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

R-LETTER

Vol. 6, No. 2, April 2015

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the April 2015 issue of the Review Letter (R-Letter) of the IEEE Communications Society Multimedia Communications Technical Committee (MMTC). This issue is brought to you by review board members who independently nominated research papers published within IEEE MMTC sponsored publications and conferences.

We hope that this issue **stimulates your research in the area of multimedia communication** featuring topics:

- smart and helmet cameras;
- adaptive video streaming in managed networks;
- mobility-aware resource allocation, screen content coding; and
- image representation.

An overview of all reviews are provided in the following:

The **first paper**, published in the *IEEE Transactions on Circuits and Systems for Video Technology* and *edited by Pradeep K. Atrey*, provides means for reducing the costs of re-identification for smart camera networks.

The **second paper**, published in the *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2014* and *edited by Frank Hartung*, describes a method for video smoothing of rough and shaky helmet camera video recordings.

The **third paper** is *edited by Roger Zimmermann* and has been published within the *IEEE Transactions on Multimedia*. It provides means for quality optimization of adaptive HTTP streaming within managed networks.

The **forth paper**, published in the *Proceedings of the IEEE GLOBECOM 2014* and *edited by Koichi Adachi*, comprises a mobility-aware resource allocation scheme under channel uncertainty.

The **fifth paper**, published in the *IEEE Transactions on Multimedia* and *edited by Bruno Macchiavello*, proposes an indexed color representation for screen content coding using HEVC.

Finally, the **sixth paper** is *edited by Jun Zhou* and published in *IEEE Transactions on Circuits and Systems for Video Technology*. It describes a marriage between conventional image representation and deep neural networks.

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: Christian Timmerer
Alpen-Adria-Universität Klagenfurt, Austria
Email: christian.timmerer@itec.aau.at

Co-Director: Weiyi Zhang
AT&T Research, USA
Email: wzhang@ieee.org

Co-Director: Yan Zhang
Simula Research Laboratory, Norway
Email: yanzhang@simula.no

Reducing Cost of Re-identification for Smart Camera Networks

A short review for "Cost-Effective Features for Re-identification in Camera Networks"
(Edited by Pradeep K. Atrey)

S.F. Tahir and A. Cavallaro, "Cost-Effective Features for Re-identification in Camera Networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 8, Aug. 2014.

Person re-identification is the problem of associating people detected in different camera views over time [1]. As the number of cameras used in camera networks increases, a high data transfer rate and a high processing power of a central node is required in the traditional multi-camera systems architecture [2]. To increase scalability *smart cameras* with their storage and computation capabilities process locally the captured videos. Appearance features such as colour, texture and shape are extracted from single or multiple images to describe people for their re-identification across the camera network [3]. Existing re-identification approaches exploit features for improving the re-identification rate without considering constraints on the resource utilization. However, smart cameras might have limited resources, which require that a minimum amount of data is generated for processing and sharing across the network. Data reduction can be achieved by feature selection.

Existing feature selection methods focus on reducing the number of features by considering their *performance*, i.e. their discriminative power in representing an object [6, 7, 8]. However, very little work has been done in considering the *cost* of features, including the computational time for their extraction and the amount of data for their storage. The cost constraint in feature selection becomes particularly important when the features type varies significantly. The cost is independent of the feature performance and leads to a constrained feature subset selection.

In this paper, the authors devise a mechanism to reduce the cost of re-identification by proposing a novel cost-and-performance-effective (CoPE) feature-selection method. The amount of data stored for each feature and the computational time for its extraction are used jointly with its performance to generate an overall feature score. Performance is quantified by measuring the similarity between two views of the same as well as different people using each feature, while the cost is the inverse of the average of storage size and computation time for feature extraction from the available objects. The most discriminative, well-performing and cost-effective features are selected by evaluating each feature individually and then by ranking the selected features based on their contribution to the task.

The primary contribution of this paper is a scalable approach for sharing data over the camera network. Individually selected features as in the case of CoPE can perform well in constrained environments when some features need to be discarded adaptively due to user requirements or application constraints; whereas in the case of feature selection based on group performance, the removal of a single feature may significantly reduce the effectiveness of the whole feature set.

Furthermore, the authors avoid the so-called performance overlapping, a measure of similarity among features, in the feature selection by iteratively removing the data points (people) from the training date that have taken part in the selection of a feature. Thus, each selected feature becomes representative of a unique subset of data: each selected new feature increases the diversity in the feature set by covering a wider range of data, which increases the discriminating ability of the feature set.

The approach proposed in the paper is developed for camera pairs and it is therefore appropriate for distributed multi-camera settings, where each camera communicate with its neighbors without a central control unit. The authors describe the deployment of the approach to the camera network as a one-time set-up off-line process. Learning for feature selection is performed once using the training data. After the setup, each camera stores locally a list of selected features for each of its neighbouring camera. Only the selected features are extracted that do not affect the subsequent on-line person re-identification task when the camera network is operational. If a new camera is added to the network, the training needs to be performed between the new camera and its

neighbouring cameras only in a pair-wise manner.

The method proposed in the paper, CoPE, is evaluated using direct distance minimization for both indoor and outdoor camera settings in the person re-identification task. CoPE is compared with two existing re-identification approaches, namely probabilistic relative distance comparison [1] and attribute sensitive feature importance [3], and five feature-selection approaches, fisher score, information gain, minimum-redundancy maximum-relevance [6], ReliefF [7] and Biclusters [8]. CoPE considerably reduces network traffic due to inter-camera feature sharing while keeping the re-identification performance at an equivalent or better level compared with the state of the art. CoPE also improves the performance of learning methods for re-identification with rankSVM [4] and AdaBoost [5] by reducing the feature dimensions, the training time and by improving their effectiveness.

In summary, CoPE proves to be a suitable feature selection approach for re-identification by identifying the most appropriate features for the task. A further reduction of the selected features is made possible to account for additional operational constraints (e.g. limited resources). CoPE decreases both the amount of data generated per feature set and the amount of time needed for the extraction of the selected feature set, up to 70% compared to using a complete feature set, without compromising the re-identification rate.

The software of CoPE is available at www.eecs.qmul.ac.uk/~andrea/software.htm

Acknowledgement:

The R-Letter Editorial Board thanks the authors of the paper for providing a summary of its contributions.

References:

[1] W.-S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," IEEE Trans. Pattern Anal. Mach. Intell., 35(3): 653–668, 2013

[2] V. Sulic, J. Pers, M. Kristan, and S. Kovacic, "Efficient feature distribution for object matching in visual-sensor networks," IEEE Trans. Circuits Syst. Video Technol., 21(7):903–916, 2011.

[3] C. Liu, S. Gong, and C.C. Loy. On-the-fly feature importance mining for person re-identification. Pattern Recognition, 47(4):1602–1615, 2014

[4] B. Prosser, W.S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in Proc. BMVC, U.K., Aug. 2010.

[5] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in Proc. ECCV, France, 2008.

[6] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," IEEE Trans. Pattern Anal. Mach. Intell., 27(8):1226–1238, 2005.

[7] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relief and relief," Mach. Learn., 53(2):23–69, 2003.

[8] Q. Huang, D. Tao, X. Li, L. Jin, and G. Wei, "Exploiting local coherent patterns for unsupervised feature ranking," IEEE Trans. Syst., Man, Cybern. B, Cybern., 41(6):1471–1482, 2011.

Pradeep K. Atrey is an Assistant Professor at the State University of New York, Albany, NY, USA. He is also an (on-leave) Associate Professor at the University of Winnipeg, Canada. He received his Ph.D. in Computer Science from the National University of Singapore. He was a Postdoctoral Researcher at the MCR Lab, University of Ottawa. His current research interests are in the area of Security and Privacy with a focus on multimedia surveillance and privacy, multimedia security, secure-domain cloud-based large-scale multimedia analytics, and social media. He has authored/co-authored over 95 research articles at reputed ACM, IEEE, and Springer journals and conferences. Dr. Atrey is on the editorial board of several journals including ACM Trans. on Multimedia, ETRI Journal and IEEE Communications Society Review Letters. He was also guest editor for Springer Multimedia Systems and Multimedia Tools and Applications journals. He has been associated with over 50 international conferences/workshops in various roles such as General Chair, Program Chair, Publicity Chair, Web Chair, Demo Chair and TPC Member.

Video Smoothing of Rough and Shaky Helmet Camera Video Recordings

A short review for "First-person hyper-lapse videos"
(Edited by Frank Hartung)

J. Kopf, M. F. Cohen, R. Szeliski, "First-person hyper-lapse videos", ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2014, Vol. 33, Issue 4, Article No. 78, July 2014.

Small helmet- or body-attached video cameras have become extremely popular and provide high-quality video at high frame rates and resolutions; in the moment, typically up to 240 frames per second and up to 4K resolution [1]. Many videos captured using such cameras can be found on YouTube and other video portals. Sports and outdoor activities are popular themes for these videos. A drawback is however that the camera paths and, thus, the captured videos are often very shaky, up to the point where it becomes unpleasant to watch them. This is even worse if the playback speed is increased for time-lapse videos. The time acceleration increases the shakiness.

While there are many frame-based video stabilization methods described in the scientific literature, and available in products like video cameras and video editing software, they are usually limited to the case where there is sufficient overlap between consecutive frames. If the video is too shaky, e.g. when a camera pans back and forth rapidly, conventional stabilization methods [2][3][4] fail.

In this publication, the authors present a complete concept for the generation of stabilized time-lapse videos from shaky and blurred video recordings using moving cameras (hence the term hyper-lapse, which refers to the case where the camera is moving through space as well as accelerated through time). The authors first construct a sparse 3D model of the scene using structure-from-motion algorithms and depth map interpolation. Then, a smooth virtual camera path through the scene is planned. Along the virtual new camera path, images are generated using depth image based rendering techniques. The images are concatenated into a video, giving a smooth time-lapse video from shaky input video material.

In the first step, the scene geometry and camera position is reconstructed for each frame of the input video. First, the video is pre-processed by removing lens distortion and converting the video to linear perspective projection. Then, the camera parameters and sparse depth maps are estimated using structure-from-motion algorithms [4][5][6]. The algorithm estimates the location and orientation of the input cameras. In addition, it computes a sparse 3D point cloud, where each point has an associated list of frames where it is visible.

In the next step, the sparse depth map is interpolated in order to get dense depth information.

Then, in the second step, a virtual camera path is planned that finds a compromise between different objectives: it should be smooth, but not too different from the original real camera path, and the virtual camera positions along the path should be oriented towards directions that can later be rendered well from the reconstructed geometry. These objectives are expressed mathematically (where the appropriateness of a camera position is measured by the texture stretch necessary to render the frame from the 3D model), combined in a weighted sum, and the resulting term is minimized to get the optimal camera path. This is done in two stages: first, the camera path is optimized, then the camera orientation is optimized. The step of optimizing the path of the virtual camera is a focus of the publication and a major new contribution.

In the third and last step, virtual images are rendered from the 3D model generated in step 1 and the virtual camera positions generated in step 2, using depth image based rendering (DIBR) techniques [7]. For each frame to be rendered, a number of original input frames –typically 3 to 5– are selected and projected into the image plane, using DIBR. These projected original images overlap and are then stitched and blended together to give a reconstructed video frame. Motion blurry original frames are excluded.

The authors have fully implemented their scheme and provide demonstration hyper-lapse video material, for example for rock climbing or bicycling video material. They also provide comparisons to traditional frame-by-frame video stabilization techniques, such as the Warp Stabilizer in Adobe After Effects, Deshaker, and the method from [8]. The results are convincing. Complexity is however a problem; the whole method is computationally very expensive, and thus currently lends itself only to non-real-time post-processing of video.

The paper combines new ideas with a practical realization. The authors nicely demonstrate how 3D scene reconstruction and DIBR based rendering, hence computer graphics based approaches, can provide better results to the problem of video stabilization, that first seems to be a 2D or 3D image processing problem.

References:

[1] A. Eisenberg, "When a Camcorder Becomes a Life Partner", New York Times, no 6, 2010.

[2] Y. Matsushita, E. Ofek, W. Ge, X. Tang, H.-Y. Shum, "Full-frame video stabilization with motion inpainting", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pp. 1150-1163, 2006.

[3] A. Goldstein, R. Fattal, "Video stabilization using epipolar geometry", ACM Transactions on Graphics (TOG), vol. 32, no. 5, 2012.

[4] M. Grundmann, V. Kwatra, I. Essa, "Autodirected video stabilization with robust L1 optimal camera paths", Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[5] N. Snavely, S. Seitz, R. Szeliski. "Photo tourism: exploring photo collections in 3D" ACM Transactions on Graphics (TOG), vol. 25, no. 3, 2006.

[6] C. Wu, "Towards linear-time incremental structure from motion", Proceedings IEEE International Conference on 3D Vision (3DV 2013), 2013.

[7] B. Macchiavelli, "Improved View Synthesis in a 3-D Camera Space - A short review for 'Expansion hole filling in depth-image-based rendering using graph-based interpolation'", IEEE Comsoc MMTC R-Letter, vol. 6, no. 1, pp. 11-12, February 2015.

[8] S. Liu, L. Yuan, P. Tan, J. Sun, "Bundled camera paths for video stabilization", ACM Transactions on Graphics (TOG), vol. 32, no. 4, pp. 78, 2013.

Frank Hartung is a full professor of multimedia technology at FH Aachen University of Applied Sciences, Aachen, Germany. He received a MSc in electrical engineering from RWTH Aachen University, Germany, and a PhD in Telecommunications from University of Erlangen, Germany. He has been working with Ericsson Research, as a research team leader in Multimedia Technologies, from 1999 to 2011. His research interests include media security, networked multimedia, immersive multimedia communication, streaming, and mobile video. He has authored or co-authored more than 50 publications in this domain, and is the co-inventor of 16 granted patents. Dr. Hartung is a member of IEEE, VDE and ITG, and was in 2003-2004 serving as chairman of the German IEEE Signal Processing Chapter.

Quality Optimization for Adaptive Video Streaming in Managed Networks

A review for "In-Network Quality Optimization for Adaptive Video Streaming Services"
(Edited by Roger Zimmermann)

Niels Bouten, Steven Latré, Jeroen Famaey, Werner Van Leekwijck, Filip De Turck, "In-Network Quality Optimization for Adaptive Video Streaming Services", IEEE Transactions on Multimedia, vol. 16, no. 8, pp. 2281–2293, December 2014.

The authors investigate the deployment of HTTP adaptive streaming techniques (HAS) in the context of so-called managed networks. These are networks where the service provider has more control over the infrastructure as compared to over-the-top (OTT) streaming, which refers to the delivery of media and other content over the regular, distributed Internet (i.e., over standard IP protocols) rather than proprietary infrastructures, such as cable networks. For OTT service providers (e.g., Netflix) there has recently been a considerable trend to move towards HAS techniques, for example by employing MPEG-DASH [1], since there exists a natural fit for using HTTP over the existing Internet infrastructure. However, even in managed networks, there is now a consolidation in progress towards using IP as the underlying protocol, since it creates flexibility and allows service providers to offer, for example, so-called triple-play services (TV, Internet and phone).

In many large-scale streaming systems resources such as bandwidth are generally limited and hence mechanisms for quality-of-experience (QoE) adaptations are necessary to optimize the overall end-user experience in the face of constantly changing demand. One of the features of HAS is that QoE adaptation is performed at the client side and not on the server. This has the advantage of simplifying the server design and distributing the computation that is needed for adaptation. However, in this study, the authors point out that purely client-driven quality-of-experience (QoE) adaptation, while having many advantages, also has some disadvantages in managed networks in that clients independently and competitively optimize their own quality and a global optimum may not be achieved. They point to other studies that have already shown that the on-off behavior of HAS in combination with high-bandwidth connections can lead to inappropriate adaptations if each client optimizes independently [2]. Hence, the authors postulate that in this case it can be beneficial to limit and manage some of the choices the client can make.

The approach that the authors propose is to apply a hybrid strategy of in-network optimizations together with client-side adaptation in order to increase the overall quality and stability in managed networks with many clients. Specifically, they model the network topology as a tree structure and impose bandwidth limitation constraints on each edge on the path between every client and the server. The proposed problem formulation then attempts to maximize the QoE over all clients while adhering to the edge bandwidth constraints. To achieve the optimization, several different goals could be targeted: (a) maximize the overall delivery rate in the system, (b) proportional bandwidth fairness among all the clients, or (c) minimize the number and the distance of quality switches. Other optimization objectives are also possible.

To solve the problem formulation the authors use linear programming approaches. An optimal solution can be achieved with a centralized Integer Linear Programming (ILP) algorithm. In this case the server (or some other centralized node) has global knowledge of all the needed information and computes the optimal bandwidth assignment for each client and quality level (hence termed *Centralized Exact*). However, the number of constraints and the information needed about the topology and clients renders such a solution complex and very time consuming to execute. To address this concern, the authors propose two other solutions, both of them decentralized. The first still uses an ILP algorithm where each node locally optimizes the allocation problem and forwards the solution to its predecessor node in the tree (a technique termed *Decentralized Exact*). The second approach not only decentralizes the optimization computation to each node in the distribution topology, but it also relaxes the integer constraints by heuristics and approximations (*Decentralized Relaxed*). Clearly this last method cannot achieve optimal assignments. However, as is shown in the experimental results, the algorithm execution time is much reduced with the decentralized relaxed approach.

One of the practical concerns of a linear programming approach is that the solution needs to be recalculated whenever there is a change in the topology, i.e., whenever a client joins or leaves. This is one of the main reasons that the centralized solution is not very practical. Not only does the optimization computation take much time, it also needs to be redone very frequently. Under these circumstances the distributed solutions have the advantage that not only the optimization space is smaller, if there are no changes in the sub-tree of a node then the optimization does not need to be rerun. In the *Decentralized Exact* case results still have to be propagated up the tree, which leads to an overall long execution time. In the relaxed case the local calculation performed is fast, but then a global optimum cannot be achieved. However, the authors show that the difference in the optimization utility between the accurate solutions and the relaxed, decentralized approach is in the low, single-digit percentages. Therefore, this seems to be a sensible compromise.

The evaluation of the proposed algorithms is performed with an NS3 based network simulation framework and the IBM CPLEX solver is used for the optimization functions. One of the important, realistic aspects of the simulation is that the execution time of the linear programming solver is accounted for such that the optimization results only become available after some delay. This effect reduces the overall optimality in a highly dynamic environment. The authors evaluate the results in terms of the number of clients, the number of bottlenecks, the delay, multiple servers and the optimization objective. Overall they are able to show that the proposed method can lead to a client quality improvement of 14% or more. The number of quality switches can also be reduced. Interestingly, because the decentralized relaxed approach has a very short execution time and hence allows for immediate installation of the new configurations, it performs better in a dynamic environment than the exact solutions.

The proposed approach targets managed large-scale networks and within that context this work is interesting. As mentioned, other studies have shown that the on-off behavior of HAS in combination with high-bandwidth connections can lead to inappropriate adaptations if each client optimizes independently. Furthermore, there seems to be a growing trend for HAS to be adopted by managed networks (e.g., cable companies) and hence such issues will attract more attention in the future. A possible limitation of the presented evaluation may be that only one type of HAS clients was tested. In real deployments the clients would likely be heterogeneous, using different adaptation logics. This would have an influence on the achievable results.

References:

[1] "Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats." ISO/IEC 23009-1, 2012.

[2] S. Akhshabi, L. Anantakrishnan, C. Dovrolis, and A. C. Begen. "What happens when HTTP Adaptive Streaming Players compete for Bandwidth" In *Proceedings of ACM NOSSDAV*, pp. 9-14, 2012.

Roger Zimmermann is an associate professor with the Department of Computer Science at the School of Computing with the National University of Singapore (NUS) where he is also an investigator with the Interactive & Digital Media Institute (IDMI). His research interests are in both spatio-temporal and multimedia information management, for example distributed and peer-to-peer systems, spatio-temporal multimedia, streaming media architectures, georeferenced video management, mobile location-based services and geographic information systems (GIS). He has co-authored a book, six patents and more than hundred-ninety conference publications, journal articles and book chapters in the areas of multimedia and databases. He has received the best paper award at IEEE ISM 2012 and was part of the team who won second place at the ACM SIGSPATIAL GIS Cup 2013. He has been involved in the organization of conferences in various positions, for example program co-chair of ACM Multimedia 2013. He co-directs the Centre of Social Media Innovations for Communities (COSMIC) at NUS and is an investigator with the NUS Research Institute (NUSRI) in Suzhou, China. Roger Zimmermann is an Associate Editor of the ACM Transactions on Multimedia journal (TOMM, formerly TOMCCAP) and the Multimedia Tools and Applications (MTAP) journal. He is a Senior Member of the IEEE and a member of ACM. For more details, see http://eiger.ddns.comp.nus.edu.sg.

IEEE COMSOC MMTC R-Letter

Mobility-Aware Resource Allocation Scheme Under Channel Uncertainty

A short review for "Robust Resource Allocation for Predictive Video Streaming Under Channel Uncertainty"
(Edited by Koichi Adachi)

R. Atawia, H. Abou-zeid, H. S. Hassanein, and A. Noureldin, "Robust Resource Allocation for Predictive Video Streaming Under Channel Uncertainty", Proceedings of the IEEE GLOBECOM 2014, pp. 4881-4886, Dec. 2014.

During several decades, the mobile traffic has been rapidly growing up due to a rich content service. Network operators are facing formidable resource management challenges to cope with such phenomenal growth. Specifically, video accounted for over 50% of the traffic in 2012, with projections of a 14-fold increase by 2018 [1]. Due to the mobility of user and continuously changing environment, the throughput performance of wireless communication significantly varies over time and location. To tackle this issue, predictive resource allocation techniques, which exploit user mobility information and channel prediction, have been proposed to improve the throughput and fairness among the users [2,3], as well as video streaming delivery [4-7]. Different from the instantaneous resource allocation (RA) and admission control strategies [8,9], base stations (BSs) schedule more resources to users during their respective peaks, and prioritize users that are headed to poor channel conditions by performing long-term RA plans over several seconds. If it is known a user is approaching a low coverage area, content can be pre-buffered to support smooth streaming. Furthermore, as this enables efficient content pre-buffering, energy is saved since transmission will not be needed during poor conditions [4,5,7].

It is commonly assumed for predictive RA (PRA) that a user's future channel states are highly predictable by utilizing radio environment map (REM) or bandwidth (BW) map. Some earlier works have shown that correlation between location and received throughput [3,10,11]. Although such approaches provide a reasonable estimate of the future throughput of a user, it cannot accurately capture the dynamics of wireless network and congestion, and environmental/geographical changes. To overcome these problems, the authors present a fuzzy-based robust RA framework for predictive video streaming (PVS), which takes into account the channel uncertainty. The main contributions of this paper are

1. Modeling of uncertainty in the REM measurements by using triangular fuzzy numbers. It has been shown a good approximation of the REM variations under certain conditions,
2. Development of a robust RA framework for PVS using the fuzzy REM, which can balance the constraint satisfaction under rate uncertainty.

It has been illustrated in [5] how BS transmission time can be minimized by leveraging future user rate knowledge. A predictive scheme will wait to make bulk transmissions at times of high channel conditions, while making the minimal transmissions that avoids video stalling at other times. The corresponding optimization problem can be formulated as linear program (LP). The solution of this LP minimizes airtime without degrading the video only if the predicted throughput are accurate. Therefore uncertainty of predicted throughput results in either video stall or longer airtime required. The authors propose a fuzzy-based robust RA framework to adaptively capture such variations.

In the original LP, one constraint is introduced to ensure that the cumulative video content requirement is not violated at each time slot. In order to take into account the channel uncertainty, that constraint is modified by introducing the new variable which is obtained by fuzzifier. This new variable is represented by a triangular membership function. This membership function is represented by a parameter, which can be selected to reflect the uncertainty in the predicted throughput based on the degree of rate violations. This parameter greatly affects the performance. Considering the importance of learning the degree of uncertainty in order to meet the desired constraint satisfaction levels without unnecessary resource consumption, the authors introduced feedback in the a fuzzy-based robust RA framework to lean and adapt to the degree of channel uncertainty in order to re-optimize the RA. Then the proposed framework learns the uncertainties via a Kalman filter and tunes the model to adapt to the current channel conditions.

The authors consider average BS airtime and video degradation (VD) as performance metrics. The computer simulation evaluation has been performed using the 3GPP long-term evolution (LTE) system configuration. Firstly the authors show the impact of parameter setting on the average BS airtime and VD in the

proposed framework. After that how the adaptive tuning of the parameter for membership function can balance the two performance metrics, i.e., average BS airtime and VD.

In the next coming years, multimedia traffic is expected to dominate the traffic volume of wireless communications systems. To ease the congestion or resource utilization of system, more efficient resource utilization strategies are of great importance. Without any doubt, predictive resource allocation (PRA) is one of them. So the approach proposed in this paper, which tries to solve the negative impact of the channel uncertainty on the performance of PRA, may be quite useful in a realistic communications environment.

References:

[1] CISCO, "Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018," 2014. Accessed Apr. 29th, 2014.

[2] H. Abou-zeid, H. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in Proc. IEEE GLOBECOM, pp. 4714–4719, 2013.

[3] R. Margolies, A. Sridharan, V. Aggarwal, R. Jana, N. K. Shankaranarayanan, V. A. Vaishampayan, and G. Zussman, "Exploiting mobility in proportional fair cellular scheduling: Measurements and algorithms," in Proc. IEEE INFOCOM, 2014, to appear.

[4] Z. Lu and G. de Veciana, "Optimizing stored video delivery for mobile networks: The value of knowing the future," in Proc. IEEE INFOCOM, pp. 2806–2814, 2013.

[5] H. Abou-zeid and H. S. Hassanein, "Predictive green wireless access: Exploiting mobility and application information," IEEE Wireless Commun., vol. 20, no. 5, pp. 92–99, 2013.

[6] H. Abou-zeid and H. S. Hassanein, "Efficient lookahead resource allocation for stored video delivery in multi-cell networks," in Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC), 2014, to appear.

[7] H. Abou-zeid, H. S. Hassanein, and S. Valentin, "Energy-efficient adaptive video transmission: Exploiting rate predictions in wireless networks," IEEE Trans. Veh. Technol., vol. 63, no. 5, pp. 2013 – 2026, 2014.

[8] M. Katoozian, K. Navaie, and H. Yanikomeroglu, "Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization," IEEE Trans. on Wireless Commun., vol. 8, no. 1, pp. 66–71, 2009.

[9] J. B. Othman, L. Mokdad, and S. Ghazal, "Performance analysis of wimax networks ac," Wireless Personal Communications, vol. 74, no. 1, pp. 133–146, 2014.

[10] J. Yao, S. S. Kanhere, and M. Hassan, "An empirical study of bandwidth predictability in mobile computing," in Proc. ACM WiNTECH, pp. 11–18, 2008.

[11] D. Han, J. Han, Y. Im, M. Kwak, T. T. Kwon, and Y. Choi, "MASERATI: Mobile adaptive streaming based on environmental and contextual information," in Proc. ACM WiNTECH, pp. 33–40, 2013.

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. Since 2010, he has been with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include cooperative communications and energy efficient communication technologies. 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. Dr. Adachi served as General Co-chair of the 10th and 11th IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS) and Track Co-chair of Transmission Technologies and Communication Theory of the 78th and 80th IEEE Vehicular Technology Conference in 2013 and 2014, respectively. He was recognized as the Exemplary Reviewer from IEEE COMMUNICATIONS LETTERS in 2012 and IEEE WIRELESS COMMUNICATIONS LETTERS in 2012, 2013, and 2014. He was awarded excellent editor award from IEEE ComSoc MMTC in 2013.

An Indexed Color Representation for Screen Content Coding using HEVC

A short review for "Screen Content Coding Based on HEVC Framework"
(Edited by Bruno Macchiavello)

Weijia Zhu; Wenpeng Ding; Jizheng Xu; Yunhui Shi; Baocai Yin, "Screen Content Coding Based on HEVC Framework", IEEE Transactions on Multimedia, vol.16, no.5, pp.1316,1326, Aug. 2014.

Due to proliferation of video applications, screen content coding has received much interest in recent years [1]-[5]. Screen content refers to video data containing computer graphics like: cartoons, captures of a typical computer screens, video with text overlay, news ticker, etc.

Screen content has several new features not previously available in camera-captured content, e.g., sharp content, large motion, unnatural motion and repeating patterns. The new video coding standard, high efficiency video coding (HEVC), includes screen content as one of its requirements [6]. However, several coding techniques available in HEVC are not adequate for screen content encoding. HEVC increases the number of angular Intra prediction modes, in comparison with H.264/AVC, from 9 to 33. This angular prediction modes are perform at block level, but the directional correlation within a screen content may vary pixel by pixel, not necessarily block by block. Motion estimation and compensation is used in HEVC, and several other video encoders, as the main technique for reduction of temporal redundancy. Nevertheless, this method is not adequate for the unnatural and non-translational motion present in a desktop screen (like text writing). Finally, HEVC adopts variable size discrete cosine transforms (DCT) in order to remove the spatial redundancy of the residue signal. However, blocks with screen content contain complex structures and sharp edges, which cannot be compactly represented in the DCT domain.

In this paper the authors propose a non-transform coding scheme that can be incorporated into HEVC. This paper is an extension of a previous work [7], which is based on the base color representation. The authors present two new encoding modes a Multi-Stage Directional Mode (MDM) and a Multi-Stage Temporal Mode (MTM). In MDM, the current block of the video frame is first decomposed into base colors and an index map. Then the base colors and index map will be encoded by the context-based adaptive binary arithmetic encoder (CABAC) [8] present in HEVC. Since the indexes in the index map are highly correlated with their spatial neighbors, the authors employed directional prediction scheme prior to entropy coding. The left, above and left-above neighbors of the current pixels are analyzed in two different stages in order to obtain two different predictions. If the absolute difference between the left and left-above neighbors is lower than the difference between the left-above and above neighbors, then the first predictor is consider to be the index of the above neighbor and the second predictor is the index of the left neighbor. Otherwise, the predictors are inverted. There are some special cases that are also treated. The index map is encoded in raster scan order. The cost of the index matched by the first prediction is one bit. Two bits are needed for the index matched by the second prediction, and the unmatched index map is coded by CABAC.

In MTM a motion compensated block is obtained. Then all pixels values in the current block are compared with the motion compensated block. If the absolute difference of certain pixel within the block is higher than a threshold then that pixel is marked as a temporal changed pixel. The temporal changed pixels are represented using base color representation and encoded in a similar manner to MDM.

Obviously, the selection and appropriate number of base colors is a very important issue, in order to obtain the best coding performance. Therefore, the encoder needs to find the optimal base colors and corresponding index map. In their implementation, the authors vary the number of base colors from 1 to 16 and used dynamic programming to solve the color quantization problem. The runtime complexity of color quantization accounts for 3.7% of the overall encoding time during simulations.

For encoding of the unmatched indexes, using CABAC, the authors proposed a context adaptive reordering in order to improve the entropy coding efficiency. Once again, the left and above neighbors, along with the number of base colors, are used to determine the context reordering.

MDM is compare with all-intra encoding in HEVC and with the previous proposal of the same authors (MBCIM). The HEVC implementation used was HM9.0_Frext, the quantization parameters were 22, 27, 32 and 37. Ten (10) different screen content sequences were used. On average the proposed scheme achieves 35.1% and 2.9% bitrate saving compare to HM and MBCIM, respectively.

Results are also presented in relation to low-delay and random-access configurations of HEVC. MDM achieves 22.2% bitrate savings on average. When both MDM and MTM are enabled the proposed scheme achieves 23.6% of bits savings. It is important to notice that when MDM is enable the encoder complexity is about 15% higher than that of HM under low-delay configuration. And with both MDM and MTM enabled, the encoder complexity is 21% higher compared to the low-delay configuration.

In conclusion the authors proposed a HEVC extension for screen content coding. Experimental results verify the effectiveness, in terms of Rate-Distortion, of the proposed scheme. Example images, for subjective analysis are also presented in the paper. However, there are several issues to be improved in the future. The proposed algorithms significantly increase the computational complexity of the encoder and, thus, fast algorithms are needed. A loop filter specifically designed for screen content can also be proposed.

References:

[1] C. Lan, X. Peng, J. Xu and F. Wu, "Intra and inter coding tools for screen contents," Document of Joint Collaborative Team on Video Coding, JCTVC-E145, Mar. 2011.

[2] S. Wang and T. Lin,"A Unified LZ and hybrid coding for compound image partial-lossless compression," in proc. IEEE Int. Congress on Image and Signal Processing, pp. 1-5, Oct. 2009.

[3] W. Ding, Y. Lu and F. Wu, "Enable efficient compound image compression in H.264/AVC intra coding," IEEE Int. Conf. on Image Processing, pp.337-340, Sep. 2007.

[4] Z. Pan, H. Shen, Y. Lu, and S. Li, "Browser-friendly hybrid codec for compound image compression," IEEE International Symposium on Circuits and Systems, pp. 101-104, May 2011.

[5] A. Zaghetto and R. L. de Queiroz, "Segmentation-driven compound document coding based on H.264/AVC-INTRA," IEEE Trans. on Image Process., vol. 16, pp. 1755-1760, Jul. 2007.

[6] G. Sullivan, J. Ohm,W. J. Han, and T.Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[7] W. Zhu, J. Xu, and W. Ding, "Screen content coding with multi-stage base color and index map representation," JCTVC-M0330. Incheon, Korea, Apr. 2013.

[8] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 620–636, 2003.

Bruno Macchiavello is an assistant 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 also was co-organizer of a special session on Streaming of 3D content in the 19th International Packet Video Workshop (PV2012). His main research interests include video and image coding, image segmentation, distributed video and source coding, multi-view and 3D video processing.

Marriage between Conventional Image Representation and Deep Neural Networks

A short review for "DEFEATnet – A Deep Conventional Image Representation for Image Classification"
(Edited by Jun Zhou)

Shenghua Gao, Lixin Duan, and Ivor Tsang. "DEFEATnet – A Deep Conventional Image Representation for Image Classification", IEEE Transactions on Circuits and Systems for Video Technology, 2015.

While deep neural networks having demonstrated superior performance in image representation and classification [1], it does require large amount of training data and powerful processors. This is contradicting with many real-world computer vision and multimedia applications, for example, medical or environmental image analysis [2], in which training data are limited due to high cost of data collection and labelling. In these cases, hand-craft image features still show their necessity. Nevertheless, it is interesting to see such conventional image representation methods be integrated with deep learning framework. The question is, how it can be done?

In this recently accepted paper by the IEEE Transactions on Circuits and Systems for Video Technology, authors choose to borrow the ideal of cascaded structure in deep neural networks, and embed in each layer conventional image representation steps. The developed approach, namely deep feature extraction, encoding, and pooling network (DEFEATnet), does not generate huge numbers of parameters like what deep neural network does, but still can preserve prior knowledge for image representation in each layer.

The framework of DEFEATnet has four parallel channels which take the same image at different scales as input. These channels contain sequential layers each of which consists of three steps, feature extraction, encoding, and pooling. The output of a layer becomes the input to the next layer, and at the same time, also contributes directly to the final image representation. Therefore, the final image vector is formed by concatenation of outputs from all layers in all channels. This framework is quite general in nature, as a number of image features, soft/hard encoding approaches, and sum/max pooling strategies can be accommodated in the processing pipeline.

An implementation example of the DEFEATnet, is given in the paper. For each channel, dense SIFT feature [3] is first extracted. These correspond to image features extracted at different scales. Then sparse coding is used to convert SIFT features into a new feature map. This allows the discovery of the structure of the local features and reducing the influence of feature noises. A rectification step is employed to convert all values into non-negative. Finally, local max pooling over 2*2 region is performed to generate the layer output [4]. This last step is similar to the pooling step on small region by neurons in the convolutional neural networks.

Note that from the second layer, the input to the encoding step is dependent on the previous steps. The cascaded pooling step allows features evolve from small region to large region, therefore, enables representation of different object parts in different granularities. Furthermore, these layers also denoise the data. After local max pooling, each output is normalized for the final output.

DEFEATnet was evaluated on three benchmark datasets that contains 1000 to 9144 images. Results show that this method has outperformed several conventional image representation and deep learning approaches. The authors show that when dataset is small, the best performance of the method does not necessarily appear with deep structure.

Like deep neural networks, DEFEATnet can also be visualized. Filters learned at the first a couple of layers contain more small scale and simple structures. The latter layers show more complex structures. Images can be reconstructed at each layer. From first to latter layers, details of objects gradually lost, and the reconstructed images show more and more property that is common to the whole class.

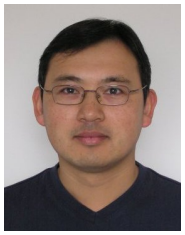
In summary, conventional image representation based on handcrafted feature extraction is still not out-dated yet, but explore deep learning structure is a path to improve this area of research. It would be interesting to see more frameworks like DEFEATnet being developed in the near future.

References:

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge". arXiv:1409.0575, 2014.

[2] Y. Chen, Z. Lin, Z. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004
[4] Y. Jia, C. Huang, and T. Darrell, "Beyond spatial pyramids: Receptive field learning for pooled image features," in *Computer Vison and Pattern Recognition*, vol. 2, 2012, pp. 2169-2178..

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.

Currently, he is 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. His research interests are in statistical pattern recognition, interactive computer vision, and their applications to hyperspectral imaging and environmental informatics.

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 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 R-Letter Editorial Board Directors:

Christian Timmerer (christian.timmerer@aau.at), Weiyi Zhang (wzhang@ieee.org), and Yan Zhang (yanzhang@simula.no).

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 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 Officers

Chair: Yonggang Wen, Singapore
Steering Committee Chair: Luigi Atzori, Italy
Vice Chair – North America: Khaled El-Maleh, USA
Vice Chair – Asia: Liang Zhou, China
Vice Chair – Europe: Maria G. Martini, UK
Vice Chair – Letters: Shiwen Mao, USA
Secretary: Fen Hou, China
Standard Liaison: Zhu Li, USA

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.