper** is *edited by Irene Cheng* and published in the *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*. This paper addresses the objective quality assessment for image retargeting based on structural similarity.

Regular Category

The regular category of this issue comprises six papers, which are introduced in the following.

The **first paper**, published in the *IEEE Transactions on Multimedia* and *edited by Tobias Hoßfeld and Christian Timmerer*, describes best practices for subjective quality assessments using crowdsourcing in order to determine the Quality of Experience of audio/visual Web services.

The **second paper**, published in the *IEEE Transactions on Image Processing* and *edited by Weiyi Zhang*, presents a state-based approach to video communication.

The **third paper** is *edited by Gene Cheung* and has been published within the *Symposium on Graph Signal Processing in IEEE Global Conference on Signal and Information Processing*. It proposes a new dictionary learning method for graph-signals, where the structure of the data kernel is embedded into the designed dictionary.

The **forth paper**, published in the *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR13)* and *edited by Hao Hu*, targets face hallucination by exploring image structures.

The **fifth paper**, published in *IEEE Transactions on Pattern Analysis and Machine Intelligence* and *edited by Jun Zhou*, describes a method for adaptive multi-frame super resolution based on the estimation of optical flow, noise level, and blur kernel, under a Bayesian framework.

Finally, the **sixth paper**, *edited by Carl James Debono* and published within the *IEEE Transactions on Multimedia*, proposes a loss-resilient coding of texture and depth for free-viewpoint video conferencing.

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

A Framework for Compressive Sensing based on Secure Watermark Detection and Privacy Preserving Storage

A short review for "A Compressive Sensing based Secure Watermark Detection and Privacy Preserving Storage Framework"

Edited by Koichi Adachi

Q. Wang, W. Zeng, and J. Tian, "A Compressive Sensing based Secure Watermark Detection and Privacy Preserving Storage Framework", IEEE Trans. Image Process., vol. 23, no. 3, pp. 1317-1328, Mar. 2014.

As cloud computing has gained popularity in recent years. In cloud computing, privacy is one of the most critical issues as the data owners outsource data storage or processing to a third party computing service. Ideally, the cloud will store the data and perform signal processing or data-mining in an encrypted domain to preserve the data privacy.

Traditional secure watermark detection techniques are designed to convince a verifier whether or not a watermark is embedded without disclosing the watermark pattern so that an untrusted verifier cannot remove the watermark from the watermark protected copy [1]-[7]. However, most of the existing secure watermark detection works assume the watermarked copy are publicly available and focus on the security of the watermark pattern, while the privacy of the target media on which watermark detection is performed has received little attention. But for some applications such as the scenario given above, it is required to protect the multimedia data's privacy in the watermark detection process.

Performing privacy preserving storage and secure watermark detection simultaneously is possible by using the existing secure watermark detection technologies such as zero-knowledge proof protocols [5]-[7] that transform the multimedia data to a public key encryption domain. However, the existing technologies require complicated algorithms, high computational and communication complexity [1], and large storage consumption in the public key encryption domain. These limitations may impede their practical applications.

Previous works [8]-[13] indicate that signal processing or data-mining in the compressive sensing (CS) domain is feasible and is computationally secure under certain conditions. Therefore, based on these backgrounds, in this paper, the authors propose a CS-based framework using secure multiparty computation (MPC) protocols to enable simultaneous secure watermark detection and privacy preserving multimedia data storage.

In the proposed framework, the following parties are assumed to be existing: data holder (DH), watermark owners (WO), cloud (CLD). Furthermore, it is assumed that there is a certificate authority (CA) to provide the public keys and CS matrix to the DH. The CA needs to issue CS matrix suites to the DH for watermarking. The CS matrix suites include the seeds and the random function used to generate the Gaussian CS matrix. In CS, it is necessary to generate CS matrix. Therefore, the CA takes the role to issue the random function to guarantee the randomness of the generated Gaussian CS matrix. The CA also needs to issue a public key pair to the DH and the DH's public key to the WO. The public key is used for the MPC based CS transformation protocol. For secure watermark detection, the watermark is transformed to the same CS domain using the secure MPC protocol and then sent to the cloud. Since the cloud only has the data in the CS domain, the cloud will perform watermark detection in the CS domain. The image data in the compressive sensing domain can be stored in the cloud and reused for detection of watermark from many other watermark owners.

It is shown that as the number of target images for secure watermark detection increases; the proposed framework (when all the target images are transformed to the same CS domain) outperforms the previous watermark detection system such as [6] in terms of communication cost. The authors show that the proposed framework has better scalability and higher efficiency when performing secure watermark detection on a larger number of images.

Through the experiments, the authors show that the proposed framework is secure under the semi-honest adversary model to protect the private data. Compared to the previous secure watermark detection protocols, our framework offers better efficiency and flexibility, and protects the privacy of the multimedia data that has not yet been considered in the previous works.

The future works include: 1) evaluation of the robustness of the proposed watermark detection in the CS domain under some other attacks, 2) development of MPC protocols for secure CS reconstruction in addition to secure CS transformation.

The proposed framework depends on which compression algorithm is used for image processing. In the paper, JPEG format, which utilizes discrete cosine transform (DCT), is considered, where some DCT coefficients can be used for watermark detection. Therefore, the exploration of the proposed algorithm to other image processing method is also one of the important and interesting future studies.

References:

[1] T. Bianchi and A. Piva, “Secure watermarking for multimedia content protection: A review of its benefits and open issues,” IEEE Signal Process. Mag., vol. 30, no. 2, pp. 87–96, Mar. 2013.

[2] Z. Erkin, A. Piva, S. Katzenbeisser, R. Lagendijk, J. Shokrollhi, G. Neven, et al., “Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing,” EURASIP J. Inf. Security, vol. 7, no. 2, pp. 1–20, 2007.

[3] J. Eggers, J. Su, and B. Girod, “Public key watermarking by eigenvectors of linear transforms,” in Proc. Euro. Signal Process. Conf., 2000.

[4] S. Craver and S. Katzenbeisser, “Security analysis of public-key watermarking schemes,” in Proc. Math. Data/Image Coding, Compress., Encryption IV, Appl., vol. 4475. 2001, pp. 172–182.

[5] A. Adelsbach and A. Sadeghi, “Zero-knowledge watermark detection and proof of ownership,” in Proc. 4th Int. Workshop Inf. Hiding, vol. 2137. 2001, pp. 273–288.

[6] J. R. Troncoso-Pastoriza and F. Perez-Gonzales, “Zero-knowledge watermark detector robust to sensitivity attacks,” in Proc. ACM Multimedia Security Workshop, 2006, pp. 97–107.

[7] M. Malkin and T. Kalker, “A cryptographic method for secure watermark detection,” in Proc. 8th Int. Workshop Inf. Hiding, 2006, pp. 26–41.

[8] O. Goldreich, The Foundations of Cryptography. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[9] K. Liu, H. Kargupta, and J. Ryan, “Random projection-based multiplicative data perturbation for privacy preserving distributed data mining,” IEEE Trans. Knowl. Data Eng., vol. 18, no. 1, pp. 92–106, Jan. 2006.

[10] W. Lu, A. L. Varna, A. Swaminathan, and M. Wu, “Secure image retrieval through feature protection,” in Proc. IEEE Conf. Acoust., Speech Signal Process., Apr. 2009, pp. 1533–1536.

[11] W. Lu, A. L. Varna, and M.Wu, “Security analysis for privacy preserving search for multimedia,” in Proc. IEEE 17th Int. Conf. Image Process., Sep. 2010, pp. 2093–2096.

[12] Y. Rachlin and D. Baron, “The secrecy of compressed sensing measurement,” in Proc. 46th Annu. Allerton Conf. Commun., Control, Comput., 2008, pp. 813–817.

[13] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, “On the security and robustness of encryption via compressed sensing,” in Proc. IEEE Military Commun. Conf., Nov. 2008, pp. 1040–1046.

Koichi ADACHI received the B.E., M.E., and Ph.D degrees in engineering from Keio University, Japan, in 2005, 2007, and 2009 respectively. From 2007 to 2010, he was a Japan Society for the Promotion of Science (JSPS) research fellow. He was the visiting researcher at City University of Hong Kong in April 2009 and the visiting research fellow at University of Kent from June to Aug 2009. Currently he is with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include green communication and cooperative communications. Dr. Adachi served as General Co-chair of the Tenth IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS) and Track Co-chair of Transmission Technologies and Communication Theory of the 78th IEEE Vehicular Technology Conference in 2013. He was recognized as the Exemplary Reviewer from IEEE Communications Letters and IEEE Wireless Communications Letters in 2012.

Evaluation of Image Retargeting: From Single to Multi-operator

A short review for "Objective Quality Assessment for Image Retargeting Based on Structural Similarity"

Edited by Irene Cheng

Yuming Fang, Kai Zeng, Zhou Wang, Weisi Lin, Zhijun Fang and Chia-Wen Lin, "Objective Quality Assessment for Image Retargeting Based on Structural Similarity", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 4, no. 1, pp. 95-105, Mar. 2014.

When big screen in Reality Centers and IMAX theatres were introduced, there was a rapid demand for high resolution content. Benefiting from the advanced wireless communication infrastructure, mobile and handheld devices are commonplace which in turn motivates researchers to explore content streaming to fit smaller handheld displays, as well as to optimize constrained resources. Although there exist in the literature numerous state-of-the-art level-of-detail algorithms for images, the outcome from linear scaling and boundary cropping cannot be satisfactory for all applications because salient objects or regions of interest will likely lose their visual importance or association in the scene as a result of the process. To address this issue, many researchers suggest non-uniform scaling. This approach fails many traditional perceptual quality evaluation metrics because they were not designed to compare non-uniformly scaled scenes. Fortunately, there are also scale-invariant techniques, which can identify structural information in images even though the structures may be deformed during the retargeting process. In this paper, the authors take advantage of two scale-invariant techniques, i.e., scale-invariant feature transform (SIFT) [1] and structural similarity (SSIM) map [2] to establish an objective metric to assess the qualities of retargeted images.

In order to understand the rationale behind the proposed metric, the authors first discuss many image retargeting methods, which include seam carving, which has been extended for video retargeting; an adaptive image retargeting algorithm based on Fourier analysis; a scale-and-stretch warping algorithm; and multi-operator operation. Observing that inadequate study has been done to evaluate the quality of retargeted images quantitatively, and also inspired by the SSIM approach, the authors introduce an Image Retargeting SSIM (IR-SSIM). A critical step in IR-SSIM is to generate an SSIM quality map, which is used to compute at each spatial location in the reference scene how the structural information changes in the retargeted scene. Fig. 1 shows an overview of the proposed approach given by the authors.

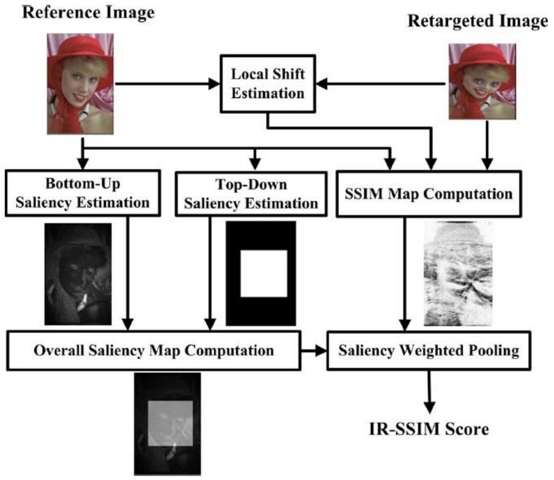

Reference Image
Retargeted Image
Local Shift Estimation
Bottom-Up Saliency Estimation
Top-Down Saliency Estimation
SSIM Map Computation
Overall Saliency Map Computation
Saliency Weighted Pooling
IR-SSIM Score

Fig. 1: Overview of the proposed framework (given in the original paper).

Some basic steps of the proposed evaluation metric are summarized below:

1. Starting from each image pixel in the reference image, the algorithm first finds its best matching pixel in the retargeted image.
2. The SSIM measure is then computed between the neighborhoods of these two pixels.
3. By repeating this process to all pixels in the reference image, an SSIM map is generated.
4. Based on the reference image, a bottom-up saliency map and a top-down saliency map are created to generate a combined overall saliency map of the scene.
5. The final IR-SSIM score of the retargeted image is derived by applying weights on the SSIM map based on the overall saliency map.

In order to illustrate the concept, faces in the scene are used as salient objects in the current implementation but it is suggested that saliency detection models can be incorporated in the framework in the future.

It is interesting to see how the authors resolve the problem associated with non-uniform scaling across the reference and retargeted images. To handle this issue, they cannot use the original SSIM algorithm directly. Instead, they employ shift estimation methods to create dense pixel correspondence between the images, and then apply SIFT flow descriptors instead of use raw pixels to match corresponding structural details.

Another interesting discussion in this paper is the multi-scale SSIM approach, which can improve image quality assessment by accounting for variation such as sampling density, viewing distance and observer's unique perceptual ability.

Some readers feel that this work has addressed an important problem of image retargeting, i.e., providing objective evaluation of the image retargeting quality, and the proposed approach is novel in combining SSIM, SIFT, visual saliency, and spatial pooling approaches, delivering good experimental results. From another perspective, there are comments saying that although this paper discusses an interesting topic for the multimedia community, the proposed technique is mostly a direct extension of the SSIM metric integrating with SIFT Flow and other existing techniques, and hence the scientific contribution is not significant. However the results are excellent compared with existing metrics. The tests and comparisons have been conducted very seriously and the authors have used an existing database as well as their own database.

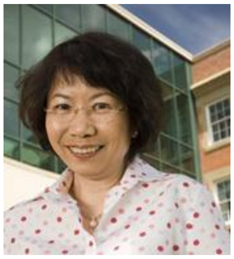
Overall, this work has initiated a helpful discussion relating to quality assessment metric for image retargeting. Furthermore, it motivates the thoughts on how algorithms should be designed to retarget images effectively in term of perceptual quality. Undoubtedly, defining the quality of retargeted images is a controversial issue. Some may focus on the preservation of salient objects sacrificing other details in the scene. There are others who prefer to keep the entire scene intact even all individual objects may appear smaller. Since application requirements and user preferences will dictate the final decision, the question is therefore: should image retargeting algorithms be semi-automatic allowing an application to choose multiple retargeting operators not only on the whole image but adaptively on different regions in the image? The multi-operator process was shown to achieve better performance than single-operator methods [3]. A modified multi-operator image retargeting process was also tested in this paper to demonstrate the potential of IR-SSIM. Image retargeting, content streaming and cross display optimization is an open and challenging research topic, and is worth more research effort in this direction.

References:

[1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150-1157.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[3] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–11, 2009.



Irene Cheng, SMIEEE is the Scientific Director of the Multimedia Research Centre, and an Adjunct Professor in the Faculty of Science, and 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. She is also the Associate Editor of IEEE Trans. on Human-Machine Systems. Her research interests include multimedia communications techniques, quality evaluation metrics and JND, visualization in scale-space, and human-Computer interaction techniques. In particular, she introduced the quality assessment metric on simplified 3D surface based on the perception of relative change.

.

Quality of Experience Assessment using Crowdsourcing

A short review for "Best Practices for QoE Crowdtesting: QoE Assessment with Crowdsourcing"

Edited by Tobias Hoßfeld and Christian Timmerer

Tobias Hoßfeld, Christian Keimel, Matthias Hirth, Bruno Gardlo, Julian Habigt, Klaus Diepold, Phuoc Tran-Gia, "Best Practices for QoE Crowdtesting: QoE Assessment with Crowdsourcing", IEEE Trans. on Multimedia, vol. 16, no. 2, pp. 541–558, Feb. 2014.

Crowdtesting. Quality of Experience (QoE) in multimedia applications is closely linked to the end users' perception and therefore its assessment requires subjective user studies in order to evaluate the degree of delight or annoyance as experienced by the users. Crowdsourcing enables new possibilities for QoE evaluation by moving the evaluation task from the traditional laboratory environment into the Internet, allowing researchers to easily access a global pool of subjects for the evaluation task. In particular, *QoE crowdtesting* refers to QoE assessment using crowdsourcing, where the crowdsourcing platforms act as an extra layer between test manager and test subject, handling the recruiting and payment of the test participants. The anonymous test subjects conduct subjective tests remotely in their preferred environment, usually with a Web-based application, that can be accessed via common Web browsers. The advantages of QoE crowdtesting lie not only in the reduced time and costs for the tests, but also in a large and diverse panel of international, geographically distributed users in realistic user settings. If the unique properties of the crowdsourcing environment and their impact on the QoE evaluation are considered appropriately, crowd-based QoE evaluation provides an efficient, simple, cheap, and representative alternative to traditional laboratory QoE studies.

Key issues. Moving from the laboratory to the crowd, however, is not as straightforward as simply generating a Web- interface for an existing test. There are significant differences between laboratory-based and crowd-based evaluation with respect to conceptual, technical, and motivational aspects due to the remote test settings that need to be considered when performing the crowd-based QoE evaluation.

Key issues arising from QoE crowdtesting include the reliability of user ratings, incentives, and task design for attracting test users, and the unknown environmental context of the tests including heterogeneity of users, used hardware, environment settings, etc. On the one hand, the actual user ratings are affected because of the QoE influence factors which are additionally emerging from the remote setting and which are not directly controlled. On the other hand, the execution of the test study and the implementation of the (Web-based) test software has to consider the crowdsourcing settings and the non-standard test equipment, e.g., software compatibility to ensure a successful execution of the test or Internet access speed for downloading the test contents which may result into undesired waiting times during the subjective study. In order to address these issues, strategies and methods need to be developed, included in the test design, and also implemented in the actual test campaign in addition to the QoE aspects to be tested, while statistical methods are required to identify reliable user ratings and to ensure high data quality. Therefore, appropriate mechanisms like reliability checks or training phases must be included in the task design.

Conceptual challenges. Conceptual challenges arise by moving the subjective user studies to the crowd due to the typically short micro-tasks in the order of a few minutes compared to long(er) lab studies. Therefore, tests designed for lab environments need to be modified for crowdtesting and a simple way is to partition the test into basic test cells. As a consequence, a crowdsourced QoE test user may only see a subset of the test conditions, which requires sophisticated statistical methods for outlier detection and quantifying reliability. Furthermore, a test moderator and direct feedback between the subjects and the moderator is missing, but the user is guided via the Web interface through the tests including an explanation about the test itself, i.e., what to evaluate and how to express the opinion. The training of subjects is mostly conducted by means of qualification tests. Nevertheless, in case of any problems with understanding the test, uncertainty about rating scales, sloppy execution of the test, or fatigue of the test user, appropriate mechanisms or statistical methods have to be applied. Therefore, it is more difficult to ensure a proper training of the subjects, specifically as no direct feedback between supervisors and subjects is possible [1]. Due to the short task duration in crowdsourcing, demo trials to familiarize the subject with the test structure and practice trials not included in the analysis significantly decrease the efficiency of a test and increase the costs. In particular, authors in [2] show that without any worker training and reliability questions the results are significantly worse than with lab or advanced crowdsourcing designs. Thus, training phases must be included in the task design.

Reliability of user ratings. There are several reasons why some user ratings are not reliable and need to be filtered out. Technical errors may occur due to errors in the Web-based test application or due to incompatibilities of the test application with the worker's hard- and software including missing video codecs or insufficient screen resolution. As a consequence, the users observe different test conditions or additional artifacts occur leading to test results which appear unreliable, but may be valid for the individual users' conditions. This requires an appropriate monitoring of the system, but also of the context. Another possible reason for unreliable user ratings is related to test instructions which may be not clear or too complex to understand. Additionally, language problems may also occur with international users.

A huge problem for QoE crowdtesting is cheating users. Commercial crowdsourcing applications suffer from workers who try to maximize their received payment while minimizing their own effort and therefore submit low-quality work to obtain such a goal. To be more precise, the actual goal is the payment to effort ratio and, therefore, tasks should be designed that provide incentives for high-quality work rather than low-quality feedback.

Best practices. In this paper, authors provide a collection of best practices addressing these issues based on a large set of conducted QoE crowdtesting studies using video quality assessment as an example. The focus is in particular on the issue of reliability, showing that the recommended two-stage QoE crowdtesting design leads to more reliable results.

The *technical implementation* of the test should take into consideration the spread of the used technology among the targeted crowd. For example, the use of widely available technologies (such as Flash player for video playback) is strongly recommended. Depending on required computational power, size of the crowd and/or geographical location, content distribution networks and cloud services can provide better service in comparison to a standalone server.

To cope with the limited reliability of the crowd and other factors influencing the rating behavior for the *Campaign and Task design*, the following steps are recommended:

1. The task should be designed to prevent cheating.
2. A pseudo-reliable crowd is created by simple, short, and cheap tests with different reliability elements. Only reliable users are then allowed to pass to the actual QoE tests with higher payments. This approach is also known as pilot task and main task.
3. Different elements need to be added in the task design to check the reliability of the users and to filter out unreliable users in the first and second stage of the QoE test. Combining these elements also leads to an improved reliability of the results. Additional reliability mechanisms include, but are not limited to:
    3.1. *Verification tests*, including captchas or computation of simple text equations: "two plus 3=?", "Which of these countries contains a major city called Cairo? (Brazil, Canada, Egypt, Japan)".
    3.2. *Consistency tests*: First, the user is asked "In which country do you live?". Later, the user is asked "In which continent do you live?".
    3.3. *Content questions about the test*: "Which animal did you see?" (Lion, Bird, Rabbit, Fish).
    3.4. *Gold standard data*: "Did you notice any stops to the video you just watched?" (Yes, No), when the actual test video did not include any stalls.
    3.5. *Application-layer monitoring*: Monitoring of response times of users and browser events to capture the focus time.

The important thing to keep in mind is not to add too many reliability items, as otherwise the assessment task will become too lengthy. Furthermore, too many of these questions may give a signal of distrust to the users. As a result, users may abort the survey. In general, incentives and proper payment schemes depending on the actual work effort are the key to high-quality work. Incentive schemes such as gamification have the potential to make crowdsourcing an even more powerful tool to deliver high-quality data in QoE assessments.

Regarding the best practices for evaluating the campaign and calculating the overall statistics of the crowdsourcing testing the use of a combination of *a)* the reliability mechanisms above which work independent of the actual user ratings, e.g., content questions and *b)* typical screening mechanisms which filter users based on the actual user ratings and common statistical assumptions. The latter methods alone cannot clearly identify unreliable users, since, for example, hidden influence factors or the variability of subjects' sensitivities to different artifacts are not determined. Therefore, those additional reliability approaches are indispensable to identify unreliable users independent of any hidden influence factor and the actual user rating. Nevertheless, reliability measures such as inter- and intra-rater reliability or Krippendorff's α should always be stated for QoE crowdtesting studies, where high values show reliable user ratings, but low values imply the presence of unreliable users or hidden influence factors in the QoE crowdtesting campaign. We further recommend including the analysis of mean and standard deviation of opinion scores (MOS and SOS) which follows a square relationship according to the SOS hypothesis [3], but also additional information about the crowdsourcing platform used for the QoE assessment.

Limitations. In principle, crowdsourcing could be used for the assessment of any stimuli and interactivity, using any type of subjective methodology. In reality, however, we are faced with several limitations on the possible scope of QoE crowdtesting. The main technical factors limiting the scope of QoE assessment are bandwidth constraints and support of the workers' devices to present the required stimuli. In particular, coding standards need to be supported by the workers devices, as it is often not feasible to provide the uncompressed stimuli to the workers due to excessive bandwidth demands. This is often in contrast to the traditional lab setting, where the aim is to avoid any additional compression of the stimuli under test.

Furthermore, the stimuli must be supported by the workers' devices. Although 2-D video and audio capabilities have become standard at most devices, 3-D video and audio capabilities or high dynamic range (HDR) displays cannot be readily assumed to be available. The support for other stimuli, for example, haptic or olfactory stimuli, is nearly non-existent in common computer hardware as used by the workers and, thus, these stimuli are currently not suitable for QoE crowdtesting. Besides these technical factors, QoE assessment methodologies requiring the interaction between different workers, e.g., for interactive video conferencing, are possible, but challenging in their execution. Taking these limitations into account, QoE crowdtesting is feasible for typical Web applications like Web browsing or file download, 2-D video, image and audio QoE assessment tasks, where the usable formats depend on the bandwidth requirement.

Future challenges of QoE crowdtesting include eliminating the limitations mentioned above. While little can be done in improving the workers' hardware configuration, an interesting approach for interactive applications is using human-based computation or game with a purpose.

References:

[1] T. Hoßfeld, "On Training the Crowd for Subjective Quality Studies," VQEG eLetter, vol. 1, Mar. 2014.

[2] T. Hoßfeld and C. Keimel, "Crowdsourcing in QoE Evaluation," in Quality of Experience: Advanced Concepts, Applications and Methods, S. Mo¨ller and A. Raake, Eds. Springer: T-Labs Series in Telecommunication Services, Mar. 2014.

[3] T. Hoßfeld, R. Schatz, and S. Egger, "SOS: The MOS is not enough!" in QoMEX 2011, Mechelen, Belgium, Sep. 2011.

Tobias Hossfeld is heading the FIA research group "Future Internet Applications & Overlays" at the Communication University of He finished his 2009 and his l thesis n) "Modeling and of Internet ns and Services" in has been visiting senior researcher at FTW in Vienna with a focus on Quality of Experience research. He has published more than 100 research papers in major conferences and journals and received the Fred W. Ellersick Prize 2013 (IEEE Communications Society) for one of his articles on QoE.

Christian Timmerer is an assistant professor at the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria. His research interests include immersive multimedia communication, streaming, adaptation, and Quality of Experience with more than 100 publications in this eneral chair of WIAMIS'08, participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, QUALINET, and SocialSensor. He also participated in ISO/MPEG work for several years, notably in the area of MPEG-21, MPEG-M, MPEG-V, and DASH/MMT. He received his PhD in 2006 from the Alpen-Adria-Universität Klagenfurt. Follow him on twitter.com/timse7 and subscribe to his blog blog.timmerer.com.

Alternative state-based approach to video communication for streaming applications

A short review for "Informative State-Based Video Communication"

Edited by Weiyi Zhang

Jacob Chakareski, "Informative State-Based Video Communication", IEEE Transactions on Image Processing, Vol. 22, No. 6, June 2013.

The authors studied the problem of streaming packetized media over an unreliable network, from a server to a client. In conventional sender-driven streaming, the client replies with an acknowledgment packet whenever a media packet arrives. The purpose of the acknowledgment packet is to inform the server that the client has received the corresponding media packet and that the server does not need to consider retransmitting that media packet again. On the other hand, in conventional receiver-driven streaming the client proactively requests the transmission of the individual media packets, by sending the server corresponding request packets on the backward channel. The server only sends data in response to arriving requests.

The concept of simultaneous acknowledgement of multiple packets has been introduced in the networking community under the name of vector acknowledgments, with the goal to allow a TCP sender to perform selective retransmission of lost data packets. Presently, this feature is not supported by the acknowledgment scheme employed by TCP, called cumulative acknowledgements (CACK) [1]. In essence, vector ACKs are binary maps that describe the correctly received or missing data in the receiver's buffer and have been adopted into several proposed feedback schemes [2]–[4]. Another alternative for providing selective retransmission in TCP is the selective acknowledgement option (SACK) [5], which allows a receiver to communicate simultaneously the identities of several contiguous blocks of successfully received data. Vector acknowledgements have been included as an option in the recently proposed Datagram Congestion Control Protocol (DCCP) [6], which implements a congestion-controlled unreliable flow of datagrams suitable for use by applications such as streaming media, Internet telephony, and on-line games. Similarly, vector acknowledgements have been defined as an optional scheme under the name of Block Acknowledgement (BA) in the IEEE 802.11e standard [7], [8] for wireless local area networks, in order to improve the efficiency of the Media Access Control (MAC) layer. The recently ratified amendment IEEE 802.11n [9], [10] enhanced the BA mechanism and made it mandatory for all wireless devices in the network.

This work proposes an alternative state-based approach to video communication for streaming applications, where the client simultaneously informs the server about the presence or absence status of multiple packets in its buffer. In particular, in sender-driven transmission, the client periodically sends to the server a single acknowledgement packet that provides information about all packets that have arrived at the client by the time the acknowledgment is sent. In receiver-driven streaming, the client periodically sends to the server a single request packet that comprises a transmission schedule for sending missing data to the client over a horizon of time.

The authors develop a comprehensive optimization framework that enables computing packet transmission decisions that maximize the end-to-end video quality for the given bandwidth resources, in both prospective scenarios. The core step of the optimization comprises computing the probability that a single packet will be communicated in error as a function of the expected transmission redundancy (or cost) used to communicate the packet. Through comprehensive simulation experiments, the authors examine the performance advances that the proposed framework enables relative to state-of-the-art scheduling systems that employ regular acknowledgement or request packets. Consistent gains in video quality of up to 2B are demonstrated across a variety of content types.

It is shown in this paper that there is a direct analogy between the error-cost efficiency of streaming a single packet and the overall rate-distortion performance of streaming the whole content. In the case of sender-driven transmission, an effective modeling approach that accurately characterizes the end-to-end performance is developed as a function of the packet loss rate on the backward channel and the source encoding characteristics.

In summary, there are two major contributions of the present paper. First, the introduction of the concept of state-based video communication and studied its merits for streaming packetized content. Second, the authors have formulated the related optimization problems of computing the packet transmission schedules that maximize the end-to-end performance, for the given network resources, in the both cases of sender-driven streaming via state-based acknowledgements and receiver-driven streaming via state-based request horizons. It is shown that the proposed framework outperforms the state-of-the-art that employs regular acknowledgements or requests in its operation. The performance gains from the proposed framework delivers against conventional content-agnostic packet scheduling systems are also decent. This work provides a comprehensive analysis and to-date experimental findings.

References:

[1]. W. Stevens, TCP/IP Illustrated, Volume 1: The Protocols. Boston, MA, USA: Addison-Wesley, 1994.

[2]. B. Doshi, P. Johri, A. Netravali, and K. Sabnani, “Error and flow control performance of a high speed protocol,” IEEE Trans. Commun., vol. 41, no. 5, pp. 707–720, May 1993.

[3]. J. C. Lin and S. Paul, “RMTP: A reliable multicast transport protocol,” in Proc. IEEE INFOCOM, vol. 3. San Francisco, CA, USA, Mar. 1996, pp. 1414–1424.

[4]. H.-S.W. So, Y. Xia, and J. Walrand, “A robust acknowledgement scheme for unreliable flows,” in Proc. IEEE INFOCOM, vol. 3. Jun. 2002.

[5]. M. Mathis, J. Madhavi, S. Floyd, and A. Romanow. (1996, Oct.). “TCP selective acknowledgment options,” [Online]. Available: http://www.ietf.org/rfc/rfc2018.txt

[6]. E. Kohler, M. Handley, S. Floyd, and J. Padhye. (2003 Oct.). “Datagram Congestion Control Protocol (DCCP),” [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-dccp-spec-05.txt

[7]. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, IEEE Standard 802.11e, Nov. 2005.

[8] X. Pérez-Costa and D. Camps-Mur, “IEEE 802.11E QOS and power saving features overview and analysis of combined performance,” IEEE Wireless Commun. Mag., vol. 17, no. 4, pp. 88–96, Aug. 2010.

[9] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 5: Enhancements for Higher Throughput, IEEE Standard 802.11n, Oct. 2009.

[10] Y. Xiao, “IEEE 802.11n: Enhancements for higher throughput in wireless LANs,” IEEE Wireless Commun. Mag., no. 6, pp. 82–91, Dec. 2005.

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.

Constructing Structured Dictionaries using Graphs

A short review for "Parametric Dictionary Learning for Graph Signals"

Edited by Gene Cheung

Dorina Thanou, David I. Shuman, Pascal Frossard*, "***Parametric Dictionary Learning for Graph Signals***," Symposium on Graph Signal Processing* in *IEEE Global Conference on Signal and Information Processing*, December, 2013.

Graph signal processing (GSP) is the study of signals that live on structured data kernels described by graphs [1]. A graph-signal ***x*** assigns a value *x(v)* to each node *v* in a graph *G = (V, E, W)*, where *V*, *E*, *W* are the nodes, edges and edge weights respectively. Weight $w_{i,j} \in W$ for an edge connected node *i* and *j* tends to have physical meaning, such as the geometric distance between the two nodes. Graph-signals exist naturally around us, in transportation, data and social networks, etc. One example would be the temperatures taken by nodes in a sensor network. Because the data kernel itself exhibits structure, known signal processing tools like transforms and wavelets developed for more traditional discrete signals like audio (1D) and images (2D) that live on regular grid are not directly applicable to graph-signals. For example, elementary operations like convolution and translation are not easily applied to graph-signals. GSP thus seeks to develop a new tool set for processing of graph-signals.

One of the recent popular representations for discrete signals is *sparse coding*: given an over-complete dictionary ***D*** of *M* atoms each of dimension *N*, where *N* < *M*, we represent a signal $\boldsymbol{x} \in R^N$ using a sparse linear combination of dictionary atoms, where the number of atoms employed are far fewer than the dimension of the signal. In other words, sparse coding finds a code vector α that minimizes the following:

$$\min_{\alpha} \|x - D\alpha\|_2 + \lambda \|\alpha\|_0 \qquad (1)$$

where λ is a parameter that trades off the fidelity term (first term) and the sparsity term (second term). Such sparse representation of signals can be very useful, for example, for regularization of ill-posed inverse imaging problems such as image denoising [2], super-resolution, etc. Given a training set of signals, dictionary training methods in the literature such as K-SVD [3] can lead to very sparse signal representation. However, the trained dictionaries are usually unstructured, which means that solving (1) can be computationally expensive. Further, if the underlying signals are actually graph-signals, then these methods are not able to exploit the underlying data kernel structure towards better dictionary design.

In this reviewed paper, the authors proposed a new dictionary learning method for graph-signals, where the structure of the data kernel is embedded into the designed dictionary. First, un-normalized graph Laplacian $\boldsymbol{L}_u$ is defined as $\boldsymbol{L}_u$=***D***-***W***, where ***D*** is the diagonal degree matrix. A normalized variant is defined as $\boldsymbol{L} = \boldsymbol{D}^{1/2} \boldsymbol{L}_u \boldsymbol{D}^{1/2}$. It can be shown [1] that the normalized Laplacian ***L*** has a complete set of orthonormal eigen-vectors $\chi = [\chi_1, \ldots, \chi_N]$ with eigen-values λ_l's between 0 and 2.

Eigen-vectors of a graph Laplacian are used to form a basis for signal transformation (called *graph Fourier transform* (GFT)), resulting in a spectral decomposition of a graph-signal ***x***:

$$\hat{x}(\lambda_l) = \langle x, \chi_l \rangle = \sum_{n=1}^{N} y(n) \chi_l^*(n) \qquad (2)$$

where $\hat{x}(\lambda_l)$ is the l^{th} graph-frequency component by computing the inner product between the signal ***x*** and the $\boldsymbol{l}^{\text{th}}$ GFT basis vector.

GFT can also be used to define the notion of *generalized convolution* for graph-signals [5]:

$$(f * g)(n) = \sum_{l=0}^{N-1} \hat{f}(l) \hat{g}(l) \chi_l(n) \qquad (3)$$

where the convolution of graph-signals ***f*** and ***g*** is computed as the multiplication of the respective GFT domain representations plus inverse transform.

Similarly, using GFT *generalized translation* operator T_i is defined as the generalized convolution with a delta centered at vertex *i* [4]:

$$\begin{aligned}(T_i g)(n) &= \sqrt{N}(g * \delta)(n) \\ &= \sqrt{N}\sum_{l=0}^{N-1}\hat{g}(l)\chi_l^*(i)\chi_l(n)\end{aligned} \tag{4}$$

Given the above development, the authors then proposed to design dictionary atoms that are localized around center node *i* in the vertex domain, by assuming the kernel $\hat{g}(\lambda_l)$ is a smooth polynomial function of degree *K*:

$$\hat{g}(\lambda_l) = \sum_{k=0}^{K}\alpha_k \lambda_l^k \tag{5}$$

Note that the *k*th the power of the Laplacian ***L*** is exactly *k*-hop localized on the graph topology. Thus combining the previous two equations, we can construct a sub-dictionary of the form:

$$D_s = \hat{g}(L) = \sum_{k=0}^{K}\alpha_{sk} L^k \tag{6}$$

The authors then learned a structured graph dictionary $D = [D_1, D_2, \ldots D_S]$ that is a concatenation of *S* sub-dictionaries. In addition, two constraints are added to ensure: i) the kernels are non-negative and uniformly bounded by a given constant c, and ii) the learned kernels should cover the entire frequency spectrum.

More precisely, given training signals $Y = [y_1, \ldots, y_M]$the optimization being solved to learn the dictionary parameters is:

$$\begin{aligned}&\arg\min_{\alpha, X}\|Y - DX\|_F^2 + \mu\|\alpha\|_2^2 \\ &s.t. \|x_m\|_0 \le T_o, \forall m \in \{1, \ldots, M\}\end{aligned} \tag{7}$$

where, again, the sub-dictionaries have to satisfy the two additional constraints mentioned earlier. In words, the constraint in (7) states that the code vector $\boldsymbol{x}_m$ corresponding to training signal $\boldsymbol{y}_m$, must be sparse, with l_0-norm no larger than a pre-defined threshold T_o.

The authors show that (7) can be solved efficiently by alternately solving for one set of variables while holding the other set fixed. Further, the computational complexity for the learned structured dictionary can be significantly smaller than unstructured counterpart when solving for sparse vectors in (1). Experimental results show that the learned structured graph dictionary offers a good tradeoff between complexity and sparse coding performance for graph-signals.

This work is significant in that it is the first attempt to learn dictionary in a structured manner, so that complexity of computing the sparse code vector in (1) can be drastically reduced. This has potential to spur on other efficient structured dictionary designs in the SP community.

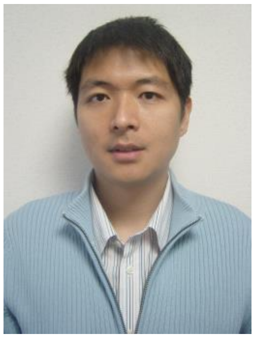
References:

[1] D. I Shuman, S. K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs: Extending High-dimensional Data Analysis to Networks and other Irregular Domains," *IEEE Signal Processing Magazine*, vol.30, no.3, pp.83-98, May 2013.

[2] W. Hu, X. Li, G. Cheung, O. Au, "Depth Map Denoising using Graph-based Transform and Group Sparsity," *IEEE International Workshop on Multimedia Signal Processing*, Pula (Sardinia), Italy, October, 2013.

[3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, November 2006.

[4] D. I Shuman, B. Ricaud, P. Vandergheynst, "A Windowed Graph Fourier Transform," *IEEE Statistical Signal Processing Workshop*, Ann Arbor, MI, August 2012.



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 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.

Face Hallucination by Exploring Image Structures

A short review for "Structured Face Hallucination"

Edited by Hao Hu

Chih-Yuan Yang, Sifei Liu and Ming-Hsuan Yang, "Structured Face Hallucination", in Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR13), pp. 1099-1106, 2013.

Face hallucination, first introduced by Baker and Kanade [1], is a domain-specific super-resolution of face images to generate high-resolution (HR) images from low-resolution (LR) inputs, clarifying details of faces. It can be viewed as an inverse task to reconstruct high-frequency details from LR images that result from HR images by a linear convolution process with down-sampling. Face hallucination has many applications in image enhancement and image compression, and in particular, it is useful in facial recognition systems or security surveillance systems where the resolution of human faces is typically low.

There are many algorithms and methods developed over the last decade to infer HR face images from LR inputs. In [1], a probabilistic framework is proposed to model the relationship between LR and HR image patches where for each query patch cropped from input image, the most similar LR patch from an exemplar set is retrieved and its corresponding HR patch is transferred with first and second order derivatives. The resulting HR images can have significantly richer details than general image processing methods, for example, bicubic interpolation. However, pure patch-based method can introduce artifacts without exploiting facial structures to resolve ambiguities between HR and LR patches. The method proposed in [2] is built on top of three constraints to ensure the quality of reconstructed HR images. It is a hybrid face hallucination method by combing a global parametric model for common faces and a local nonparametric model that learns local textures from example faces. One limitation of this method is the linear subspace representation, so that it performs well only when the images are precisely aligned at fixed poses and expressions. A similar method proposed in [3] can also suffer from ghostly artifacts as a result of using subspace representation.

In general, a face image contains several structural layers, i.e., facial components, contours and smooth regions. The key idea of this paper and what makes it unique is that the proposed method tries to treat those structural components differently and eventually fuse them together to generate the HR image. This method overcomes some drawbacks of the methods in the literature. Compared with pure-patch based method, it is more flexible and can potentially alleviate many artificial errors. In addition, it is not constrained by well-aligned pose and facial expression.

In order to generate a high quality HR image, the method takes four basic steps:

1) Finding Gradient Map for Facial Components (U_c): from a LR input, an intermediate HR image is generated via bicubic interpolation, from which, the pose and landmark points for facial components and contours can be determined using algorithms in [4]. Based on the landmarks and a given exemplar image from the dataset with the same pose, the optimal parameters of rotation, scaling and in-plane shift of each individual facial components can be calculated for the best alignment. Finally, the best gradient map for each facial component can be found from the exemplar set from a matching process. Additionally, glass tag side-information is used to improve the matching accuracy.
2) Finding Gradient Map for Facial Contours (U_e): in order to preserve the structure of facial edges, a direction-preserving upsampling method is designed where small patches and bilinear interpolation are used rather than pixels and bicubic interpolation. The edges produced by this method will not be sharp enough, so a further processing of non-parametric statistical learning based on training image set is applied to find the average magnitude of gradients for facial edges, which are used as weighting factors to determine the final gradient map from the upsampled contour map.
3) Finding Gradient Map for Smooth Regions (U_b): a simple patch-based matching algorithm is used to generate initial HR image, then, back projection algorithm [5] is applied on the HR image to ensure the down-sampled LR image matches the input. Finally, the gradients for background can be calculated.
4) Integrating Gradient Maps: the gradient map U for producing output HR image is determined via weighted sum over all three subcomponents, which eventually leads to reconstructing the HR image for the input LR image.

The performance evaluation of this proposed method is conducted over Multi-PIE dataset [6]. The dataset contains pose labels and landmarks and the authors manually generate the glasses labels for training images. The MATLAB implementation takes about 1 minute to process one 60x80 LR image on a machine with 2.8 GHz Quad Core CPU.

For performance comparison, some state-of-the-art methods are considered. In terms of objective metrics (PSNR, SSIM), the classic back-projection method [6] performs the best, but the generated HR images contain jaggy edges and fewer details. The method of [2] performs well but it suffers from noisy and blocky effects and over-smooth regions. Other methods tend to produce obscure facial details as well. For input images at different pose, the subspace-based methods do not perform well due to lack of precise face alignment. On the contrary, both the back-projection method and the proposed method show satisfying results, and the proposed method performs better in constructing facial details.

This work shows the benefits of decomposing face hallucination problem into finding gradient maps for each image components. The experimental results show that the method can perform well visually with more facial details. It can inspire other researches to explore domain-specific knowledge and derive novel image/components processing algorithms to improve not only the face hallucination but also super-resolution method in general.

References:

[1] S. Baker and T. Kanade, “Hallucinating faces,” *in Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, March 2000.

[2] C. Liu, H.-Y Shum and W. T. Freeman, “Face hallucination: Theory and Practice,” *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[3] X. Wang and X. Tang, “Hallucinating face by eigentransformation,” IEEE trans. On Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 35, no. 3, pp. 425–434, 2005.

[4] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild”, in *Proc. IEEE CVPR*, 2012.

[5] M. Irani and S. Peleg, “Improving resolution by image registration”, *CVGIP: Graphical Models and Image Processing, vol 53, no. 3, pp 231-239, May 1991*.

[6] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE”, *in Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.

Hao Hu received the B.S. degree from Nankai University and the M.S. degree from Tianjin University in 2005 and 2007 respectively, and the Ph.D degree from Polytechnic Institute of New York University in Jan. 2012.

[illegible] ENG Labs, Cisco Systems, San Jose, CA. He interned in the Corporate Research, Thomson Inc., NJ in 2008 and Cisco Systems, CA in 2011. His research interests include video QoE, video streaming and adaptation, and distributed systems. He has served as reviewer for many journals and magazines, including IEEE JSAC, IEEE TCSVT, IEEE TMM, IEEE TON, IEEE TPDS; he also served as TPC member for conferences, including ICNC, ICME.

Estimating Important Factors in Multi-frame Super Resolution

A short review for "On Bayesian adaptive video super resolution"

Edited by Jun Zhou

C. Liu and D. Sun, "On Bayesian adaptive video super resolution", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, vo. 2, pages 346-360, February, 2014.

Multi-frame super resolution methods aim at reconstructing a sequence of low-resolution video frames into a high resolution image. Such video enhancement technique is highly demanded with the rapid development of high-definition television and monitors.

During the past decades, many super-resolution approaches have been proposed, for example, using interpolations, patch prior learning, statistical edge information-based reconstruction [1], and more convenient single frame up-sampling [2]. Among these methods, common tasks often include estimation of motion, noise level, and blur kernel. However, seldom has any approaches performed these three tasks simultaneously.

In this paper, Liu and Sun proposed a method for adaptive multi-frame super resolution that integrates three key factors, i.e., the estimation of optical flow, noise level, and blur kernel, under a Bayesian framework. The relationship between these three key factors, aliasing signals, and quality of super resolution has been analyzed under the assumption of both perfect motion estimation and unknown motion. The authors pointed out that an optimal-size blur kernel leads to a balance performance in suppressing aliasing in image formation and boosting noise in image reconstruction.

Given a high-resolution frame, the corresponding low resolution frame is considered as a smoothed, down-sampled image with noise. Therefore, the optimal reconstruction solution can be solved by a Bayesian MAP method in which the posterior is the product of priors of high-resolution image, smoothing kernel, optical flow field, and noise given the low-resolution frame sequence. With the current estimation of other parameters handy, the proposed approach first performs image reconstruction by solving a linear system of all parameters using an iterated re-weighted least squares method. Then the reconstructed image and the blue kernel are used to jointly estimate the flow field and noise level in a coarse-to-fine fashion. Finally the blur kernel can be recovered by fixing other parameters in the same manner as the image reconstruction step.

In order to analyze the performance bound for motion estimation with aliasing and noise, the authors performed a two-step analysis using the Cramer-Rao bounds [3]. The first step studies how blur kernel and noise influence motion estimation with aliasing signals, and how noise affects super resolution with perfect motion. The second step studies the performance bound for image reconstruction with errors in motion, in particular, maximum likelihood estimator with perfect motion, and the performance of the estimator with motion error. Two important conclusions can be drawn from such analysis. First, a higher noise level makes super resolution more difficult. Second, a small blur kernel boosts less imaging noise during image reconstruction but suppresses less aliasing during image formation, and vice versa.

The effectiveness of the proposed method was validated on several real-world video sequences [4]. Experiments have been performed to demonstrate that the proposed video super resolution system is robust to noise. When level of added synthetic white noise increases, the peak signal to noise ratio of the reconstructed image does not degrade much. The results also show that the estimation of the point spread function of the blur kernel is accurate. In order to show the advantage of the proposed Bayesian method, it is compared against several state-of-the-arts methods. It has generated better reconstructed images in terms of peak signal to noise ratio and structural similarity.

The take home message from this paper is that jointly parameter estimation can lead to better multi-frame super resolution performance. During the system development, aliasing shall be treated in a prudent manner as low level aliasing means little information can be propagated from adjacent frames for reconstruct high-frequency details but too strong aliasing causes problems in motion estimation. Therefore, an optimum smoothing kernel shall be estimated.

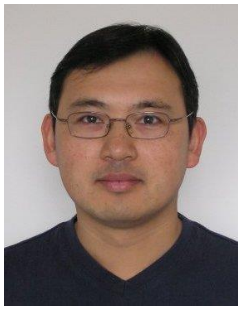
References:

[1] S. Park, M. Park, and M. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21-36, May 2003.

[2] Q. Shan, Z. Li, J. Jia, and C-K Tang, "Fast Image/Video Upsampling," ACM Transactions on Graphics, vol. 27, no. 5, article 153, 2008.

[3] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[4] C. Liu and D. Sun, "A Bayesian approach to adaptive video super resolution," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 209-206, 2011.



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.

Coding of Free-Viewpoint Video in Error-prone Channels

A short review for "Loss-Resilient Coding of Texture and Depth for Free-Viewpoint Video Conferencing"

Edited by Carl James Debono

B. Macchiavello, C. Dorea, E.M. Hung, G. Cheung and W.-T. Tan, "Loss-Resilient Coding of Texture and Depth for Free-Viewpoint Video Conferencing," IEEE Transactions on Multimedia, vol. 16, no. 3, pp. 711-725, April 2014.

Improvements in multimedia technology are already providing stereo video transmission to the homes. Further development is expected to present more realistic 3D immersive solutions. Amongst these applications is Free-viewpoint video transmission [1]. This allows the viewer to arbitrarily select his/her point of observation of the 3D scene. This demands the transmission of a huge amount of views to satisfy all the possible users. To reduce the number of views needed, texture video can be accompanied by the respective depth map video [2] and virtual views reconstructed at the receiver from two such streams using depth-image-based rendering (DIBR) [3]. Transmission errors and packet loss on the channel has an adverse effect on the quality of the synthesized video. Techniques are therefore needed to reduce the impact of packet losses on the quality of experience of the viewer.

The key technical contribution of the paper being reviewed is twofold, both based on the following observation: the texture-plus-depth format of free viewpoint video—texture and depth videos captured from multiple closely spaced viewpoints—is itself a redundant representation. In particular, a voxel in the 3D scene visible from two different viewpoints is represented twice as one or more pixels in the two viewpoint images. Assuming the surface reflectance is Lambertian, the double representation of a 3D voxel does not provide any more information than the single representation of the same voxel. Therefore, when streaming two texture / depth image pairs from two viewpoints (which are coded independently) over loss-prone and bandwidth-limited networks, the virtual view synthesis process at the receiver can judiciously mix the two representations for rendering of a virtual voxel, where a more reliably delivered representation is weighted more heavily in the mixture. In other words, any pixel errors in one of the views can be minimized during virtual view synthesis by relying on the pixels representing the same 3D object in another view that has been more reliably delivered, during view synthesis. This receiver-side optimization as is referred to as adaptive blending.

Correspondingly, at the transmitter, judiciously protection against channel errors can be dedicated to only one of the two representations. That is one of the views is more heavily error resilient encoded to ensure that at least one of the representations is correctly delivered with high probability. Instead of using an automatic retransmission request (ARQ) that would introduce end-to-end retransmission delay or forward error correction (FEC) which does not perform well in burst-loss environments, the authors of the paper accomplish unequal loss protection via reference picture selection (RPS), where an important pixel block in a viewpoint image deemed important is either intra-coded or predicted using a reference block in a previous frame that is further into the past—one that has already been acknowledged (ACKed) by the receiver to have been correctly decoded, for example. For the same quantization parameter (QP), employing a reference frame further into the past results in a large temporal distance between the reference and the target frames, leading to a coding overhead. The optimization is then to find the optimal set of motion vectors (MV) for blocks in the current frame to minimize the virtual view synthesis distortion, taking into consideration the adaptive blending process at the decoder. This was implemented using the Lagrangian relaxation and an alternating two-step algorithm.

Finally, the authors of the paper note that the adverse effects on the synthesized view image given errors in the texture and depth images are different: a pixel error in a texture image leads directly to a proportional error in the synthesized view image, while a pixel error in a depth image leads to a geometric error, resulting in wrong mapping of the texture pixels to the virtual view. This effect of depth pixel errors is modeled using quadratic penalty functions, so that texture and depth maps can be separately optimized for transmission, leading to a faster, parallel optimization procedure. Experimental results conducted by the authors of the paper show that the solution produces better quality results in error-prone channels.

Transmission of free-viewpoint services is still far from becoming a mainstream type of transmission.

Work is currently being done in the standardization arena to develop the 3D High Efficiency Video Coding (3D-HEVC) extension that will be based on texture plus depth coding and is expected to exploit also the correlation between texture and depth other than the traditional correlations in space, time and between views. This will definitely reduce the amount of data that needs to be transmitted making it more viable in bandwidth-limited channels. Further work is also needed in the area of error correction and control, such that the reconstructed video data is presented with better quality, given that the original data is recovered. This has to be coupled with better error concealment methods that exploit all the data available from other views and depth data to inpaint any missing data blocks. The realization of free-viewpoint television and 3D television depends on the success of such solutions that can provide the user of the technology with a high quality immersive and personal experience.

References:

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, “Multi-view imaging and 3DTV,” in *IEEE Signal Processing Magazine*, vol. 24, no.6, November 2007.

[2] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, “Multi-view video plus depth representation and coding,” in *Proceedings of the IEEE International Conference on Image Processing*, October 2007.

[3] D. Tian, P.-L. Lai, P. Lopez, and C. Gomila, “View synthesis techniques for 3D video,” in *Applications of Digital Image Processing XXXII*, *Proceedings of the SPIE*, vol. 7443 (2009), 2009, pp. 74 430T–74 430T–11.

Carl James Debono (S’97, M’01, SM’07) received his B.Eng. (Hons.) degree in Electrical Engineering from the University of Malta, Malta, in 1997 and the Ph.D. degree in Electronics and Computer Engineering from the University of Pavia, Italy, in 2000.

Between 1997 and 2001 he was employed as a Research Engineer in the area of Integrated Circuit Design with the Department of Microelectronics at the University of Malta. In 2000 he was also engaged as a Research Associate with Texas A&M University, Texas, USA. In 2001 he was appointed Lecturer with the Department of Communications and Computer Engineering at the University of Malta and is now an Associate Professor. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta.

Prof. Debono is a senior member of the IEEE and served as chair of the IEEE Malta Section between 2007 and 2010. He is the IEEE Region 8 Vice-Chair of Technical Activities for 2014. He has served on various technical program committees of international conferences and as a reviewer in journals and conferences. His research interests are in wireless systems design and applications, multi-view video coding, resilient multimedia transmission, and modeling of communication systems.

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

Multimedia Communications Technical Committee (MMTC) Officers

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.