JOINT GAZE-CORRECTION AND BEAUTIFICATION OF DIBR-SYNTHESIZED HUMAN FACE VIA DUAL SPARSE CODING

Xianming Liu[1,2], Gene Cheung[2], Deming Zhai[1], Debin Zhao[1], Hiroshi Sankoh[3], Sei Naito[3]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, 150001
[2]National Institute of Informatics, Tokyo, Japan
[3]KDDI Laboratories, Tokyo, Japan

ABSTRACT

Gaze mismatch is a common problem in video conferencing, where the viewpoint captured by a camera (usually located above or below a display monitor) is not aligned with the gaze direction of the human subject, who typically looks at his counterpart in the center of the screen. This means that the two parties cannot converse eye-to-eye, hampering the quality of visual communication. One conventional approach to the gaze mismatch problem is to synthesize a gaze-corrected face image as viewed from center of the screen via depth-image-based rendering (DIBR), assuming texture and depth maps are available at the camera-captured viewpoint(s). Due to self-occlusion, however, there will be missing pixels in the DIBR-synthesized view image that require satisfactory filling.

In this paper, we propose to jointly solve the hole-filling problem and the face beautification problem (subtle modifications of facial features to enhance attractiveness of the rendered face) via a unified dual sparse coding framework. Specifically, we first train two dictionaries separately: one for face images of the intended conference subject, one for images of "beautiful" human faces. During synthesis, we simultaneously seek two code vectors—one is sparse in the first dictionary and explains the available DIBR-synthesized pixels, the other is sparse in the second dictionary and matches well with the first vector up to a restricted linear transform. This ensures a good match with the intended target face, while increasing proximity to "beautiful" facial features to improve attractiveness. Experimental results show naturally rendered human faces with noticeably improved attractiveness.

Index Terms— Video conferencing, gaze correction, face beautification, sparse coding

1. INTRODUCTION

With the advent of imaging and network technologies enabling inexpensive yet high-quality video capture and reliable high-bandwidth transmission, video conferencing connecting two parties separated by a large physical distance is now ubiquitous with tools such as Skype[1] and Google Hangouts[2]. A common problem with these tools is the *gaze mismatch* problem: because the capturing camera typically resides above or below the display monitor while the human subject tends to look at his/her counterpart in the center of the screen, the conversing parties cannot talk eye-to-eye, hampering the quality of visual communication.

Leveraging on recent advances in 3D imaging [1], one common approach to the gaze mismatch problem is to synthesize the viewpoint image as observed from the center of screen (*virtual view*) via *depth-image-based rendering* (DIBR) [2], assuming texture and depth viewpoint images[3] are available from the capturing camera(s) [3, 4, 5]. The key problem to this approach is the *disocclusion hole-filling* problem: spatial regions in the virtual view that were occluded by foreground elements in the camera-captured viewpoint contain missing pixels and need to be completed satisfactorily. While hole-filling for DIBR-synthesized images has been addressed formally in the literature [6, 7, 8], to the best of our knowledge no one has yet tailored a hole-filling algorithm specifically for human face for rendering of a natural gaze-corrected view.

In this paper, we propose to jointly solve the disocclusion hole-filling problem for human faces together with the *face beautification* problem in a unified sparse coding framework. Face beautification [9] is the process of subtly modifying facial features to enhance attractiveness of the rendered face. Unlike previous hole-filling approaches [6, 7, 8] that attempt to fill in missing pixels using only available data in the current captured images, we assume the availability of a large corpus containing images of *both* the rendered subject *and* a lot of "beautiful faces" when training two corresponding dictionaries offline[4]—a fair assumption in the age of big data where voluminous databases of images are easily accessible through social networks and search engines. During view synthesis, we simultaneously seek two code vectors—one is sparse in the first dictionary and explains the available DIBR-synthesized pixels, the other is sparse in the second dictionary and matches well with the first vector up to a restricted linear transform. The restriction of the linear transform to a pre-determined set ensures a good balance between recognizability of the captured subject and proximity to ”beautiful” facial features to improve attractiveness. Experimental results show naturally rendered human faces with noticeably improved attractiveness.

The outline of the paper is as follows. We first overview related work in Section 2. We then overview our proposed joint gaze correction / face beautification system in Section 3. We describe in precise mathematical terms our dual sparse coding framework for joint hole-filling / face beautification in section 4. Finally, we discuss experimentation and conclusion in Section 5 and 6, respectively.

[1] http://www.skype.com/

[2] http://www.google.com/hangouts/

[3] A depth image is a per-pixel distance map measuring the physical separation between objects in the 3D scene and capturing camera. It can be captured using depth sensing cameras such as Microsoft Kinect.

[4] It is argued in [10] that *big data* (explosion of collected data across networks that are made available via efficient cloud-based retrieval services) can fundamentally alter how traditional signal processing problems are tackled. In this work, we demonstrate how the gaze mismatch problem can benefit from big data via offline training of appropriate dictionaries.

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

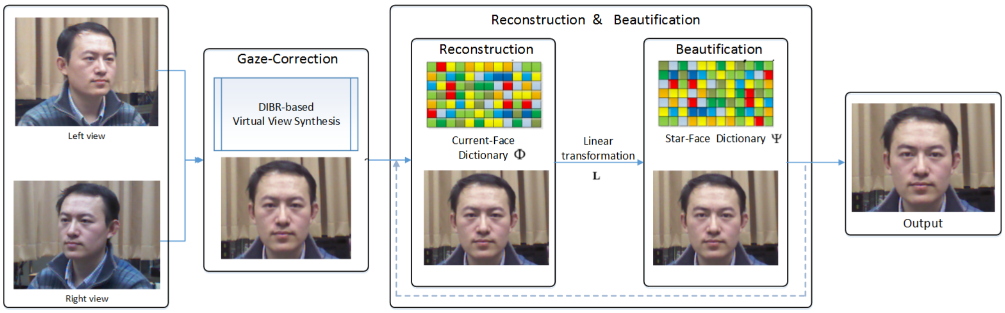

Fig. 1. Block diagram of the proposed joint gaze correction and face beautification system

2. RELATED WORK

Gaze mismatch is a well known problem for video conferencing [11], and there exist numerous solutions in the literature [12, 13, 3, 4, 5]. Early solutions [12, 13] performed *image-based rendering* (IBR) given stereo-captured viewpoint images, which tends to be computation-intensive. Leveraging on the recent advance in depth sensing technologies, more recent proposals [3, 4, 5] assumed a captured camera provides both texture and depth images[5] for DIBR synthesis of the gaze corrected view. Towards real-time implementation, these proposed solutions to the disocclusion hole-filling problem tend to be simple and ad-hoc, which work well if the camera viewpoint(s) and the target virtual viewpoint are not far apart. However, for enhanced immersiveness during video conferencing, a large display is often used, and the capturing cameras can be located far from the screen center. Thus the resulting holes are large, and our more sophisticated dual sparse coding scheme using dictionaries tailored for the specific human subject can be more beneficial.

Note that in our approach the needed dictionaries are trained offline using an available corpus of images, so the only complexity during conferencing is the computation of appropriate sparse vectors. It is conceivable that this can be done in real-time for reasonably sized dictionaries using the latest optimization tools such as Lasso.

While beautification has been studied in the literature for static individual images [9], doing so for video of human face in a conferencing situation brings new challenges of temporal consistency and real-time implementation. Though not demonstrated in this paper, we believe our sparse coding approach (first in the beautification literature) can achieve both goals by initializing the sparse vectors of new time instant using optimized vectors of previous instant. Further investigation of these two issues are left for future work.

3. SYSTEM OVERVIEW

The architecture of our proposed joint gaze correction and face beautification system is illustrated in Fig. 1. The system employs two cameras facing the user on the left and right side of a display. DIBR is first used to synthesized a gaze corrected view, assuming texture and depth maps are available at the camera-captured viewpoints. Due to self-occlusion, there will be missing pixels in the DIBR-synthesized view image that require filling[6].

The key component is the reconstruction and beautification module, which jointly solves the self-occlusion hole-filling problem for human faces together with the face beautification problem in a unified dual sparse coding framework. This module relies one two dictionaries, which are learned using two offline training sets of images, which we assume can be collected *a priori*.

The dictionary used for face reconstruction is learned by using specific face photos of the current person in video conferencing. It has been observed that the space spanned by the appearance of faces is relatively small, and the space spanned by a specific face is significantly smaller. Therefore, by using of such a dictionary, we can derive useful priors to obtain high-quality face reconstruction.

The dictionary used for face beautification is trained by using a set of general beautiful faces, which in our case are collected beautiful faces of Asian film stars from the Internet. Through it, we can exploit the discriminative nature of sparse representation to perform face beautification. The atoms that yield the most compact representation should be preferred as the candidates of beautification. That is, a beautified face can be represented as a linear combination of these atoms. Furthermore, a restricted linear transformation $\mathbf{L}$ is employed to constraint the level of beautification, that is, the modification should be subtle. The reconstruction and beautification process can be performed iteratively to obtain better results.

4. JOINT HOLE-FILLING AND FACE BEAUTIFICATION

We now discuss the details of the hole-filling / beautification module. We propose that a successful transformation must meet the following three criteria:

- Fidelity: holes in the DIBR-synthesized facial image must be completed, resulting in a natural-looking human face.
- Attractiveness: The rendered face should have enhanced facial attractiveness, with "beautiful" local features.
- Identifiability: Only subtle modifications should be performed to the original face, so that the target human subject is unmistakably recognizable.

The first and second criteria can be formulated as a dual sparse coding problem, where a *patch*-based over-complete dictionary is learned from example faces of the current person to characterize the structure domain of the current face, and a *feature*-based over-complete dictionary is learned from a set of beautiful faces of movie stars to characterize the feature domain of beautiful faces. The third criteria can be met by restricting the types of linear transformations performed for different facial features to limit the degree of beautification. All of these criteria can be cast into a unified optimization framework to obtain final results.

[5] This texture-plus-depth video format requires compression of depth images at sender, which has been intensively studied in recent years [14, 15].

[6] Filling of pixels not belonging to the human face is an orthogonal problem solvable using generic hole-filling techniques [6, 7, 8], and thus is not considered in this paper.

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

4.1. Dual Dictionary Learning

How to derive appropriate dictionaries for dual sparse coding is a critical issue in our proposed scheme. Two separate dictionaries are learned for face reconstruction and beautification respectively.

4.1.1. Dictionary Learning for Reconstruction

For face reconstruction, our purpose is to fill in holes exposed by DIBR in the synthesized face image. In the literature, image restoration using local patches has become very popular and was shown to be highly effective [17, 18]. Hence, for face reconstruction, we propose a patch-based local adaptive method to learn the dictionary. For given example face images of the current person, we extract all patches with overlaps. Then we classify the collected patches into clusters with similar geometric structure by using the K-means clustering method [19], and model each cluster by learning a compact sub-dictionary. Specifically, for a certain cluster i which includes n_i image patches to be coded, we stack the vectors of patches into a matrix denoted by $\mathbf{X}_i$. Then, we learn an adaptive sub-dictionary $\mathbf{\Phi}_i$ that is most relevant to $\mathbf{X}_i$ by applying *principal component analysis* (PCA) on $\mathbf{X}_i$. PCA generates the dictionary $\mathbf{\Phi}_i$ whose atoms are the eigenvectors of the covariance matrix of $\mathbf{X}_i$.

Since the patches in a cluster are similar to each other, the sub-dictionary is not necessary to be over-complete, but all sub-dictionaries are combined together to construct a large over-complete dictionary $\mathbf{\Phi}$ to characterize all possible local structures of the face [19].

4.1.2. Dictionary Learning for Beautification

Even though faces are objects with large varieties, they are made up of several basic features, such as eyes, eyebrows, nose and mouth. Studies have demonstrated the importance of facial features in beauty judgment tasks [16]. For face beautification, we take a feature-based dictionary learning approach. Given a frontal face as input, we first identify a set of facial landmarks [21]. Then, according to the locations of landmarks, we extract six classes of facial features including: left eye, right eye, left eyebrow, right eyebrow, nose and mouth. In practical implementation, we do not beautify the mouth, since it typically moves fast and often during the conferencing video. We construct a sub-dictionary for each feature.

Specifically, for a feature $\mathbf{x}_i$ of the current face, we collect sufficient example features $A_i = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{n_i}]$ from the beautiful face dataset as training samples. Light normalization is performed on the training samples to cancel the influence of different lighting condition. We propose the following model for light normalization:

$$\widetilde{\mathbf{v}}_i = \frac{(\mathbf{v}_i - \min(\mathbf{v}_i))\,(\max(\mathbf{x}_i) - \min(\mathbf{x}_i))}{\max(\mathbf{v}_i) - \min(\mathbf{v}_i)} + \min(\mathbf{x}_i) \quad (1)$$

where $\min(\cdot)$ and $\max(\cdot)$ represent the minimization and maximization operators. After light normalization, the training samples themselves are used as base elements of the beautification dictionary $\mathbf{\Psi}$.

4.2. Joint Sparse Coding

We achieve the overall objective by finding sparse code vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with respect to the two trained dictionaries, as well as the linear transformation $\mathbf{L}$, via minimization of the following energy function:

$$\begin{aligned}\arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{L}} & \|\mathbf{x} - \mathbf{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_1 \\ & + \|\mathbf{\Phi}\boldsymbol{\alpha} - \mathbf{L}\mathbf{\Psi}\boldsymbol{\beta}\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_1 .\end{aligned} \quad (2)$$

The first two terms in the objective function are the standard sparse coding to reconstruct the face vector $\mathbf{x}$ with respect to person-specific dictionary $\mathbf{\Phi}$. The third and fourth terms perform sparse coding to represent the reconstructed face $\mathbf{\Phi}\boldsymbol{\alpha}$ with respect to the beautiful-face dictionary $\mathbf{\Psi}$ for beautification purpose, where the relationship between the two sparse code vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is established through a linear transform $\mathbf{L}$, which denotes the transformation from the reconstructed face to the synthesized beautiful face. By constraining $\mathbf{L}$, we can limit the degree and types of beautification performed to meet the identifiability criteria.

The objective function is not jointly convex in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\mathbf{L}$, but is convex in one variable if others are fixed. Therefore, we can employ the alternating procedure to optimize these variables iteratively. Specifically, to tackle the objective function in Eq. (2), we separate the objective function into three sub-problems, namely sparse coding for the current face, sparse coding for beautiful face, and mapping function updating. This procedure is repeated until convergence or after a maximum number of iterations T has been reached. In what follows, we will describe the initialization process, three sub-problems and their optimal solutions.

4.2.1. Initialization Process

We obtain the initial sparse representation coefficients $\boldsymbol{\alpha}$ by solving the following problem:

$$\arg\min_{\boldsymbol{\alpha}} \left\{ \|\mathbf{x} - \mathbf{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_1 \right\}. \quad (3)$$

The optimization solution of $\boldsymbol{\alpha}$ can be effectively and efficiently solved by a fast ℓ_1-minimization algorithm, known as *Augmented Largrangian Methods* (ALM) [20]. And the linear transformation $\mathbf{L}$ is first initialized as a scalar matrix for the purpose of a global weighted linear combination of current-face feature and the star-face feature. More specifically, we initialize $\mathbf{L}$ as $\mathbf{L} = 0.4\mathbf{I}$ to limit the level of beautification, where $\mathbf{I}$ is the identity matrix.

4.2.2. Optimization with respect to $\boldsymbol{\beta}$

With the initialization of $\mathbf{L}$ as a diagonal matrix and the derived $\boldsymbol{\alpha}$ in Eq. (3), the optimization problem with respect to $\boldsymbol{\beta}$ becomes:

$$\arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{\Phi}\boldsymbol{\alpha} - \mathbf{L}\mathbf{\Psi}\boldsymbol{\beta}\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_1 \right\}. \quad (4)$$

Finally, the sparse representation coefficient $\boldsymbol{\beta}$ can also be obtained via ALM as done previously for $\boldsymbol{\alpha}$.

4.2.3. Optimization with respect to $\mathbf{L}$

With the derived coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we update the linear transformation $\mathbf{L}$ as the following minimization problem:

$$\arg\min_{\mathbf{L}\in\mathcal{L}} \|\mathbf{\Phi}\boldsymbol{\alpha} - \mathbf{L}\mathbf{\Psi}\boldsymbol{\beta}\|_2^2 . \quad (5)$$

In practical implementation, we restrict $\mathbf{L}$ to be a few simple known linear transformations, such as rotation, scaling, shifting, which construct the set $\mathcal{L}$. Then, the problem of optimizing $\mathbf{L}$ converts to simply search through the defined transformations. In this way, the degree and types of beautification is constrained, thus, recognizability is maintained in the beautification process.

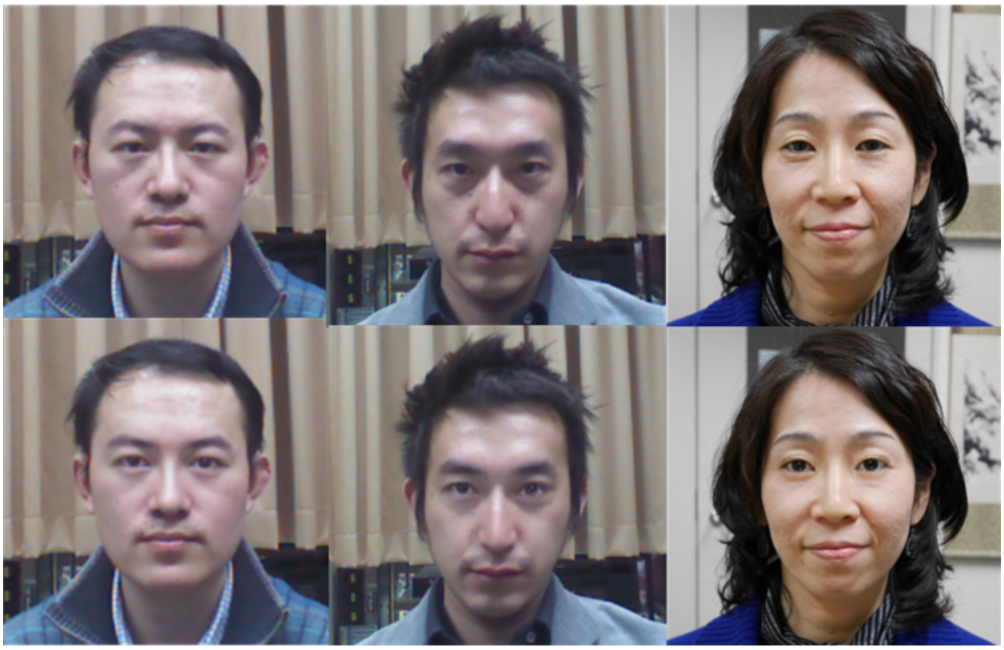

Fig. 2. Results of gaze-corrected face reconstruction and beautification. The first row: reconstructed faces, the second row: beautified faces.

4.2.4. *Optimization with respect to $\boldsymbol{\alpha}$*

Fixing $\boldsymbol{\beta}$ and $\mathbf{L}$, the sub-problem for optimizing code vector $\boldsymbol{\alpha}$ becomes:

$$\arg\min_{\boldsymbol{\alpha}} \left\{ \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{L}\boldsymbol{\Psi}\boldsymbol{\beta}\|_2^2 \right\} \quad (6)$$

With simple algebra and deleting the constant, the above objective function can be reformulated as a standard sparse coding problem:

$$\arg\min_{\boldsymbol{\alpha}} \left\{ \left\| \frac{1}{2} \left(\mathbf{x} + \mathbf{L}\boldsymbol{\Psi}\boldsymbol{\beta} \right) - \boldsymbol{\Phi}\boldsymbol{\alpha} \right\|_2^2 + \lambda_1 \|\boldsymbol{\alpha}\|_1 \right\}. \quad (7)$$

which also can be efficiently solved by the ALM algorithm.

5. EXPERIMENTATION

In this section, experimental results are presented to demonstrate the performance of our proposed joint gaze-correction and beautification scheme. In our experiments, the specific face dataset is constituted by face regions of the camera-captured viewpoints and other frontal photos of the current person. The star faces dataset is collected from the web, including 600 samples for male and female, respectively. To reduce the influences of age, skin color and other irrelevant factors, we only select young East Asian male and female film stars in frontal view, without glasses, to construct star face training set. All star faces are normalized by the distance between pupils of the test face, and are aligned to bring the eyes to the same location. Other preprocessing procedures are needed, such as face detection and alignment. Two different dictionaries are trained for female and male star faces. For simplicity, to simulate DIBR disocclusion holes, we manually inject random impulse noises into camera-captured frontal-view images as input to our face reconstruction / beautification algorithm. In practical experiments, about 30% of all pixels are corrupted by impulse noise.

The results of reconstructing and beautifying the gaze-corrected face generated by DIBR are shown in Fig.2, including two male test samples and one female sample. Visually, it can be seen that the first male test subject has enlarged eyes and relatively thicker eyebrows. The second male test subject has more beautiful shapes for the eyes, and the wrinkles around the eyes are eliminated. The female test subject has enlarged eyes, which improve attractiveness. Note that in each of these subjects, the difference between the original face and the beautified one is quite subtle, and thus the resemblance between the two faces is unmistakable. Yet, the subtle changes clearly have a substantial impact on the attractiveness of these faces.

6. CONCLUSION

Gaze mismatch is a known problem in video conferencing. In this paper, we propose to jointly perform hole-filling in DIBR-synthesized gazed-corrected view and face beautification in one unified sparse coding framework. Assuming the availability of a large corpus of images *a priori*—a reasonable assumption in the era of big data—the key idea is to learn two dictionaries offline (one for target human subject and one for "beautiful" faces), so that in real-time two sparse code vectors can be sought: one explains the synthesized pixels from camera-captured view(s), and one approximates features of beautiful faces. A linear transform that maps from one reconstruction to another is carefully chosen to ensure recognizability of the conferencing subject. Experimental results show naturally rendered faces with enhanced attractiveness.

7. ACKNOWLEDGEMENT

This work is supported in part by National Science Foundation of China under Grants 61300110 and 61272386, and by KDDI Laboratories, Tokyo, Japan.

8. REFERENCES

[1] M. Tanimoto, M. P. Tehrani, T. Fujii and T. Yendo, “Free-Viewpoint TV,” *IEEE Signal Processing Magazine*, vol.28, no.1, January 2011.

[2] D. Tian, P.-L. Lai, P. Lopez and C. Gomila, “View Synthesis Techniques for 3D Video,” *Applications of Digital Image Processing XXXII, Proceedings of the SPIE*, 7443 (2009), 2009.

[3] S.-B. Lee, I.-Y. Shin and Y.-S. Ho, “Gaze-corrected View Generation Using Stereo Camera System for Immersive Video-conferencing,” *IEEE Transactions on Consumer Electronics*, vol.57, no.3, pp.1033-1040, August 2011.

[4] C. Kuster, T. Popa, J.-C. Bazin, C. Gotsman and M. Gross, “Gaze Correction for Home Video Conferencing,” *ACM SIGGRAPH Asia*, vol.31, no.6, November 2012.

[5] W. Eng, D. Min, V.-A. Nguyen, J. Lu and M. Do, “Gaze Correction for 3D Tele-immersive Communication System,” *11th IEEE IVMSP Workshop: 3D Image/Video Technologies and Applications*, Seoul, Korea, June 2013.

[6] K.-J. Oh, S. Yea and Y.-S. Ho, “Hole-Filling Method Using Depth Based In-Painting Fore View Synthesis in Free Viewpoint Television FTV and 3D Video,” *Picture Coding Symposium*, Chicago, IL, May 2009.

[7] S. Reel, G. Cheung, P. Wong and L. Dooley, “Joint Texture-Depth Pixel Inpainting of Disocclusion Holes in Virtual View Synthesis,” *APSIPA ASC*, Kaohsiung, Taiwan, October 2013.

[8] B. Macchiavello, C. Dorea, E. M. Hung and G. Cheung and I. Bajic, “Low-Saliency Prior for Disocclusion Hole Filling in DIBR-Synthesized Images,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, May 2014.

[9] T. Leyvand, D. Cohen-Or, G. Dror and D. Lischinski, “Data-Driven Enhancement of Facial Attractiveness,” *ACM SIGGRAPH*, vol.27, no.3, August 2008.

[10] H. Yue, X. Sun, J. Yang and F. Wu, “Cloud-based Image Coding for Mobile Devices–Towards Thousands To One Compression,” *IEEE Transactions on Multimedia*, vol.15, no.4, pp.845-857, June 2013.

[11] J. Gemmell, k. Toyama, C. L. Zitnick, T. Kang, and S. Seitz, “Gaze awareness for video conferencing: A software approach, *IEEE Multimedia*, vol.7, no.4, pp. 2635, Oct-Dec 2000.

[12] A. Criminisi, J. Shotton, A. Blake, and P. Toor, “Gaze manipulation for one-to-one teleconferencing, *IEEE International Conference on Computer Vision*, Nice, France, October 2003.

[13] R. Yang and Z. Zhang, “Eye gaze correction with stereovision for video-teleconferencing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.26, no.7, pp.956960, July 2004.

[14] G. Cheung, J. Ishida, A. Kubota, A. Ortega, “Transform Domain Sparsification of Depth Maps using Iterative Quadratic Programming,” *IEEE International Conference on Image Processing*, Brussels, Belgium, September 2011.

[15] W. Hu, G. Cheung, X. Li, O. Au, “Depth Map Compression using Multi-resolution Graph-based Transform for Depth-image-based Rendering,” *IEEE International Conference on Image Processing*, Orlando, FL, September 2012.

[16] Yael Eisenthal, Gideon Dror, Eytan Ruppin, “Facial Attractiveness: Beauty and the Machine,” *Neural Computation* vol.18, no.1, pp.119-142, 2006.

[17] M. Aharon, M. Elad, and A.M. Bruckstein, “The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for S-parse Representation,” *IEEE Transactions On Signal Processing*, vol.54, no.11, pp.4311-4322, November 2006.

[18] X. Liu, D. Zhai, D. Zhao, W. Gao, “Image Super-Resolution vi-a Hierarchical and Collaborative Sparse Representation,” *Proceeding of Data Compression Conference*, pp.93-102, 2013.

[19] W. Dong, L. Zhang, G. Shi, and X. Wu,, “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization,” *IEEE Transactions On Image Processing*, vol.20, no.7, pp. 1838-1857, July 2011.

[20] Allen Y. Yang, Z. Zhou, A. Ganesh, S. Sastry, and Y. Ma, “Fast l1-minimization algorithms for robust face recognition,” *IEEE Transactions On Image Processing*, vol.22, no.8, pp. 3234-3246, Aug. 2013.

[21] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2879-2886, June 2012.