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

R-LETTER

Vol. 6, No. 4, August 2015

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the August 2015 issue of the Review Letter (R-Letter) of the IEEE Communications Society Multimedia Communications Technical Committee (MMTC). This issue is a special issue as – for the first time – it comprises **R-Letters for the best papers awarded at IEEE ICME 2015** in Torino (June 29 – July 3, http://www.icme2015.ieee-icme.org/) which are preceded by an guest editorial provided by ICME'15 general co-chair *Enrico Magli* from Politecnico di Torino, Italy. Additionally, we received **two regular R-Letter** 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**.

An overview of all R-Letters is provided in the following:

The **first paper**, published in the *Proceedings of IEEE ICME'15 as best paper* and *edited by Frank Hartung*, evaluates music recommendations in a real-world setting focusing on data splitting and evaluation metrics.

The **second paper**, published in the *Proceedings of IEEE ICME'15 as best paper runner-up* and *edited by Bruno Macchiavello*, describes means for a characteristic Number Regression for Facial Feature Extraction.

The **third paper** is *edited by Lifeng Sun* and has been published within the *Proceedings of IEEE ICME'15 as best student paper*. It provides an annotated dataset and model for salient object detection in complex scenes.

The **forth paper**, published in the *IEEE Transactions on Multimedia* and *edited by Carl James Debone*, provides a parallel in-loop filtering hardware implementation for HEVC decoders.

Finally, the **fifth paper**, published in the *IEEE Journal of Selected Topics in Signal Processing* and *edited by Gwendal Simon*, is about understanding an exciting new feature in HEVC, i.e., tiles!

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

IEEE ICME 2015 Best Paper Awards

Guest Editorial Introduction by Enrico Magli (IEEE ICME'15 General Co-Chair)

As a general co-chair of IEEE ICME 2015, which was held in Torino (Italy) from June 29 to July 3, 2015, it is my pleasure to report about the best work that has been presented during the conference. The technical program chairs and the awards chairs have selected three papers representing the best contributions that have been presented during the conference. These papers have respectively been given a **best paper**, **best paper runner-up**, and **best student paper** award. It is my pleasure to introduce them here on behalf of the organizing committee of ICME 2015.

The best paper award has been assigned to paper "**Evaluating Music Recommendation in a Real-World Setting: On Data Splitting and Evaluation Metrics**", by *Szu-Yu Chou, Yi-Hsuan Yang, Yu-Ching Lin*. This paper addresses a very important problem, namely how to assess the performance of music recommendation systems. In order to overcome the limitations of existing systems, the authors propose a novel evaluation framework using various time-based data splitting methods. Tests are performed over millions of commercial listening records, comparing the performance of collaborative filtering and content-based models with low-level audio features and semantic audio descriptors. The result of this analysis is that the content-based models turn out to be extremely good at trading off accuracy, novelty, diversity, freshness and popularity.

The best paper runner-up award has been assigned to paper "**Robust Interactive Image Segmentation with Weak Supervision for Mobile Touch Screen Devices**", by *Tinghuai Wang, Huiling Wang, Lixin Fan*. We all know how essential image segmentation is to many image processing tasks. This paper addresses the segmentation problem from a specific perspective, namely its application on touchscreen devices. Employing techniques such as geodesic methods and level sets, the authors develop a new algorithm that maximizes accuracy of boundary placement and region connectivity, while minimizing the amount of user interaction needed. The outcome is a practical and innovative solution for visual object cutout on mobile touchscreen devices, which facilitates several new applications such as creation of oil painting effects.

Finally, the best student paper award has been assigned to paper "**Characteristic Number Regression for Facial Feature Extraction**", by *Yuntao Li, Xin Fan, Risheng Liu, Yuyao Feng, Zhongxuan Luo, Zezhou Li*. Facial feature extraction has long been studied, including regression methods; these latter lack explicit face shape constraints, and require a large training set. This paper introduces a novel invariant, named "characteristic number", that explicitly characterizes the intrinsic geometry of human facial points. Based on this, a shape-to-gradient regression framework is also developed, leading to an efficient and robust feature extractor for facial images in the wild, which does not need large numbers of training images.

These three papers are only a sample of the work that has been presented during ICME. I invite all of you to check out the conference proceedings and contribute yourselves to the next ICMEs!

Enrico Magli is currently an Associate Professor at Politecnico di Torino, Italy. His research interests are in the field of compressive sensing, image and video coding, computer vision and security. He has published over 50 journal papers and 140 conference papers in these areas, several book chapters, and two patents. He is an associate editor of IEEE T-CSVT, IEEE T-MM, and EURASIP Journal on Image and Video Processing. He is a co-recipient of the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, and has received the 2010 Best Reviewer Award of IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing, and the 2010 and 2014 Best Associate Editor Award of IEEE Transactions on Circuits and Systems for Video Technology. He is a distinguished lecturer for CAS during 2015-2016. He has been general co-chair of IEEE MMSP 2013 and IEEE ICME 2015, and TPC co-chair of IEEE MMSP 2011, IEEE ICME 2012, VCIP 2012 and VCIP 2014. He is a senior member of the IEEE.

ICME'15 Best Paper: Improved Recommendations on Music Streaming Platform

A short review for "Evaluating Music Recommendation in a Real-World Setting: On Data Splitting and Evaluation Metrics" (Edited by Frank Hartung)

Szu-Yu Chou, Yi-Hsuan Yang, and Yu-Ching Lin, "Evaluating Music Recommendation in a Real-World Setting: On Data Splitting and Evaluation Metrics", Proceedings of the IEEE International Conference on Multimedia and Expo, June/July, 2015.

Music streaming services are getting more and more popular, and are increasingly replacing the traditional purchase of music media like CDs. Popular services for music streaming include Spotify, Grooveshark, KKBOX, and others, and recently Apple has entered the arena with Apple Music.

Besides the audio experience itself, these streaming services also have a strong social media component. Users can share their playlists, let other users follow their user behavior, and vice versa subscribe to others' playlists. The social component also includes recommendations. Users receive recommendations or proposals for music they might like. Such recommendations can be passive, by just showing popularity or rankings for music in the offered catalogue, or they can be more active, by actively suggesting music to a user that the system thinks he might like. This requires automatic music recommendation algorithms. Such algorithms have been the subject of research for more than 20 years [1][2], and today's services all use them. Still, improvements are possible.

The topic of the paper that received the IEEE International Conference on Multimedia and Expo 2015 (ICME 2015) best paper award are recommendation algorithms for music, and their performance evaluation.

As a basis, the authors have access to a database from a real music streaming service, KKBOX. The database contains 100 million music listening records from over 28,000 users, over one year, from October 2012 to September 2013. The database contains playback timestamps and other metadata for over 124,000 songs. This data can be used to train, and to evaluate different recommendation algorithms. In that context, the authors investigate different approaches for splitting the available database into training data, for training the algorithms, and test data, for evaluating the algorithms. The splitting logic and its impact on the recommender system is one focus of the paper. Traditionally, the available database is split randomly into a training set and a test set. This does not consider the time component. In reality, the task is to produce music recommendations for the future, based on the history. The authors thus propose a different splitting strategy where the "older" parts of the data set are used for training, and the "newer" parts are used for evaluation and test. This is closer to reality, and the authors call this the "real-life split strategy" (RLS).

The authors also investigate and propose different evaluation metrics that verify the effectiveness of recommendation algorithms and systems. They propose evaluation metrics for accuracy (how are predicted rating of music for a user and real rating by the user correlated), novelty (the top recommended songs are new to the user or not), diversity (diversity of the genres of recommended songs), freshness (whether the top songs recommended to the user are composed of new or old songs), and popularity (based on the number of users that have listened to that song).

Further, a drawback of many recommendation algorithms is that they use only information about music already rated or played by many users [5]. Music that is popular in a certain user group, is likely to be attractive to other users that belong to that group. Such algorithms, that are based on past user behavior, discriminate new songs. It would however be good to also identify new songs that could be attractive to users, based on their preferences in the past. The authors call that the cold-start problem. Since for new music no usage data is available, this requires analysis of the new music, in terms of semantic descriptors. The authors propose different criteria and descriptors, and evaluate them. This is another focus of the publication. The descriptors include low-level audio features like Mel-frequency cepstral coefficients (MFCC) and audio word (AW) based features, and high-level semantic audio descriptors that capture genre, acoustics and emotion and are trained with suitable data bases [4].

The authors have evaluated different combinations of data splitting logic, recommendation algorithm [3], and descriptors/features. The evaluation was done with respect to the evaluation metrics mentioned above.

The authors conclude that content-based (CB) factorization machine (FM) models [3] perform well if combined with realistic data splitting concepts that address the cold-start problem of new music added to the catalogue. Further, the use of high-level semantic features (like the genre and emotion features) in content-based models achieves a good tradeoff between the accuracy, diversity, novelty and other measures of recommendation.

The value of the paper lies in its complete assessment of music rating evaluation. The authors examine different aspects (database splitting concepts, evaluation metrics, recommendation algorithms, audio features) and their combinations, and thus provide an evaluation framework that can also be applied to other recommendation aspects and problems.

References:

[1] Upendra Shardanand, *Social information filtering for music recommendation*. Diss. Massachusetts Institute of Technology, 1994.

[2] Òscar Celma, *Music Recommendation and Discovery,* Springer, 2010

[3] R. Steffen, "Factorization machines with libfm," ACM Transactions on Intelligent Systems and Technology, vol. 3, no. 3, pp. 57:1–57:22, May 2012.

[4] Y.-H. Yang and J.-Y. Liu, "Quantitative study of music listening behavior in a social and affective context,", IEEE Transactions on Multimedia, vol. 15, no. 6, pp. 1304–1315, Oct 2013

[5] X. Wang and Y. Wang, "Improving content-based and hybrid music recommendation using deep learning," in Proceedings ACM Multimedia, 2014, pp. 627–636.

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.

ICME'15 Best Paper Runner-Up: A Novel Scribble-based Image Segmentation Method for Mobile Devices

A short review for "Robust Intereactive Image Segmentation with Weak Supervision for Mobile Touch Screen Devices"
(Edited by Bruno Macchiavello)

Tinghuai Wang; Ang Huiling; Fan Lixin, "Robust Interactive Image Segmentation with Wak Supervision for Mobil Touch Screen Devices," Proceedings of the IEEE International Conference on Multimedia and Expo, June/July, 2015.

Since the early days of computer vision and image processing, image segmentation has received a lot of attention. Image segmentation refers to the act of dividing and image into non-overlapping regions, where all pixels within a certain region must be similar with respect to some characteristic or property. Due to the difficulty of obtaining a fully automatic segmentation method that works for a wide range of different images and applications, several interactive image segmentation algorithms have emerged in recent years [1-3]. One mayor motive for the increasing interest on interactive image segmentation methods is the popular use of mobile touch screen devices. Such devices can facilitate spatially localized media manipulation, since prior knowledge about the desired object and/or background can be easily defined with simple user interactions. One of the most common interaction modes is loosely drawing scribbles on certain objects. Despite of the significant advances in recent years, two issues prevent the scribble-based methods from being widely used on mobile devices. First, most methods require to draw scribbles on both foreground and background objects, which can be troublesome for compact touch screens. Moreover, the need to switch between foreground and background scribbles can complicate the user interface. Second, some methods require the user to correct mis-segmentations, like disjoint regions, which can be cumbersome for most users.

In this paper the authors tackle these two compelling issues by proposing an interactive segmentation method that: *(i)* only requires scribbles on foreground objects and *(ii)* automatically promotes accurate boundary placements and fills contiguous regions. The proposed method combines geodesic distance information with the level set method.

The level set method for capturing dynamic interfaces and shapes was introduced in [4]. The main idea of the level set method is to represent a contour as the zero level set of a higher dimensional function. Evolving the contour is achieved by evolving the embedding function, which is defined as a signed distance measure. The distance of all pixel inside the desired region, with respect to the initial seeds, should be positive, while the distance outside the contour should be negative. The evolution of the Level Set Function (LSF) is governed by minimizing its energy. Similar to a previously proposed method [5], the authors incorporate geodesic distance information into the energy function of the LSF. This geodesic active component is minimized when the zero level contour is located at the object boundaries. However, in order to reduce the sensitivity to seeds placement, the authors propose a novel geodesic region term which measures the statistics of the geodesic distance of all pixels instead of the individual geodesic distance. The proposed geodesic region term provides region-based information which encodes both color distribution and spatial information.

For seeds generation, the image is first over-segmented into small coherent regions, i.e. superpixels, using a fast graph-based segmentation method [6]. Each point of the hand-drawn scribbles is associated with the segmented regions that intersects it. This is set as the initial seeds for foreground objects. The set of initial background seeds consists of segmented regions along the image border. Note, that it is possible that some background seeds were initially mislabeled, however, a label pruning step is adopted to eliminate mislabeled seeds. The pruning step is performed by computing a modified geodesic distance from each initial background seed to its closets foreground seed. Intuitively, true background seeds are identified as those that their nearest foreground seed is as far as possible. Therefore, all initial background seeds are ranked and the bottom 40% are eliminated.

Finally, a seed propagation process is also proposed. The authors propose a cascade classification process

which can be structured into two stages. In the first stage, linear support vector machines are used to generate a candidate pool of new background seeds based on appearance similarity. In the second stage geodesic distance is use to enforce a spatial constraint and select the appropriate new seeds from the candidate pool.

The authors present both quantitative and qualitative evaluations. For the quantitative evaluation the segmentation overlap score, proposed in [7], is used as an objective metric. Six state-of-the art scribble-based segmentation methods were selected for comparison [8-13]. Two schemes of the proposed geodesic level set framework were compared, i.e., the geodesic level set approach without using the seed propagation process (GDLS), and the one augmented with seed propagation (GDLS-SG). The paper shows that the proposed GDLS achieves the highest average segmentation overlap score (91.10%) and also yields the tightest standard deviation (0.0674) across the whole test dataset. For qualitative evaluation the authors present 7 images segmented with all methods. In those images, the GDLS-SG appears to subjectively outperform other methods.

As future work a complete subjective test with several subjects can be performed. Moreover, different applications for the proposed interactive segmentation method can be proposed, specifically for mobile devices.

References:

[1] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive gmmrf model," in ECCV, 2004, pp. 428–441.

[2] J.Wang, M. Agrawala, and M. F. Cohen, "Soft scissors: an interactive tool for realtime high quality matting," in SIGGRAPH. 2007, pp. 585–594, ACM.

[3] D. Liu, K. Pulli, L. G. Shapiro, and Y. Xiong. Fast interactive image segmentation by discriminative clustering. In Proceedings of the 2010 ACM multimedia workshop, MCMC '10, New York, NY, USA.

[4] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," J. Comput. Phys., vol. 79, no. 1, pp. 12–49, Nov. 1988.

[5] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," IEEE Trans. Image Process., vol. 19, no. 12, pp. 3243–3254, 2010.

[6] P. Felzenszwalb and D. Huttenlocher, "Efficient graph based image segmentation," IJCV, vol. 59, no. 2, pp. 167–181, 2004.

[7] M. Everingham, L. J. Van Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," IJCV, vol. 88, no. 2, pp. 303–338, 2010.

[8] N. Paragios and R. Deriche, "Geodesic active regions for supervised texture segmentation," in ICCV, 1999, pp. 926–932.

[9] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," in ICCV, 2007, pp. 1–8.

[10] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in ICCV, 2001, pp. 105–112.

[11] Olga Veksler, "Star shape prior for graph-cut image segmentation," in ECCV, 2008, pp. 454–467.

[12] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in CVPR, 2010, pp. 3129–3136.

[13] B. L. Price, B. S. Morse, and S. Cohen, "Geodesic graph cut for interactive image segmentation," in CVPR, 2010, pp. 3161–3168.

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.

ICME'15 Best Student Paper: Projective Invariant Characteristic Number for Facial Analysis

A short review for "Characteristic Number Regression for Facial Feature Extraction"
(Edited by Lifeng Sun)

Yuntao Li, Xin Fan, Risheng Liu, Yuyao Feng, Zhongxuan Luo and Zezhou Li, "Characteristic Number Regression for Facial Feature Extraction Proceedings of the IEEE International Conference on Multimedia and Expo, June/July, 2015

Facial feature extraction plays an important role in many multimedia applications, including face tracking, pose estimation and expression analysis. Significant progress has been made on the extraction under controlled environments [1]. However, facial feature extraction is still a challenging problem, due to complex environment in the wild, like illumination, pose or viewpoint and partial occlusions. In order to overcome those challenges, geometrical constraints are always vital for facial analysis as human faces are highly structured.

For facial feature extraction, the successful shape prior modeling techniques are active shape models (ASM) [2] and active appearance models (AAM) [3]. They construct parametric shape models by performing principal component analysis (PCA) on a set of labeled face shapes. Numerous variants from ASM and AAM have been emerging, including those in [4, 5, 6] that improves the shape models in various aspects. In recent years, researchers develop implicit shape regression, generating mappings from texture to shape. Two-level Fern regression for coarse-to-fine face alignment was introduced in [7]. The work in [8] further improves the performance by combining joint regressors with local features instead of global ones in order to avoid over fitting. However, these regressions highly depend on availability of training examples. Hence, the training data availability highly curbs their abilities to handle a wider range of pose changes.

Taking into account this characteristic of sensitive to training samples for implicit shape regression, the authors firstly devise explicit shape priors from a newly developed projective invariant, named characteristic number (CN), that is a projective invariant and it can also reflect intrinsic geometries. These priors reflect the common geometry of fiducial points share by human face across age, gender and race, and they do not rely on training examples as those in regression methods. For example, all the CN values of three collinear points are -1. The CN values of a subset of five points are also consistent. They also find the CN values of several six-point subsets showing common values for different faces. In order to verify this invariant property, the authors calculate all CN values the Labeled Faces in the Wild (LFW) database [9]. They find some point combinations having consistent CN values on all these images with great variations in LFW. This invariant property naturally brings robustness to facial viewpoint changes when they apply these shape priors from CN to the localization.

The authors combine the geometric constraint from CN with simple texture constraint to extract facial feature points. The objective function is a nonlinear least square, which can be solved by gradient descent or second order Newton method. But unfortunately, the computation of gradients or hessian matrix is slow. Instead, the authors assume that the location offsets can be linearly determined by CN differences. To reduce the complexity of the computation, they build the regression from point configurations to target CN values (geometry to geometry). This method learns a linear regression for the relationship between CN and location in order to enhance the speed of location.

Shape to shape regression can explicit utilize shape priors from CN to extract facial feature points. These shape constraint is also a complementary to implicit regression. Specially, the authors impose their shape-to-shape regression based on implicit regressions as additional constraints. The eye corners are colinearity by 3-point CN regression and the accuracy of other corners like mouth or nose are also improved. To the best of the authors' knowledge, this is the first paper that put forward shape-to-shape regression with combining CN, which is a absolute

complementary to implicit regressions.

To summarize, the main contributions of their paper are:

- Proposing a novel projective invariant, named characteristic number (CN), to extract general facial geometries.
- Developing an efficient shape-to-shape regression framework to calculate CN on real world facial images.
- Various experiments on different benchmark databases demonstrate that our CN regression framework generally performs much better than existing facial feature extractors for real-world uncontrolled facial images.

Acknowledgement:

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

References:

[1] H. Dibeklioglu, A. Salah, and T. Gevers, "A statistical method for 2-d facial landmarking," IEEE TIP, vol. 21, no. 2, pp. 844–858, 2012.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-Their training and application," CVIU, vol. 61, no. 1, pp. 38–59, 1995.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE TPAMI, vol. 23, no. 6, pp. 681–685, 2001.

[4] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in ECCV, 2008, pp. 504–513. 2008.

[5] P.N. Belhumeur, D.W. Jacobs, D. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in CVPR, 2011, pp. 545–552.

[6] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer, "Robust and accurate shape model fitting using random forest regression voting," in ECCV, pp. 278–291. 2012.

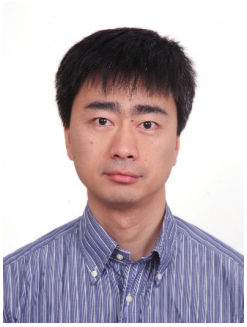
[7] P. Doll′ar, P. Welinder, and P. Perona, "Cascaded pose regression," in CVPR, 2010, pp. 1078–1085.

[8] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in CVPR, 2014, pp. 1685–1692.

[9] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, 2007.

[10] X. Xiong and F. De La Torre, "Supervised descent method and its applications to face alignment," in CVPR, 2013, pp. 532–539.

[11] Z. Luo, X. Zhou, and D.X. Gu, "From a projective invariant to some new properties of algebraic hypersurfaces,"Science China Mathmatics, vol. 57, no. 11, pp. 2273–2284, 2014.



Lifeng Sun received the B.S. and Ph.D. degrees in System Engineering from National University, of Defense Technology China, in 1995 and 2000, respectively. He joined the Department of Computer Science and Technology (CST), Tsinghua University (THU), Beijing, China, in 2001. Currently, he is a Professor in CST of THU. His research interests include networked multimedia, video streaming, 3D/multi-view video coding, multimedia content analysis, multimedia cloud computing and social media. He has authored or coauthored about 80 high quality papers, one chapter of a book. He got the Annual Best Paper Award of IEEE TCSVT (2010), Best Paper Award of ACM Multimedia (2012) and Best Student Paper Award of Multimedia Modeling (2015).

Parallel In-loop Filtering Hardware Implementation for HEVC Decoders

A short review for "Efficient In-loop Filtering Across Tile Boundaries for Multi-core HEVC Hardware Decoders with 4 K/8 K-UHD Video Applications" (Edited by Carl James Debono)

S. Cho, H.M. Kim, H.Y. Kim, and M. Kim, "Efficient In-loop Filtering Across Tile Boundaries for Multi-core HEVC Hardware Decoders with 4 K/8 K-UHD Video Applications," IEEE Transactions on Multimedia, vol. 17, no. 6, pp. 778-791, June 2015.

High Efficiency Video Coding (HEVC) is the latest standard for efficient video coding of high resolution video content presenting similar subjective visual quality to H.264/AVC [1-2]. Due to its increased complexity, the standard allows parallelization of processes to make it usable in low latency applications. This is done through the support of tiles and wavefront-based parallel processing. To make HEVC useable in resource limited devices, hardware implementations that exploit parallel algorithms are necessary. Thus, techniques that minimize data dependencies between the parallel processing blocks must be developed.

Some efforts in implementing HEVC in hardware have already been done. A single-core decoder chip was developed in [3] where a seven-stage pipeline was used. This managed to decode 4K Ultra High Definition (UHD) video at 30 frames/second. However, for higher resolutions and lower power consumptions, more parallelizable solutions are needed. A parallel intra prediction structure was proposed in [4] to minimize encoding time. Other works found in literature focus on the implementation of specific components, such as [5 – 6].

Tiles and Wavefront Parallel Processing (WPP) are amongst the picture partitioning schemes introduced in HEVC. These allow parallel processing of the high resolution video content with some loss in coding performance [7-8]. Using tiles and WPP allows a large image to be split into a number of partitions that can be processed on different processing cores in parallel. These tools allow different configurations of the encoder and thus its implementation. However, tiles and WPP cannot be used simultaneously on a video sequence.

The work of the authors of the paper focusses on the In-Loop Filtering (ILF). The ILF demands additional resources or induces latency due to dependencies on the neighboring tiles at the tile boundaries. On the other hand, ILF processing across WPP boundaries can be implemented through sharing of the ILF samples between the interested processing cores using a common buffer. This is possible because a partition in WPP is a Coding Tree Unit (CTU) row, which is always regular and is constraint to be decoded in order.

To implement ILF processing across the tile boundaries on parallel processing structures demands a complex control algorithm, as the tiles within a frame can have different sizes. Thus in a parallel implementation, this implies that new tiles might need to start decoding before earlier tiles have been decoded. This means that in such cases, the decoder would need either addition filters to cope with the load or impose idle periods to allow synchronization of the parallel processing cores.

The authors of the original paper propose a hardware solution that does not demand additional in-loop filters to perform any post processing at the tile boundaries. Furthermore, it also avoids loose synchronization between the parallel processing units. The memory space needed for storing and loading of the tile boundary data is just half that needed for the post processing scheme.

ILF across tile boundaries is needed to avoid cross-shaped blocking artifacts at these points which spread through coding dependencies in predicted and bi-predicted frames. The control algorithm presented by the authors of the paper starts by checking the ILF status of the CTUs that are at the tile boundaries, where an index having two bits is used to track a boundary adjacent CTU's pipeline processing status, and confirms that they are updated. This data is used to determine the samples that are needed for processing in each decoder processing core. ILF is then performed on these sample areas for the currently decoded CTU in each processing core independently of the other processing cores decoding the neighboring tiles. The samples that are required for boundary CTUs and the filter parameters are loaded from external buffers that are shared between the decoder processing units. This implies that no additional in-loop filters are necessary to process the CTUs at the boundaries and tile decoding can be processed independently of the tile scan order. The authors of the original paper verify the control algorithm through its implementation on a quad-core HEVC decoder for 4

K-UHD video. The prototyping was done on a Field Programmable Gate Array (FPGA) with reported results indicating ILF processing speedups which are proportional to the number of processing units and negligible increase in overhead due to the control algorithm.

In order to get HEVC closer to the market, especially in mobile and resource limited devices, better parallel and faster hardware processing units are needed to minimize encoding and decoding latency. Demand for high resolution video services and applications is increasing both on the downlink and the uplink, as users generate and share their content. Power consumption remains an issue for battery powered devices and thus the video encoder and decoder play an important role in keeping this low with increased multimedia use.

References:

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, December 2012.

[2] P. Hanhart, M. Rerabek, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation of the upcoming HEVC video compression standard," in *Proc. SPIE Applications of Digital Image Processing XXXV*, vol. 8499, pp. 84990 V-1–84990 V-13, October 2012.

[3] M. Tikekar, C.-T. Huang, C. Juvekar, V. Sze, and A. Chandrakasan, "A 249-mpixel/s HEVC video-decoder chip for 4 K ultra-HD applications," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 61–72, January 2014.

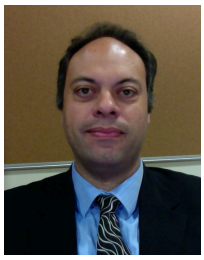
[4] W. Lu, N. Yu, J, Nan, and D. Wang, "A hardware structure for HEVC intra prediction," in *Proceedings of the 2nd International Conference on Information Science and Control Engineering,* pp. 555-559, April 2015.

[5] C.M. Diniz, M. Shafique, S. Bampi, and J. Henkel, "A reconfigurable hardware architecture for fractional pixel interpolation in high efficiency video coding," *IEEE Transactions on Computer-Aided Design of Integrate Circuits and Systems*, vol. 34, no. 2, pp. 238 – 251, February 2015.

[6] Y. Chen, and C. Liu, "Area-efficient video transform for HEVC applications," *Electronics Letters,* vol. 51, no. 14, pp. 1065 – 1067, July 2015.

[7] C.C. Chi, M. Alvarez-Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schierl, "Parallel scalability and efficiency of HEVC parallelization approaches," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no.12, pp. 1827 – 1838, December 2012.

[8] *On Tiles and Wavefront Tools for Parallelism*, doc. JCTVC-I0198, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), May 2012.



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 was the IEEE Region 8 Vice-Chair of Technical Activities between 2013 and 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.

Understanding an Exciting New Feature of HEVC: Tiles

A review for "An Overview of Tiles in HEVC"
(Edited by Gwendal Simon)

Misra, K.; Segall, A.; Horowitz, M.; Shilin Xu; Fuldseth, A.; Minhua Zhou, "An Overview of Tiles in HEVC," Selected Topics in Signal Processing, IEEE Journal of , vol.7, no.6, pp.969,977, Dec. 2013

The High Efficiency Video Coding (HEVC) standard significantly improves coding efficiency (gains reported as 50% when compared to the state-of-the-art MPEG-4 AVC H264), and thus is expected to become popular despite the increase in computational complexity. HEVC also provides various new features, which can be exploited to improve the delivery of multimedia systems. Among them, the concept of tiles is in my opinion a promising novelty that is worth attention. The paper "An Overview of Tiles in HEVC" provides an excellent introduction to this concept.

The goal of a video decoder (respectively encoder) is to convert a video bit-stream (respectively the original sequence of arrays of pixel values) into a sequence of arrays of pixel values (respectively a bit-stream). The main idea that is now adopted in video compression is the hierarchical structure of video stream data. The bit-stream is cut into independent Group of Pictures (GOP), each GOP being cut into frames, which have temporal dependencies with regards to their types: Intra (I), Predicted (P) or Bidirectional (B) pictures. Finally, each frame is cut into independent sets of macroblocks, called *slices* in the previous encoders.

The novelty brought by HEVC is the concept of *tile*, which is at the same "level" as slice in the hierarchical structure of video stream data. The motivations for both slices and tiles are, at least, twofold: error concealment and parallel computing. First, having an independently parsable unit within a frame can break the propagation of errors. Indeed, due to the causal dependency between frames, an error in a frame can make the decoder unable to process a significant portion of the frames occurring after the loss event. Slices and tiles limit, at least from a spatial perspective, the propagation of an error on the whole frame. Second, the complexity of recent video and the requirements of high-speed CPU speed (which unfortunately requires power and generates heat) can be partially addressed by parallelizing the decoding computation task across multiples computing units, regardless of whether these are cores in many-cores architectures or computing units in Graphics Processing Units (GPUs). The independency of slides and tiles is expected to facilitate the implementation of video decoder on parallel architectures.

Unfortunately, the concept of slices suffers in practice from serious weaknesses, which tiles are expected to fix. In the paper, the authors introduce the main differences between tiles and slices, which are two concepts that, at a first glance, can be confused. They focus on the motivation for parallel computation.

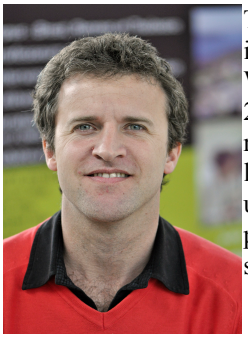
The first part of the paper explains in details the main principles between both approaches, in particular the fact that tiles are aligned with the boundaries of Coded Tree Blocks (CTD), which provides more flexibility to the partitioning. This brings several benefits: a tile is more compact, which leads to a better correlation between pixels within a tile when compared to the correlation between pixels in a slice. Tiles also require less headers, among other advantages.

The authors also introduce the known constraints to be taken into account when one wants to use tiles today. The whole Section 3 is about the tile proposal in HEVC, and the main challenges to be addressed for a wide adoption. Next, the authors present some examples when tiles are useful. Both parts are written so that somebody being just familiar with the concepts can understand both the limitations behind the concept of tiles and how these weaknesses have been addressed in practice.

The last part of the paper, in Section 5, deals with some experiments, which demonstrate the efficacy of HEVC for lightweight bit-streams and parallel architectures. At first authors assess the parallelization and the sensibility of network parameters, including the Maximum Transmission Unit (MTU), on the performances of slices versus tiles. They finally measure the performances of stream rewriting for both approaches.

In short, the paper shows that tiles appear to be more efficient than slices on a number of aspects. The paper proposes a rigorous, in-depth, introduction of the main advantages of tiles. This can foster research on

the integration of tiles into next-generation multimedia delivery systems.



Gwendal Simon is Associate Professor at Telecom Bretagne. He received his Master Degree in Computer Science in 2000 and his PhD degree in Computer Science in December 2004 from University of Rennes 1 (France). From 2001 to 2006 he was a researcher at Orange Labs, where he worked on peer-to-peer networks and social media innovations. Since 2006, he has been Associate Professor at Telecom Bretagne, a graduate engineering school within the Institut Mines-Telecom. He has been a visiting researcher at University of Waterloo from September 2011 to September 2012. His research interests include large-scale networks, distributed systems, optimization problems and video delivery systems.

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 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.