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

MMTC Communications – Review

Vol. 8, No. 5, October 2017

IEEE COMMUNICATIONS SOCIETY

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the October 2017 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises four reviews that cover multiple facets of multimedia communication research including medical video analysis, cloud-based 3D modeling and big data analysis, and deep image learning. These reviews are briefly introduced below.

The first paper, published in IEEE Transactions on Multimedia and edited by Bruno Macchiavello, investigated heart rate and rhythm analysis using 3D motion tracking in depth video.

The second paper is published in IEEE Transactions on Image Processing and edited by Carsten Griwodz. It proposed 3D-3D registration algorithms to align point clouds generated by different means.

The third paper, published in IEEE Transactions on Multimedia and edited by Xiaohu Ge, studies the data-driven delay estimation in a practical cloud media with heavy traffic, and proposes an accurate estimation strategy only with a small amount of dataset.

The fourth paper is published in IEEE International Conference on Multimedia and Expo and edited by Frank Hartung. The paper examined the use of deep learning for single image super-resolution (SISR).

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

IEEE ComSoc MMTC Communications – Review Directors

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

Qing Yang
University of North Texas, USA
Email: qing.yang@unt.edu

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

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

Measuring Heart Rate through Head Motion in Depth Video

A short review for "Estimating Heart Rate and Rhythm via 3D Motion Tracking in Depth Video"

Edited by Bruno Macchiavello

C. Yang, G. Cheung and V. Stankovic, "Estimating Heart Rate and Rhythm via 3D Motion Tracking in Depth Video," in *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1625-1636, July 2017.

Measuring heart rate can help monitor a person's fitness level; moreover, it can help spot developing health problems. There are several ways that can be used to measure heart rate. Lately, image-based health monitoring systems have attracted attention due to the advantage of being completely non-contact [1]. In 2008, Verkrysse et al. [2], estimated heart rate from an RGB video recorded using a single conventional digital camera, by analyzing subtle changes in the skin color caused by blood circulation. A similar approach has been used by other studies [4-6]. Even smartphone applications have been proposed [7]. In this work, the authors present an image-based heart rate measuring technique that only uses depth video. Depth images are generated by computing the per-pixel distance between the physical objects in the 3D scene and the sensing device. The advantage of using depth images from a texture RGB signal is that the illumination variation is not a factor. Previous depth-images-based systems have demonstrated that certain biometrics, like respiratory rate, can be accurately estimated [8]. Nevertheless, due to limitations of today's depth sensing technologies, capture depth videos typically suffer from low bit-depth and acquisition noise. Thus, it is difficult to process acquired depth images to estimate biometrics that require tracking of subtle 3D motion of a human subject.

The authors of this work strive to overcome this difficulty and propose to capture depth video of a human subject using Kinect 2.0 to estimate his/her heart rate and rhythm. As blood is pumped from the heart to the head for circulation, the head will oscillate slightly due to Newtonian mechanics, and tracking this tiny oscillatory movement can lead to an estimate of heart rate. The proposed method can be divided into three steps. The first is restoration of capture depth images via a joint bit-depth enhancement and denosing procedure. In order to do so, the authors derive a suitable noise model for Kinect 2.0 captured pixels in a depth frame. For model derivation, a flat static board was placed on a table and video of T frames was recorded. The vertical and horizontal auto-correlation of all pixels for $T = 15{,}000$ was computed. The authors verified that the correlation with immediate neighboring pixels is strong but weakens considerably thereafter. Using the computed correlations, the authors proposed a Gaussian Markov Random Field noise model. With the proposed model, the authors developed a joint-bit depth enhancement/denoising for both spatial and temporal noise. For spatial denoising, the idea is to estimate the original vector of depth values $\mathbf{x}$, from $\mathbf{y}$, which is the quantized version of $\mathbf{x}+\mathbf{n}$, where $\mathbf{n}$ is the added Gaussian noise. They use a graph-signal smoothness prior in an optimization problem, using a *maximum a posteriori* (MAP) formulation. The nodes in the graph correspond to pixels in $\mathbf{x}$, each node is connected to its horizontal and vertical neighbors to yield a 4-connected graph. It is assumed that $\mathbf{x}$ is piecewise smooth if the graph Laplacian regularizer $\mathbf{x}^{\mathbf{T}}\mathbf{L}\mathbf{x}$ is small, where $\mathbf{L}$ is the generated graph. The derived optimization function has two inter-dependent variables that are optimized alternately, fixing the other, until the solution converges. A similar approach is performed for temporal denoising. In this case given motion vector v points to a matching block $\mathbf{x}_{t\text{-}1}$ in the previous restored frame t-1, the optimization becomes the search for motion vector v and denoised patch $\mathbf{x}_t$ that minimize a graph-signal smoothness term $\mathbf{x}_t^{\mathrm{T}}\mathbf{L}\mathbf{x}_t$, a fidelity term with respect to observation y_t and a motion estimation term.

In the second step, the head region of the human subject is identified in the first frame of the video sequence. And then, head tracking is performed in the following frames. The detection is made through a template matching method. First, Canny edge detection is applied on the restored depth frames, followed by a scale-invariant Chamfer matching [51]. Next, a circular region is extracted around each detected location and fit with a hemisphere model to locate the probable head position. Once the ROI is selected, a *kernelized correlation filter* (KCF) is used for tracking [52]. The authors proposed a consistency check for motion information in order to improve the tracking results. They use the derived formulation for joint bit-depth enhancement/ temporal denoising and the motion vectors obtained from the first run of the tracker are used to re-formulate the optimization problem. The motion vectors from the tracker are used as a motion prior and the optimization is performed again to improve the temporal denoising. Then, the tracker is used one more time over the denoised signal, and the obtained motion vectors are used for estimating heart rate and rhythm.

Given the deduced 3D motion, the information is projected via PCA in order to perform 1D analysis. PCA performs Eigen decomposition to determine three eigenvectors of the motion information and arranges them in descending order of magnitude. The authors then project the motion information onto each eigenvector separately, resulting in three 1D projected trajectories. Next, it is verifies which trajectory is the most periodic one. Then, a band-pass filter is used to isolate a normal-heart-rate frequency band of interest, which the authors set to be 0.7 Hz to 4 Hz. Prior to heart rate estimation, wavelet denoising is also performed.

Finally, the heart rate is computed Welch PSD estimation [64]. Given the wavelet reconstructed signal, the frequency with maximum PSD is designated as the heart-beat frequency. The authors also compute the heart rhythm via peak detection. Experimental results show robustness to different views, and accurate estimation of the heart rate and rhythm using the proposed algorithm compared to the values estimated by a portable finger pulse oximeter.

References

[1] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, “A survey on ambient intelligence in healthcare,” Proc. IEEE, vol. 101, no. 12, pp. 2470–2494, Dec. 2013.

[2] W. Verkruysse, L. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” Opt. Express, vol. 16, no. 26, pp. 21434–21445, Dec. 2008.

[3] J. Krause, H. Jin, J. Yang, and L. Fei-Fei, “Fine-Grained Recognition without Part Annotations,” CVPR, 2015.

[4] K. Y. Lin, D. Y. Chen, and W. J. Tsai, “Face-based heart rate signal decomposition and evaluation using multiple linear regression,” IEEE Sensors J., vol. 16, no. 5, pp. 1351–1360, Mar. 2016.

[5] H. Y.Wu et al., “Eulerian video magnification for revealing subtle changes in the world,” ACM Trans. Graphics, vol. 31, no. 4, Jul. 2012.

[6] H. E. Tasli et al., “Remote PPG based vital sign measurement using adaptive facial regions,” in Proc. IEEE Int. Conf. Image Process., Paris, France, Oct. 2014.

[7] R. Stricker et al., “Non-contact video-based pulse rate measurement on a mobile service robot,” in Proc. IEEE Int. Symp. Robot Human Interactive Commun., Edinburgh, U.K., Aug. 2014, pp. 1056–1062.

[8] C. Yang, Y. Mao, G. Cheung, V. Stankovic, and K. Chan, “Graphbased depth video denoising and event detection for sleep monitoring,” in Proc. IEEE Int.Workshop Multimedia Signal Process., Jakarta, Indonesia, Sep. 2014, pp. 1–6.

[9] H. G. Barrow et al., “Parametric correspondence and chamfer matching: Two new techniques for image matching,” in Proc. Int. Joint Conf. Artif. Intell., Cambridge, MA, USA, Aug. 1977, pp. 659–663.

[10] P. D. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” IEEE Trans. Audio Electroacoust., vol. AU-15, no. 2, pp. 70–73, Jun. 1967.

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

A New Point Cloud Registration Method to Align a Pair of 3D Models

A short review for

"A Systematic Approach for Cross-Source Point Cloud Registration by Preserving Macro and Micro Structures"

Edited by Carsten Griwodz

Xiaoshui Huang; Jian Zhang; Lixin Fan; Qiang Wu; Chun Yuan, "A Systematic Approach for Cross-Source Point Cloud Registration by Preserving Macro and Micro Structures", in *IEEE Transactions on Image Processing* 26(7), July 2017, pp. 3261-3276

The method discussed in the paper aims at overcoming the severe problems that other known algorithms face concerning their robustness to a variety of problems. These are:

- Varying point densities
- Missing data
- Variation due to different viewpoints
- Varying scale factors
- Noise and outliers

All of these challenges appear obvious from the perspective of the developer of a real-world application. Point density vary simply because of the reconstruction technique used, where different methods will most likely not only fail to produce the same point densities, they will fail to produce strong regional variations in the density differences; for example, structured-from-motion (SfM) has high density in textured regions, low density in badly texture regions, whereas LiDAR provides a uniform angular density.

Missing data can easily be the result of an SfM reconstruction that fails on untextured regions of a scene, whereas different viewpoints can frequently hide parts of a scene entirely from an immobile LiDAR while a camera used for SfM must move around to create structure and will often see sides of an object that are not exposed to the LiDAR. The scales of reconstructions as well as the presence of outliers are the final challenges listed by the authors. In SfM, for example, scale is intrinsically unknown, scale is always a degree of freedom in its the reconstructions unless additional data is provided; the accuracy of distance estimations depend to a large degree on the angle between matched pixel in camera positions, even when the matching process gives a perfect result. LiDARs, on the other hand, have always been designed for accurate distance measurements.

All of these challenges make it hard for 3D-3D registration algorithms to align point clouds generated by different means.

After an excellent discussion of the state of the art, which includes the most successful methods to date, the authors propose a method that matches on large and small scales, considering both the global structure as well as local details. These two sides are called the macro and micro structures, and they are use iteratively in the propose registration process. The authors point out that the macro structure focuses entirely on geometric structure, while the micro structure looks at small regions of stable geometry and may include additional information.

The proposed algorithm has 5 steps: pre-processing, structure extraction, graph construction, optimization, and transformation estimation.

Pre-processing is based on existing work, but makes the assumption that both point clouds are covering at least approximately the same regions. This implies that the algorithm is not suited for identifying the position of one point cloud in a larger space reconstructed for a second point cloud. This may have interesting consequences for matching SfM and LiDAR reconstructions, but should be easily overcome by manual intervention.

For the extraction of structure, the authors make use of a variation of Papon et al.'s super voxel extraction [1] to group voxels with consistent properties. These super voxel are interpreted a micro structure, while the geometry relation of the super voxel centres is considered the macro structure.

In graph construction, the authors extract and use ESF descriptors [2] to match super voxel with each other, and then create a new descriptor from distance ratios and Euler angles between super voxel centres that are connected due to their adjacency. This descriptor can be matched in the following step (optimization) without attention to scale and rotation and can robustly compare adjacency graph structure.

The optimization step revisits the generated graphs and tries to match them. They are, however, very likely to differ in number of nodes due to the expected lack of overlap of the reconstructed region, and the matching must be robust to this. The authors make use of factorized graph matching (FGM) [3] to match these graph, but modify it to avoid local minima in the optimization more efficiently for their cause. The end result of the optimization step is the adjacency matrix for the super voxels that are present in both representations.

With these macro structures now known to match, the authors compute the rigid transformation that aligns the one graph to the other, using a 3D RANSAC to find approximate alignment, followed by iterative closest point refinement as a last step.

In conclusion, the authors have dug deeply into the existing state of the art and found a brand-new combination of steps to base the global structure of point clouds on knowledge about matching local structures, and the application of robust iterative matching technique to gain a highly accurate end result.

The authors demonstrate the performance of their method on examples from three databases of objects that have been reconstructed by a variety of means. They can demonstrate the performance of their technique by examples that range from well-known and simple to highly complex. They can even show a better matching of the Pisa Cathedral model used by Mellado et al.'s method [4].

With the propose algorithm, the authors are currently leading the state-of-the-art in point cloud alignment, and multimedia researchers are recommended to take a close look at their work. However, there is still a considerable number of limitations, and general real-world use of 3D-3D alignment in arbitrary scenes is still an open problem that should be tackled by the multimedia community.

Unfortunately, at the time of this writing, the authors cannot share exact details of the algorithm or share its code due to patent procedures. But this does only go to show that commercial interest in this topic is waking and should not be ignored by the research community.

References

[1] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel Cloud Connectivity Segmentation - Supervoxels for Point Clouds," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2027–2034.

[2] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3D object classification," in *2011 IEEE International Conference on Robotics and Biomimetics*, 2011, pp. 2987–2992.

[3] F. Zhou and F. De la Torre, "Deformable Graph Matching," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2922–2929.

[4] N. Mellado, M. Dellepiane, and R. Scopigno, "Relative Scale Estimation and 3D Registration of Multi-Modal Geometry Using Growing Least Squares," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 9, pp. 2160–2173, Sep. 2016.

Carsten Griwodz is chief research scientist in the Media Department at the Norwegian research company Simula Research Laboratory AS, Norway, and professor at the University of Oslo.
His research interest is the performance of multimedia systems. He is concerned with streaming media, which includes all kinds of media that are transported over the Internet with temporal demands, including stored and live video as well as games and immersive systems. To achieve this, he wants to advance operating system and protocol support, parallel processing and the understanding of the human experience. He was area chair and demo chair of ACM MM 2014, and general chair of ACM MMSys and NOSSDAV (2013), co-chair of ACM/IEEE NetGames (2011), NOSSDAV (2008), SPIE/ACM MMCN (2007) and SPIE MMCN (2006), TPC chair ACM MMSys (2012), and systems track chair ACM MM (2008). He is currently editor-in-chief of the ACM SIGMM Records. More information can be found at http://mpg.ndlab.net.

Combining Big Data with Intelligent Analysis for Media Cloud

A short review for "On Data-Driven Delay Estimation for Media Cloud"

Edited by Xiaohu Ge

Liang Zhou, "On Data-Driven Delay Estimation for Media Cloud," in *IEEE Transactions on Multimedia*, vol. 18, no.5, pp. 905-915, 2016.

With the tremendous developments of the multimedia service in the mobile internet, multimedia traffics have dominated the network contents. Media cloud, as a promising and efficient multimedia processing and application platform, has become the main solution of the large-scale multimedia services in the big data era [1]–[3].

Previous works have demonstrated that waiting time is the core metric since usually the multimedia services are delay sensitive [4]–[5]. Thus, how to reduce the waiting time has become one of the most important topics in the framework of media cloud.

For the ordinary data service in the cloud (for example, short message service, etc.), delay is negligible due to the enormous data processing capacity of the cloud, so there is relatively little incentive to provide delay announcement. However, for the large-scale multimedia services in the cloud (e.g., video game, video communications, etc.), the delay becomes significant and non-ignorable [6]. In this case, delay announcement is a very useful and inexpensive tool to improve the user satisfaction.

Generally speaking, this work studies the data-driven delay estimation in a practical cloud media with heavy traffic, and proposes an accurate estimation strategy only with a small amount of dataset. Importantly, it explicitly models the subjective announcement dependent user response via an objective response function through the elaborate data analysis and model. The main contributions of this work are as follows.

On the theoretical end, through designing an appropriate window function to realize the data-cleaning, a simple but efficient objective expression for the subjective user response with respect to delay announcement is established for the heavy traffic regime. In particular, an updated data driven delay estimation method is proposed in the framework of fluid model, which enables the cloud designer to use only a small amount of data set according to the user demand and user satisfaction, and to analyze its impact in terms of the user response. Importantly, this updated estimation method is able to adapt dynamic cloud to various user subjective responses.

On the technical end, this work analyzes the conditions for data driven delay estimation, and investigates the impact of the data set on the subjective performance. This work proves that the proposed updated data-driven method is more applicable than the existing data-driven methods by taking into account both the performance and cost. In particular, the process of service dynamics can be well modeled via an appropriate window function, and the steady state distribution of the waiting time can be described by the user response as well. Interestingly, the updated data set is able to obtain a near-optimal solution within a finite time period, which demonstrates that approximate optimal delay estimation exists in the framework of fluid model with a finite time window.

In conclusion, this work mainly focuses on studying the data-driven delay estimation for heterogeneous multimedia services in the cloud. On one hand, the advantages and characteristics of the data-driven method have been thoroughly analyzed. On the other hand, an updated data-driven method has been proposed to achieve the balance between the computational complexity and estimation accuracy. In particular, through transforming the subjective user response into the objective response expressions, this study designs a general and adaptive data set acquirement in the framework of fluid model.

References

[1] Y. G. Wen, X. Q. Zhu, J. P. C. Rodrigues, and C. W. Chen, "Mobile cloud Media: Reflections and outlook," IEEE Trans. Multimedia, vol. 16, no. 4, pp. 885–902, Jun. 2014.

[2] X. Wang, et.al, "AMES-cloud: A framework of adaptive mobile video streaming and efficient social video sharing in the clouds," IEEE Trans. Multimedia, vol. 15, no. 4, pp. 811–820, Jun. 2013.
[3] C. Lai, Y. Lai, H. Chao, and J. Wan, "Cloud-assisted real-time transrating for HTTP live streaming," IEEE Wireless Commun., vol. 20, no. 3, pp. 62–70, Jun. 2013.
[4] D. Wu *et al.*, "Quality-of-protection-driven data forwarding for intermittently connected wireless networks," IEEE Wireless Commun., vol. 22, no. 4, pp. 66–73, Aug. 2015.
[5] X. Ge, L. Pan, Q. Li, G. Mao and S. Tu, "Multi-Path Cooperative Communications Networks for Augmented and Virtual Reality Transmission," IEEE Transactions on Multimedia, Vol 19, Issue 10, pp. 2345-2358, 2017.
[6] M. Chen, Y. Zhang, Y. Li,M. Hassan, and A. Alamri, "AIWAC: Affective interaction through wearable computing and cloud technology," IEEE Wireless Commun., vol. 22, no. 1, pp. 20–27, Feb. 2015.

Xiaohu Ge is currently a Professor with the School of Electronic Information and Communications at Huazhong University of Science and Technology (HUST), China. He is the director of China International Joint Research Center of Green Communications and networking. His research interests are in the area of mobile communications, green communications. He has published about 160 papers in refereed journals and conference proceedings and has been granted about 15 patents in China. He is servicing as an editor of IEEE Wireless Communications.

Image Detail Can Be Learned: Super-Resolution Using Machine Learning

A short review for "Deep Hybrid Residual Learning With Statistic Priors For Single Image Super-Resolution"
Edited by Frank Hartung

Risheng Liu, Xiangyu Wang, Xin Fan, Haojie Li, Zhongxuan Luo, "Deep Hybrid Residual Learning With Statistic Priors For Single Image Super-Resolution", *In proceedings of the IEEE International Conference on Multimedia and Expo (ICME)* 2017, pp. 1111-1116.

In multimedia communications, machine learning, deep learning and convolutional neural networks are now intensively investigated. Some examples include the application of machine learning for compression [1], de-blocking [2], transcoding [3], stream scheduling [4], and video segmentation.

The authors of the publication under discussion have investigated the use of deep learning for single image super-resolution (SISR), i.e., the estimation of high-resolution images from corresponding low-resolution observations. Applications for SISR include video surveillance, face recognition, video streaming, medical image processing and long-distance imaging. There has already been significant research into SISR. Previous methods assumed that images can be represented by a limited set of representative patterns in a dictionary, and focused on the construction of such dictionaries, and the mapping between low-res and high-res patterns [5].

Recently, deep learning, specifically convolutional neural networks (CNNs) [6], has proven to be an efficient data-driven tool for many low-level vision tasks, providing state-of-the art results. In [7], Dong et al. proposed the first CNN based concept that addresses the SISR problem. This as well as subsequent publications [8, 9] fully relies on the quality of the data used for training the CNN; no other prior knowledge about the intrinsic image structure is used.

In the publication discussed here, the authors propose a CNN based concept for SISR where so-called domain-knowledge like image priors and data fidelity information is used in the design of the CNN. Thus, the network combines statistic priors of natural images and trained data structures. The problem to be solved is to determine the high-resolution version of an image from its blurred, down-sampled and noisy low-resolution version. It is easy to understand that this is a very ill-posed problem; many different high-resolution images could result in the same noisy / blurred / down-sampled low-resolution picture. This is why super-resolution is indeed a hard problem.

The authors develop their SISR system as a CNN with two parallel branches (Figure 1 in the actual paper provides visualization) that together form a maximum a posteriori (MAP) inference framework.

The first branch is called the prior network. It is in fact not a trained network in the machine learning sense, although it is realized as a three-layer CNN with corresponding convolutional layers. However, it contains a Markov random field (MRF) that models the statistics of high quality natural images. The input of this prior network is the low-resolution image; the output is already an enhanced (higher-resolution) version. However, that version is not yet satisfying.

The second branch is called the fidelity network. This is a 20-layer CNN network trained on test data sets. It has been trained to yield the difference between the MRF-enhanced output version of the prior network, and the target high-resolution version of the low-resolution input image. This can be trained based on the ground truth contained in the test data sets. The training is conducted as to minimize the mean squared error. The authors use a nonlinear activation function based on the Radial Basis Function. In order to have good control of the training process, a greedy learning scheme is employed where each subnetwork is optimized separately. Possible overfitting during training is reduced by a control mechanism that minimizes a loss function. The output of the fidelity network is an enhancement image that is added to the output of the prior network, giving the final estimated high-resolution version of the input low-resolution image.

The scheme has been validated using three different test datasets (Set5, Set14, BSD100), and has been compared in terms of PSNR and SSIM to five methods from the prior art including [7, 8]. In almost all cases, the proposed method outperforms the prior art; only [8] provides similar yet slightly worse results. The proposed methods outperforms the reference prior art by up to 0.7 dB PSNR. In terms of subjective visual impression, the method is clearly better than the prior art; the method provides images that appear sharper. Computational cost is not discussed in the paper, but is likely to be higher than for prior art that that does not use machine learning techniques and CNNs.

Overall, this method provides a convincing progress of the state-of-the-art in single-image super-resolution, with applications mentioned above.

References

[1] Philipp Helle, Heiko Schwarz, Thomas Wiegand, Klaus-Robert Müller, "Reinforcement learning for video encoder control in HEVC", Proceedings 2017 International Conference on Systems, Signals and Image Processing (IWSSIP), 2017, pp. 1-5.

[2] Leida Li, Weisi Lin, Hancheng Zhu, "Learning Structural Regularity for Evaluating Blocking Artifacts in JPEG Images", IEEE Signal Processing Letters, 2014, vol. 21, no. 8, pp. 918-922.

[3] Xiangyu Li, Rong Xie, Li Song, "Machine learning based VP9-to-HEVC video transcoding", Proceedings 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2017, pp. 1-6.

[4] Yu-Lin Chien, Kate Ching-Ju Lin, Ming-Syan Chen, „Machine learning based rate adaptation with elastic feature selection for HTTP-based streaming“, Proceedings 2015 IEEE International Conference on Multimedia and Expo (ICME), 2015.

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

[6] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," Proceedings Advances in Neural Information Processing Systems 25 (NIPS), 2012, pp. 1097–1105.

[7] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks.," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295–307, 2015.

[8] Z. Wang, D. Liu, J. Yang, W. Han, and T.S. Huang, "Deep networks for image super-resolution with sparse prior," Proceedings International Conference on Computer Vision, 2015, pp. 370–378.

[9] J. Kim, J. Lee, and K. Lee, "Accurate image superresolution using very deep convolutional networks," Proceedings Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

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. In 2008, he was a visiting researcher at Stanford University, Palo Alto, USA. In 2015, he was a visiting professor at Namibia University of Science and Technology, Windhoek, Namibia, and in 2016, he was a visiting researcher at Eurecom, Sophia-Antipolis, France. His research interests include media security and forensics, video analysis, 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 23 granted patents.

Paper Nomination Policy

Following the direction of MMTC, the Communications – Review platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia includes, but is not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication. Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings, or other distinguished journals/conferences within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to Review Board Directors: Pradeep K. Atrey (patrey@ albany.edu), Qing Yang (qing.yang@unt.edu), Wei Wang (wwang@mail.sdsu.edu), and Jun Wu (wujun@tongji.edu.cn). The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper, when possible.

Review Process

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

Best Paper Award

Accepted papers in the Communications – Review are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board). For more details, please refer to http://mmc.committees.comsoc.org/.

MMTC 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.