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

R-LETTER

Vol. 5, No. 6, December 2014

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from the Review Board Directors

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

We hope you that this issue stimulates your research in the area of multimedia communication and an overview of all reviews are provided in the following:

The **first paper**, published in the *IEEE Transactions on Wireless Communication* and *edited by Koichi Adachi*, deals with the potential gain of multiuser multiple input multiple output (MIMO) scheduling for mobile video applications.

The **second paper**, published in the *IEEE Transactions on Multimedia* and *edited by Carsten Griwodz*, presents an improved method for adding depth to 2D images and movies.

The **third paper** is *edited by Jun Zhou* and has been published within Proceedings of the European Conference on Computer Vision. It highlights means to reconstruct the world across time.

Finally, the **forth paper**, published in the *IEEE Transaction on Image Processing* and *edited by Bruno Macchiavello*, describes an efficient feature descriptors encoding for mobile augmented reality.

We would like to thank all the review board members for their time and efforts. In particular, we would like to thank Vladan Velisavljevic who resigned.

IEEE ComSoc MMTC R-Letter
Director: Christian Timmerer
Alpen-Adria-Universität Klagenfurt, Austria
Email: christian.timmerer@itec.aau.at
Co-Director: Weiyi Zhang
AT&T Research, USA
Email: wzhang@ieee.org
Co-Director: Yan Zhang, Simula, Norway
Email: yanzhang@simula.no

The Potential Gain of Multiuser MIMO for Mobile Video Applications

A short review for "Multiuser MIMO Scheduling for Mobile Video Applications"
(Edited by Koichi Adachi)

W. Ni, R. P. Liu, J. Biswas, X. Wang, I. B. Collings, and S. K. Jha, "Multiuser MIMO Scheduling for Mobile Video Applications", IEEE Trans. Wireless Commun., vol. 13, no 10, pp. 5382-5395, Oct. 2014.

A relentless growth in the demand for wireless data transmission is expected due to bandwidth-demanding mobile video applications. The increase in the demand for wireless data traffic is predicted to be 18 times between 2011 and 2016 [1].

To meet such a huge increase of data traffic, multiuser multiple-input multiple-output (MIMO) is one of the potential physical layer technologies [2]. In downlink multiuser MIMO, simultaneous data transmission to multiple users is enables by utilizing beamforming at the base station (BS) [3]. When a large number of users are present, the BS schedules its transmission to users that have favorable channel conditions, exploiting multiuser diversity which is a form of selection diversity and improves throughput. This physical layer technology greatly enhances the system throughput.

Shifting one's attention to mobile video applications, it has very distinct Quality of Service (QoS) requirements. To take into account such QoS requirements, several application layer mechanisms adjust the video qualities in order to adapt to the wireless channel fluctuation. For example, each video stream is encoded into a necessary base layer and a number of supplementary enhancement layers at the video application server (e.g., scalable video coding, SVC) [4]. Although multiuser MIMO can enhance the physical layer system throughput, most existing multiuser MIMO scheduling methods do not consider aforementioned QoS requirements holistically [5]-[10]. In other previous works, authors have attempted to address some QoS aspects of multiuser MIMO; nevertheless, they are unable to take advantage of the flexibility of multi-layer streaming videos [11]-[14]. Therefore, even though multiuser MIMO has been applied to system with mobile video application, the potential gain provided by multiuser MIMO has not been fulfilled.

To overcome the above issues, in this paper, a new cross-layer multiuser MIMO scheduling algorithm is proposed, which leverages the two key factors of the overall system throughput and the quality of individual video applications.

Particularly, the quality is stabilized with dynamically balanced wireless transmission data rates in the base and enhancement layers. The key idea is to select multiuser MIMO users based on the priority carefully assigned to indicate the achievable wireless data rates of the users, the arrival rates of individual video flows, as well as the queue status of the BS. Another important aspect is that the authors propose a new parallel technique to precisely calculate the data rates and subsequently the priorities of the individual users with low complexity, which is crucial in a practical sense. The data rates of the users are necessary for the video server to predetermine the video qualities (i.e., the balanced wireless transmission data rates in the base and enhancement layers) that are stable in the long term. To enable the analytical prediction of the data rates of the users, closed-form expressions are derived. The accuracy of the derived expressions is validated by conducting simulations.

In the proposed scheduling algorithm, namely downlink user/bearer and queue scheduling (DUBQS) algorithm, user selection and packets scheduling are performed based on both the queue states and channel information. DUBQS algorithm takes into account not only the channel correlation between the selected users but the intrinsic characteristics of video applications. The priority of the bearers and queues is carefully designed to indicate the key aspects of mobile video traffic, i.e., throughput, packet loss, as well as the ratio between the packets of the two layers (i.e., the base and enhancement layers). The authors first design the priority of bearers by introducing the bearer weight of each bearer. The purpose of the bearer weight is to maintain timely delivery of the critical base layer bitstreams, and avoid the MAC queues building up and even overflowing at the BS. The proposed problem formulation enables the users whose

effective channel gains are high or whose MAC queues start to build up to be selected. Then, the authors design the priority of queues within each bearer, which is referred to as dynamically balanced queuing (DBQ). The proposed design can balance the transmit rates of the base and enhancement layers at any instant. This is critical to the video quality. Predicting the data rate of multiuser MIMO is crucial for the video server to predetermine the encoding parameters (e.g., setting up the ratio of the base and enhancement layers) to support sustained QoS. The proposed algorithm can also analytically predetermine the ratio between the base and enhancement layers, which is the key to stabilize the video quality.

The authors conduct simulation campaign to evaluate the effectiveness of the proposed algorithm. Simulation results show that the proposed algorithm allows video applications to achieve almost the throughput upper bound of multiuser MIMO systems. The proposed approach also improves the video quality by reducing the loss of enhancement packets by an order of magnitude and reducing the delay by 35%, compared to the prior art.

By jointly considering the physical layer transmission rate and MAC layer queuing, the proposed algorithm succeeded in fulfilling the potential gain of multiuser MIMO for mobile video transmission.

References:

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016 2012. [Online]. Available: http://cisco.com

[2] J. Duplicy et al., "MU-MIMO in LTE systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2011, no. 1, pp. 1–13, 2011.

[3] W. Ni, Z. Chen, H. Suzuki, and I. B. Collings, "On the performance of semi-orthogonal user selection with limited feedback," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1359–1361, Dec. 2011.

[4] X. Zhu et al., "Layered Internet video adaptation (LIVA): Network-assisted bandwidth sharing and transient loss protection for video streaming," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 720–732, Aug. 2011.

[5] T. Yoo and A. J. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[6] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[7] A. Razi, D. J. Ryan, I. B. Collings, and J. Yuan, "Sum rates, rate allocation, and user scheduling for multiuser MIMO vector perturbation precoding," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 356–365, Jan. 2010.

[8] E. Conte, S. Tomasin, and N. Benvenuto, "A comparison of scheduling strategies for MIMO broadcast channel with limited feedback on OFDM systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, pp. 1–12, 2010.

[9] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[10] S. Lee, I. Pefkianakis, S. Choudhury, S. Xu, and S. Lu, "Exploiting spatial, frequency, and multiuser diversity in 3GPP LTE cellular networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 11, pp. 1652–1665, Nov. 2011.

[11] H. Shirani-Mehr, G. Caire, and M. J. Neely, "MIMO downlink scheduling with non-perfect channel state knowledge," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2055–2066, Jul. 2010.

[12] M. Torabzadeh andW. Ajib, "Packet scheduling and fairness for multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1330–1340, Mar. 2010.

[13] C. Wang and R. D. Murch, "Optimal downlink multiuser MIMO crosslayer scheduling using HOL packet waiting time," *IEEE Trans. Wireless Commun.*, vol. 5, no. 10, pp. 2856–2862, Oct. 2006.

[14] X. Zhang and J. Lee, "Low complexity MIMO scheduling with channel decomposition using capacity upper bound," *IEEE Trans. Commun.*, vol. 56, no. 6, pp. 871–876, Jun. 2008.

Koichi ADACHI received the B.E., M.E., and Ph.D degrees in engineering from Keio University, Japan, in 2005, 2007, and 2009 respectively. From 2007 to 2010, he was a Japan Society for the Promotion of Science (JSPS) research fellow. Since 2010, he has been with the Institute for Infocomm Research, A*STAR, in Singapore. His research interests include cooperative communications and energy efficient communication technologies. He was the visiting researcher at City University of Hong Kong in April 2009 and the visiting research fellow at University of Kent from June to Aug 2009.

Dr. Adachi served as General Co-chair of the 10th and 11th IEEE Vehicular Technology Society Asia Pacific Wireless Communications Symposium (APWCS) and Track Co-chair of Transmission Technologies and Communication Theory of the 78th and 80th IEEE Vehicular Technology Conference in 2013 and 2014, respectively. He was recognized as the Exemplary Reviewer from IEEE Communication Letters in 2012 and IEEE Wireless Communication Letters in 2012 and 2013. He was awarded excellent editor award from IEEE ComSoc MMTC in 2013.

An Improved Method for Adding Depth to 2D Images and Movies

A short review for "Robust Semi-Automatic Depth Map Generation in Unconstrained Images and Video Sequences for 2D to Stereoscopic 3D Conversion"
(Edited by Carsten Griwodz)

Phan, R.; Androutsos, D., "Robust Semi-Automatic Depth Map Generation in Unconstrained Images and Video Sequences for 2D to Stereoscopic 3D Conversion," IEEE Transaction on Multimedia, vol. 16, no. 1, pp 122-136, Jan. 2014.

Filmmakers in recent times have the opportunity to record their films in 3D, budget permitting. For a large amount of pre-existing content, or content created within device and budget constraints, this option does not exist and it is kept only as flat, 2D videos and images. And although the latest hype of 3D movies has passed, audiences have increasingly access to 3D projection devices and it is commercially viable to satisfy audiences' request for 3D content. Since depth encoding is not always available because the content has never been recorded in 3D, post-processing is required to add depth information to existing 2D content.

The process that is currently at the state-of-the-market for adding depth information to 2D content is called rotoscoping. It is a very labor-intensive manual process that requires human intervention for every frame, and this is where the paper by Phan and Androutsos starts. In contrast to other papers that constitute the state-of-the-art [1]–[3], and which rely on the fully automatic addition of depth information, the authors of this paper aim at a half-automatic approach that would be commercially viable at this time. Other research work precedes their papers ([4],[5]), and the important contribution of the paper lies in the authors understanding that they can approach the goal as an image segmentation problem rather than an object tracking problem.

The authors provide a concise overview of the state-of-the-art; mentioning motion-based techniques [6], scene understanding-based techniques [7], and finally, techniques that provide guesses for depths based on image retrieval [8] or neighboring frames [9].

Phan and Androutsos's semi-automatic approach is more conservative, probably slower, but because of an increased accuracy of the resulting segmentation, also the resulting depth map is more easily applicable in industrial solutions.

Their approach starts out like the commercially available rotoscopy. A user labels every object in a frame as either foreground or background. In contrast to the classical approach that requires pixel-accurate masking, the user is already supported in this step by an object segmentation approach

Phan and Androutsos express this in there paper by stating that human depth labeling doesn't have to be accurate, just consistent. This is consistent with 3D movie delivery, which keeps depth variation below realism anyway. The second important detail that the authors point out is that strict object segmentation and flat depth labeling leads to cardboard cutouts with fixed depth values assigned to objects. While this is sometimes the correct decision, it tends to look unrealistic for most real-world objects, and smoothed depth labels across edge are actually more realistic representations of the real world.

With these conditions in mind, they combine a series of four major steps:

- Let users create a depth prior by labeling arbitrary elements of a keyframe with a grey value or a color that represents depth.
- A graph-cut algorithm determines for every depth label K from a limited set of labels whether a pixel belongs to it or not. This likelihood is based on color space separation between a foreground and a background color, and separates along the most likely edge that maximizes distinction between a fair of colors. Repeating for every K, this provides a limited set of regions of identical depth value.
- Represent all pixels of the image as a labeled grid of points containing both depth value and the image's edge values as relative color differences between neighboring pixels. Let a Random Walk process assign depth values to all unlabeled point in up to K possible labels such that the most likely label is added to all points without user-defined labels. This assigned optimal label can connect different depth labels using fractions of original weights relative to the weight of each edge. This is repeated at multiple resolutions and subsequently merged. The resulting depth map is a blurry continuum of values.
- Finally, they modify the random walk approach by adding the weight acquired in the graph-cut step to the labels used in step 3. A heuristic

weighting factor is introduced to moderate the random walk probabilities with the edge probabilities of the graph cut. Executing the random walk process under this condition leads to considerably more realistic depth labels than each of the algorithms by themselves.

After such a creation of the keyframe labeling, Phan and Androutsos turn to extending their approach to provide depth labels for images between the keyframes. They track objects across frames and are able to assign labels to them, but refer to ongoing work for tracking depth that requires adjustment.

The paper is most noteworthy for its ability to assign realistic depth to keyframe object from a very small number of manually provided labels. The approach yield very promising results for realistic scenes and doesn't suffer from the cardboard cutout effect seen in clearly identified depth labels. An extension to sequences of frames is proposed in the remainder of the paper, but by admission of the authors, there is more work to be done. The paper should be read for its practical applicability on keyframe labeling.

References:

[1] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Transactions on Graphics*, vol. 29. p. 1, 2010.

[2] T. Yan, R. W. H. Lau, Y. Xu, and L. Huang, "Depth Mapping for Stereoscopic Videos," *Int. J. Comput. Vis.*, vol. 102, no. 1–3, pp. 293–307, 2013.

[3] C. Fehn and R. S. Pastoor, "Interactive 3-DTV-Concepts and Key Technologies," *Proc. IEEE*, vol. 94, 2006.

[4] M. Guttmann, L. Wolf, and D. Cohen-Or, "Semi-automatic stereo extraction from video footage," in *Proceedings of the IEEE International Conference on Computer Vision*, 2009, pp. 136–142.

[5] O. Wang, M. Lang, M. Frei, a Hornung, a Smolic, and M. Gross, "StereoBrush : Interactive 2D to 3D Conversion Using Discontinuous Warps," *Symp. Sketch*, vol. 1, pp. 47–54, 2011.

[6] Y. Chen, R. Zhang, and M. Karczewicz, "Low-complexity 2D to 3D video conversion," *Proc. SPIE*, vol. 7863. p. 78631I–78631I–9, 2011.

[7] T. Y. Kuo, Y. C. Lo, and C. C. Lin, "2D-to-3D conversion for single-view image based on camera projection model and dark channel model," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2012, pp. 1433–1436.

[8] J. Konrad, M. Wang, and P. Ishwar, "2D-to-3D image conversion by learning depth from examples," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 16–22.

[9] M. Park, J. Luo, A. Gallagher, and M. Rabbani, "Learning to Produce 3D Media from a Captured 2D Video," in *Proceedings of the 19th ACM International Conference on Multimedia*, 2011, pp. 1557–1560.

Carsten Griwodz is senior researcher 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 a 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

Reconstruct the World Across Time

A short review for "Scene Chronology"
(Edited by Jun Zhou)

*Kevin Matzen and Noah Snavely. "*Scene Chronology*", Proceedings of the European Conference on Computer Vision, pp. 615-630, 2014.*

With enormous amount of images available on the Internet, 3D scene modeling and visualization methods have been intensively studied and developed, which lead to successful practice of commercialization. From early work of automatic image stitching [1] and photo tourism [2] to the evolving product of Google Street View [3], approaches such as structure from motion [4] have been used to reconstruct scenes, streets, cities, and even larger scales [5].

Nonetheless, existing efforts have been mainly put on reconstruct static scenes, while the temporal aspects embedded in the collected images have seldom been addressed. It would be interesting to see how a scene evolves over time. To make this become true, authors of this paper proposed to bring temporal components back into the reconstruction process, so that the reconstructed scenes can be segmented into time spanned regions. Such spatial-temporal modeling enables analysis of areas with changes of different momentum, such as road signs, display boards, and street arts, and can be used to correct wrong time-stamp of photos.

This method is based on two assumptions. First, it requires that an initial 3D model can be reliably reconstructed using existing structure from motion approach. Then there is the assumption that both static and dynamic components shall be in the scene and can be observed in the collected images. The static components provide geometry information of the scene, which are the key for 3D reconstruction. The dynamic data are more related to the scene appearance, and shall remain static for certain period of time before changes happen.

The proposed spatial-temporal reconstruction approach has several steps. The first step is dense 3D point cloud generation which follows the standard routine of structure from motion and multi-view stereo. The main contribution of this paper comes from the following steps which focus on reasoning the time interval of each 3D point, and segment the scene into spatial-temporally coherent point sets. A major concern here is that these steps shall be able to cope with large amount of images captured from the Internet.

Given a dense set of 3D patches, estimation of the visibility of each patch in images is performed. To do so, a visibility graph is constructed to define positive edges for which a patch shall be observed in a view, or negative edge for which a patch is not observed in a view. The latter defines the period that this patch is occluded or does not exist at all. The next step is time interval estimation. Original image time stamps have provided the very first information for this estimation. However, due to the errors in the time stamps, inconsistent time intervals have hindered the accurate time interval estimation. To this end, time interval selection is treated as a classification problem with F-score calculated for the optimal span selection.

Finally, the scene is segmented into spatial-temporal consistent point sets. This is done by a modified RANSAC algorithm which only considers points that meet spatial, temporal, and common view requirements. The reduction of computation cost makes this method more suitable to dealing with large-scale applications than a couple of models in the literature [6].

In this paper, authors showed two applications of the proposed method. The first application is viewing and rendering scenes across time. The second one is photo time stamping correction. Experiments were performed on three online datasets that contains large amount of images. Some interesting results are reported which show that the proposed method is quite effective in detecting time stamp errors. Authors also discussed the limitations of this approach. These are mainly due to the violation of assumptions, such

as mis-registration of images and structural changes of buildings in a scene.

References:

[1] M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, Vol. 74, No. 1, pp. 59-73, 2007.

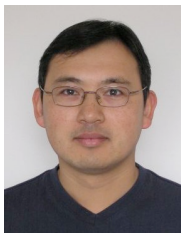
[2] N. Snavely, S. Seitz, and R. Szeliski. "Modeling the World from Internet Photo Collections". *International Journal of Computer Vision,* Vol. 80, No. 2, pp. 189-210, 2008.

[3] B. Klingner, D. Martin, and J. Roseborough "Street view motion-from-structure-from-motion," *Proceedings of the International Conference on Computer Vision*, pp. 950-960, 2013.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[5] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. Seitz, and R. Szeliski. "Building Rome in a Day", Communications of the ACM, Vol. 54, No. 10, pp. 105-112, 2011.

[6] G. Schindler and F. Dellaert. "Probabilistic temporal inference on reconstructed 3D scenes," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1410-1417, 2010.



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.

Efficient Feature Descriptors Encoding for Mobile Augmented Reality

A short review for "Interframe Coding of Feature Descriptors for Mobile Augmented Reality"
(Edited by Bruno Macchiavello)

Makar, M.; Chandrasekhar, V.; Tsai, S.S.; Chen, D.; Girod, B., "Interframe Coding of Feature Descriptors for Mobile Augmented Reality," IEEE Transaction on Image Processing, vol.23, no.8, pp.3352,3367, Aug. 2014.

Mobile augmented reality is a relatively new technology. The goal is to combine the advantages of mobile computing, i.e. human-computer interaction irrespective of user movement, with augmented reality (AR). An AR system aims to superimpose computer-generated information on a user's acquire digital signal of the real world, providing a composite output signal. On a mobile device the main idea is to blend the additional elements into the video stream of the mobile device's camera.

Besides computer games, several applications are emerging like medical imaging [1], industrial design [2], head-mounted displays [3] and image-base retrieval on mobile devices [4]. Streaming of mobile augmented reality applications require real-time recognition, segmentation and tracking of the objects of interest. In order to perform such operations it is common to utilize robust local features. Examples of local features include Scale-Invariant Feature Transform (SIFT) [5], Speeded Up Robust Features (SURF) [6] and Compressed Histogram of Gradients (CHoG) [7]. Furthermore, since the selected features or their corresponding patches are going to be transmitted over the network, it is desirable that the amount of data sent is as low as possible. Hence, previous studies have been presented for feature descriptor compression [8].

In this work, the authors propose a temporally coherent keypoint detector, and strive to transmit each patch or its equivalent feature descriptor with as few bits as possible by modifying a previously transmitted patch. Independent detection of keypoints and encoding of feature descriptors in a image does not exploit the temporal redundancy present in a video sequence. Therefore, The proposed method has the potential to achieve a better performance of feature descriptors encoding for mobile augmented reality.

During feature detection the video frames are divided into two categories: Detection frames (D-frames) and Forward Propagation frames (FP-frames). In the D-frames a conventional keypoint detection algorithm is applied, in this case SIFT. Each detected keypoint will be associated to a image patch, referred as D-patch. In FP-frames, instead of re-computing the keypoints in the entire image, patch matching is performed in order to connect each patch with a patch in a previous frame. Since each patch is characterized by its location, orientation and scale, not only the spatial location but all mentioned parameters vary during the proposed patch matching process. It is important to notice that the number of frames between two consecutive D-frames can significantly affect the accuracy of patch propagation. Therefore, the authors provide and algorithm to adaptively modified the number of FP-frames between D-frames. Whenever the average minimum distortion between D-Patches and their closest corresponding FP-Patches increases drastically, a new D-frame is inserted.

In the scenario where the patches need to be transmitted, the authors use a previously proposed encoder with a similar structure to JPEG. A Gaussian blurring and mean removal on the patches is performed as a pre-processing step. Then, a 2D Discrete Cosine Transform (DCT) is applied in each patch. New encoding modes are introduced, which use predictive coding for both D and FP patches. Instead of encode each patch independently, only the difference between a previous encode patch and the current patch is sent. For FP-patches the authors observed that the energy between two consecutive patches is usually very low. This suggests that it is not necessary to transmit the prediction residuals in FP-frames. In those frames, the keypoints location, orientation and scale can be updated using the descriptors extracted from the patches of the previous D-frame.

The authors also propose an inter-descriptor coding scheme to be applied when all feature descriptors need to present at the transmitter and there is no need for patch transmission. The encoder uses lossy lattice encoding combined with arithmetic entropy coder. Temporal information is used as context to improve the performance of the entropy coder. A Differential Pulse Code Modulation (DPCM) mode is also studied, however best results in term of bit-rate and matching performance are achieve when D-frame descriptors are encoded using independent lattice coding and the FP-frames descriptors are not transmitted.

The experimental results provided by the authors show that the proposed temporally coherent detection mechanism results in an image matching and retrieval performance comparable to the oblivious detection of new keypoints every video frame with a significant bit-rate reduction.

As mentioned earlier, streaming of mobile AR systems commonly need to perform efficient real-time object recognition and tracking at low bit-rates. This work provides tools for the low bit-rate aspect of the system. As future work, a combination of the proposed encoding techniques with a fast feature extraction algorithm can improve the system response within strict time constraints.

References:

[1] P. Mountney, S. Giannarou, D. Elson, GZ. Yang., "Optical biopsy mapping for minimally invasive cancer screening", in Medical Image Computing and Computer-assisted Intervention, n 12, pp 483–90, 2009.

[2] S. Noelle, "Stereo augmentation of simulation results on a projection wall" in Proc. IEEE International Symposium on Mixed and Augmented Reality, 2002.

[3] (2013). Google Glass [Online]. Available: http://www.google.com/glass/start/

[4] (2011). Amazon Flow [Online]. Available: http://a9.amazon.com//company/flow.jsp

[5] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[6] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: Speeded up robust features,” in Proc. Eur. Conf.. Comput. Vis., Graz, Austria, May 2006.

[7] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, “CHoG: Compressed histogram of gradients—A low bitrate feature descriptor,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Florida, FL, USA, Jun. 2009, pp. 2504–2511.

[8] V. Chandrasekhar et al., “Survey of SIFT compression schemes,” in Proc. 2nd Int. Workshop Mobile Multimedia Process. (WMMP), Istanbul, Turkey, Aug. 2010.

Bruno Macchiavello is an assistant professor at the Department of Computer Science of the University of Brasilia (UnB), Brazil. He received his B. Eng. degree in the Pontifical Catholic University of Peru in 2001, and the M. Sc. and D.Sc. degrees in electrical engineering from the University of Brasilia in 2004 and 2009, respectively. Prior to his current position he helped develop a database system for the Ministry of Transport and Communications in Peru. He also was co-organizer of a special session on Streaming of 3D content in the 19th International Packet Video Workshop (PV2012). His main research interests include video and image coding, image segmentation, distributed video and source coding, multi-view and 3D video processing.

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:

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

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

Review Process

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

R-Letter Best Paper Award

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

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

MMTC R-Letter Editorial Board

Multimedia Communications Technical Committee Officers

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

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