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

E-LETTER

Vol. 6, No. 11, November 2011

IEEE COMMUNICATIONS SOCIETY

CONTENTS

SPECIAL ISSUE ON "3D DATA EVALUATION, RETRIEVAL AND PRIVACY"

3D Data Evaluation, Retrieval and Privacy

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

In pace with the advancements in interactive 3DTV technology, the content of 3D data repositories has been evolved rapidly not only quantitatively but also qualitatively. Emerging 3D data in particular those associated with multi-views and depth information facilitates the creation of perceptually appealing displays and animations. The trend in virtual reality (VR) and augmented reality (AR) is towards the integration of traditional 2D and latest 3D methodologies to meet the demands in many industrial and research applications like advertising, movies, entertainments, social web such as Second Life, GOOGLE map and so on. To enhance users' perceptual experiences beyond the 2D horizon, many rendering techniques, such as immersion, stereo and multi-views, have been adopted not only in IMAX theatres, but also in household commodity systems. State-of-the-art 3D viewing techniques have become hot topics in multi-disciplinary research. Nevertheless, a challenging issue continues to be improving user perceptual experience and measuring user satisfaction, based on which computational model(s) can be developed for content assessment. The August 2011 E-Letter presented an excellent discussion on Quality of Experience (QoE). This special issue is dedicated to the 3D aspect of data evaluation, retrieval and privacy.

In order to deliver perceptually impressive content, solely relying on rendering techniques is not sufficient. Data degradation can be introduced at any stage along the processing pipeline, which includes acquisition, synthesis and processing, compression and transmission before rendering. The first two papers in this special issue entitled "Depth estimation using a single camera-based computational imaging" and "Depth and multi-view image and video coding: algorithms and quality issues beyond the ones of standard image and video compression," respectively provide a comprehensive discussion on how the quality of 3D images can be improved by using accurately computed depth information and the development of 3D camera technology – multiple color-filter aperture (MCA) camera. While the first paper presents a novel approach on depth estimation, the second paper focuses on the efficient coding of depth information for 3D view synthesis and multi-views video. Increasingly, researchers have come to the consensus that conventional quantitative approach is not adequate to evaluate data quality. This is especially so when human viewing satisfaction is a decisive factor when evaluating stereo images in a multi-views setting. The second paper highlights a number of issues related to the quality assessment of multi- to free viewpoint representations, and the recall of some standard approaches which were believed inapplicable to the new 3DTV platform. While quantitative measures are still useful, integration with qualitative metrics, taking perceptual factors into consideration, will certainly enhance the outcome in achieving better viewing experience.

While the development of 3DTV technology is promising, its popularity and general acceptance brings an enormous creation of 3D content. Searching target objects in 3D repositories real-time is comparatively more challenging than searching in its 2D counterpart. Although similarity match techniques effective for 2D content retrieval can continue be applied, more promising methods need to be introduced for addressing issues associated with the high dimensional data. The third paper entitled "3D face recognition: where we are" discusses a special case of 3D matching – face recognition, which is a hot research topic in security and surveillance type of applications. The fact that the anticipated matching algorithms have to be "expression" invariant, and ideally facial mask transparent, poses great difficulties in the recognition task. This paper points out the advantage of analyzing faces in 3D, which is increasingly successful with the latest launch of depth camera technology. As well, the paper presents a number of ongoing face recognition projects and revolution in 3D-vision related research.

There is no unique solution for all similarity match problems. The fourth paper on “3D object retrieval: challenges and research directions” discusses the various approaches employed in 3D object retrieval. While specific feature descriptors can be catered for selected applications, integrating several descriptors can likely generate better results in general. This scenario has been proved to be true in the recent 3D shape retrieval contests. Reduced dimensionality is a desired property for shape descriptors in order to be computational efficient real-time. Furthermore, rotation invariance is also important in order to discover similar objects captured from different viewing perspectives. The fifth paper reviews some algorithms in generating model skeleton (medial axis) which is a compact and semantically rich abstraction of the original model, and is useful in applications where the animated poses of the 3D objects need to be distinguished. Robust skeletons can also guide anatomically preserved decomposition for body part matching.

With digital technology integrated into our daily live and society, the trade-off is loss of privacy. Contrary to the discovery of faces described in the third paper, the sixth paper suggests how faces or crucial identity cues can be hidden using adversary knowledge modeling. While current research pays more attention on hiding identity, e.g. by blurring faces, the future trend should be more focused on how to perform automatic detection and removal of sensitive information as well as its adverse impact on reduced data utility.

Undoubtedly, there are many challenging R&D work ahead. By coordinating a discussion forum among researchers through this E-Letter, we intend to encourage further discussions on 3DRPC related topics. In this regard, I am glad to announce that our Interest Group is organizing an IEEE Multimedia Magazine special issue on “3D imaging techniques and applications,” and another special issue for the Advances in Multimedia Journal on “Communication of Human Affect using Smart-Multimedia.” We would like to invite your high quality submissions.

On behalf of the 3DRPC IG, special thanks to all contributors and the E-Letter Editorial Board in coordinating this special issue.

Irene Cheng
Chair, MMTC 3DRPC Interest Group.

Depth Estimation Using a Single Camera-Based Computational Imaging

Sangjin Kim, Eunsung Lee, Wonseok Kang, and Joonki Paik
Department of Image, Chung-Ang University, South Korea
paikj@cau.ac.kr

1. Introduction

The recovery of depth information of a scene from 2D images is a fundamental problem in many image processing and computer vision applications such as object recognition, scene interpretation, 3D reconstruction, and multifocusing. Although various depth estimation algorithms have been proposed, accurate estimation of depth information with a moderate amount of computation is still a challenging problem.

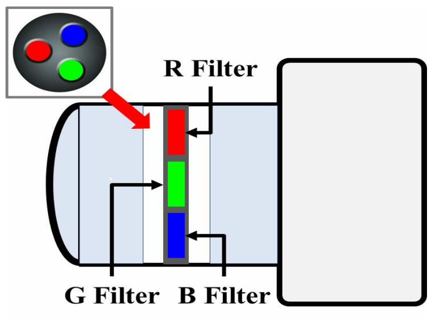
Recently, A computational camera uses a combination of unconventional optics and digital post-processing to produce new forms of information that may be used to solve a variety of imaging problems, such as increasing the field of view of the camera, refocusing a captured image, increasing the dynamic range of an imaging system, extracting depth information to determine a scene's 3-D structure, depth-based editing, and relighting [1][2][3][4].

A camera that uses multiple apertures in the imaging process falls into the category of a computational camera. This letter presents a novel approach to depth estimation using a multiple color-filter aperture (MCA) camera. An image acquired by the MCA camera contains spatially varying misalignments between the RGB color channels, where the direction and length of the misalignment is a function of the distance of an object from the camera. Therefore, if the misalignments can be estimated, then depth estimation may be performed.

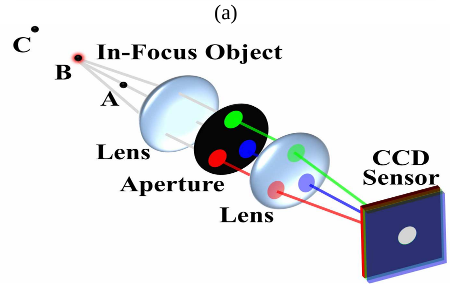
2. Multiple Color-Filter Aperture Camera

The aperture of an optical system is the opening device which can adjust the amount of incoming light. Generally, the center of an aperture is aligned to the optical axis of the lens, and the convergence pattern on the imaging plane forms either a point or a circular region. On the other hand, if the center of the aperture is not aligned to the optical axis, the image formation position on the sensor is deviated from the optical axis. As shown Fig. 1(a), the MCA system consists of three apertures with R, G, and B color-filters.

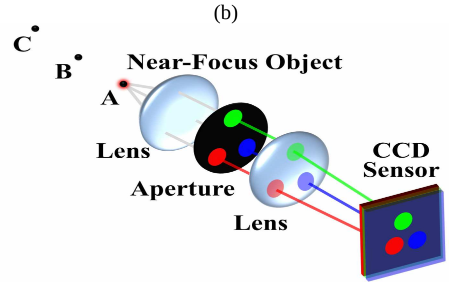
The main advantage of the MCA system is that it provides additional depth information of objects at different distances as shown in Figs. 1 (b)-(d).



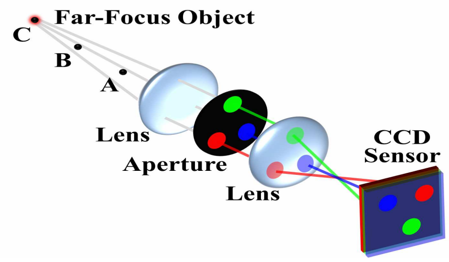
(a)



(b)



(c)



(d)

Fig. 1. (a) The MCA camera, (b) an object is on the in-focus position b, (c) an object is on the near-focus position c, and (d) an object is on the far-focus position a.

3. Depth Estimation Using the MCA Camera

We first segment the image into multiple clusters, each of which contains approximately uniform misalignment using a color-based region classification method,. and then produce the corresponding region-of-interest (ROI). The proposed cluster-based segmentation algorithm employs K-means clustering and block partitioning. K seeds in the clustering algorithm are automatically determined by using the hill-climbing method in the three-dimensional (3D) CIE Lab histogram of an input image [5][6]. After labeling salient regions, the input image is partitioned into same sized blocks. These partitioned blocks generate the ROI map for the color shift.

Among three corresponding ROIs in RGB color channels, color-shift vectors (CSVs) are estimated using a phase-only correlation method [7][8]. The CSV of an MCA camera image provides depth information of the corresponding cluster of each ROI. More specifically, the farther an object goes away from the focusing plane, the larger the CSV becomes. In addition, the angle between the red-to-green (R-G) CSV and blue-to-green (B-G) apertures form a regular triangle. These properties of the CSV are used to robustly estimate the distance of a cluster from the camera. Fig. 2 shows the step-by-step results of the proposed depth estimation workflow. Figs. 2(a) and 2(b) respectively show ROIs and the corresponding clusters with superimposed ROIs. Fig. 2(c) shows the CSV distribution of the corresponding clusters of ROIs, where upper and lower dotted circles contain far- and near-focused clusters, respectively. The result of depth estimation in conjunction with the proposed cluster-based image segmentation result is shown in Fig. 2(d).

4. Conclusions

In this letter, we presented a novel approach to depth estimation using the MCA camera. An image acquired by the MCA camera contains spatially varying misalignment among RGB color channels, where the direction and length of the misalignment is uniquely determined by the distance of an object from the camera. We can estimate depth of each region using the direction and length of the corresponding CSV. The estimated depth information may be used to various depth-based vision applications.

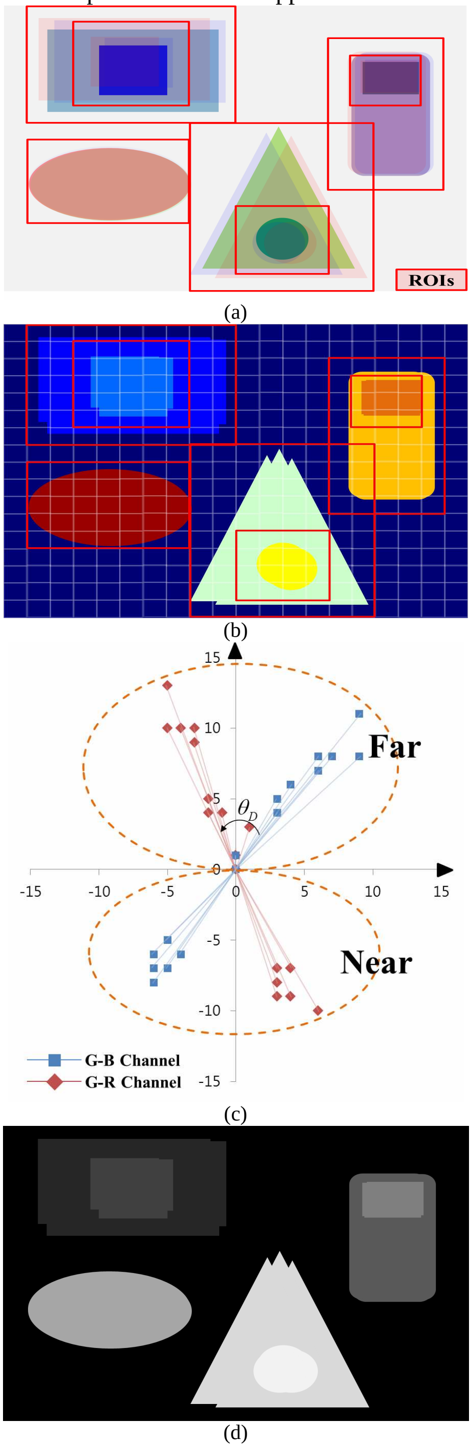


Fig. 2. The proposed depth estimation using CSVs; (a) ROIs in the image, (b) the clustered image with superimposed ROIs, (c) distribution of CSVs for the corresponding clusters of ROIs, and (d) the result of depth estimation in

conjunction with the proposed cluster-based segmentation result.

References

[1] S. Nayar, "Computational Cameras: Redefining the image," IEEE Computer Magazine, Special issue on Computational Photography, vol. 39, no. 8, pp.30-38, August 2006.
[2] V. Maik, D. Cho, J. Shin, and J. Paik, "Color shift model-based segmentation and fusion for digital autofocusing," Journal of Imaging Science and Technology, vol. 51, no. 4, pp. 368-379, July/August 2007.
[3] Y. Bando, B. Chen, and T. Nishita, "Extracting depth and matte using a color-filtered aperture," ACM Transactions on Graphics, vol. 27, no. 5, pp. 134:1-134:9, December 2008.
[4] T. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution", IEEE Int. Conf. Computational Photography, April 2009.
[5] Z. Aghbari and R. Al-Haj, "Hill-manipulation: an effective algorithm for color image segmentation," Image and Vision Computing, vol. 24, no. 8, pp. 894-903, February 2006..
[6] T. Ohashi, Z. Aghbari, and A. Makinouchi, "Hill-climbing algorithm for efficient color-based image segmentation," IASTED Int. Conf. on Signal Processing, Pattern Recognition and Applications (SPPRA), pp. 404-409, June 2003.
[7] H. Foroosh, J. Zerubia, and M. Berthod, "Extension of phase correlation to sub-pixel registration," IEEE Transactions on Image Process, vol. 11, no. 3, pp. 188-200, 2002.
[8] M. Hagiwara, M. Abe, and M. Kawamata, "Estimation method of frame displacement for old films using phase-only correlation," Journal of Signal processing, vol. 8, no. 5, pp. 421-429, 2004.

Sangjin Kim was born in Seoul, Korea in 1978. He received the B.S. degree in electronic engineering from Kangnam University, Korea, in 2003. He received the M.S and Ph.D. degrees in image engineering from Chung-Ang University, Korea, in 2005 and 2009 respectively. Dr. Kim is currently Post-Doctorial researcher in Image Processing and Intelligent Systems Laboratory at Chung-Ang University. His research interests include image restoration, compressive sensing, computational camera, and real-time object tracking.

Eunsung Lee was born in Seoul, Korea in 1982. He received the B.S. degree in electronic engineering from Chung-Ang University, Korea, in 2009.
Currently, he is pursuing M.S. degree in image processing at Chung-Ang University. His research interests include image restoration, computational camera, and image enhancement.

Wonseok Kang was born in Jeju, Korea in 1983. He received the B.S. degree in electronic engineering from Korea Aerospace University, Korea, in 2010.
Currently, he is pursuing M.S. degree in image processing at Chung-Ang University. His research interests include image restoration, computational camera and data fusion.

Joonki Paik was born in Seoul, Korea in 1960. He received the B.S. degree in control and instrumentation engineering from Seoul National University in 1984. He received the M.S. and the Ph.D. degrees in electrical engineering and computer science from Northwestern University in 1987 and 1990, respectively. From 1990 to 1993, he joined Samsung Electronics, where he designed the image stabilization chip sets for consumer's camcorders. Since 1993, he has joined the faculty at Chung-Ang University, Seoul, Korea, where he is currently a Professor in the Graduate school of Advanced Imaging Science, Multimedia and Film. From 1999 to 2002, he was a visiting Professor at the Department of Electrical and Computer Engineering at the University of Tennessee, Knoxville. Dr. Paik was a recipient of Chester-Sall Award from IEEE Consumer Electronics Society, Academic Award from the Institute of Electronic Engineers of Korea, and Best Research Professor Award from Chung-Ang University. He has served the Consumer Electronics Society of IEEE as a member of the Editorial Board. Since 2005, he has been the head of National Research Laboratory in the field of image processing and intelligent systems. In 2008, he has worked as a full-time technical consultant for the System LSI Division in Samsung Electronics, where he developed various computational photographic techniques including an extended depth of field (EDoF) system. From 2005 to 2007 he served as Dean of the Graduate School of Advanced Imaging Science, Multimedia, and Film. From 2005 to 2007 he has been Director of Seoul Future Contents Convergence (SFCC) Cluster established by Seoul Research and Business Development (R&BD) Program. Dr. Paik is currently serving as a member of Presidential Advisory Board for Scientific/Technical policy of Korean Government and a technical consultant of Korean Supreme Prosecutor's Office for computational forensics.

Depth and multi-view image and video coding: algorithms and quality issues beyond the ones of standard image and video compression

Pietro Zanuttigh, Simone Milani and Guido M. Cortelazzo
University of Padova, Italy
{zanuttigh, simone.milani, corte}@dei.unipd.it

1. Introduction and challenges of 3D image and video compression

In recent years the growth of new media representations, such as stereoscopic movies, free viewpoint television, virtual browsing of 3D scenes and online 3D games, has created a great interest for compression and transmission schemes able to efficiently handle these kinds of data. In this letter we will focus on the coding of two representations that are currently attaining a big interest: i.e., the representation of geometry information through depth maps and the multi-view images and videos that are currently exploited by stereoscopic, multi-view and free-viewpoint applications.

Current solutions for the coding of these representations still rely on the extension of approaches originally developed for 2D image and video compression. For instance the well-known H.264 MVC standard [1] is just an extension of the H.264 video coding standard where the temporal prediction scheme has been applied also to the inter-view dimension. Even if these approaches can achieve good performances due to the highly performing image and video coding schemes behind them, they are not able to fully exploit the three dimensional nature of these data. This is the reason that in the long term makes them unlikely to be the best compression choice.

Specific solutions targeted to the compression of 3D representations that exploit the peculiar characteristics of these kind of data and the three dimensional structure of the underlying scene are currently being investigated by many research groups and in the near future are expected to deliver the best performances.

2. Coding of depth information for 3D video and view-warping applications

Depth information plays an important role in many 3D video representations: in fact depth data allows warping the available color views to different viewpoints thus allowing free viewpoint visualization. Depth maps typically feature large smooth regions separated by sharp edges. Edges play a key role in 3D warping applications: in particular their low-pass filtering, usually introduced by standard image compression techniques based on frequency domain approaches, introduces severe artifacts in the warped views obtained with compressed depth data. Fig. 1 shows a simple example of how compression artifacts on depth edges can affect the results of image warping.

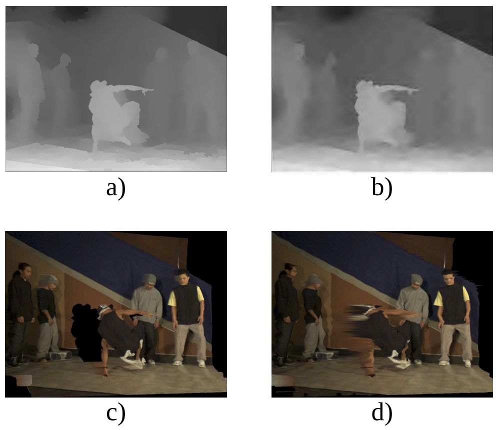


Fig. 1. a) Original depth map from the *breakdancers* sequence; b) the same depth map compressed with JPEG 2000 at 0.01bpp; c) warped view computed with the original depth map; d) warped view computed with the compressed depth map.

Compression methods specifically conceived for depth information usually rely on an explicit modeling of edges in order to improve compression performance and to avoid edge blurring. The method of [2], for instance, is based on an explicit representation of edges and a shape-adaptive wavelet transform. In this way wavelet coefficients in proximity of edges are greatly reduced thus reducing data entropy.

In the approach of [3] a subsampled depth map is used together with edge information order to build a prediction of the full resolution depth data that is then used for compression purposes. Fig. 2 (referring to the approach of [3]) shows how from very low resolution data and edge information it is possible to build a very accurate estimation of the depth map thanks to the already described peculiar structure of such data, essentially made by smooth regions and sharp edges.

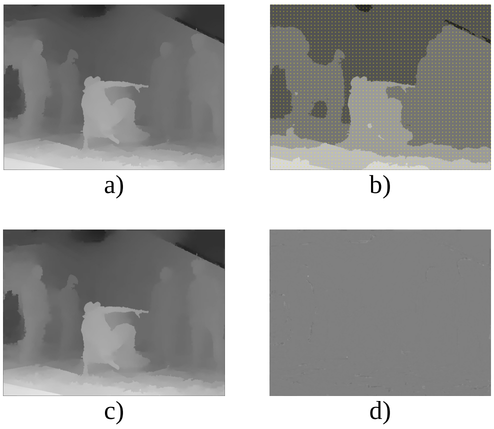


Fig. 2. a) Depth map from the *breakdancers* sequence; b) segmented depth map (yellow dots represent the position of low resolution samples); c) predicted depth map; d) difference between the predicted and the actual depth map.

In [4] a quadtree decomposition is first performed. Subsequently each region is approximated by way of platelets [5], i.e. a piecewise linear approximation is done for each region. In [6] instead the depth information of single objects is characterized by progressively overlapping different masks that refine the quality of depth values.

It is worth noting how all these approaches, although very different, share two basics ideas, namely that edges should be explicitly represented and that edge regions need special handling while smooth regions can just be represented only by few coefficients. However edge signaling is rather expensive in terms of bitrate. The edges obtained by standard edge detection and segmentation approaches represent what is relevant from a computer vision perspective, which does not necessarily correspond with what is most effective for coding purposes. One of the key research issues for efficiently exploiting this family of approaches is the development of rate-distortion optimization schemes capable to select which edges are worth to be explicitly signaled.

The quality evaluation of depth data is a fundamental issue. Representing the depth as an image and applying common quality measures such as MSE and PSNR does not give a trustworthy estimation of the results, especially with respect to edge regions. A possible alternative is to exploit the compressed depth data for view warping and then measure the quality of warped views. Unfortunately this approach introduces a dependency of the quality measure on both the employed warping algorithm and the color data that is being warped. The proper evaluation of depth data quality remains a key future research issue, especially in order to suitably tune rate-distortion optimization algorithms for depth and color-plus-depth compression schemes.

3. Coding of multi-view data

Since most of the 3D visualization displays require the availability of multiple video streams acquired from different viewpoints (e.g. stereoscopic displays with active glasses require two video streams, while autostereoscopic monitors simultaneously process up to 16 views), multi-view data is a widely used format for 3D video. It clearly requires effective compression strategies in order to reduce the huge amount of data involved in multi-view representations.
To this purpose, the extension MVC [1] of the layered coding standard H.264 SVC aims at jointly coding multiple video sequences of a common scene captured from several cameras at different viewpoints. The H.264 MVC coder tries to exploit the correlation existing between sequences taken from neighboring viewpoints, but the adopted prediction and decorrelation strategies derive from traditional 2D predictive video coders [1]. Similarly other approaches, like [7], extend the traditional 2D wavelet decomposition to the multi-view case. As a matter of fact, the performance of these schemes proves to be limited by the performance bounds of the previous techniques.

Better compression ratios can be obtained by tailoring the coding strategy to the geometrical relations among the different views. The availability of depth maps allows establishing pixel by pixel correspondences between views by a warping transform. This fact enables the possibility of building warping-based multi-view codecs (Fig. 3 shows the general architecture of this kind of approaches). One of the views (*reference*) is specified using a traditional 2D coding standard, while the others are compressed by predictive coding schemes. The predictions are obtained by warping the reference on the different viewpoints [8,9]. Other approaches operate an additional transform along the warped views, like the approach based on 3D-DCT of [10].

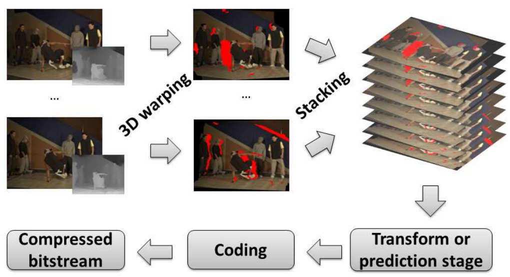


Fig. 3. Architecture of a warping-based multi-view encoder.

Experimental results show that the performance of warping-based solutions is strictly dependent on the quality of the reconstructed depth map. As a matter of fact, the effective and quality-preserving compression of depth information plays a crucial role in such techniques. Moreover, motion in depth sequences is strongly correlated with its corresponding view sequences and joint compression can significantly benefit from this. These issues will be considered with more details in the next section.

Let's also recall that recent research studies have also highlighted the need for multi-view quality metrics with a strong correlation with human perception of 3D scenes in order to allow the optimization of coding parameters. Several works in literature infer the quality of the reconstructed 3D scene by averaging the quality of the different views. However, perceptual analysis on stereoscopic and multi-view vision shows that this is not enough to grant a comfortable and pleasant experience to the end user. Indeed, the 3D position of the rendered objects with respect to the viewer significantly affects the level of comfort of the viewer, and as a matter of fact, all the rendered details need to lie in the stereoscopic comfort zone [11]. Furthermore, this region might not guarantee a pleasant 3D vision in case of fast objects moving across the depth range.

As a matter of fact, the most recent 3D quality evaluation strategies estimate depth information from the available views and try to weight standard 2D quality metrics with it [12]. Depth information is also used to recalibrate the video signal for the different views in order to make it compatible with respect to the comfort zone bounds [13]. As a result, many of these strategies can be adapted to evaluate the quality of compression for depth signals whenever it is associated to standard video sequences.

4. Joint coding of depth and color information

The previous section highlighted how the video signals associated to different views of a multi-camera system are significantly correlated and how such correlation can be exploited in order to obtain better compression ratios. Since co-located pixels in color and depth information correspond to a common 3D point, color signal motion is highly correlated with depth signal motion. Moreover, object boundaries and shapes prove to be the same on both sequences.

The approach of [14] exploits the correlation of motion in order to reduce the amount of coded information and to make more robust the motion vectors (MVs) estimation. In this case, motion can be computed from the temporal correlation estimated from both depth and color signals, and MVs can be signaled just once in the bit stream, for both components.

The solution of [15] exploits the correlation existing between object boundaries and it infers stationary regions in the depth information from the segmentation of the color components.

It is possible to push further such possibilities by designing a video+depth codec that relies on the mutual correlation also in the segmentation stage. The method of [16] identifies object shapes in both color and depth signals by a segmentation strategy operating on both components and uses them for accurate joint motion estimation.

Joint compression schemes can be optimized and evaluated from the depth-related 3D quality metrics, where depth information is estimated or assumed to be available [17]. Other evaluation strategies consider warping the reference view on different viewpoints using first the original uncompressed depth and color signals and then their reconstructed versions. It is then possible to compute standard 2D quality metrics between the warped views by measuring the effect of color and depth compression on the resulting signal. As a matter of fact, it is possible to design joint rate allocation [14] and joint optimization algorithms that maximize the perceptual quality of the rendered scene [19].

5. Conclusions

This paper addresses the issues related to the coding of two representations that play a key role in novel 3D video applications, i.e., depth data and multi-view images and videos. We showed how the peculiar nature of these representations and the underlying 3D structure of the scene can be exploited in order to develop specific coding schemes that outperform standard image and video coding solutions. Furthermore new 3D video applications usually require both depth and multi-view data. As a matter of fact, the exploitation of the redundancy between color and geometry information is an intriguing research issue in the design of joint compression schemes. Finally several issues related to the quality evaluation of multi-view and free-viewpoint representations were recalled because standard image and video evaluation approaches do not give trustworthy measures of the perceived 3D quality.

References

[1] A. Vetro, T. Wiegand, and G. J. Sullivan, Overview of the Stereo and Multiview Video Coding Extensions of the H.264/AVC Standard, Proceedings of the IEEE, Special Issue on "3D Media and Displays", vol. 99, issue 4, pp. 626-642, April 2011.

[2] M. Maitre, M.N. Do, "Depth and depth-color coding using shape-adaptive wavelets", Journal of Visual Comm. and Image Representation, vol. 21, no. 5-6, Special issue on Multi-camera Imaging, Coding and Innovative Display, pp. 513-522, Jul.-Aug. 2010.

[3] P. Zanuttigh, G.M. Cortelazzo, "Compression of depth information for 3D rendering," 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, pp.1-4, May 2009

[4] Y. Morvan, D. Farin, P.H.N. de With, "Depth-Image Compression based on an R-D Optimized Quadtree Decomposition for the Transmission of Multi-view Images", IEEE International Conference on Image Processing, Sept. 2007

[5] R. M.Willett and R. D. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," IEEE Trans. on Medical Imaging, vol. 22, pp. 332–350, Mar. 2003.

[6] S. Milani, G. Calvagno, "A Depth Image Coder Based on Progressive Silhouettes," IEEE Signal Process. Lett., vol. 17, no. 8, pp. 711–714, Aug. 2010.

[7] J.U. Garbas, B. Pesquet-Popescu, A. Kaup, "Methods and Tools for Wavelet-Based Scalable Multiview Video Coding", IEEE Trans. On CSVT, vol. 21, no. 2, feb 2011, pp. 113-126.

[8] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, Y. Yashima, "View Scalable Multiview Video Coding Using 3-D Warping With Depth Map," IEEE Trans. On CSVT , vol.17, no.11, pp.1485-1495, Nov. 2007

[9] M. Zamarin, P. Zanuttigh, S. Milani, G. M. Cortelazzo, S. Forchhammer, "A Joint Multi-View Plus Depth Image Coding Scheme Based on 3D-Warping", Proc. of ACM Workshop on 3D Video Processing, Florence, Italy, pp. 7-12, Oct. 2010.

[10] M. Zamarin, S. Milani, P. Zanuttigh, G.M. Cortelazzo, "A novel multi-view image coding scheme based on view-warping and 3D-DCT," Journal Vis. Comm. and Image Representation, vol. 21, no. 5-6, pp. 462–473, 2010.

[11] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: a review," *Journal of Imaging Science and Technology,* no. 53, pp. 030201-1-030201-14, 2009.

[12] A. Benoit, P. Le Callet, P. Campisi and R. Cousseau, "Quality assessment of stereoscopic images", *EURASIP Journal on Image and Video Processing, special issue on 3D Image and Video Processing*, vol. 2008, Article ID 659024, 13 pages, doi:10.1155/2008/659024, 2008.

.[13] M. Lang. A. Hornung, O. Wang. S. Poulakos, A. Smolic, and M. Gross, "Non-linear Disparity Mapping for Stereoscopic 3D", *ACM Transactions on Graphics (SIGGRAPH 2010)*, Jul. 2010.

[14] I. Daribo, C. Tillier, and B. Pesquet-Popescu, "Motion Vector Sharing and Bitrate Allocation for 3D Video-Plus-Depth Coding", EURASIP Journal on Advances in Signal Processing, 2009.

[15] S. Milani, P.Zanuttigh, M. Zamarin, S. Forchhammer, "Efficient depth map compression exploiting segmented color data", Proc. of the 2011 IEEE International Conference on Multimedia and Expo, Barcelona, Spain, Jul. 2011

[16] S. Milani, G. Calvagtno, "3D Video Coding via Motion Compensation of Superpixels", Proc. of EUSIPCO 2011, Barcelona, Spain, Sep. 2011

[17] M. Maitre, Y. Shinagawa, M.N. Do, "Wavelet-Based Joint Estimation and Encoding of Depth-Image-Based Representations for Free-Viewpoint Rendering," IEEE Trans. Image Process., vol. 17, no. 6, pp. 946–957, 2008.

[18] A. Smolic, K. Mueller, N. Stefanoski, et al., "Coding algorithms for 3DTV - A survey," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, pp. 1606–1620, 2007.

[19] S. Milani, G. Calvagno,"A Cognitive Approach for Effective Coding and Transmission of 3D Video", Proc. of ACM Multimedia 2010, Florence, Italy, pp. 581-590. Oct. 2010.

Pietro Zanuttigh was born in 1978. He graduated in Computer Engineering at the University of Padova (Italy) in 2003 and got the Ph.D. degree from the same university in 2007. He has been visiting researcher at the University of New South Wales (Australia) in 2006 and 2010. In 2007 he became an assistant professor at the University of Padova. Now he works in the multimedia technology and telecommunications group and his research activity focuses on 3D data processing. His main research interests are the transmission and remote visualization of scalably compressed 3D scenes, compression of depth maps and multi-view videos and dynamic 3D scenes acquisition with multiple sensors. He was a guest editor of the International Journal of Digital Multimedia Broadcasting. He also holds a patent on interactive visualization on mobile phones.

Simone Milani was born in Camposampiero (PD), Italy, in 1978. From the University of Padova, Italy, he received the Laurea degree in Telecommunication Engineering in 2002, and the Ph.D. degree in 2007. In 2006 he was a visiting Ph.D. Student at the University of California-Berkeley under the supervision of prof. K. Ramchandra. From 2008, he has been post-doc researcher collaborating with prof. Calvagno and Cortelazzo (University of Padova), prof. R. Rinaldo (University of Udine), and prof. S. Tubaro (Politecnico di Milano). He has also worked with STMicroelectronics, Agrate Brianza, Italy. His main research topics are image and video source coding, distributed source coding, robust video transmission, 3D video acquisition and compression, and multimedia forensics. He has been reviewer for many transactions and conferences supported by IEEE and EURASIP.

Guido M. Cortelazzo was born in Este (Italy) in 1952. He holds a "Laurea in Ingegneria Elettronica" degree from the University of Padova (1976) and a Ph.D. Degree in E.E. from the University of Illinois at Urbana-Champaign (1984). From 1983 to 1986 he worked with M/A-COM Linkabit, a satellite communication company of San Diego, California. In 1986 he joined the Department of Information Engineering (DEI) of the University of Padova where he is now full professor. In 1986 he funded the Image Processing Laboratory of DEI and in 1993 the Multimedia Telecommunications and Technology Laboratory of DEI, which is still directing. In 1998 he was visiting associate with the California Institute of Technology. The professional interests of Guido M. Cortelazzo concern telecommunications, image processing and computer vision. His current research concentrates on the automatic construction of 3D models of still and dynamic scenes and on their progressive transmission. He organized various special sessions and meetings, among which in 2002 he was co-chair of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT) and a founding member of the steering committee. Guido M. Cortelazzo is the author of about sixty journal papers, two book chapters and one book. In 2004 he was Co-Guest Editor of a special issue of Computer Vision and Image Understanding and in 2006 of a special issue of Signal Processing-Image Communication. He holds two industrial patents.

3D Face Recognition: Where we are.

Jean-Luc Dugelay, Nesli Erdogmus and Neslihan Kose
EURECOM, France
{ jean-luc.dugelay, nesli.erdogmus, neslihan.kose}@eurecom.fr

1. Short overview of 3D Face Recognition

Although face recognition (FA) has drawn increasing interest for decades and that significant advances have been achieved, intra-class variations due to various factors in real-world scenarios such as illumination, pose, expression, occlusion and age remain a challenge in practical face recognition (FR) systems.

In the Face Recognition Vendor Test (FRVT) 2002 systems based on 2D intensity or color images were shown to deliver recognition rates of over 90% under controlled conditions [1]. However, with the introduction of aforementioned variations, in particular differences in expression, performance deteriorates significantly. These results led to acceleration in studies of alternative modalities, especially invoving three-dimensional (3D) face images, since it is an a priori natural approach to mitigate difficulties related to pose and illumination. Consequently, through the use of 3D, FRR levels of 0.01 and FAR levels of 0.001 were reported at FRVT 2006.

Today, a number of early commercial systems based on diverse approaches have appeared on the market. For instance, the Geometrix Face Vision 3D Sensor outputs tightly calibrated stereo image pairs that are processed into a precise 3D shape while the Konica Minolta Vivid series exploits laser-beam light sectioning technology. More recently, with the arrival of Microsoft's affordable Kinect has given rise to a depth camera revolution in vision-related research and enables a wide range of applications such as robotics and HCI.

Even with the a priori inherent advantages of 3D FR, challenges to make this technology directly usable and reliable in real applications are still numerous with regards to fingerprint tecnologies for example.

Obviously, 3D FR has greater potential than its 2D counterpart but, in practice, it is not straightforward to obtain accurate 3D image input data. Systems that extract shape information from 2D images, e.g. passive stereo approaches, rely on the knowledge of extrinsic parameters of the scene and intrinsic parameters of the camera to obtain a certain degree of accuracy. With other acquisition technologies such as active sensors, i.e. laser scanners, a scan can take several seconds, and addionally requires the subject to remain still during the acquistion process. The reconstruction of depth is furthermore possible only at short distances. Additionally, while a 3D facial shape is intrinsically independent of illumination variations, the accuracy of current 3D acquisition systems is still sensitive to differences in lighting conditions. Finally, while pose and illumination variations can be managed by utilizing 3D data, expressions which alter the facial surface characteristics remain an open issue.

2. Eurecom's contributions to 3D FR

Currently, Eurecom participates in two 3D face recognition related research projects: a national French project entitled Face Analysis and Recognition in 3D (FAR 3D) [7], and a European project entitled Trusted Biometrics under Spoofing Attacks (TABULA RASA) [8].

FAR 3D: This project harmnesses the effort of four teams (Eurecom, USTL, ECL and THALES) to explore 3D face recognition techniques. Eurecom's contributions mainly focus on the concept of asymmetrical protocols in 3D face recognition where enrollment performed in both 2D and 3D (i.e. shape and texture of faces) whereas test face images are captured only in 2D or from video. The problem of expressions in 2D probe images is addressed by simulating facial expressions using the 3D neutral model of each subject during enrollement. In order to obtain realistic simulations of facial expression, an automatic procedure is proposed to generate MPEG-4 compliant animatable face models from the 2.5D facial scans (range images) of the enrolled subjects based on a set of automatically detected feature points (see Fig. 1).

TABULA RASA: This European project, addresses some of the issues of direct (spoofing) attacks against trusted biometric systems, Eurecom is leading efforts to develop countermeasures against 3D FR spoofing attacks. Since the recent introduction and development of

3D FR systems, the main issues which have been addressed are related to expression variations or occlusions, i.e. more generally speaking, robustness against intra- versus inter- class variability. However, spoofing attacks against 3D FR systems, where a malicious user might pretend to be someone else at the sensor or acquistion point of the system, still remain under investigated. Eurecom's efforts within this project aim to fill this gap.

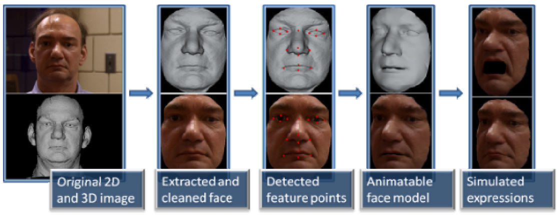


Fig 1. Simulation of a facial expression is illustrated.

3. Current 3D Face Products

Security is a very important topic, and new technologies may help in proving better solutions. There exists an important and increasing demand for more effective and less intrusive security systems. Therefore, in addition to numerous 2D FR technologies, several well-established or also new companies started to produce 3D FR products to meet this demand. Below, we briefly present some early products thus far developped.

- **3D FR Technology of L-1 Identity Solutions (now part of Safran MORPHO):**

The *3D FastPass Face Reader* produced by L1 Identity Solutions provides high-speed access to offices and restricted areas. It authenticates users in less than a second. The *3D FastPass* system increases security while maintaining high throughput for access to buildings while making the process effortless for employees.

- **SureMatch 3D Suite of Technest:**

SureMatch 3D is a suite of facial recognition tools that provides more robust detection capabilities through 3D enhancement [3]. The suite consists of 3D Face Map, 3D Enroll, 3D FaceCam, 3D Sketch Artist, 3D Face Match and integrated systems.

- **3D FR System of Godrej Group:**

The Face Reader produced by Godrej group can be connected to a LAN and remotely managed through the Administration software. With an access controller it can be effectively used with other biometrics and card based systems [4]. The reader also authenticates users in less than a second and is optimized for high traffic areas.

- **Vision Access 3D FR of A4 Vision (Applications for Vision), Inc.:**

A4Vision develops advanced identification systems for tracking and targeting camera systems using 3D FR technology [5]. *The Vision Access 3D face reader* is commonly used to control physical access to buildings. An integrated LCD and audio feedback assist the user to adopt a suitable position in order for the reader to quickly capture an optimal image [6].

References

[1] P. J. Phillips, P. Grother, R. J Micheals, D. M. Blackburn, E Tabassi, and J. M. Bone, "FRVT 2002: Evaluation Report," March 2003.

[2] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale results," National Institute of Standards and Technology, Tech. Rep. NISTIR 7408.

[3] http://www.genextech.com/

[4] http://www.godrej.com/godrej/SecuritySolutions/

[5] http://www.bioapi.org/products.asp

[6] http://www.meyer.com.tr/ftp/A4VISION.pdf

[7] http://www-rech.telecom-lille1.eu/far3d/

[8] http://www.tabularasa-euproject.org/

Jean-Luc Dugelay received his Ph.D. degree in 1992 from the Univ. of Rennes, France. Doctoral research was carried out, from 1989 to 1992, at the France Telecom Research. He is currently a professor in the Dept. of Multimedia Com. at Eurecom. He is the founding EiC of the EURASIP J. on image and video processing (SpringerOpen).

Nesli Erdogmus graduated from Middle East Technical University (METU), Ankara, Turkey in 2005. She received her master's degree in 2008. Since then, she is a Ph.D. student at Eurecom multimedia.

Neslihan Kose graduated from Middle East Technical University (METU), Ankara, Turkey in 2007. She received her master's degree in 2009. Since then, she is a Ph.D. student at Eurecom multimedia.

3D Object Retrieval: Challenges and Research Directions

Apostolos Axenopoulos and Petros Daras
Informatics & Telematics Institute, Greece {axenop,daras}@iti.gr

1. Introduction

Recent advances in 3D scanning mechanisms, computer-aided modeling tools, display and rendering devices have led to an explosion of the number of available 3D models over the Internet. It is now possible to easily construct complete 3D geometric models with relatively low cost and time consumption. This widespread availability of 3D models intensified the need for effective search through the various online media databases. Towards this direction, extensive research has been conducted in the area of 3D content-based search and retrieval.

3D object retrieval methods extract automatically the 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 discriminating 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 existing 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. Results of the latest 3D Shape Retrieval Contests (SHREC'11 [1], SHREC '10 [2]), a worldwide benchmarking initiative for 3D shape retrieval, have demonstrated that the combination of already existing methods can achieve higher retrieval accuracy than a single, highly complex algorithm.

This paper will provide an overview of the most well-known approaches in 3D object retrieval. It will highlight the strengths and weaknesses of the various approaches and it will introduce potential directions and challenges in the field of content-based 3D shape retrieval. The rest of the paper is organized as follows: Section 2 provides a state-of-the-art review of the most common methodologies. In Section 3, some promising research directions and challenges are summarized.

2. Recent Progress in 3D Object Retrieval

The first step in a typical 3D object retrieval procedure is a preprocessing phase, where the 3D model is translated so that its center of mass coincides with the center of a coordinate system 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 second step involves feature extraction, where low-level descriptors are extracted from the 3D object and uniquely represent its shape. During search and retrieval, the low-level descriptors of the query object are matched with the low-level descriptors of the objects in a database, using an appropriate dissimilarity metric, and the most similar objects are returned to the user. Apart from the above procedures, a 3D object retrieval procedure can be enhanced by the following actions: a) feature selection applied on the extracted low-level descriptors; b) appropriate selection of the optimal distance measure; c) combination of more than one low-level descriptors; d) re-ranking of the search results using manifold learning approaches and d) relevance feedback techniques for refining the retrieved results. These actions, although optional, can significantly improve the performance of an existing 3D object retrieval system, therefore, they should by no means be underestimated.

Rotation normalization is an essential step before almost every descriptor extraction method. The most popular solution for rotation estimation relies on Principal Component Analysis (PCA) [3], which is based on the computation of the 3D objects' inertia moments. PCA has been proven to be inaccurate, therefore, new methods have been recently introduced, such as Continuous PCA (CPCA) [4], Rectilinearity-based rotation estimation [5] and a Combined Pose Estimation method (CPE) [6]; the latter provides the highest accuracy to the best of our knowledge.

Concerning low-level feature extraction from 3D objects, the existing methods can be classified into four main categories [7]: histogram-based, transform-based, graph-based, view-based and, finally, combinations of the above. The first category includes methods that integrate the local or global features extracted from 3D objects into histograms [8]. These methods are, in general, easy to implement but usually they are not discriminating enough to make subtle distinctions between classes of shapes. In transform-based methods, the 3D object is usually described as a function in the 3D space. Then, a mathematical transformation is applied to this function, which captures specific geometrical features of it [9], [10]. 2D view-based methods consider the 3D shape as a collection of 2D projections taken either from canonical or non-canonical viewpoints. Each projection is then described by standard 2D image descriptors [7], [11], [12]. The results of the last three 3D Shape Retrieval Contests [13], [2], [1] have proven that 2D view-based methods achieve the highest performance among all other categories of 3D object retrieval methods. Their only drawback is that they discard valuable 3D information (due to the self-occlusion).

All the above categories of descriptor extraction methods are suitable for rigid, global-shape 3D object retrieval. For non-rigid or partial 3D object retrieval, graph-based methods [21] provide the optimal solution, since they encode geometrical and topological shape properties in a more faithful and intuitive manner. Recent advances in non-rigid 3D object retrieval can be found in SHREC'11 Track: Retrieval on Non-rigid 3D Watertight Meshes [14], where the best methods worldwide are compared. Similar to articulated shape matching, partial 3D object retrieval [15] focuses on identifying salient points on the 3D objects' meshes, followed by a local shape descriptor, which is applied around each interest point. Then, matching subsets of salient points are found so that the local shape dissimilarity and deformation is minimal. The drawbacks of graph-based methods are that they do not generalize easily to all 3D shape representation formats and they require dedicated matching schemes, thus, they are not suitable for general-purpose 3D object retrieval tasks.

The dimensionality of the descriptor vectors can vary from a few hundreds to tens of thousands real or integer values. As the descriptor size increases, search and retrieval in very large databases becomes prohibitive, since similarity matching of large descriptor vectors requires high response times. This problem can be overcome by using feature selection methods [16], [17], which have been widely used in pattern analysis in order to select the most significant features of a given descriptor. 3D object retrieval accuracy can be further improved if more than one low-level descriptors are combined in order to produce one unified descriptor vector. Combination of different descriptors is usually expressed as a weighted sum of their individual dissimilarities [18]. In most of existing methods, these weights are determined heuristically, which, however, does not guarantee the maximum possible improvement.

Moreover, manifold ranking has been adopted to improve the performance of 3D search and retrieval methods. The use of manifold ranking was 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 SHREC 2010 contest [2], the first two ranked methods were view-based methods using manifold ranking as a post-processing step to improve the retrieval accuracy.

Despite their significant progress in terms of retrieval accuracy, even the most accurate 3D shape retrieval methods may fail to return results that fully satisfy the end-users. This is due to the fact that the above methods do not take into account user's subjectivity. In order to enable personalized retrieval, Relevance Feedback (RF) techniques [22] have been introduced. Some of the most common RF approaches involve query refinement, where the initial query is moved so as to get closer to the relevant objects, and techniques based on query expansion, which usually replace the query point with multiple query points. Another category of RF approaches use different weights on the objects' low-level features, when computing their pairwise dissimilarity measure.

3. Feature Directions and Challenges

Based on the results of the latest 3D Shape Retrieval Contests [1], [2], it can be easily concluded that the combination of multiple 3D object descriptors can achieve better retrieval accuracy than a single descriptor vector alone, no matter how efficient this descriptor is. Thus, future research in 3D object retrieval should focus not only on the investigation of the optimal descriptor but also on the appropriate combination of low-level descriptors as well as on the selection of the best features and matching metrics. This means that as soon as a new promising method appears, the next step is to combine it with other state-of-the-art approaches and incorporate it into a unified 3D shape matching framework. Provided that the new method belongs to transform-based or view-based methods and describes the global features of a 3D object, it can be easily integrated, as follows: find the optimal dissimilarity metric for this descriptor and normalize it to a common range of values (e.g. between 0 and 1); the overall dissimilarity is the weighted sum of dissimilarities of multiple descriptors, where the weights should be appropriately optimized.

Apart from transform-based and view-based methods, graph-based methods [21] can be also used for global-shape 3D object retrieval tasks. However, graph-based methods may usually require complex matching schemes, which makes their combination with transform and view-based methods problematic. Another limitation is that most of the graph-based methods, although they work well with articulated objects, they have very limited accuracy in categories of 3D models that have no articulation (e.g. buildings, furniture, cars, cups, etc.). On the other hand, in the case of non-rigid and partial 3D shape retrieval tasks, graph-based (topology-based) methods should be preferred instead of transform-based and view-based ones. It is worth to mention that in SHREC'11 Track: Retrieval on Non-rigid 3D Watertight Meshes [14], the best retrieval performance is achieved by a framework that combines two different methods, the Spectral Decomposition of the Geodesic Distance Matrix [19] and the Scale Invariant Feature Transform for meshes (meshSIFT) [20].

Further improvements in the retrieval performance are also expected, if appropriate schemes for Relevance Feedback are investigated. Relevance Feedback has been extensively used in text-based and image-based retrieval, while in 3D object retrieval there is still a lot of work to be done. What is of utmost importance is that an optimal relevance feedback approach should try to keep the low-level descriptors of the dataset objects unaltered. Otherwise, the efficiency of the method will significantly decrease, when it is applied to large-scale 3D model datasets, due to scalability issues.

The current work is expected to provide a useful reference for further research, as well as to contribute to shaping the future research directions in 3D object retrieval.

Acknowledgements

This work has been supported by the EC-funded project I-SEARCH (http://www.isearch-project.eu/isearch/).

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", *4th Eurographics Workshop on 3D Object Retrieval (3DOR 2011)*, April 10, 2011, Llandudno (UK).

[2] T.P. Vanamali, A. Godil, H. Dutagaci,T. Furuya, Z. Lian, R. Ohbuchi, In: M. Daoudi, T. Schreck, M. Spagnuolo, I. Pratikakis, R. Veltkamp (eds.), "SHREC'10 Track: Generic 3D Warehouse", *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on 3D Object Retrieval*, 2010.

[3] E. Paquet and M. Rioux, "Nefertiti: a query by content system for three-dimensional model and image databases management". *Image and Vision Computing*, 17(2), 157–166, 1999.

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

[5] Z. Lian, P. L. Rosin, and X. Sun, "Rectilinearity of 3D meshes", *International Journal of Computer Vision* Volume 89, Numbers 2-3, 130-151, DOI: 10.1007/s11263-009-0295-0.

[6] A. Axenopoulos, G. Litos, P. Daras, "3D Model

Retrieval using Accurate Pose Estimation and View-based Similarity", *1st ACM International Conference on Multimedia Retrieval* (ICMR2011), April 17-20, Trento, Italy.

[7] P. Daras, A. Axenopoulos, "A Compact Multi-View Descriptor for 3D Object Retrieval*" IEEE 7th International Workshop on Content-Based Multimedia Indexing* (CBMI 2009), Chania, Greece, Jun 2009.

[8] H.-P. Kriegel, P. Kroeger, Z. Mashael, M. Pfeifle, M. Poetke and T. Seidl, "Effective similarity search on voxelized cad objects". *In Proc. of IEEE eighth international conference on database systems for advanced applications*. Washington, DC, USA, 2003.

[9] P. Papadakis, I. Pratikakis, S. Perantonis, T. Theoharis, "Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation", *Pattern Recognition* 40 (9), pp. 24372452, 2007.

[10] D. Zarpalas, P. Daras, A. Axenopoulos, D. Tzovaras and M. G. Strintzis, "3D model search and retrieval using the spherical trace transform," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 23912, p. 14, 2007.

[11] R. Ohbuchi, K. Osada, T. Furuya and T. Banno, "Salient local visual features for shape-based 3D model retrieval", *In Proc. og the IEEE International Conference on Shape Modeling and Applications*, (SMI 2008), pp. 93-102, 2008.

[12] M. Chaouch and A. Verroust-Blondet, "3D model retrieval based on depth line descriptor", *IEEE International Conference on Multimedia & Expo (ICME'07)*, Beijing, China, July 2007.

[13] C. Akgul, A. Axenopoulos, B. Bustos, M. Chaouch, P. Daras, H. Dutagaci, T. Furuya, A. Godil, S. Kreft, Z. Lian, T. Napoleon, A. Mademlis, R. Ohbuchi, P. L. Rosin, B. Sankur, T. Schreck, X. Sun, M. Tezuka, Y. Yemez, A. Verroust-Blondet, M.Walter, "SHREC 2009 - Generic Shape Retrieval Contest" *30th International conference on EUROGRAPHICS 2009, workshop on 3D object retrieval*, Munich, Germany, Mar 2009.

[14] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura, Y. Kurita, G. Lavoué, H.V. Nguyen, R. Ohbuchi, Y. Ohkita, Y. Ohishi, F. Porikli, M. Reuter, I. Sipiran, D. Smeets, P. Suetens, H. Tabia, D. Vandermeulen: "SHREC'11 Track: Shape Retrieval on Non-rigid 3D Watertight Meshes". *Eurographics Workshop on 3D Object Retrieval* (2011), pp. 79-88, 2011.

[15] T. Funkhouser and P. Shilane, "Partial matching of 3D shapes with priority-driven search", *Proceedings Eurographics Symposium on Geometry Processing 2006*, pp. 131–142.

[16] B. Bonev, F. Escolano, D. Giorgi and S. Biasotti, "High-Dimensional Spectral Feature Selection for 3D Object Recognition Based on Reeb Graphs", *Lecture Notes in Computer Science*, 2010.

[17] S. Marini, G. Patane, M. Spagnuolo and B. Falcidieno, "Feature Selection for Enhanced Spectral Shape Comparison", *Eurographics Workshop on 3D Object Retrieval*, 2010.

[18] B. Bustos, D. Keim, D. Saupe, T. Schreck, D. Vranic, "Automatic selection and combination of descriptors for effective 3D similarity search". *In: Proc. IEEE international workshop on multimedia content-based analysis and retrieval*, 2004.

[19] D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen, P. Suetens: "Isometric Deformation Modelling for Object Recognition". *In Proc. CAIP'09* (2009), pp. 757–765.

[20] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens, D. Vandermeulen: "Feature detection on 3D face surfaces for pose normalisation and recognition". *In Proc. BTAS'10* (2010).

[21] A. Mademlis, P. Daras, A. Axenopoulos, D. Tzovaras, M. G. Strintzis, "Combining Topological and Geometrical Features for Global and Partial 3D Shape Retrieval", *IEEE Transactions on Multimedia*, Volume 10, Issue 5, Aug 2008.

[22] E. Onasoglou, P. Daras, "Semantic Force Relevance Feedback, Content-Free 3D Object Retrieval and Annotation Propagation: Bridging the Gap and Beyond", *SPRINGER Multimedia Tools and Applications Journal* (MTAP), Special Issue on Multimedia Semantics, DOI 10.1007/s11042-008-0216-1, Jul 2008

Apostolos Axenopoulos was born in Thessaloniki, Greece, in 1980. He received the Diploma degree in electrical and computer engineering and the M.S. degree in advanced computing systems from Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003 and 2006, respectively. He is an Associate Researcher at the Informatics and Telematics Institute, Thessaloniki. His main research interests include 3D object indexing, content-based search and retrieval and bioinformatics. Currently, he is a PhD candidate in Computer & Communication Engineering Department, University of Thessaly.

Petros Daras (M'07) was born in Athens, Greece, in 1974. He received the Diploma degree in electrical and computer engineering, the M.Sc. degree in medical informatics, and the Ph.D. degree in electrical and computer engineering, all from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1999, 2002, and 2005, respectively. He is a Researcher Grade C, at the Informatics and Telematics Institute (ITI) of the Centre for Research and Technology Hellas (CERTH). His main research interests include search, retrieval and recognition of 3D objects, 3D object processing, medical informatics applications, medical image processing, 3D object watermarking and bioinformatics. He serves as a reviewer/evaluator of European projects and he is a member of IEEE and key member of the IEEE MMTC 3DRPC IG.

3D Model Retrieval Using Semantically Rich Skeleton: A Review

R. Shen, T. Wang, L. Shi, X.D. Yang and I. Cheng
University of Alberta
{rshen, tw, lshi2, xingdong, locheng}@ualberta.ca

1. Introduction and Challenges of 3D Models Retrieval

Many Internet applications require real-time access to online 3D object repositories. An example application is similarity match and retrieval of 3D objects. 2D image retrieval attracted significant interests in the early ninetieths, and because of the deficiency in text annotation, new algorithms were introduced including the comparison of patch colors, region textures, object shapes, features extraction and so on. An important property of these techniques is transformation (in particular scale and rotation) invariant. While these techniques can also be applied on 3D objects, the advancement of 3DTV technology has added a new dimension to the requirements, bringing issues associated with higher dimensionality to the matching and retrieval problem. Due to the huge number of possible viewing perspectives of a 3D model in probably uncontrolled lighting conditions, similarity matching between two 3D models is more challenging than computations in 2D space.

To address the above issue, a dimension reduction technique – skeletonization, has been studied by researchers with immense interests. In this paper, we will focus on how a skeleton, which provides a semantically rich compact abstraction of the original 3D model, can be utilized for model similarity match, starting with an overview of skeleton generation in the next section.

2 Skeleton Generation

Skeletons can adopt different forms, for the purpose of 3D model retrieval, we focus on connected unit-width curve-skeleton, which defines the medial axis (loci of some centered points) of the 3D model. To generate a skeleton, the input format can be a mesh, point cloud or solid (voxels). Depending on the input format, a corresponding skeletonization algorithm can be employed. A few of these algorithms are discussed here.

Shen et al. [1] proposed a mesh-based technique to generate a hierarchy of skeletons from dense to coarse. An application can choose an appropriate level depending on the details required. The approach can also be extended to handle volume data, *e.g.* DICOM. Fig. 1 (b) to (d) shows three levels of details of a horse object (Fig. 1 (a)) from dense to coarse.

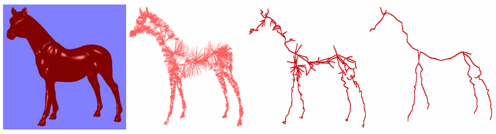


Fig. 1: *(a) original model, and (b) to (d) are three levels of detail from dense to coarse.*

This technique, inspired by the Sphere Sampling approach [2], uses overlapping maximal spheres to cover the free space inside the 3D object defined by the following steps:

1) Starting from a user-defined face on the mesh surface, compute its normal and project the reverse normal inside the object on the opposite surface.
2) Fit a maximal sphere along the path inside the object. Record the sphere center as a skeleton point.
3) A fixed number of sample points (children) are placed evenly on the circumference starting from a predefined orientation, based on a user-defined separation angle θ [3]. For each child, a maximal sphere is drawn. If the sphere touches at least two vertices on the surface, the child becomes a skeleton point and is linked to the center of the parent sphere. Also, a sphere is valid only if its radius is larger than a predefined threshold *R*.
4) The valid spheres are pushed into a priority queue in descending radius order.
5) A sphere is popped from the queue and Steps (3) to (5) are repeated until the queue is empty.

Yang et al. [4] proposed a voxel-based technique. Fig. 2 shows (a) the original model, (b) model covered with voxels on the surface after three iterations, and (c) the tightly bound voxels after running six iterations, giving a good approximation of the horse. This approach is similar to sphere tree construction [5], but the spheres are replaced with cubes (voxels). An octree structure is used and the idea is to generate cubes sufficiently small to cover the mesh surface obtaining a good approximation of the shape. The motivation is that using this approximation of the surface, we can locate the medial axis of the 3D model based on the regular geometry and placement of the cubes, and use the cube centers to compute a skeleton

.

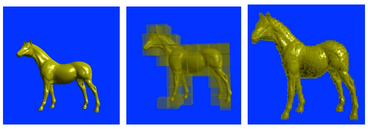


Fig. 2: *(a) original model, (b) surface approximated by cubes after three iterations, and (c) after six iterations.*

Among all connectivity preservation skeleton-ization methods, 3D thinning algorithms are generally faster than others. However, most 3D thinning algorithms cannot guarantee to generate unit-width curve skeleton, which is required by many applications. Wang et al. [6] presented an approach to generate fully parallel unit-width curve skeleton. A 3D thinning algorithm applies in a local neighborhood of an object point and iteratively removes object points that satisfy some pre-defined deleting masks to generate skeletons in a 3D binary image. A 3D binary image can be converted from a 3D mesh by voxelization (Fig. 3 (left)). The basic idea of voxelization is to make the hollow 3D mesh solid. The thinning operation iteratively deletes (removes) some object points (i.e., changes some black points to white) based on the deleting masks (Fig. 3 (middle)) until only some restrictions prevent further operation. Although 3D thinning algorithms have two major advantages (connectivity preservation and efficiency) over other skeletonization techniques, many 3D thinning algorithms [7] fail to generate unit-width curve skeletons (i.e., the skeleton is not one-voxel thick). For instance, in Fig. 3 (Right), there are some regions on the skeleton swarmed with many points. This is an undesired property [8] for many applications such as animation and virtual navigation that require unit-width curve skeleton as an input.

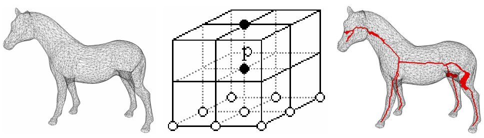


Fig. 3: *(Left) A 3D binary image (converted from and covered by a 3D mesh); (Middle) one deleting mask; (Right) the skeleton in the middle of the model.*

In order to preserve unit-width in the crowded region, there are four steps. In the first step, the degree of each object point is calculated. Then, the so-called crowded regions are identified and the "exits" of the each region are located. After that, the center of each crowded region is computed based on the proposed degree-based spatial median (DBSM) algorithm. In the last step, Dijkstra shortest path finding algorithm is applied to connect the exits with the center in each crowded region, and remove other object points that are not on any shortest path between the center and an exit in each region, to generate unit-width curve skeleton (Fig. 4)

.

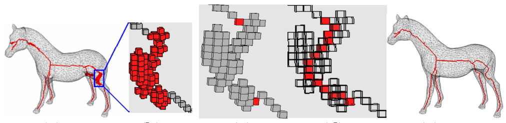


Fig. 4: *(a) Non-unit-width curve skeleton; (b) a crowded region; (c) two exits of this crowded region; (d) the shortest path; and (e) unit-width skeleton generated after performing algorithm.*

However, the presence of noise is a general concern in skeletonization algorithms including thinning. It is crucial to eliminate both global noise and local noise in skeleton branches, as well as tiny redundant sub-branches, in order to improve the overall skeleton representation. Shi et al. [9] applied scale-space filtering (SSF) in 3D skeletonization generating a spectrum of level-of-details.

3 Skeleton Matching Using Chain Code Expression

Once a semantically rich skeleton has been generated, representing a compact abstraction of the 3D object, 3D model matching is reduced to the problem of skeleton matching. Comparing with the highly dense 3D mesh surface, scattered points set and voxels volume, a unit-width curve skeleton provides more efficient real-time similarity match and retrieval. Lopez et al. [10] described using 3D chain expressions in encoding curve skeletons for measuring shape dissimilarity. By integrating a unit-width skeletonization technique and a graph-based approach, the proposed algorithm can compare not only the skeleton topology but also the curvature of skeleton segments. The method can distinguish between relatively similar objects.

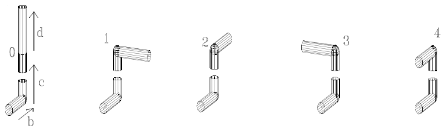


Fig. 5: *Chain code elements for 3D curve skeleton*

There are five chain code elements (Fig. 5) to encode a 3D curve skeleton using the values 0 to 4 forming a numeric string hierarchy. Suppose we take two orthogonalized (diagonals modified) skeletons S and P of lengths l_s and l_p, where $l_s \leq l_p$, and start to compare them one descriptor at a time. If the two descriptors are the same, we move to the next pair of descriptors. If they are not the same, we perform one of the following operations:

$\int_{s \to p}$ - Stretch S by inserting the corresponding descriptor from P.

$\int_{p \to s}$ - Stretch P by inserting the corresponding descriptor from S.

$\vartheta_{s \to p}$ - Bend the descriptor in S to match the descriptor in P.

$\vartheta_{p \to s}$ - Bend the descriptor in P to match the descriptor in S.

The algorithm tests each of the above by looking one step ahead. That is, after performing the operation to check whether the next pair of descriptors matches, the operation that generates the next match will be chosen. If more than one operation generates the next match, the algorithm checks one more pair of descriptors ahead, and so on. If no operation generates the next match, then the default operation is $\int_{s \to p}$. In the worst case, the maximum number of stretch and bend operations combined is bounded by $l_s + l_p$. At each step, there are 4 operation checks resulting in a maximum of $4 \times (l_s + l_p)$ total operations. The look-ahead test is bounded by $l_s + l_p$. In order to evaluate the dissimilarity, we count the following at each step:

$\#S$ - Increase by 1 if the pair of descriptors is the same.

$\#\int$ - Increase by 1 if either $\int_{s \to p}$ or $\int_{p \to s}$ is performed.

$\#\vartheta$ - Increase by 1 if either $\vartheta_{s \to p}$ or $\vartheta_{p \to s}$ is performed.

The dissimilarity score in the range [0, 1], with 0 representing a complete match, is given by the following function:

$$D(S,P) = \frac{\#\vartheta + \#\int}{\#S + \#\vartheta + \#\int} \qquad (1)$$

4. Conclusions

This paper reviews the generation and some development of skeletons (medial axis), which is a compact abstraction of the original 3D model, for similarity match and object retrieval. Challenges in robust skeleton generation are discussed with suggested solutions. Readers who are interested in this topic can obtain details in the literature and explore further. We believe unit-width curve skeleton can be valuable in many aspects, which include quality of experience (QoE) modeling, but there are still technical issues to be resolved before skeletons can be effectively utilized as a semantically rich compact abstraction of 3D objects.

References

[1] R. Shen and I. Cheng, "Tracking Shape Change using a 3D Skeleton Hierarchy," SIGGRAPH Jul/Aug 2006, Abstract & Poster Session, Boston, USA.

[2] Y. Yang, O. Brock and R. Moll, “Efficient and Robust Computation of an Approximated Medial Axis,” In *Proc. of the ACM Symposium on Solid Modeling and Applications*, pp.15-24, Jun2004.

[3] M. Foskey, M. Lin and D. Manocha, “Efficient Computation of a Simplified Medial Axis,” In *Proc. of the 8th ACM Symposium on Solid Modeling and Applications*, pp 96-107, 2003.

[4] X. Yang and **I. Cheng**, "3D Skeletonization using an Enhanced Voxel Tree," SIGGRAPH Jul/Aug 2006, Abstract & Poster Session, Boston, USA.

[5] G. Bradshaw and C. O’Sullivan, “Sphere-Tree Construction using Medial-Axis Approximation,” In *Proc. of SIGGRAPH* 2002.

[6] T. Wang and I. Cheng, “Generation of Unit-width Curve Skeletons based on Valence Driven Spatial Median (VDSM),” Int’l Symposium on Visual Computing, Dec 2008, Las Vegas, USA (LNCS Series, 10 pages, 2008, Springer-Verlag Berlin Heidelberg).

[7] C. Lohoua and G. Bertrand. A 3D 6-subiteration curve thinning algorithm based on P-simple points, Discrete Applied Mathematics, Vol. 151, pp 198–228, 2005.

[8] N. D. Cornea. Curve-Skeletons: Properties, Computation And Applications, Ph.D. Thesis, The State University of New Jersey, May 2007.

[9] L. Shi, I. Cheng and A. Basu, “Anatomy Preserving 3D Model Decomposition Based on robust Skeleton-Surface Node Correspondence,” IEEE Int. Conf. on Multimedia (ICME), 6 pages, 2011.

[10] V. Lopez, I. Cheng, E. Bribiesca, T. Wang and A. Basu, “Twist-and-Stretch: A Shape Dissimilarity Measure based on 3D Chain Codes,”

SIGGRAPHASIA Dec 2008 (Abstract & Poster Session), Singapore.

Rui Shen (S'07) received the B.Eng. degree in computer science and technology from Beihang University, Beijing, China, and the M.S. degree in computing science from the University of Alberta, Edmonton, AB, Canada. He is currently working toward the Ph.D. degree in computing science at the University of Alberta. He is the author or coauthor of eight papers published in international conferences and journals. His current research interests include probability models, sensor fusion, stereo vision, and image processing. Mr. Shen is the recipient of the iCORE Graduate Student Scholarship in the Information and Communications Technology and the IzaakWalton Killam Memorial Scholarship.

Tao Wang (PhD) is working at the Data Mining Group at the SAS Institute, North Carolina. He received B.Eng. and M.Eng. from the Department of Computer Science and Engineering, Zhejiang University, Hangzhou, Zhejiang Province, China. He finished his Ph.D. in the Department of Computing Science at the University of Alberta, Edmonton, Alberta, Canada in 2009.

He worked as a software engineer in the broadband department at UTStarcom, Hangzhou, Zhejiang Province, China, from 2003 to 2004. His current research interests include active contour techniques and 3D skeletonization. He has many publications in international conferences and journals. Mr. Wang has won a number of international, national, provincial, institutional, and departmental awards for his academic achievement.

Liang Shi received the master degree in 2011 from the Computing Science Department of the University of Alberta, Canada. Her bachelor is in Communication Engineering from School of Information and Telecommunication at Beijing University of Posts and Telecommunication. She was a research intern at the Institute of Automation Chinese Academy of Science from 2008 to 2009, focusing on image summarization, object recognition and content adaptation. Currently, she is the Lead Engineer at Scope Technologies, developing brand-new gaming platform on mobile phones. In the recent Start-up Weekend Edmonton, she established Sizmio, a geo-based audio sharing site.

Xing-Dong Yang is a Ph.D. student in Human-computer Interaction at the University of Alberta, Edmonton, AB, Canada. He received his M.Sc. Degree in Haptics user interfaces from the same university. Xing-Dong Yang's research interests are novel human-computer interaction techniques, user performance modeling, and haptic user interfaces. His Ph.D. thesis focuses on designing and evaluating 3D gesture-based interactions on various platforms, including mobile devices, desktop systems, tabletop systems, and wall-size displays.

Irene Cheng (M'02–SM'09) received her Ph.D. degree in Computing Science from the University of Alberta, Canada. She is now the Scientific Director of the Multimedia Research Group, and an Adjunct faculty in the same department. She also holds an Adjunct position in the Faculty of Medicine and Dentistry, and

Research Affiliate position in Glenrose Rehabiliation Hospital. She was the Chair of the IEEE NCS EMBS Chapter (2009-11), the Chair of the Communications Society MMTC 3DRPC Interest Group and a board member of the SMC TC on Human Perception in Vision, Graphics and Multimedia. Dr. Cheng was a General Chair in IEEE ICME 2011 and is awarded visiting professorship at INSA Lyon in 2011. She has two books and more than 100 papers published in international journals and peer-reviewed conferences. Her research interests include incorporating human perception—Just-Noticeable-Difference—following psychophysical methodology, to improve multimedia transmission techniques. She is also engaged in research on 3-D TV and perceptually motivated technologies in multimedia, including online education.

Considering Implicit Channels in Privacy Analysis of Video Data

Mukesh Saini[*], Pradeep K. Atrey*[†], Sharad Mehrotra*[ϕ] and Mohan S. Kankanhalli*[*]*
[*]*National University of Singapore, Singapore*
†University of Winnipeg, Canada
[ϕ]*University of California, Irvine, USA*
**{mksaini,mohan}@comp.nus.edu.sg, †p.atrey@uwinnipeg.ca, [ϕ]sharad@ics.uci.edu*

1. Introduction

With increasing number of security threats, people's safety is of utmost importance. Video surveillance is considered very useful tool for ensuring safety of citizens. Video surveillance systems employ multiple cameras to monitor activities over large areas [1]. However, this additional safety and security comes at the cost of privacy loss of citizens who are not involved in any illicit activities [2]. Due to its invasive nature, it is very hard to perform effective video surveillance, while still preserving the privacy of individuals.

Privacy is a big concern in current video surveillance systems. Due to privacy issues, many strategic places remain unmonitored leading to security threats. With respect to surveillance video, there are mainly two places where privacy loss could occur, when security personnel are watching the video currently being captured by the cameras; and when the recorded video is disseminated for forensics and other research purposes. For both the cases, the first step towards privacy protection is to analyze the characteristics of the video which cause privacy loss (privacy modeling) and second step is to modify the video data (data transformation), such that the privacy is preserved, yet intended surveillance goals can be achieved. To accomplish these steps, we need to model and quantify privacy loss and utility of the video data.

2. Privacy Loss

Privacy loss occurs when an adversary is able to obtain information about individuals in the video, which they otherwise do not want to share with others. Thus there are two contributing factors of privacy loss: 1) Identity leakage – the certainty with which an adversary can identify individuals in the video; 2) Sensitivity Index – a quantitative measure of sensitive information in the video, which generally depends on individual's habits, preferences, and moral views [4]. For example some people consider the places they visit as sensitive, while others are sensitive to their companion information when they are visiting particular places. Figure 1 shows the overview of the privacy loss calculation for a given video. Privacy loss occurs when the sensitivity index and identity leakage both have significant values.

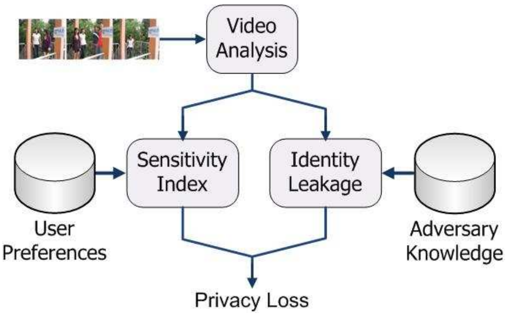


Fig.1. Contributing factors in privacy loss calculation.

Privacy loss can be reduced by either removing sensitive information or hiding identity. In many applications such as surveillance, it is very hard to remove sensitive information; therefore, main emphasis is given to preserving identity information. The identity leakage can occur due to *explicit inference channels* such as facial information or *implicit inference channels* that exist due to contextual knowledge of the adversary.

3. Implicit Inference Channels and Identity Leakage

Traditionally, identity leakage is modeled as a binary variable. If facial/bodily information is present, the identity leakage is considered unity; otherwise it is considered zero. The past works [5-8] account only for explicit inference channels of identity leakage (e.g. face). We identify that while blocking these explicit channels is necessary for hiding identity information, it is not sufficient. The adversary could still infer the identity by observing the idiosyncrasies of the individuals. The adversary can look at the places visited, observe the behavior, and analyze this information on time axis to infer identities [9].

For example, a person entering Jaya's office in early morning is very likely to be Jaya. Similarly, a person spending sufficient time around a shop is likely to be one of the staff members. Further, if the adversary has prior knowledge that only one person is short among the staff; s/he can almost certainly infer the identity of the person. In this way, the *implicit inference channels* due to *what* (activities), *when* (time), and *where* (location) information can cause significant identity leakage. Figure 2 shows an example of explicit and implicit inference channels.

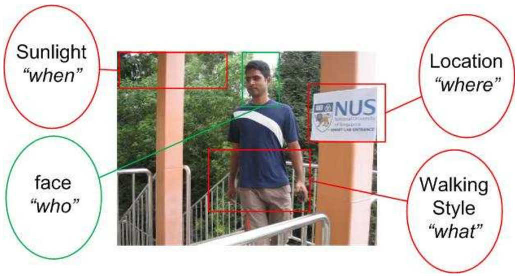


Fig. 2. The green areas show explicit inference channels while red areas are implicit inference channels.

While facial information uniquely identifies a person, the implicit inference channels generally associate identity to a group of people. The size of an association group is determined based on the video content and adversary's prior knowledge. The size of the association group gives *anonymity*, while its inverse is *implicit identify leakage* i.e. the certainty with which the adversary can associate the identity to the individual in the video.

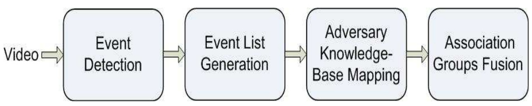


Fig. 3. The framework for measuring *implicit identity leakage.*

The framework for anonymity based identity leakage detection is shown in Figure 3. The video is analyzed to detect events. The events are divided into event lists corresponding to each individual in the video. These events are then compared with the adversary's knowledge-base to find matching event patterns. These event patterns equip the adversary to associate the identity of each individual to a group of people. Finally, these association groups are fused to obtain the overall anonymity and identity leakage.

4. Applications

The implicit identity leakage highlights privacy breach in current privacy protection methods for video data. It indicates that the usually ignored background information can cause significant privacy loss. Therefore, for robust privacy protection, we need to transform regions belonging to both background and foreground. However, this can affect the utility of the data and we need to have computational model for utility to study the tradeoff between the two quantities.

Video Data Publication

A prevailing application of the ever increasing video data is to publish it for research purposes. The challenge here is to identify all the image regions that can cause identity leakage, and subsequently transform such that both privacy loss and utility loss are minimized. Analogous to the relational data publication scenario, the utility can be computed by comparing the accuracy of the intended tasks before and after data transformation [10]. Further, different transformation methods, such as quantization, blurring, pixelization etc., can be compared to obtain the best tradeoff between privacy and utility.

CCTV Surveillance and Monitoring

CCTV monitoring scenario imposes additional challenges on privacy protection. Due to familiarity with the monitoring site, the CCTV operator can easily infer identities of the individuals. If the operator acts as an adversary, s/he can obtain sensitive information about the habitants and cause privacy loss. This restricts the placement of cameras at many strategic locations such as private premises, offices, etc. It is important to decouple the strong context knowledge of the adversary in order to hide the identities [11]. However, this context knowledge is required for effective surveillance; therefore, this is a very hard problem.

5. Conclusions & Future Challenges

In this letter we have investigated privacy issues associated with video data from a novel perspective. We have discussed that *implicit identity leakage* channels due to *what, when,* and *where* information can cause privacy loss and need to be blocked for robust privacy protection. It is discussed that to study the tradeoff between privacy and utility, appropriate computational models for privacy loss and utility loss are required. Two application scenarios of video data application and CCTV monitoring are discussed as case studies.

Whether the information in the video can cause privacy loss or not depends heavily on the adversary's knowledge. Therefore an accurate model of the adversary's knowledge is very useful to provide more precise privacy loss. Following are the research issues related to the adversary knowledge modeling: 1) How to measure the adversary's knowledge? 2) How does the knowledge evolve over time? For privacy loss calculation, it is important to know what is already known to the adversary.

Another crucial factor in privacy measurement is accuracy of the event detectors. Given that state-of-the art detectors are not precise enough, the privacy loss needs to be calculated following pessimistic approaches. It should be noted that this can increase the number of false alarms; therefore, it is important to study its effect on the utility of the data. Current data transformation methods mainly focus on hiding the identity; however, privacy loss can also be reduced by controlling the sensitivity index. It would be interesting to study automatic detection and removal of sensitive information and its impact on data utility. Finally, the privacy protection methods may cause additional computational load on the surveillance systems; which needs to be dynamically distributed [12] in order to pursue effective real-time surveillance.

References

[1] M. Valera and S. Velastin, "Intelligent distributed surveillance systems: a review," *In IEE Proceedings on Vision, Image and Signal Processing*, 2005, pp. 192–204.

[2] A. Cavallaro, "Privacy in video surveillance," *Signal Processing Magazine*, IEEE, 24(2):168 –166, 2007.

[3] C. Dwork, "Differential privacy," *In International Colloquium on Automata, Languages and Programming*, pages 1–12, 2006.

[4] M. Langheinrich, "Privacy by design - principles of privacy-aware ubiquitous systems. *In Proceedings of the 3rd International Conference on Ubiquitous Computing*, pages 273–291. Springer-Verlag, 2001.

[5] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacyprotecting data collection in media spaces," *In The 12th ACM International Conference on Multimedia*, pages 48–55, New York City, USA, 2004.

[6] F. Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," *In Proceedings of SPIE*, 2011.

[7] J. Schiff, M. Meingast, D.K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: Detecting visual markers in real-time to address privacy concerns," *Protecting Privacy in Video Surveillance*, pages 65–89, 2009.

[8] M. Barhm, N. Qwasmi, F. Qureshi, and K. el Khatib, "Negotiating privacy preferences in video surveillance systems," *Modern Approaches in Applied Intelligence*, pp. 511–521, 2011.

[9] M. Saini, P. Atrey, S. Mehrotra, S. Emmanuel, and M. Kankanhalli, "Privacy modeling in video data publication," *In IEEE International Conference on Multimedia and Expo*, pp. 60–65, Singapore, 2010.

[10] B. Fung, K. Wang, R. Chen, and P. Yu, "Privacy-preserving data publishing: A survey on recent developments," *In ACM Computing Surveys*, volume 42, 2010.

[11] M. Saini, P. Atrey, S. Mehrotra, and M. Kankanhalli, "Anonymous surveillanc," *In IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety with IEEE International Conference on Multimedia & Expo*, Barcelona, Spain, July, 2011.

[12] M. Saini, W. Xiangyu, P. Atrey, and M. Kankanhalli, "Dynamic workload assignment in video surveillance systems," *In Proceedings of the IEEE International Conference on Multimedia and Expo*, Barcelona, Spain,2011.

Mukesh Saini is a PhD candidate at School of Computing, National University of Singapore (NUS). He has submitted his thesis, and currently he is working as a Research Associate on the project "NExT Search Center" at NUS. Prior to PhD, he obtained Master of Technology (M. Tech) in Electronics Design and Technology from Indian Institute of Science, Bangalore. During his PhD stint, he visited University of Winnipeg for six months (Sep 2009 to Feb 2010) under the Canadian Commonwealth Exchange Program. His PhD dissertation is on the cutting-edge topic "Privacy-aware multimedia surveillance". His other research interests include generic architecture for observation systems and system level performance evaluation of multimedia systems. He has played roles of Reviewer, TPC member, Tutorial organizer, and Panelist for various reputed conferences and journals.

Pradeep K. Atrey is an Assistant Professor and a Senator at the University of Winnipeg, Canada. He received his Ph.D. in Computer Science from National University of Singapore, M.S. (Software Systems) and B.Tech. (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 Web. He has authored/co-authored over 55 research articles at reputed ACM, IEEE, and Springer journals and conferences. Dr. Atrey is on the editorial board of ETRI Journal and Journal of Convergence (Web and Multimedia). He is actively involved in his research community and he has been associated with over 20 international conferences in various roles such as General Chair, Program Chair, Publicity Chair, Web Chair, and TPC Member. Dr. Atrey was 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".

Sharad Mehrotra is a Professor in the School of Information and Computer Science at University of California, Irvine (UCI) and Director of the Center for Emergency Response Technologies (CERT) at UCI. Mehrotra's research expertise is in data management and distributed systems areas in which he has made many pioneering contributions. Two such contributions include the concept of "database as a service" and "use of information retrieval techniques, particularly relevance feedback, in multimedia search". Mehrotra is a recipient of numerous best paper nominations and awards including SIGMOD Best Paper award (2001), Best of VLDB submissions (1994), and best paper award in DASFAA (2004). Mehrotra's current research focuses on building sentient spaces using multimodal sensors, data privacy, and data quality. Mehrotra's recent research, particularly, in the context of CERT has focused on situational awareness from multimodal input including conversational speech data. Many of his research contributions have been incorporated into software artifacts which are now in use at various first responder partner sites. He holds a patent on privacy protection of data collection in pervasive environments.

Mohan Kankanhalli is a Professor at the Department of Computer Science of the National University of Singapore. He is also the Associate Provost for Graduate Education at NUS. Before that, he was the Vice-Dean for Academic Affairs and Graduate Studies at the NUS School of Computing during 2008-2010 and Vice-Dean for Research during 2001-2007. Mohan obtained his BTech (Eletrical Eng.) from the Indian Institute of Technology, Kharagpur, in 1986 and his MS and PhD (Computer $ Systems Eng.) from the Rensselaer Polytechnic Institute in 1998 and 1990, respectively. He was a researcher at the Institute of Systems Science at NUS during 1990-1997. He then became a faculty member at the Department of Electrical Engineering of the Indian Institute of Science, Bangalore during 1997-1998 after which he joined NUS again. He visited the University of California at Berkeley during Jan-Jun 2004. He is actively involved in the Multimedia Systems community and he is currently the Director of Conferences for ACM SIG Multimedia. He is on the editorial boards of several journals including the ACM Transactions on Multimedia Computing, Communications, and Applications, Springer Multimedia Systems Journal, Pattern Recognition Journal and Multimedia Tools & Applications. His current research interests are in Multimedia Systems (content processing, retrieval) and Multimedia Security (surveillance, digital rights management and privacy).

TECHNOLOGY ADVANCE COLUMN

Multimedia Communications for Ubiquitous Healthcare

Jianfeng Wang
Philips Research North America, USA
jianfeng.wang@philips.com

Healthcare industry is undergoing extraordinary revolution. The use of electronic medical records (EMR), telehealth and home healthcare would become norm. The goal to provide anytime, anywhere, and personalized healthcare services challenge network and service mobility, availability, reliability, and security/privacy from transmission, storage to data mining of patient data.

The article "A Look at Future Mobile Healthcare Applications" describes various healthcare scenarios from home to hospital which require mobility support and operation over heterogeneous networks. The author presents an overview of the application requirements in each scenario. As different networking technologies have vastly different communication characteristics, such as coverage area, data rate, delays and power consumption, different networks and protocol stacks are expected to be used for different applications at different places. This article then proposes an architecture which simplifies mobile healthcare applications to operate in such environments.

The article "E-Doctor: A Real-time Home Monitoring and Mobile Healthcare Platform" introduces an integrated real-time home monitoring and mobile healthcare platform, called E-Doctor. E-doctor can monitor patients' body condition continuously, transmit the health data in real-time, automatically generate alarm signal and allow the doctors diagnose remotely. It also maintains a comprehensive healthcare database for every patient. Various monitors on one patient are connected by one smart phone forming a star-topology, where the smart phone is the gateway and monitors are clients. Bluetooth technique is adopted as the transmission protocol between the monitors and smart phones.

In the article "Privacy-preserving Wireless Data Transmission for e-Healthcare Applications", the privacy issues are presented and some state-of-art solutions are reviewed. The patient privacy issues are categorized into two types: data privacy and communication privacy. Patients must be able to select the access policy for their own personal information and apply distinct access policies to their medical records in different situations. In addition, to preserve identity privacy, location privacy, and contextual privacy, the communication protocols must resist an attack which aims to link the source and the destination of a transaction. The article also discusses the trade-off between safety, privacy, and access control in different contextual cases.

While EHR-based healthcare systems offer convenience and flexibility such as anytime anywhere medical record access, they also put patients' medical data privacy into great danger. Privacy breaches can be incurred in the handling of the EHRs, e.g., storage, retrieval, transfer, update, where secure storage plays an important role because it is the foundation for all the other operations to be carried out securely.The article "Data Privacy and Secure Storage for Electronic Health Record" presents a cloud based solution for protecting EHR data privacy while making it convenient for secure storage and emergency treatment. Patients can obtain dedicated service from the cloud service provider to set up their personal clouds. The personal clouds process patients' EHRs (e.g., encryption), store the result on the public cloud, and retrieve the EHRs upon physicians' requests after proper authentication.

Jianfeng Wang is a Senior Member of Research Staff at Philips Research North America. He received his Ph.D. in electrical and computer engineering from University of Florida in 2006.

His research interests include cognitive radio, medical body area networks, wireless telehealth, M2M, and smart grid communications. His work led to over thirty publications in journals and conferences. In his role as editor and task group leader, he actively contributed to Ecma 392 and IEEE 802.22. He and his team received the Frost and Sullivan North American Cognitive Networks Excellence in Research of the Year Award in 2007.

A Look at Future Mobile Healthcare Applications

Amjad Soomro, Philips Research North America, USA
amjad.soomro@philips.com

1. Introduction

Healthcare practice is undergoing extraordinary and rapid changes now. For example, use of electronic medical records (EMR), telehealth technologies and home healthcare would become norm rather than an exception in future. Provisioning of anytime, anywhere healthcare services imply mobile patients, mobile caregivers and mobile applications. Patient borne sensors and caregivers will be integrated into an EMR everywhere. As different networking technologies have vastly different communication characteristics, such as coverage area, data rate, delays and power consumption, different networks will be used for different applications at different places. More than one networking technology will be used by enabling operation over heterogeneous networks to provide seamless connectivity in a reliable manner to form a tightly integrated system for patient safety.

2. Healthcare Delivery Scenarios

In this section we categorize different healthcare scenarios, list wireless networking technologies and class of applications used in each.

Hospitals: In hospitals web-based applications are used to provide: a) administrative functions such as billing and discharge; b) medical applications such as lab reports and electronic medical records. Also, there are custom IT-like applications such as PACS which may not be web based. IT type applications are characterized by high bandwidth requirements with high tolerance of delays and immunity to some data loss in network layers. In contrast real-time life critical applications such as remote monitoring of physiological parameters, such as ECG, have low data rate with low tolerance to packet loss in network layers. Other applications such as remote control of medical devices, such as infusion pumps, require very low date rate capabilities and very high data reliability with no data loss at the application layer. Some of the applications require continuity of sessions as patients are allowed to ambulate or when they are moved around in hospitals.

In hospitals, traditionally proprietary dedicated wireless networks were used to carry life-critical real-time streaming data, however, more recently Wi-Fi networks are being used either exclusively to carry life critical data or in wireless spectrum configured to be shared with IT applications by using IEEE 802.11e QoS features to provide priority to life critical data. In addition to Wi-Fi, WPAN technologies such as Zigbee is used for shorter range communication, for example, to connect physiological sensors on the body to a body worn concentrator which further transmits the data to the wireless infrastructure.

Homes: In the coming decades it is expected that the proportion of elderly population will grow. Shortage of clinicians coupled with rising healthcare costs will lead to many non-critical healthcare services being provided at home, for example elderly care and diabetes management. Typical applications include web-based transfer of data to caregivers or clinicians, potential voice calls and video conferencing sessions with them, automatic alerts to caregivers for emergency conditions, for example, by a fall detector. To provide greater flexibility for elderly, they may be also monitored while away from home, thus making use of cellular technologies and mobility support a requirement for these scenarios.

In home traditional land line telephones (POTS) has been used to transfer patient data to clinicians. Availability of Wi-Fi residential gateways in homes is increasing and more home monitoring systems will have the capability to use Wi-Fi as well. Cellular 3G/4G systems are being increasingly used for data transfer activities and, correspondingly, it is expected that home monitoring will evolve to use this capability as well. Availability of wider bandwidths in Wi-Fi and cellular 4G technologies will make video conferencing sessions more prevalent in home care scenarios. Most of the applications in this scenario are used to push data to caregivers, EMRs (electronic medical records) or to send alerts. However, some may require session initiation from remote end, for example, in response to a fall detector alert received from an elderly patient, a call center may initiate a voice callback to check on his or her status.

Mobile Caregiver: In this scenario mobile caregiver receives medical alerts, accesses patient EMRs, talks or participates in teleconferences with colleagues and patients anytime, anywhere. For example, in one scenario an emergency responder responds to an emergency call and provides immediate care to patient with chest pain at home. A heart monitor is attached to the patient and the ECG and other vitals are sent to another remote specialist. An audio or video session with the remote specialist is established over a portable device for remote consultation. During the session the patient EMR is shared between the caregivers. Also, patient's next of kin are informed about the patient condition by text messages.

In this scenario, wireless technologies such as cellular 3G/4G or Wi-Fi technologies are expected to provide connectivity to the internet. The medical data is exchanged using web-based applications or via remote collaboration applications. There are sessions for real-time streaming, video and voice. The sessions could be initiated from either end. Presently, voice calls are typically carried through a protocol stack provided by cellular companies which is different and distinct from TCP/IP stack used by computer applications.

Telehealth: In this emerging scenario, healthcare services are provided by clinical practitioners to patients who are in remote location; typically staffed by generalists or trained technicians. A high definition video teleconferencing session is used to have interactive session with patients and, optionally, it is connected with diagnostic tools to provide real-time audio or video data to the remote clinician to aid in diagnosis. Most of the time telehealth facilities are stationary; however, they could be mobile, for example, in mobile clinics which tour remote sparsely populated areas.

Two-way high definition web conferencing and data transfer are more important applications in this scenario. High bandwidth wired networks, or wireless networks such as Wi-Fi or 4G cellular networks could satisfy the application requirements in this scenario.

3. Mobility in Different Scenarios

The application requirement and capabilities of different types of wireless networks are discussed in [1, 2]. In this article we will discuss an application architecture which lends itself to operation over heterogeneous networks regardless of underlying protocols used. For example, in the current architecture, as shown in Figure 1, voice and messaging applications are carried through cellular stack but video, data and real-time streaming applications are carried over TCP/IP stack. The current architecture requires applications to be aware of the available client network connections to use appropriate protocol and network interface to establish a voice session. This increases application complexity.

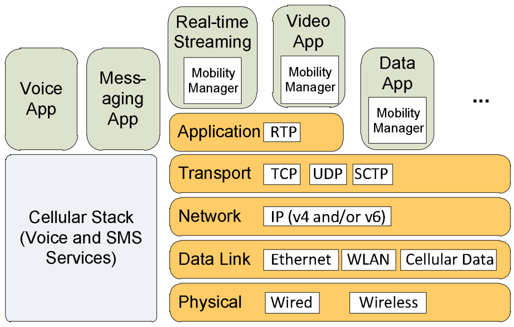


Figure 1. Different networks use different protocols to transfer multimedia data.

Future wireless networks are expected to be all IP-based, therefore, an application framework conforming to this architecture will simplify applications which are expected to be mobile and operate over heterogeneous networks. We propose to use a middleware to include mobility management, session management and real-time transport in one intermediate layer. The intermediate layer is based on IETF SIP [3] protocol and it provides a common API to applications regardless of the type of the underlying network. The proposed architecture is shown in Fig. 2.

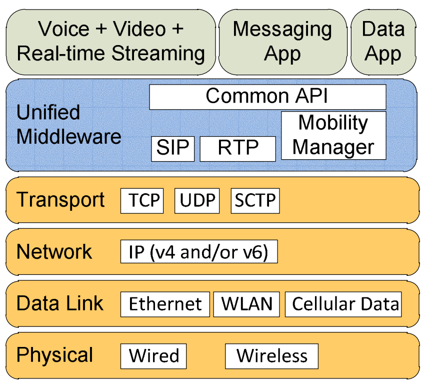


Figure 2. Applications use a common API to transfer multimedia data regardless of the underlying network type.

In this architecture, the end-points are addressed by a SIP URI, which is similar to an email address. By using a SIP URI an application's identity is dis-associated from its physical location. This architecture allows additional benefits such as handover over heterogeneous networks and sessions which are initiated from a central location. The reader is referred to [4] for more details.

4. Conclusions

In this article we described different healthcare scenarios which require mobility support and operation over heterogeneous networks. We present an overview of the application requirements in each scenario and the associated wireless networks which are expected to be used. Mobility requires operation over heterogeneous networks, and an architecture is proposed which simplifies mobile healthcare applications to operate in such environments.

References

[1] R. Istepanian, E. Jovanov and Y. Zhang, "Guest Editorial Introduction to the Special Section on M-Health: Beyond Seamless Mobility and Global Wireless Healthcare Connectivity", IEEE Trans. Info. Tech. Biomedicine, vol. 8, no. 4, Dec. 2004.

[2] A. Soomro and D. Cavalcanti, "Opportunities and Challenges in Using WPAN and WLAN Technologies in Medical Environments", IEEE Communications Magazine, pp. 114-122, Feb. 2007.

[3] J. Rosenberg *et. al.*, "SIP: session initiation protocol", Internet Engineering Task Force, RFC 3261, Jun. 2002, [Online], Available: http://www.rfc-editor.org/rfc/rfc3261.txt

[4] A. Soomro and R. Schmitt, "A framework for mobile healthcare applications over heterogeneous networks", *Proceedings of IEEE International Conference on e-Health Networking, Applications and Services* (IEEE Healthcom 2011), Columbia, MO, Jun. 2011.

Amjad Soomro (amjad.soomro@philips.com) (M'88–SM'05 IEEE) is Senior Member Research Staff at Philips Research North America, Briarcliff Manor, NY since July, 1999. Prior to that he worked as R&D Engineer in digital signal processing group at Hewlett Packard, Spokane Division, Spokane, WA. He received his Ph.D. degree in electrical engineering from U. of Maryland at College Park, College Park, MD in 1994. From 1987 to 1990, he worked as Communications Systems Programmer at Riyad Bank, Saudi Arabia. He received his B.S. and M.S. in electrical engineering from King Fahd U. of Petroleum and Minerals, Dhahran, Saudi Arabia in 1984 and 1987, respectively. His research interests are wireless communications, computer communications and digital signal processing.

E-Doctor: A Real-time Home Monitoring and Mobile Healthcare Platform

Jin Zhang and Qian Zhang
Hong Kong University of Science and Technology, Hong Kong
{jinzh, qianzh}@ust.hk

1. Introduction

With the newly development of machine-to-machine (M2M) technique and biomedical sensing technique, remote health care application is becoming a hot topic recent years. Both academic and industrial researchers become interested in this field because of remote healthcare's significant social impact and benefit. Using advanced biomedical sensors, the body condition of patients and aging persons can be monitored at home. Through wireless communication, their doctors and families can monitor their body condition in real time remotely and take suitable treatment and clinical care. It is estimates that more than 200 million people in the EU and the US suffer from one or several diseases where home monitoring can become a treatment option. In that case, they don't need to visit the hospital regularly, which saves a lot of time and money. Meanwhile, as continuous monitoring can discover symptoms earlier, it helps preventing the disease from growing severely. Therefore, home monitoring and remote healthcare can significantly improve people's life quality, especially for chronic patient and aging people.

There have been several existing studies on remote monitoring [2-8], however, most of them focus on the wireless body area network technique, without providing an integrated end-to-end home monitoring and mobile healthcare platform. Disease management and diagnosis after the signal transmission are not considered. However, this component is as important as the medical signal monitoring in remote healthcare.

In this paper, we will introduce our novel integrated real-time home monitoring and mobile healthcare platform, called E-Doctor. E-doctor can monitor patients' body condition continuously, transmit the health data in real-time, automatically generate alarm signal and allow the doctors diagnose remotely. It also maintains a comprehensive healthcare database for every patient.

2. System Architecture

The E-Doctor platform consists of three key components, real-time medical data collection, remote disease management and healthcare data center.

Medical data are observed and collected in real time by various biomedical monitors, which are placed on top of patients' body. These medical data are monitored at home and are transmitted to the remote healthcare data center. Remote disease management system then read the data from the healthcare data center. Leveraging the remote disease management system, the doctors and nurses can learn the patients' condition and diagnose remotely. They can also interact with the patient or his/her family by text, audio or video communication. The healthcare data center is a central database, where all the patients' monitoring data are stored and the treatment and medication records are saved. The healthcare data center also provides various operations on the database such as search, sort, and other statistic tools. Leveraging which, patients or doctors can easily trace one's health condition by viewing the medical condition during a long duration or derive certain relationships between certain symptom and the disease.

3. Real-Time Medical Data Collection

Real-time medical data collection component takes charge of the original medical data collection and transmission.

Various kinds of monitors are developed to monitor different biomedical parameter. For example, blood pressure monitor, electrocardiography (ECG) monitor, blood oxygen monitor, air flow monitor, glucose monitor observes and collects the blood pressure, ECG, blood oxygen, respiration and glucose signals respectively. All these monitors are small-size and easy to wear. Patients can wear them or use them any time as they work in office or stay at home for medical signal monitoring.

E-Doctor delivers the medical data from the monitors to the data center based on machine-to-machine (M2M) technique. Each healthcare monitor is equipped with a wireless communication module, which can communicate with the corresponding module in a smart phone.

Various monitors on one patient are connected by one smart phone forming a star-topology, where the smart phone is the gateway and monitors are clients. All the monitoring data are transmitted out of the body area network through the gate way. Many monitors can communicate with and transmit data to the gateway simultaneously.

In E-doctor, Bluetooth [1] technique is adopted as the transmission protocol between the monitors and smart phones. Compared with other techniques such as WiFi or Zigbee, Bluetooth consumes lower transmission power while supporting a high enough data rate. Low energy consumption and real-time data transmission are two basic requirements for healthcare monitor design. The monitor should work as long as possible in one power charging cycle. It should also be able to support sufficient bandwidth to transmit the monitor data in real-time. As WiFi has large power consumption, and Zigbee cannot support high enough transmission rate, Bluetooth is a suitable choice for monitor data transmission. As for the transmission from the gateway smartphone to the remote healthcare data center, GPRS or 3G technique are adopted. As mobile 2G/3G network is ubiquitous, such technique provide everywhere coverage of the remote monitoring system.

4. Remote Disease Management Platform

While the real-time medical data collection system transmitted the signals from the patient at home to the remote healthcare data center, the remote disease management platform provides auto-alarm function as well as diagnosis and real-time interaction function between doctors and patients.

Remote disease management platform can send auto-alarm signal when abnormal conditions happened. The platform reads original healthcare signals from the data center and analyzes them in real time. When abnormal conditions happened, it sends auto-alarm signals to both doctors and patients' family. It also provides both a web application and a smart phone application for them to check the patient's current status either on desktop computer or smart phone, helping to diagnose the disease.

The remote disease management platform provides various kinds of interaction functions between doctors and patients. Patients can communicate with doctors about their symptom and the doctors can tell them the medication treatment. The communications are conducted either in text, audio or video phone.

By tracing the trends of the patients' physical condition, the platform can automatically provide disease management function and propose suitable intervention scheme for the patients. This is effective especially for chronic diseases. In that case, the patients no longer need to drive to the hospital time and time again. The platform can trace their physical condition in real-time and send to the doctors. The advices from the doctors can also be sent back to the patients in time. The patients only need to follow the doctor's advice at home and visit the hospital when his/her condition goes severe.

5. Healthcare Date Center

E-doctor Platform maintains a healthcare data center for all the patients. This is a distributed cloud-based database. For one patient, all its real-time monitoring data as well as the medication records, treatment records are stored. All the data can be accessed from cloud client such as smart phone, laptop or desktop anywhere.

The data center provides not only basic database operations such as search and sort, it also function as a data warehouse and provide huge amount of healthcare data for data mining and machine learning. Based on the information in this data center, using data mining technique such as Bayesian network, reasoning technique, certain unknown relationship between certain symptom and the disease can be found, which can facilitate future biomedical research.

5. Conclusions

In summary, E-Doctor is an integrated real-time monitoring and remote healthcare platform. It provides various continuous body condition parameter monitoring and remote disease management function. It maintains a comprehensive healthcare database for every user and analyzes the body conditions automatically. With the wide deployment of E-Doctor system, more and more patients can obtain benefit from its superiority.

References

[1] IEEE std 802.15.1. Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPANs). http://www.ieee802.org/15/Bluetooth/802-15-1_Clause_05.pdf

[2] Jovanov E. Wireless technology and system integration in body area networks for m-health

applications. In Proc. of 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS'05), pp. 7158-7160, 2005.

[3] Gao Y., Pun S. H. and Du M. A preliminary two dimensional model for intra-body communication of body sensor networks. In International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP'08), pp. 273-278, 2008.

[4] Yun D., Kang J. and Kim J E. A body sensor network platform with two-level communications. In IEEE International Symposium on Consumer Electronics (ISCE'07), pp. 1-6, 2007.

[5] Willig A., Hauer J H. and Karowski N. The ANGEL WSN architecture. In 14th IEEE International Conference on Electronics, Circuits and Systems (ICECS'07), pp. 633-636, 2007.

[6] Lorincz, K., et al. Sensor networks for emergency response: Challenges and opportunities. In IEEE Pervasive Computing, 3(4), pp. 16–23, 2004.

[7] van Halteren, et. Al. Mobile patient monitoring: The mobihealth system. In the Journal on Information Technology in Healthcare, 2(5), pp. 365–373, 2004.

[8] DeVaul R., Sung M. and Gips J. MIThril 2003: Applications and architecture. In Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC '03), 2003.

Jin Zhang received her B.S. degree and M.S. degree in Electronic Engineering from Tsinghua University, China, in 2004 and 2006, respectively. Then, she received her Ph.D. degrees in Computer Science and Engineering from Hong Kong University of Science and Technology in 2009.

Currently, she is a research assistant professor in department of computer science and engineering, HKUST. Her research interests include wireless sensor network in healthcare application, cooperative communication and networking, network coding, dynamic spectrum access in cognitive radio networks, interference alignment and cancelation in multi-antenna networks, cross-layer design and implementation in software-defined radio testbeds.

Dr. Zhang joined Hong Kong University of Science and Technology in Sept. 2005 where she is a full Professor in the Department of Computer Science and Engineering. Before that, she was in Microsoft Research Asia, Beijing, from July 1999, where she was the research manager of the Wireless and Networking Group. Dr. Zhang has published more than 200 refereed papers in international leading journals and key conferences in the areas of wireless/Internet multimedia networking, wireless communications and networking, wireless sensor networks, and overlay networking. She is the inventor of about 30 pending International patents. Her current research is on cognitive and cooperative networks, dynamic spectrum access and management, as well as wireless sensor networks.

Dr. Zhang has received MIT TR100 (MIT Technology Review) world's top young innovator award. She also received the Best Asia Pacific (AP) Young Researcher Award elected by IEEE Communication Society in year 2004. She received the Best Paper Award in Multimedia Technical Committee (MMTC) of IEEE Communication Society in 2005 and Best Paper Award for QShine 2006, IEEE Globecom 2007, IEEE ICDCS 2008, and IEEE ICC 2010. She received the Oversea Young Investigator Award from the National Natural Science Foundation of China (NSFC) in 2006. She has been elected as IEEE Communication Society Distinguished Lecture from Jan. 2010 to Dec. 2011.

Dr. Zhang was CHAIR of the Multimedia Communication Technical Committee of the IEEE Communications Society from 2008 to 2010.

Dr. Zhang received the B.S., M.S., and Ph.D. degrees from Wuhan University, China, in 1994, 1996, and 1999, respectively, all in computer science.

Privacy-preserving Wireless Data Transmission for e-Healthcare Applications

Xiaohui Liang, University of Waterloo, Canada
Mrinmoy Barua, University of Waterloo, Canada
Rongxing Lu, University of Waterloo, Canada
Xuemin (Sherman) Shen, University of Waterloo, Canada
{x27liang, mbarua, rxlu, xshen}@bbcr.uwaterloo.ca

1. Introduction

Recent advances in body sensors and wireless communications have revealed the possibility of providing remote healthcare monitoring and fast emergency services to patients via a smart e-healthcare system. The e-healthcare system pervasively adopts electronic and portable devices, such as smart phones, to monitor, transmit, and store patient medical records, and shifts healthcare tasks from a traditional clinical environment to a pervasive patient-centered setting. Such system has been widely regarded as a potential solution to reduce healthcare expenses through more efficient use of clinical resources and earlier detection of medical conditions. However, the e-healthcare system leads to more challenging privacy issues especially when wireless data transmission from patients to healthcare service providers is employed. To increase public acceptability from privacy-sensitive patients, it is critical to investigate their privacy requirements for various e-healthcare applications and maximally preserve patient privacy based on these requirements.

2. Patient Privacy for e-Healthcare

Patient privacy for healthcare applications has been extensively explored among government officers, research scholars and industrial investors in the past years. Government legislations, such as the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH), introduce penalty rules towards the privacy violations by healthcare service providers and prompt these providers to find privacy enhancing technologies to protect medical records. However, privacy policies are often dynamic and should be defined according to system and environmental changes [1]. The current legislations are disconnected to the ever-increasing patient privacy requirements and cannot fit the emerging e-healthcare applications. For example, healthcare monitoring devices become smaller, light-weighted and smarter such that they can be easily deployed around, on or in human body. Patients are able to read their medical records directly from the devices and they require to primarily controlling the access to their privacy-sensitive records. Patient self-controllable privacy preservation has been proposed in [2,3,4], aiming to grant any healthcare service providers an access to patient medical records following patient access policies. Meanwhile, certain privacy requirements in the traditional wireless communication design also need to be extended for e-healthcare applications. These privacy requirements can be the protection of identity information, location information, and transmitted data etc. In the following, we categorize the privacy issues for e-healthcare applications into two types: data privacy and communication privacy, and present some state-of-art solutions [5,6,7,8].

3. Data Privacy

The e-healthcare applications require fine-grained access control for privacy-sensitive data (i.e., medical records). The access control must be patient-centric, since patients are able to monitor, store and transmit the data using portable devices for most cases. In other words, patients must be able to select the access policy for their own personal information and apply distinct access policies to their medical records in different situations. For example, in a normal situation they may only send their data to a remote trusted authority; whereas, in a life-threatening scenario they may want to disseminate the data around to find a helper while disclosing minimal personal information to the public. The trade-off between safety, privacy, and access control in different contextual cases is shown in Figure 1(a). The research efforts [5,6] have been devoted to attribute-based encryption (ABE) schemes for fine-grained access control without lengthy user authorization process. Such schemes encrypt a patient's medical record with an access tree, where each leaf node represents an attribute (or access role) and each non-leaf node represents a threshold value shown in Figure 1(b). An access tree can be semantically transformed to a Boolean function. If the healthcare service provider has an attribute set that satisfies the Boolean function, he/she is able to decrypt the patient's medical record. In this way, the access to patient medical records can be controlled by patients themselves. Patients are able to choose appropriate access policies for different e-healthcare applications and only allow other users they trust to access their privacy-sensitive medical records.

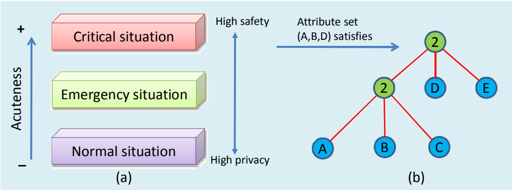


Figure 1. Trade-off and access policy

4. Communication Privacy

The communication privacy for e-healthcare is to protect multiple types of information related to communication entities, e.g., patients and healthcare service providers. Patient identity and location information are most critical and they must be protected from unauthorized entities by default unless patients allow the disclosure for special purposes. In [7], the multiple pseudonym technique is adopted to attain identity privacy preservation and location privacy preservation. The basic idea is to require patients to frequently change the pseudonyms that are used for authentication. Since other users cannot link the pseudonyms to the real identities of patients, patient' identities cannot be retrieved from the communications. Moreover, as the pseudonyms are generated independently, the transactions from one patient are not linkable and any other users cannot trace the patient's behaviors.

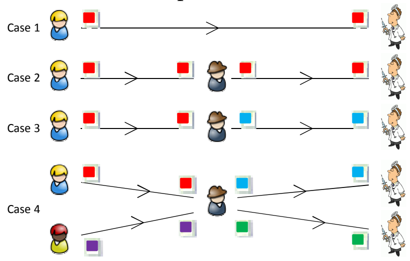


Figure 2. Several contextual privacy cases

Another contextual privacy [8] has also been addressed for the wireless data transmission of e-healthcare applications. To preserve contextual privacy, the communication protocols must resist an attack which aims to link the source and the destination of a transaction. If an adversary is able to do so, the relations between patients and healthcare service providers will be exposed and the patient medical conditions might be revealed. In [8], a mix technique is utilized to resolve this problem. The main idea is to cut off the relation between the input and the output. As shown in Figure 2, the contextual privacy of Cases 1 and 2 is easily violated since an adversary can link the patient with the healthcare services provider by simply comparing the contents in the transactions. For Case 3, a relay user re-randomizes the transaction but it cannot preserve contextual privacy since the input number and the output number are both one, and an adversary can still link them. The contextual privacy is successfully preserved in Case 4, where 2 patients and 2 healthcare service providers connect to the same relay user. The probability that the adversary has a correct observation is only 1/2. Therefore, the above mix technique can be used to preserve contextual privacy if at least a relay user has *m* inputs and *n* outputs. The parameters *(m,n)* should be as large as possible. Otherwise, the contextual privacy level will decrease quickly.

5. Conclusion

E-healthcare systems are very promising and attractive to both healthcare service providers and patients since they can provide long-term, real-time monitoring and fast emergency services anywhere at any time. As the transmitted data in such system is often privacy-sensitive to patients, patient privacy concerns must be well addressed. In this article, we have presented the privacy issues specifically for wireless data transmission in e-healthcare systems, and provided potential solutions to resolve these problems.

Reference

[1] A. AL Faresi, D. Wijesekera, and K. Moidu, "A Comprehensive Privacy-aware Authorization Framework Founded on HIPAA Privacy Rules," IHI'10, Arlington, Virginia, pp. 637-646, 2010.

[2] N. Huda, N. Sonehara, and S. Yamada, "A Privacy Management Architecture for Patient-controlled Personal Health Record System," Journal of Engineering Science and Technology, Vol. 4, No. 2, pp. 154-170, 2009.

[3] M. Li, W. Lou, and K. Ren, "Data Security and Privacy in Wireless Body Area Networks," IEEE Wireless Communications Magazine, Vol. 17, Issue 1, pp. 51-58, 2010.

[4] J. Sun, Y. Fang, and X. Zhu, "Privacy and Emergency Response in E-Healthcare Leveraging Wireless Body Sensor Networks," IEEE Wireless Communications Magazine, Vol. 17, Issue 1, pp. 66-73, 2010.

[5] X. Liang, R. Lu, X. Lin, and X. Shen, "Patient Self-controllable Access Policy on PHI in eHealthcare Systems", AHIC 2010, Kitchener, Ontario, Canada, 2010.
[6] M. Barua, X. Liang, R. Lu, and X. Shen, "PEACE: An Efficeint and Secure Patient-centric Access Control Scheme for eHealth Care System," IEEE Infocom Workshop on Security in Computers, Networking and Communications, pp. 970-975, Shanghai, China, 2011.
[7] X. Liang, R. Lu, L. Chen, X. Lin, and X. Shen, "PEC: A Privacy-preserving Emergency Call Scheme for Mobile Healthcare Social Networks", Journal of Communications and Networks, Vol. 13, No.2, pp. 102-112, 2011 .
[8] X. Lin, R. Lu, X. Shen, Y. Nemoto, and N. Kato, "Sage: a strong privacy-preserving scheme against global eavesdropping for ehealth systems," IEEE Journal on Selected Areas in Communications, Volume 27, Issue 4, pp. 365-378, 2009.

Xiaohui Liang [S'10] is currently working toward a Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a research assistant with the Broadband Communications Research (BBCR) Group, University of Waterloo. His research interests include network security and privacy, applied cryptography, and e-healthcare system.

Mrinmoy Barua is now pursuing the Ph.D. degree in Electrical and Computer Engineering at the University of Waterloo, Canada. His research interests include applied cryptography, wireless network security, and security and privacy in eHealth and cloud computing.

Rongxing Lu [S'09, M'11] is currently working toward a Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a research assistant with the Broadband Communications Research (BBCR) Group, University of Waterloo. His research interests include wireless network security, applied cryptography, and trusted computing.

Xuemin (Sherman) Shen [M'97, SM'02, F'09] received a B.Sc. (1982) degree from Dalian Maritime University, China, and M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey, all in electrical engineering. He is a professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo. His research focuses on mobility and resource management, UWB wireless networks, wireless network security, and vehicular ad hoc and sensor networks. He served as an Area Editor for IEEE Transactions on Wireless Communications and Editor-in- Chief for Peer-to- Peer Networks and Applications. He is a Fellow of Engineering Institute of Canada, a registered Professional Engineer of Ontario, Canada, and a Distinguished Lecturer of the IEEE Communications Society and Vehicular Technology Society.

Data Privacy and Secure Storage for Electronic Health Record

Jinyuan Sun, University of Tennessee, Knoxville, USA
Xiaoyan Zhu, Xidian University, China
{jysun@eecs.utk.edu, xyzhu@mail.xidian.edu.cn*}*

1. Introduction

Advances in wireless communications and computing technologies have lent great forces to migrating healthcare systems from the paper-based to the EHR (electronic health record)-based, giving rise to increased efficiency in human operations, reduced storage costs and medical errors, improved data availability and sharing, etc. While EHR-based healthcare systems offer convenience and flexibility such as anytime anywhere medical record access, they also put patients' medical data privacy into great danger. Patients' concerns on their medical data privacy thus become one of the major barriers in the deployment of EHR-based healthcare systems. Privacy breaches can be incurred in the handling of the EHRs, e.g., storage, retrieval, transfer, update, where secure storage plays an important role because it is the foundation for all the other operations to be carried out securely.

2. EHR Data Privacy

Data privacy of electronic health record refers to the confidentiality and access restrictions of patients' protected health information (PHI) which contains sensitive and personal information such as disease history and undergoing treatment. There are good reasons for keeping the records private and limiting the access to only minimum-necessary information: an employer may decide not to hire someone with psychological issues, an insurance company may refuse to provide life insurance when aware of the disease history of a patient, a person with certain types of disease may be discriminated by the healthcare provider, and so on. However, fundamental developments of healthcare systems have threatened the confidentiality of medical records and patient privacy [1], one of which is the exponential increase in the use of computers and automated information systems for health records. Computers connected to a network are now commonly used by healthcare providers to store and retrieve patients' electronic EHRs.

Protecting patients' medical data privacy in EHR-based healthcare systems is a highly challenging task. According to the Health Insurance Portability and Accountability Act (HIPAA) that was established in the US to regulate EHR related operations, patients should have full control over the creation, modification, release and access of their EHRs. This privacy policy may cause conflicts with or inconvenience for some other EHR related operations, i.e., EHR storage, emergency treatment. Detailed discussions on preserving patient privacy and facilitating emergency treatment can be found in [2]. Patients nowadays tend to choose third-party services (e.g., Microsoft HealthVault) that store their EHRs for easy access. Storing medical data on an untrusted or semi-trusted server increases the risks of medical data leakage and illegal access. In emergencies when patients are unable to grant access to or physically release their EHRs, security mechanisms should be in place to ensure timely and proper access to the medical data. In what follows, we discuss the issues around secure EHR storage while keeping emergency treatment in mind.

3. Secure EHR Storage

The third-party storage site is commonly assumed to be honest-but-curious in the sense that it does not maliciously delete or modify the stored EHRs for no gain. Instead, it is curious about the content of the EHRs which contain sensitive information and may sell the stored EHRs for monetary gain. A few research works [3-6] address the issue of secure EHR storage. Tan et al. [3] specify the algorithms for storing and retrieving health records. The scheme in fact failed to achieve privacy protection in that the storage site will learn the ownership of the encrypted records (i.e., which records are from which patient) in order to return the desired records to the querying doctor. Such leakage will compromise patients' privacy by violating the unlinkability requirement. Both [3] and [4] assume the existence of wireless body sensor networks (WBSNs) where the PDA can act on the patient's behalf for emergency treatment. The drawback of these solutions is that they cannot be applied when a trusted device (e.g., PDA) is not available. In [5] and [6], patients encrypt their own EHRs and store them on a third-party server to preserve EHR data privacy. Both proposals are capable of coping with emergency situations where [5] relies on attribute-based encryption to finely control physicians' access rights and [6] leverages some backup mechanisms that act on patients' behalf upon emergencies. The fine-grained access control in [5] assumes personal attribute authorities who grant access rights to emergency physicians. It is not clear who will take the role of these authorities in practice. The backup mechanisms in [6] rely on someone or something the patient trusts whose availability cannot be guaranteed at all times.

A promising solution for protecting EHR data privacy while making it convenient for secure storage and emergency treatment, is to leverage personal clouds, as shown in Fig. 1. Patients can obtain dedicated service (represented by the shaded servers in Fig. 1) from the cloud service provider to set up their personal clouds.

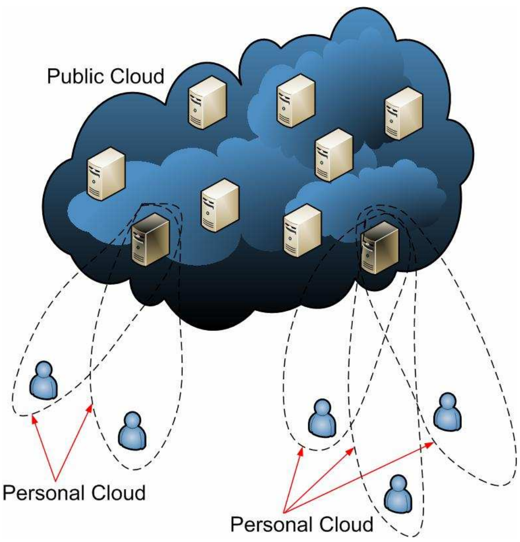


Figure 1: Secure Storage of EHR Leveraging Personal Cloud.

The personal clouds process patients' EHRs (e.g., encryption), store the result on the public cloud, and retrieve the EHRs upon physicians' requests after proper authentication. Patients can be relieved from computational tasks such as encryption and outsource these tasks to their personal clouds. In emergencies, personal clouds are always available to act on patients' behalf to decrypt the requested EHRs for emergency physicians. However, unlike the trusted personal device, personal clouds are still third-party servers that are honest-but-curious. It brings out the challenging question: Given the power of decrypting the EHRs, how to guarantee the personal clouds do not engage in activities such as illegally distributing or selling these medical data? We propose to use the secret sharing technique, where the personal clouds need to maintain certain proof (i.e., a signature from the physician who requested the patient's EHRs). Since the personal cloud is the only entity (besides the patient himself) that is able to decrypt the EHRs, any illegal leakage of EHRs can be traced back to the personal cloud unless it can provide the proof for a genuine request. The personal cloud based solution for secure storage while ensuring EHR data privacy and emergency treatment is part of our ongoing research.

4. Conclusion

Medical data privacy is of paramount importance to both patients and the deployment of EHR-based healthcare systems that offer great features. Protecting data privacy is becoming an increasingly challenging issue since careless designs will hinder other important EHR-related operations such as secure storage and emergency treatment. In this article, we reviewed the requirements for EHR data privacy and secure storage. We also discussed their relationship and presented a promising solution to achieve the requirements.

References

[1] G. M. Stevens, "A brief summary of the medical privacy rule," *CRS Report for Congress*, 2003.

[2] J. Sun, X. Zhu, and Y. Fang, "privacy and emergency response in e-healthcare leveraging wireless body sensor networks," *IEEE Wireless Communications*, 2010.

[3] C. C. Tan, H. Wang, S. Zhong, and Q. Li, "body sensor network security: an identity-based cryptography approach," *The ACM Conference on Wireless Network Security (WiSec'08)*, Apr. 2008.

[4] J. Sun, X. Zhu, and Y. Fang, "preserving privacy in emergency response based on wireless body sensor networks," *Proc. of IEEE Globecom'10*, Dec. 2010.

[5] M. Li, S. Yu, K. Ren, and W. Lou, "Securing Personal Health Records in Cloud Computing: Patient-centric and Fine-grained Data Access Control in Multi-owner Settings," *in Proc. of The 6th International ICST Conference on Security and Privacy in Communication Networks (SecureComm2010)*, Singapore, Sept. 2010.

[6] J. Sun, X. Zhu, C. Zhang, and Y. Fang, "HCPP: cryptography based secure EHR system for patient privacy and emergency healthcare," *Proc. of IEEE Int'l Conf. on Distributed Computing Systems (ICDCS'11)*, June 2011.

Jinyuan Sun received the BSc degree in computer information systems from Beijing Information Technology Institute, China, in 2003, the MASc degree in computer networks from Ryerson University, Canada, in 2005, and the PhD degree in electrical and computer engineering from the University of Florida, in 2010. She was a Network Test Developer at RuggedCom Inc., Ontario, Canada, 2005-2006. She has been an assistant professor in the Department of Electrical Engineering and Computer Science at University of Tennessee Knoxville since August 2010. Her research interests include the security protocol and architecture design of wireless networks.

Xiaoyan Zhu received her BE degree in information engineering from Xidian University, Xian, China, in July 2000, and her ME degree in information and communications engineering from Xidian University, Xian, China, in March 2004. She is now working towards her Ph.D. degree at Xidian University. Her research interests include wireless security and network coding.