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

R-LETTER

Vol. 4, No. 1, January 2013

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from Review Board

Farewell and Welcome

Since October 2012, the new Review Board as a team contributes collective effort to recommend articles which discuss and present the latest state-of-the-art techniques, systems and applications. Regrettably, due to unexpected circumstances, Dr. Xianbin Wang has to resign as the Co-Director. On behalf of the Board, I gratefully thank Dr. Wang's contribution and support. I would also like to take this opportunity to welcome Dr. Christian Timmerer and Dr. Weiyi Zhang to join as Co-Directors of the Review Board. Dr. Christian Timmerer has been a member of the Review Board since its establishment. His experience is an important asset of the team. Dr. Zhang has been an editor of the MMTC E-Letter Board before accepting this post. His expertise will certainly help advancing our Review Board activities.

New Initiatives from Interest Groups

As announced earlier by Dr. Kai Yang, Vice-Chair (Letter & Communications), MMTC Interest Groups will actively participate in R-letter production by recommending distinguished articles authored by IEEE/ACM Fellows. The new R-letter will thus contain two types of articles: distinguished and regular respectively. The new format is expected to appear in the March 2013 issue.

Highlight of This Issue

Over more than a decade, **online videos** have been providing entertaining and educational contents to Internet users, but at the same time have unveiled many challenges to the multimedia communications research community. To deliver high quality scalable videos on heterogeneous displays online, in particular over wireless networks, is still an ongoing research which involves the study of various techniques, which include efficient video streaming, coding standard and signal interference handling. In this issue, we recommend seven papers. While the first five focus on online videos including virtual views generation for multi-views video, the last two focus on the **3D media**, *i.e.*, 3D head animation and online 3D content retrieval.

The first paper, published in the *IEEE Transactions on Multimedia*, introduces a time and space efficient Lasso framework to address the issue of automatic analysis or indexing of video contents on the Internet. **The second paper**, in *IEEE Transactions on Multimedia*, exploits the techniques to handle occluded objects in synthesized views for 3D videos. **The third paper**, from the *IEEE Transactions on Multimedia*, proposes a congestion control framework for real-time video streaming over wireless ad-hoc networks, as opposed to the traditional rate distortion optimization techniques. **The fourth paper**, from *IEEE Transactions on Multimedia*, investigates an energy-efficient strategy for heterogeneous clients to transmit scalable video over wireless multicast channels. **The fifth paper**, published in the *IEEE Journal on Selected Areas in Communications*, discusses the advantage of using Femtocell networks to address the interference problems in wireless transmission and deliver good quality video streaming. **The sixth paper**, from the *IEEE Transactions on Multimedia*, reviews the different methods including semantically-based and geometrically-based techniques for 3D object retrieval from online media databases. **The last paper**, published in *IEEE ICME 2011*, presents a complete pipeline for speech-driven animation of generic 3D head models. The proposed system is adaptable to any language.

I would like to thank all the authors, reviewers, nominators, editors and others who contribute to the release of this issue.

IEEE ComSoc MMTC R-Letter

Director,
Irene Cheng, University of Alberta, Canada
Email: locheng@ualberta.ca

Distributed Video Concept Detection with Parallel Lasso

A short review for "Parallel Lasso for Large-Scale Video Concept Detection"

Edited by Jiang Zhu

B. Geng, Y. Li, D. Tao, M. Wang, Z. Zha and C. Xu, "Parallel Lasso for Large-Scale Video Concept Detection", IEEE trans. On Multimedia, Vol. 14, No. 1, pp. 55-65, Feb. 2012.

With the explosive multimedia contents, especially the video contents, on the Internet, content analysis or indexing becomes more and more important. Manual labeling for each video is of utterly high cost, if not impossible, due to the enormous size of today's video database. Therefore, researchers have been actively working on systems and algorithms to automatically yet efficiently detect video concepts such as object, people and event. Based on multimedia signal processing and machine learning technologies, a mapping function can be constructed to infer high-level semantic concepts from low-level video features, e.g., color, texture, histogram, etc.

Common machine learning techniques for video concept detection include support vector machines [1], regularized least squares [2], and logistic regression [3]. However, they may suffer from scalability issue [4]. Data driven learning approaches usually require a huge amount of data to generalize proper mappings and the learning complexity is in the order of super-quadratic. On the other hand, combining a large set of video features can boost the detector's performance; therefore, it is desirable to have sparse learning technique to tackle high-dimension as well as large-scale settings.

In this paper, the authors propose a novel distributed video concept detection method, i.e., Plasso, to overcome the aforementioned scalability and high-dimension issues. This method, which is based on lasso[5], utilizes parallel computing so that both time and space complexities are significantly reduced. For a dataset with n samples in a d-dimensional space, the original lasso has time complexity $O(d^3)$ and space complexity $O(d^2)$, whereas Plasso has $O(h^2 d/m)$ and $O(hd/m)$ respectively (h is the reduced dimension from incomplete Cholesky factorization and m is number of processors). A kernel extension of Plasso is also developed with time complexity $O(h^2 n/m)$ and space complexity $O(hn/m)$.

Generally, the optimization solution to lasso problem is hard to parallelize. Following standard primal-dual interior point method (IPM)[6], the authors identify two linear equations to iteratively approach the optimum. Directly applying the two linear equations requires $O(d^2)$ space complexity and $O(d^3)$ time complexity to solve for the optimal mapping functions, where d denotes feature dimension. Such costs are still too high for practical large-scale multimedia applications.

One of the bottlenecks is really on the data covariance matrix XX^T in the linear regression equations, where $X = [x_1, x_2, \dots, x_n]$ with $x_i \in \mathbb{R}^d$ being a data sample padded with unitary element. Similar to parallel SVM[28], the authors adopt Parallel Incomplete Cheolesky Factorization (PICF) to approximate the covariance matrix, i.e., $HH^T \approx XX^T$, by selecting the h most pivotal vectors, wherein h is determined according to $tr(HH^T - XX^T) \leq \varepsilon$ and $h < d$. This step reduces the complexity from $O(d^3)$ and $O(d^2)$ to $O(dh^2)$ and $O(dh)$, respectively. Then, based on Sherman-Morrison-Woodbury (SMW) formula, explicit step update equations (PIPM) are obtained, which can be distributed across m different processors. Complexity analysis shows that the new Parallelized lasso method has overall time complexity $O(dh^2/m)$ and space complexity $O(dh/m)$. The communication overhead for PICF and PIPM is $O(h^2 logm)$.

Kernel-based learning algorithms usually outperform the corresponding linear algorithms. So, a Kernel version Plasso is also developed in the paper. Following the ideas in [7], the optimization problem is reformulated which is then tackled with PICF and PIPM. Instead of distributing features of each sample to different processors in the linear counterpart, the kernel Plasso distributes samples to different processors. Therefore, the time and space complexity for kernel Plasso is $O\left(\frac{nh^2}{m}+h^3\right)$ and $O(nh/m)$ respectively.

The proposed Plasso algorithm is evaluated on both TRECVID-05 and TRECVID-07 datasets with two performance metrics, average precision and time consumed. SVM and PSVM are chosen as comparison schemes. From the results, Plasso achieves comparable detection performance to that of SVM and perform slightly better than PSVM. Interestingly, SVM performs better in concepts “Car”, “Entertainment” and “Road”, while Plasso is better in concepts “Crowd”, “Weather” and “Urban”. Overall, the performance degradation for Plasso is 1.35% worse than SVM with full rank and is 5.16% worse with reduced rank (h=0.3n). On the other hand, SVM consumes significantly higher compute time than Plasso whose compute time is on par with PSVM. On average, the time saving from Plasso is 76.1%. Other important observations from the results are a) the speedup of Plasso is almost linear w.r.t. the increment of processor numbers; b) the speedup of Plasso is super-linear w.r.t. the decrement of h; c) when rank ratio is larger than 0.3, the performance degradation is negligible.

With the advance of hardware technology, processors and memories are becoming low cost. Hundreds or thousands processors can be clustered together and communicate with each other through shared memory and I/O. Distributed computing is well suited for solving problems with increasing dimensions and scales given that they can be decomposed into (independent) sub problems. Learning with distributed computing, or the so-called parallel learning, is a hot research topic in machine learning and data mining. Although this paper is focusing on applying PICF and PIPM to video concept detection, the methodology can be easily extended to other scenarios that can be formulated in the lasso framework.

Acknowledgement:

The R-Letter Editorial Board thanks Dr. Hao Hu from Cisco Research, Cisco System Inc. for his editorial support and Dr. Jun Zhao for nominating this work.

References:

[1] V. N. Vapnik, Statistical Learning Theory, New York: Wiley-Interscience, 1998.

[2] M. Wang, Y. Song, and X.-S. Hua, “Concept representation based video indexing,” in Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 2009, pp. 654-655.

[3] R. Yan and A. G. Hauptmann, “Probabilistic latent query analysis for combining multiple retrieval sources,” in Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, 2006, pp. 324-331.

[4] Q. Liu, X. Li, A. Elgammal, X.-S. Hua, D. Xu, and D. Tao, “Introduction to the special issue on video analysis,” *Compututer Visison and Image Understanding*, vol. 113, no. 3, pp. 317–318, 2009.

[5] R. Tibshirani, “Regression shrinkage and selection via the lasso,” Journal of the Royal Statistical Society, serial B, vol. 58, pp. 267–288, 1994.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[7] J. Bi, K. Bennett, M. Embrechts, C. Breneman, and M. Song, “Dimensionality reduction via sparse support vector machines,” The Journal of Machine Learning Research, vol. 3, pp. 1229–1243, Mar. 2003.

Jiang Zhu is a senior technical leader in Advanced Architecture and Research group at Cisco Systems, Inc. He has over 15 years of industrial experience building large-scale distributed media systems. His research focuses on adaptive content networking, large-scale data systems, software defined networking (SDN), cloud service orchestrations and applications of data mining and machine learning in these fields. He did his doctoral study focusing on SDN and OpenFlow in High Performance Networking Group at Stanford University. He also received his M.S. degrees in Electrical Engineering and in Management Science & Engineering from Stanford University. Before Stanford, he obtained his M.S. in Computer Science from DePaul University and B.Eng in Automation from Tsinghua University.

A New Approach for Handling Uncovered Areas in Synthesized Views for 3D Video

A short review for "Depth Image-based Rendering with Advanced Texture Synthesis Methods"

Edited by Lifeng Sun

Patrick Ndjiki-Nya, Martin Köppel Dimitar Doshkov, Haricharan Lakshman, Philipp Merkle, Karsten Müller and Thomas Wiegand , "Depth Image-based Rendering with Advanced Texture Synthesis Methods," IEEE Transactions. on Multimedia, vol. 13, No.3 1, June. 2011.

The popularity of 3-D video applications is rapidly growing as many stereo video products are currently entering the mass market. Autostereoscopic multi-view displays provide 3-D depth perception without the need to wear additional glasses by showing many different views of a scene from slightly different viewpoints simultaneously. Nevertheless, only a limited number of original views can be recorded, stored and transmitted. Consequently, the need to render additional virtual views arises, in order to support autostereoscopic multi-view displays.

Depth image-based rendering (DIBR) [1] is an appropriate technology for synthesizing virtual images at a slightly different view perspective, using a textured image and its associated depth map (DM). A critical problem is that regions occluded by foreground (FG) objects in original views may become visible in synthesized views. This is particularly problematic in case of extrapolation beyond the baseline of the original views, as there is no additional information from another original camera.

In the literature, two basic options are described to address the disocclusion problem. Either the missing image regions are replaced by plausible color information or the DM is preprocessed in a way that no disocclusions appear in the rendered image. The color filling approaches insert known background (BG) samples into the disoccluded areas utilizing interpolation methods [2], simple inpainting [3], line-wise filling [4] or basic texture synthesis [5]. However, these methods tend to introduce blur into the unknown areas or to suffer from severe artifacts in the event of structured backgrounds and dominant vertical edges. Depth pre-processing methods apply a low-pass filter [1] to smoothen out large depth gradients in the virtual view. Nevertheless, BG textures and FG objects can be considerably distorted by this approach. Another disadvantage of existing solutions is that they support only small baselines for rendering and yield very annoying artifacts for larger baseline rendering. Furthermore, most approaches render the new images frame-by-frame, ignoring the temporal correlation of the filled areas, and therefore causing typical flickering artifacts in the virtual view.

In this paper, the authors present a new approach for handling uncovered areas in synthesized views for 3-D video. Hereby, a solution for one of the most challenging problems in DIBR is presented. It is shown that disocclusions, which occur especially in extrapolated views with large baselines, can be successfully reconstructed in a visually plausible and temporally consistent manner. Therefore, statistical dependencies between different pictures of a sequence are taken into consideration via a background (BG) sprite. Furthermore, a new robust initialization procedure of the unknown image regions, a novel photometric correction algorithm and an enhanced texture synthesis approach are utilized in the DIBR system.

The proposed framework takes images and the associated DMs of a 3-D video sequence as input. Initially, the original view and the associated DMs are warped to the virtual position. In a second step, the disoccluded areas in the DM are filled with a new method that clusters the spatial neighborhood of the unknown samples to enable appropriate depth value selection. This ensures a consistent recovery of the uncovered areas in the DM. Additionally, an adaptive threshold to discriminate BG from FG information is computed automatically during the DM reconstruction. Next, the holes in the current picture are filled in from the BG sprite if image information is available. In order to account for illumination variations, the covariant cloning method [6] is utilized to fit the BG sprite samples to the color distribution in the relevant neighborhood of the current picture.

Furthermore, the remaining disocclusions in the virtual frame are treated by first initializing the areas from spatially adjacent original texture thus providing an estimate of the missing information. In the next step, patch-based texture synthesis is used to refine the initialized areas. The method proposed in [7] is enhanced in two ways in this work to account for the proposed DIBR scenario. First, the filling priority is substantially improved and additionally steered such that the synthesis starts from the BG area towards the FG objects. Second, to ensure smooth transitions between adjacent patches, an efficient new post-processing method, based on covariant cloning [6] is utilized. Finally, the BG sprite is updated with BG and synthesized image information.

Experimental results have shown that the presented approach yields considerable subjective and objective gains compared to state-of-the-art view synthesis techniques. However, depth estimation inconsistencies especially at FG-BG transitions may lead to degradation of the rendering results. The examinations also showed that the lack of an adequate perceptual measure for 3-D content hampers a fully optimized configuration of the view synthesis algorithm.

The authors' primary contribution is to propose a new hole filling approach for DIBR with advanced texture synthesis methods. The algorithm works for large baseline extensions with spatiotemporally consistent rendering results. Disocclusions in each virtual view are compensated using image information from a causal picture neighborhood via a BG sprite. Residual uncovered areas are initially coarsely estimated and then refined using texture synthesis.

Acknowledgement:

The R-Letter Editorial Board thanks Dr. Vladan Velisavljevic for nominating this work.

References:

[1] C. Fehn et al., "Interactive 3-DTV-concepts and key technologies," IEEE Trans. on IEEE, vol. 94, no. 3, pp. 524–538, 2006.

[2] S. Zinger et al., "Free-Viewpoint Depth Image Based Rendering," in *Journal Visual Communication and Image Representation*, vol. 21, issues 5-6, pp. 533-541, 2010.

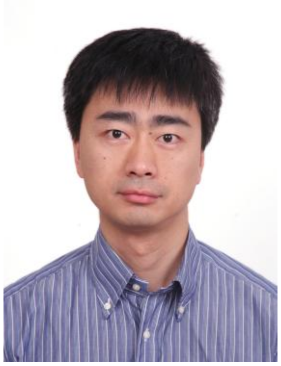
[3] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," in *International Journal of Graphic Tools*, vol. 9, no. 1, pp. 25-36, 2004.

[4] K. Müller et al., "View Synthesis for advanced 3D Video Systems,“ in *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 438148, 11 pages, 2008

[5] X. Jiufei et al., "A New Virtual View Rendering Method Based on Depth Image," in *Proc. APWCS*, Shenzhen, China, April 2010.

[6] T. G. Georgiev, "Covariant Derivates and Vision," in *Proc. Europ. Conf. on Comp. Vision*, Graz, Austria, 2006.

[7] A. Criminisi et al., "Region Filling and Object Removal by Exemplar-based Inpainting," in *IEEE Trans. on Image Proc.*, vol. 13, no. 9, pp. 1200-1212, January 2004.



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, Tsinghua University (THU), Beijing, China, in 2001. Currently, he is an Associate Professor in CST of THU, Beijing, China. His research interests include video streaming, 3D video processing and Social Media. He is a Member of IEEE.

Congestion Control for Low-Latency Video Streaming in Mobile Ad-Hoc Networks

A short review for "Low-Latency Video Streaming with Congestion Control in Mobile Ad-Hoc Networks"

Edited by Xiaoli Chu

C. Greco, M. Cagnazzo and B. Pesquet-Popescu, "Low-Latency Video Streaming with Congestion Control in Mobile Ad-Hoc Networks," IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 1337-1350, Aug. 2012.

Mobile ad-hoc networks (MANETs) are formed of mobile devices connected by wireless links, self-organised in a mesh topology [1]. This kind of setup offers desirable properties – flexibility, ease of deployment, robustness, etc. – that make MANETs appealing in environments without a pre-existing infrastructure.

One of the applications of major interest in ad-hoc networking is peer-to-peer video streaming. Due to the inherently lossy nature of MANETs, this challenge can be efficiently addressed using schemes based on Multiple Description Coding (MDC) [2].

In MDC, the source signal (e.g., an image, a video clip, or an audio signal) is encoded in independent streams, referred to as *descriptions*. Each description can be decoded independently from another, and each further description improves the quality of the reconstructed signal. When all descriptions are available, the highest quality is attained. The price of this beneficial property is paid in terms of a higher bit rate needed to encode a video sequence for a given quality.

Previous studies on content delivery over MANETs showed that structured protocols, i.e., protocols that construct and maintain an overlay, have a reduced delay and a better scalability if compared to unstructured protocols, but their performance degrades in more dynamic networks [3]. To overcome the limits of structured protocols in MANETs, the authors consider a cross-layer approach, which has been pointed out as a needed shift of perspective in protocol design for wireless transmission [4].

In this spirit, the authors proposed *A Broadcast Content Delivery Protocol* (ABCD) [5], a protocol inherently designed for the MDC ad-hoc streaming case, exploiting the natural broadcast property of the medium in a cross-layered fashion. ABCD performs well in terms of availability, robustness and scalability, and presents a low and stable latency.

In this work, the authors propose a congestion control framework for real-time multiple description video multicast over wireless ad-hoc networks based on the idea of congestion-distortion optimisation (CoDiO) – as opposed to the traditional rate distortion optimisation used in video streaming – which was introduced to model the effects that self-inflicted network congestion has on video quality [6]. The congestion control framework is designed for multi-tree overlays, consistent with the use of MDC, in mobile ad-hoc environments with high node density and stringent constraints on delay. Namely, we introduced a dynamic adjustment of the MAC layer transmission parameters, optimised w.r.t. a congestion-distortion criterion.

The model for congestion and distortion takes into account both the video stream coding structure as well as the unavoidable redundancy of the topology of the overlay network; it also provides the MAC layer with video-coding awareness, thus making it possible to perform an optimisation of congestion and distortion. In particular, for distortion estimation, the authors introduced an efficient way to propagate information about possible paths alternative to the multi-tree overlay, in order to classify the nodes in groups differently affected by the loss of a packet. This information is formed by refining precise but local information collected by the single nodes, then aggregating it as the information is propagated up-tree. The total distortion is then estimated by weighting the expected distortions of each group with the estimated number of receiving nodes in each group. This allows making a reliable prediction on the consequences of sending a packet with a particular retry limit, thus optimising video transmission in a CoDiO sense. For the congestion model, we consider the number of packets that cannot be sent while the channel is occupied by the currently transmitting node, times the duration of the transmission. This number of packets corresponds to the length of a virtual packet queue, distributed among the current node and its neighbours. This model differs from the assumption usually made in the literature that *all* the neighbours of a node are always willing to transmit at any time, and is based on the knowledge of the neighbours' state, namely, the number of packets they have to send.

This framework can be integrated into any tree-based video streaming protocol for MANETs to improve its performance; we show here that, if integrated into the ABCD protocol, both congestion and distortion are optimised in dense networks imposing a conversational pattern on the video stream, attaining a significant reduction of both average (over time and nodes) and maximum end-to-end delay, maintaining a delivery rate close to 100%. In terms of video quality, the nodes of the network decode with a high average PSNR, barely 0.24 dB below the central decoding quality. Most of them actually decode at central quality, and all of them at least 3 dB *above* the average side decoding quality.

The main challenge in this CoDiO framework is the distributed estimation of both the network topology, in order to capture the multiple paths that a video packet may follow, and the channel conditions, in order to estimate the effects on end-to-end delay. This information is propagated in an efficient and compact way through the network, leading to significant improvements in terms of both delay and objective video quality, as demonstrated by the simulations.

Acknowledgement:

The R-Letter Editorial Board thanks Dr. Ing. Carl James Debono for nominating this work.

References:

[1] M. Frodigh, P. Johansson and P. Larsson, "Wireless ad-hoc networking: the art of networking without a network," *Ericsson Review*, vol. 4, pp. 248-263, 2000.

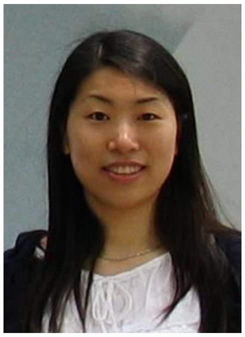
[2] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Proc. Mag.*, vol. 18, no. 5, pp. 74-93, 2001.

[3] D. N. da Hora, *et al*, "Enhancing peer-to-peer content discovery techniques over mobile ad-hoc networks," *Elsevier J. on Computer Commun.*, vol. 1, pp. 1-15, 2009.

[4] M. Van der Schaar and N. Sai Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 50-58, 2005.

[5] C. Greco and M. Cagnazzo, "A cross-layer protocol for cooperative content delivery over mobile ad-hoc networks," *Inderscience IJCNDS.*, vol. 7, no. 1-2, pp. 49-63, 2011.

[6] E. Setton and B. Girod, "Congestion-distortion optimized scheduling of video over a bottleneck link," in *Proc. IEEE Multimedia Sign. Proc.*, 2004.



Xiaoli Chu is a lecturer in the Dept. of Electronic and Electrical Engineering at the University of Sheffield. She received the B.Eng. degree from Xi'an Jiao Tong University in 2001 and the Ph.D. degree from Hong Kong University of Science and Technology in 2005. From Sep 2005 to Apr 2012, she was with the Centre for Telecom Research at King's College London. Her research interests include heterogeneous networks, interference management, cooperative communications, cognitive communications, and green radios.

An Energy-Efficient Solution to Transmit Scalable Video over Wireless Networks

A short review for "Energy-efficient resource allocation and scheduling for multicast of scalable video over wireless networks"

Edited by Carl James Debono

S.-P. Chuah, Z. Chen, and Y.-P. Tan, "Energy-efficient resource allocation and scheduling for multicast of scalable video over wireless networks," IEEE Transactions on Multimedia, vol. 14, no. 4, August 2012.

Recent advances in wireless radio technology and video coding techniques have enabled pervasive and ubiquitous access to multimedia contents via mobile devices. The different display sizes and technologies used in these devices demand transmission of scalable video for better visual quality. Moreover, limited bandwidths and resources in the wireless environment require a reduction in data transmission and energy-efficiency. Multicast is a possible solution since it needs less bandwidth compared to a number of unicasts. To allow its use, efficient resource allocation and scheduling are necessary for scalability.

Among the next generation wireless networking techniques, wireless mesh networks [1] are a popular research topic within the research community. Wireless mesh networks are characterized as self-organizing and self-configuring, whereby mesh routers in the network automatically establish an ad hoc network. These constitute the backbone of the mesh network by acting as a gateway to the clients which locate within mesh cells. The major advantage of wireless mesh networking is quick and dynamic deployment of network coverage. In addition, the cost of such network infrastructure is relatively low. However, the network service of mesh networks is often best-effort with limited connectivity. In portable routers, power availability is a limited resource.

On the other hand, the Scalable Video Coding (SVC) extension of H.264/AVC [2] has been recently finalized. This allows for partial transmission and decoding of a video bitstream. The decoded video has lower resolution, either spatial or temporal resolution, or reduced fidelity. SVC offers a good solution for video streaming with heterogeneous clients whose devices have capabilities with different resolutions. For video transmission over wireless channels, SVC enables robust transmission with graceful degradation in the presence of error and loss. SVC deploys the layered coding technique which results in unequal priority among video data. The robust features of SVC can only be fully demonstrated if a communication network is media-aware in its streaming adaptation.

In their paper, the authors considered the wireless multicast [3] of scalable video within a mesh cell. Heterogeneous clients subscribe to various video programs precoded in SVC with different resolutions. The broadcast nature of the wireless medium was exploited in delivering the video data to multiple receivers by establishing video multicast groups, to which the clients joined according to their subscriptions. For example, if two video programs are provided, each having two resolutions: CIF 30fps and QCIF 15fps respectively, there will be four multicast groups. The authors considered the video multicast service to be a best-effort service, constrained by a limited energy budget and air-time. Subject to these limited network resources, they optimize the multicast strategy of the mesh router such that the video qualities of the subscribing clients are maximized. The multicast strategy here includes the Modulation and Coding Scheme (MCS) and transmission power.

Optimizing the multicast strategy for each multicast group requires a reference wireless channel. However, wireless multicast channel features one-to-many where different clients experience different channel fading and loss. The authors select the worst channel condition among subscribing clients as reference to ensure video quality for all admitted clients. To determine the worst channel condition, the probing protocol proposed in [4] was adopted. To find this channel, the mesh router begins by sending a probing frame to all clients. Non-subscribing clients ignore the frame and do nothing, whereas subscribing clients initiate a random backoff processes before replying. The backoffs are randomly selected from the ranges depending on their channel Signal-to-Noise Ratio (SNR).

Lower channel SNR corresponds to a lower backoff range, and vice versa. As a result, clients with lower channel SNR will have shorter backoff period and the chance to reply earlier. Similar probing is done for all multicast groups.

The resource allocation problem uses the information from the worst channel conditions to maximize the video utility of clients experiencing these conditions subject to the limited energy budget and air-time. This is a mixed-integer programming problem whose efficient solution is not readily available. The authors demonstrated that the problem can be reduced to binary-integer programming problem which can be solved iteratively using the dynamic programming approach [5]. To determine the scheduling order among video layers, the quality impact is computed as the impact of a video layer with respect to the highest resolution in an SVC bitstream. This gives an indicator on how important a video frame is in the SVC structure. Video frames with higher impact are scheduled before others in the iterative way. The authors showed that the optimal solution can be obtained by searching for the solution candidate which consumes the least resources in each iteration.

The authors evaluate the performance of the proposed solution framework using the network simulator ns-2.33 and the SVC reference software JSVM 9.8. IEEE 802.11g was deployed as the wireless standard in the mesh cell. Their simulation results indicate notable improvement in average PSNR of clients under various levels of energy and air-time constraints. In particular, when the energy budget or air-time is severely limited, the proposed quality impact scheduling plays the crucial role in video quality gain. As the energy budget or air-time is increased, quality gains by the proposed resource allocation in high resolution videos are more significant since they need high bitrates and consume more resource to transmit. Furthermore, simulations confirm that the proposed methods work well for various circumstances, such as multicast of various video programs, video programs of various resolutions, and various numbers of subscribing clients. In addition to the video quality improvement, the method allows for fairness in ensuring video qualities among various video programs. The variance in the average PSNR is smaller than the reference.

Multicast together with SVC are a promising solution for improved visual quality for heterogeneous devices. However, further work in algorithms to control and maintain the fragile wireless links is needed to keep uninterrupted video transmission. Furthermore, cross-layer solutions can help in increasing the media-awareness within the network allowing for better use of resources.

Acknowledgement:

The R-Letter Editorial Board thanks Dr. Xiaoli Chu for nominating this work.

References:

[1] I. Akyildiz, and X. Wang, “A survey on wireless mesh networks,” *IEEE Communication Magazine*, vol. 43, pp. S23-S30, September 2005.

[2] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol*., vol. 17, no. 9, pp. 1103–1120, September 2007.

[3] H. Gossain, C. M. Cordeiro, and D. P. Agrawal, “Multicast: Wired to Wireless,” *IEEE Communication Magazine*, vol. 40, pp. 116-123, June 2002.

[4] J. Villalon, P. Cuenca, L. Orozco-Barbosa, Y. Seok, and T. Turletti, “Cross-layer architecture for adaptive video multicast streaming over multirate wireless LANs,” *IEEE J. Select Areas Communications*, vol. 25, no. 4, pp. 699–711, May 2007.

[5] D. P. Bertsekas, Dynamic Programming and Optimal Control. Belmont, MA: Athena Scientific, 2007.

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 a Senior Lecturer. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta.

Dr Debono is a senior member of the IEEE and served as chair of the IEEE Malta Section between 2007 and 2010. He is the IEEE Region 8 Conference Coordination sub-committee chair for 2011. 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.

Quality-Aware Video Streaming over Femtocell Networks

A short review for "On medium grain scalable video streaming over cognitive radio femtocellnetwork"

Edited by Pradeep K. Atrey

D. Hu and S. Mao, "On Medium Grain Scalable Video Streaming over Femtocell Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, vol. 30, vo. 3, April 2012.

A Cisco study has predicted that the global mobile data traffic will increase 26-fold between 2010 and 2015, and almost 66% of the mobile data will be video related. "Smart-phone revolution" is one of the main contributors to this [1]. Such drastic growth and change in the composition of wireless data have greatly stressed the capacity of traditional wireless access networks. Since the wireless networks use open space as transmission medium, they are usually limited by interference. This problem is aggravated when mobile users move away from the base station and a considerably larger transmit power is needed to overcome attenuation. This causes interference to other users and deteriorates network capacity.

It has been recognized that Femtocell will be one of the most important and effective technologies for off-loading wireless data, by bringing infrastructure closer to mobile users. A femtocell is a small cellular network, with a femto base station (FBS) connected to the owner's broadband wired network [2]. A critical challenge in the design and deployment of femtocell network is how to manage the interference between macrocells and the femtocell themselves. To this end, cognitive radios (CR) represent an effective solution by exploiting spectrum opportunities [3].

The objective is the work by Hu and Mao is to maximize the capacity of the femtocell CR network to carry real-time video data, while bounding the interference to primary users. This has several challenges such as maintaining the Quality of Service requirements of real-time videos [4] and dealing with the new dimensions of network dynamics (i.e., channel availability) and uncertainties (i.e., spectrum sensing and errors) found in CR networks [5]. Also, various design factors that necessitates cross-layer optimization, such as spectrum sensing and errors, dynamic spectrum access, interference modeling and primary user protection, channel allocation, and video performance also need to be considered [6, 7].

Hu and Mao have addressed the core problem of supporting video streaming in two-tier femtocell networks, while incorporating CR as an effective means for interference mitigation. Although the high potential of cognitive femtocells has been well recognized, video over cognitive femtocell networks has not been much studies so far and the Hu and Mao's work is among the first ones that addresses this important problem with a systematic solution, and provides a refreshing perspective on how robust multi-user video streaming can be achieved in cognitive femtocell networks.

In the proposed approach, the authors addressed the problem of resource allocation for medium grain scalable videos over femtocell CR networks. First the case of a single FBS is examined, and then the more general case of multiple non-interfering or interfering FBSs is considered. The algorithm for the single and non-interfering FBS case is distributed one and optimal, while the algorithm for the interfering FBS case is a centralized one that can be executed at the macro base station. In the case of a single FBS, the authors applied dual decomposition to develop a distributed algorithm that can compute the optimal solution. Furthermore, in the case of multiple non-interfering FBS's, the authors showed that the same distributed algorithm can be used to compute optimal solutions. For multiple interfering FBSs, a greedy algorithm that can compute near-optimal solutions is developed and a closed-form lower bound for its performance based on an interference graph model is provided. The key design issues and trade-offs have been captured with a multistage stochastic programming problem simulation. The performance evaluation is thorough and clearly demonstrates the superior performance of the proposed approach over two heuristic algorithms and the method proposed in [8]. Also, it has been shown that this approach is able to efficiently solve complex optimization problems involving cross-layer interactions.

In summary, the finding of the paper throws new light on the feasibility of cognitive femtocell networks in supporting quality-aware transmission of real-time videos.

Acknowledgement:

The R-Letter Editorial Board thanks Dr. Chonggang Wang, InterDigital Communcations USA, for nominating this work and for providing a summary of its contributions.

References:

[1] Cisco, “Visual Networking Index (VNI),” July 2010. [On-line]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/

[2] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, “Femtocell networks: a survey,” IEEE Commun. Mag., vol. 46, no. 9, pp. 59–67, Sept. 2008.

[3] S.-M. Cheng, S.-Y. Lien, F.-S. Chu, and K.-C. Chen, “On exploiting cognitive radio to mitigate interference in macro/femto heterogeneous networks,” IEEE Wireless Commun., vol. 18, no. 3, pp. 40–47, 2011.

[4] Z. Wang, L. Lu, and A. C. Bovik, “Video quality assessment using structural distortion measurement,” Signal Processing: Image Commun., no. 2, pp. 121–132, Feb. 2004.

[5] R. Kim, J. S. Kwak, and K. Etemad, “WiMAX femtocell: requirements, challenges, and solutions,” IEEE Commun. Mag., vol. 47, no. 9, pp. 84–91, Sept. 2009.

[6] I. Gven, S. Saunders, O. Oyman, H. Claussen, and A. Gatherer, “Femtocell networks,” EURASIP J. Wireless Comm. and Networking, 2010, article ID 367878, 2 pages, doi:10.1155/2010/367878.

[7] F. Kelly, A. Maulloo, and D. Tan, “Rate control in communication networks: shadow prices, proportional fairness and stability,” J. Oper. Res. Soc., vol. 49, no. 3, pp. 237–252, Mar. 1998.

[8] A. Hsu, D. Wei, and C. Kuo, “A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks,” in Proc. IEEE WCNC’07, Hong Kong, P.R. China, Mar. 2007, pp. 105–110.

Pradeep K. Atrey is an Associate Professor at the University of Winnipeg, Canada. He received his Ph.D. in Computer Science from the National University of Singapore, M.S. in Software Systems and B.Tech. in Computer Science and Engineering from India. He was a Postdoctoral Researcher at the Multimedia Communications Research Laboratory, University of Ottawa, Canada. His current research interests are in the area of Multimedia Computing with a focus on Multimedia Surveillance and Privacy, Image/Video Security, and Social Media. He has authored/co-authored over 70 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 Computing, Communications and Applications, ETRI Journal and IEEE Communications Society MMTC R-letters. He has been associated with over 25 international conferences in various roles such as General Chair, Program Chair, Publicity Chair, Web Chair, and TPC Member. Dr. Atrey was a recipient of the ETRI Journal Best Reviewer Award (2009) and the University of Winnipeg Merit Award for Exceptional Performance (2010). He was also recognized as “ICME 2011 - Quality Reviewer”.

Towards More Accurate Search and Retrieval of 3D Objects

A short review for "Investigating the Effects of Multiple Factors towards More Accurate 3-D Object Retrieval"

Edited by Jun Zhou

Petros Daras, Apostolos Axenopoulos, Georgios Litos, "Investigating the Effects of Multiple Factors towards More Accurate 3-D Object Retrieval", IEEE Transactions on Multimedia, vol. 14, no. 2, pages 374-388 , 2012.

The widespread availability of 3D models has intensified the need for effective 3D contents-based search through various online media databases. 3D object retrieval methods automatically extract low-level features (e.g. shape) from 3D objects in order to retrieve semantically similar objects. Starting from simple, heuristic approaches that detect a few generic geometric features in the surface or the volume of a 3D model, 3D content-based search has evolved over the years to include highly complex algorithms that apply sophisticated mathematics so as to detect fine discriminative details and achieve the highest possible accuracy. Recently though, it has become apparent that further research towards investigating even more complex algorithms can hardly improve the performance of the state-of-the-art methods. Therefore, much effort has been put into combining existing 3D shape descriptors into one unified similarity matching framework, which has been proven to be much more effective than using each descriptor separately [1]. This paper investigates all those factors that may affect the accuracy of 3D object retrieval and proposes potential research directions in this field.

The first step in a typical 3D object retrieval procedure is preprocessing, in which the 3D model is translated and scaled in order to lie within a bounding sphere of radius 1. Rotation normalization is also desired when a rotation-dependent descriptor extraction method is used. The most popular solution for rotation estimation relies on Principal Component Analysis (PCA), while several improvements have been introduced, such as Continuous PCA (CPCA) [2] and Rectilinearity-based rotation estimation [3]. In this paper, a Combined Pose Estimation method (CPE) is introduced, which combines the features of both CPCA and Rectilinearity to achieve more accurate rotation normalization.

The second step involves feature extraction, where low-level descriptors are extracted from the 3D object and uniquely represent its shape. Existing 3D object descriptor extraction methods can be classified into four main categories [4]: histogram-based, transform-based, graph-based, view-based and, finally, combinations of the above. According to the results of the SHREC 2011 contest [1], 2D view-based methods have achieved the highest performance. In this paper, descriptors are extracted using three different methods: the Compact Multi-View Descriptor (CMVD) [4], the Spherical Trace Transform (STT) [6] and the Depth Silhouette Radical Extent descriptor (DSR) [8], which are view-based, transform-based and a combination of both view- and transform-based methods respectively. The combination of the above methods produces a descriptor vector of high dimensionality. As the descriptor size increases, search and retrieval in very large databases become prohibitive, since similarity matching of large descriptor vectors requires high response time. In this paper, this problem has been overcome using feature selection, which has been widely used in pattern analysis in order to select the most significant features of a given descriptor. More specifically, Correlation-based Feature Selection (CFS) [5] has been selected to reduce the dimensionality of the descriptor vector and improve the retrieval accuracy of the descriptors.

Combination of different descriptors is usually expressed as a weighted sum of their individual dissimilarities. This involves the following actions: i) find the best dissimilarity metric for each descriptor and ii) determine the optimal weights for the weighted sum. In this work, the optimal dissimilarity metrics for each selected descriptor have been experimentally found to be the Squared L-1 distance for CMVD, the X-distance for STT and the X^2-distance for DSR. The weights that determine the contribution of each descriptor are optimized using Particle Swarm Optimization.

Moreover, manifold learning has been adopted to improve the performance of the proposed 3D object retrieval method. The use of manifold learning is based on the concept that low-level descriptors, in general, follow a nonlinear manifold structure, which makes classical Euclidean metrics inappropriate. By properly unfolding this manifold structure, a more representative feature space of lower dimension is achieved. In the proposed framework, a manifold learning technique based on Laplacian Eigenmaps [7] is applied for non-linear dimensionality reduction.

Several experiments have been conducted to measure the impact of all factors described above in 3D retrieval accuracy, from which it is proven that all of them should be taken into account in order to produce an accurate descriptor. In the SHREC 2011 contest [1], the proposed 3D retrieval framework was ranked first among the most well-known 3D retrieval techniques. This work is expected to provide a useful reference for further research.

Acknowledgement:

The R-Letter Editorial Board thanks Dr. Petros Daras for helping the preparation and Dr. guillanume Lavoue for the nomination of this article.

References:

[1] H. Dutagaci, A. Godil, P. Daras, A. Axenopoulos, G. Litos, S. Manolopoulou, K. Goto, T. Yanagimachi, Y. Kurita, S. Kawamura, T. Furuya, R. Ohbuchi, B. Gong, J. Liu, X. Tang, "SHREC'11 Track: Generic Shape Retrieval", *Proceedings of the 4th Eurographics Workshop on 3D Object Retrieval (3DOR 2011)*, pp. 65-69, 2011.

[2] D. V. Vranic, D. Saupe and J. Richter, "Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics". *In Proceedings of the IEEE fourth workshop on multimedia signal processing*, pp. 293–298, 2001.

[3] Z. Lian, P. L. Rosin, and X. Sun, "Rectilinearity of 3D meshes", *International Journal of Computer Vision* Volume 89, Numbers 2-3, 130-151, 2010.

[4] P. Daras, A. Axenopoulos, "A compact multi-view descriptor for 3D object retrieval" *IEEE 7th International Workshop on Content-Based Multimedia Indexing* (CBMI), pp. 115-119, 2009

[5] M.A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning", In *Proceedings of the 7th International Conference on Machine Learning*, pp. 359–366, 2000.

[6] D. Zarpalas, P. Daras, A. Axenopoulos, D. Tzovaras and M. G. Strintzis, "3D model search and retrieval using the spherical trace transform," *EURASIP Jouornal of Applied Signal Process*., vol. 2007, no. 1, p. 207, 2007.

[7] R. Ohbuchi, J. Kobayashi, "Unsupervised Learning from a Corpus for Shape-Based 3D Model Retrieval", *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (MIR)*, pp. 163-172, 2006.

[8] D. Vranic, DESIRE: a composite 3D-shape descriptor, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) 2005, pp. 962–965.

Jun Zhou received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Canada, in 2002, and the Ph.D. degree in computing science from University of Alberta, Canada, in 2006.

He joined the School of Information and Communication Technology in Griffith University as a lecturer in June 2012. Prior to this appointment, he had been a research fellow in the Australian National University, and a researcher at NICTA. His research interests are in statistical pattern recognition, interactive computer vision, and their applications to hyperspectral imaging and environmental informatics.

Speech Driven Animation of Complex 3D Head Models

A short review for "Animation of Generic 3D Head Models driven by Speech"

Edited by Guillaume Lavoué

L. Terissi, M. Cerda, J.C Gomez, N. Hitschfeld-Kahler, B. Girau, and R. Valenzuela. "Animation of Generic 3D Head Models driven by Speech", in Proceedings of IEEE International Conference on Multimedia & Expo (ICME), pp. 1-6, Barcelona, Spain, 2011.

Animation of virtual characters is playing an increasingly important role due to the widespread use of multimedia applications such as computer games, online virtual characters, video telephony, and other interactive human-machine interfaces. Several approaches for facial animation have been proposed in the literature [1], where the animation can be data-driven (e.g. by video, speech or text data), manually controlled, or a combination of both. Usually, the animation techniques require a tedious and time consuming preparation of the head model to be animated, in order to have control of the animation by a reduced set of parameters.

For the particular case of speech-driven facial animation, the animation is controlled by a set of visual features estimated from speech, by means of a trained audio-visual model. Among the main approaches proposed for computing the facial movements from a given speech input, the ones based on rules, Vector Quantization, Neural Networks, Nearest Neighbor, Gaussian Mixture Models, and Hidden Markov Models, can be mentioned. Regardless of the method employed to model the audio-visual data and to compute facial movements from speech, the different approaches usually divide the training audio-visual data into classes. This clustering is performed taking into account the acoustic data, the visual data, or a combination of both. For each cluster, an audio-visual model is trained using the corresponding audio-visual data. The first step in the synthesis stage is the computation of the class that the novel audio data belongs to. The trained audio-visual model associated with the selected class is then used to estimate the corresponding visual parameters. In these approaches, the performance of the whole system is directly affected by errors derived by these classification procedures.

A complete pipeline for speech-driven animation of generic 3D head models is proposed in this paper. The proposed approach attempts to simplify the annoying and time-consuming preparation of the head model to be animated, and to avoid the prior segmentation or classification of the audio-visual data, preventing in this way the errors derived from these classification procedures. The proposed system allows the animation of complex head models in a simple way, and since it makes use of raw speech data (no phoneme (word) segmentation is required), it is adaptable to any language.

This speech-driven animation system can be divided in two main processing blocks, one related to the audio-visual modeling, and the other one related to the animation process.

In the audio-visual model training stage, audio-visual information is extracted from videos of a real person, and then used to train a single Audio-Visual Hidden Markov Model (AV-HMM). Visual features are represented in terms of a parameterized simple 3D face model, namely Candide-3 [2], which can be adapted to different real faces, and allows the representation of facial movements with a small number of parameters. In particular, the method proposed in [3] is used to extract visual features related to mouth movements during speech. As already mentioned, the proposed training method does not require prior classification of the audio-visual data. This has the advantage of avoiding the errors introduced by these classification procedures. Given the trained AV-HMM and a novel speech signal, an extension of the method described in [4], based on Hidden Markov Model Inversion (HMMI), is proposed to estimate the associated visual features.

The animation block takes as input the estimated visual features to produce the animation of arbitrary complex head models. This animation procedure is based on the fact that the visual features are the animation parameters of Candide-3 model, i.e., these parameters directly animate the Candide-3 model. The animation of the complex head model is obtained by mapping and interpolating the movements of the simple model to it. To perform this task, a semi-automatic registration procedure is presented in this paper. Using a graphical interface, the user has to place both models in approximately the same position and orientation, and to adapt the anatomy of the Candide-3 to the target head model using 10 sliders. This adaptation procedure takes only a couple of minutes and it is sufficient to check head size/shape and lips/nose correspondences. Once the complex head model is adapted, it can be animated using the input visual features. Although this work is focused on speech-driven facial animation, it must be noted that the proposed animation method allows the animation of complex head models from visual features extracted from videos, i.e., it can be also used for video-driven facial animation.

The animations generated by the proposed system were perceptually evaluated in terms of intelligibility of visual speech through perceptual tests, carried out with a group of 20 individuals. The perceptual quality of the animation proved to be satisfactory, showing that the visual information provided by the animated avatar improves the recognition of speech in noisy environments.

As a conclusion, this paper presents a complete system for speech-driven animation of 3D head models. Its performance has been assessed by a very nice subjective experiment, which has raised the importance of the avatar animation in the recognition of speech by human observers.

Future works include the incorporation of new audio features such as prosody information, and additional visual features related to the movements of other regions of the face, such as eyebrows and cheeks.

Acknowledgement:

The R-Letter Editorial Board thanks the authors for assisting the preparation of this article.

References

[1] N. Ersotelos and F. Dong, “Building highly realistic facial modeling and animation: a survey,” Visual Computer, vol. 28, pp. 13–30, 2008.

[2] J. Ahlberg, “An updated parameterized face,” Technical Report LiTH-ISY-R-2326, Department of Electrical Engineering, Linkoping University, Sweden, 2001.

[3] L. D. Terissi and J. C. Gomez, “3D head pose and facial expression tracking using a single camera,” Journal of Universal Computer Science, vol. 16, no. 6, pp. 903–920, 2010.

[4] K. Choi, Y. Luo, and J. Hwang, “Hidden Markov Model inversion for audio-to-visual conversion in an MPEG-4 facial animation system,” Journal of VLSI Signal Processing, vol. 29, no. 1-2, pp. 51–61, 2001.

Guillaume Lavoué received his engineering degree in signal processing and computer science from CPE Lyon (2002), his M.Sc. degree in image processing from the University Jean Monnet, St.-Etienne (2002), and his Ph.D. degree in computer science from the University Claude Bernard, Lyon, France (2005). Since September 2006 he is associate professor at the French engineering university INSA of Lyon, in the LIRIS Laboratory (UMR 5205 CNRS).

He is author or co-author of over 50 publications in international journals and conferences. he is member of the IEEE Technical Committee on Human Perception in Vision, Graphics and Multimedia (SMC society), key member of the 3DRPC Interest Group of the IEEE Multimedia Communication Technical Committee (ComSoc society) as well as associate editor of ISRN Computer Graphics journal. His research interests include 3D mesh analysis and retrieval, 3D data transmission and streaming (including compression and watermarking), Web 3D, Perception and Human factors for computer graphics.

Paper Nomination Policy

Following the direction of MMTC, the R-Letter platform aims at providing research exchange, which includes examining systems, applications, services and techniques where multiple media are used to deliver results. Multimedia include, but are not restricted to, voice, video, image, music, data and executable code. The scope covers not only the underlying networking systems, but also visual, gesture, signal and other aspects of communication.

Any HIGH QUALITY paper published in Communications Society journals/magazine, MMTC sponsored conferences, IEEE proceedings or other distinguished journals/conferences, within the last two years is eligible for nomination.

Nomination Procedure

Paper nominations have to be emailed to R-Letter Editorial Board Directors:

Irene Cheng locheng@ualberta.ca ,
Weiyi Zhang maxzhang@research.att.com and
Christian Timmerer
christian.timmerer@itec.uni-klu.ac.at

The nomination should include the complete reference of the paper, author information, a brief supporting statement (maximum one page) highlighting the contribution, the nominator information, and an electronic copy of the paper when possible.

Review Process

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

R-Letter Best Paper Award

Accepted papers in the R-Letter are eligible for the Best Paper Award competition if they meet the election criteria (set by the MMTC Award Board).

For more details, please refer to http://committees.comsoc.org/mmc/rletters.asp

MMTC R-Letter Editorial Board

Multimedia Communications Technical Committee (MMTC) Officers

MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.