Camera network using MIMO Y channel and its application to sports activity analysis

Seung-Won Jung*, Byungju Lee**, Byonghyo Shim**, and Sung-Jea Ko*
* Department of Electrical Engineering, Korea University, Anam-dong, Sungbuk-gu, Seoul, 136-713, Korea
** School of Information and Communication, Korea University, Anam-dong, Sungbuk-gu, Seoul, 136-713, Korea
{jswwwww, docdream, bshim, sjko}@korea.ac.kr

Abstract. In this paper, we propose our work on camera network using the MIMO Y channel. In this setup, each camera delivers information for connected cameras via the intermediate relay while receiving side information from other cameras. As an example for demonstrating the validity of proposed scheme, we introduce an interesting application of our camera network to sports activity analysis, where a virtual view synthesis is chosen as our case study. By exploiting our camera network, we can produce an intermediate virtual-view image with high accuracy over conventional virtual view synthesis schemes.

Keywords: MIMO Y channel**,** capacity, view synthesis, human body modeling.

1 Introduction

Recently, researchers have focused on developing camera network for many practical applications. In wireless network, multiuser MIMO schemes have drawn a lot of attention due to their potential to achieve high throughput [1]. More recently, Lee et al. [2] proposed the signal space alignment on the MIMO Y channel to investigate the network information flow problem. The capacity improvement over multiuser MIMO scheme is achieved by the efficient utilization of the signal space.

In this paper, we present a camera network using the MIMO Y channel model. In this network, each camera conveys side information to the other cameras via an intermediate relay. By exploiting signal space alignment which makes the information align at the same signal dimension of the relay, each camera can utilize side information coming from the other cameras as side information, and thus can produce more improved outputs. Among many applications which can be video chatting and smart observation system using camera-installed robot, we focus on generating a virtual-view image using our camera network.

The virtual view synthesis plays an important role to facilitate an analysis of players [3][4] or visualize a sports game more plausibly. This is because the number of installed cameras is often limited in most local and amateur sporting. The proposed virtual view synthesis system consists of kernel-based player tracking, silhouette extraction, ellipse fitting to the silhouette, and player interpolation. First, the locations

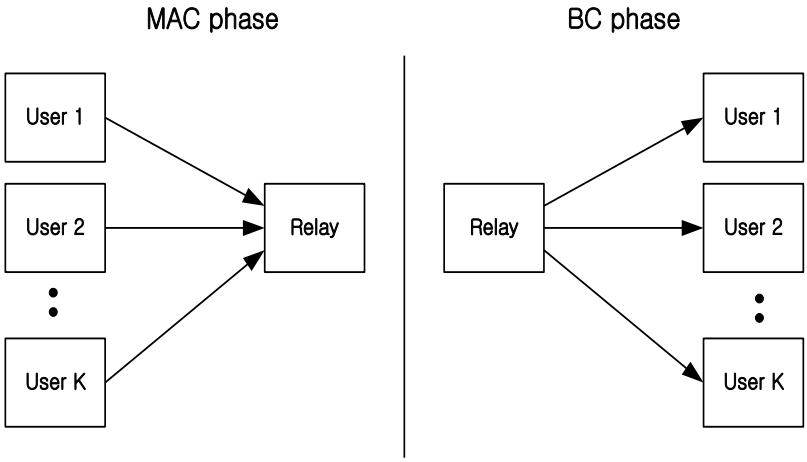

Figure 1. *K*-user MIMO Y channel network

of two players are found by tracking players from the top-view video. The positions of the players at other viewpoint images are simply obtained by projecting 3-D location of the players into each image. After extracting the silhouette of each player by using the level-set segmentation algorithm [5], player body is modeled by multiple ellipses [6]. Finally, the player at the intermediate viewpoint is reconstructed by interpolating the modeled players from two different viewpoint images.

The rest of this paper is organized as follows: In section 2, we describe the camera network using MIMO Y channel. The proposed virtual synthesis system is presented in Section 3. The experimental results are provided in Section 4, and the conclusions are given in Section 5.

2 Camera network using MIMO Y Channel

In this section, we present camera network using the MIMO Y channel. Recent work by Lee et al. referred to as signal space alignment for network coding on the MIMO Y channel [2] investigates a MIMO Gaussian wireless network with three users and a single intermediate relay. In this network, each user receives independent information from the other two users via the intermediate relay while transmitting independent information for two different users. The key feature of this scheme is to achieve large capacity over other conventional multiuser MIMO schemes, where the capacity denotes the measure of the information's performance in the wireless network. In Figure 1, we describe the *K*-user MIMO Y channel network. The MIMO Y channel network consists of K users with M antennas and an intermediate relay with N antennas. Each user sends K-1 independent information to the other corresponding K-1 users via the intermediate relay.

In the MAC phase, all users simultaneously transmit the information to the relay. The received signal at the relay Y_r can be expressed as

$$Y_r = H_{r,1}x_1 + H_{r,2}x_2 + \cdots + H_{r,K}x_K + n_r \quad (1)$$

where x_j is the transmitted information vector for *j*-th user with the power constraint $E[\|x_j\|^2] \le SNR$, $H_{r,j}$ is the $N \times M$ channel matrix for *j*-th user, and n_r is a complex Gaussian noise at the relay. In the BC phase, the relay broadcasts information to all users. The received signal at the *j*-th user is given by

$$Y_j = H_{j,r}x_r + n_j \quad \text{for} \quad j=1,2,\cdots,K \quad (2)$$

where x_r denotes the transmitted information vector at the relay with the power constraint $E[\|x_r\|^2] \le SNR$, $H_{j,r}$ is the $M \times N$ channel matrix from the relay to *j*-th user, and n_j is the AWGN noise vector at the *j*-th user.

The capacity of this signal space alignment on MIMO Y channel network achieves $3M \log(SNR) + o(\log(SNR))$ while the conventional multiuser MIMO scheme on the MIMO Y channel obtains $\min\{3M, N\}\log(SNR) + o(\log(SNR))$ [2], which means that the former scheme can transmit more information over the latter scheme. If we set M=2, N=3, the sum rate of the signal space alignment on the MIMO Y channel achieves twice as many as the sum rate of the conventional multiuser MIMO scheme in the high SNR regime.

To bring the advantage of capacity achievement, we apply the signal space alignment on the camera network using MIMO Y channel which synthesizes the virtual view image by interpolating captured images. In this network, each camera exchanges side information via the relay to produce robust virtual view image. The previous modeling information referred to as side information can help the other cameras to reconstruct the virtual view image. In the next section, we describe the detail of virtual view synthesis on the MIMO Y channel.

3 Application to Sports Activity analysis

3.1 System Environment

In this section, we apply the MIMO Y channel model to sports activity analysis problem. Figure 2 shows an example of our system, where a tennis player is captured by three cameras mounted in top and two side positions. Although we consider this simple set for simplicity, we mention that our model can be applied to more general scenarios.

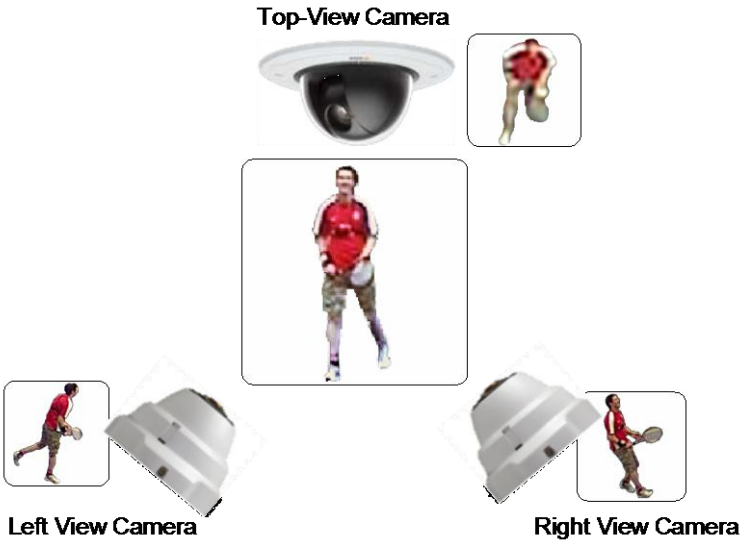

Figure 2. Camera environment

The top-view camera plays a role of the relay in the MIMO Y channel. As discussed in the previous section, the relay can send data to two side-view cameras through a wireless channel, and the two cameras can exchange data via the relay. By using this camera network, we can combine different pieces of information from three cameras to facilitate additional work such as human pose estimation, human modeling, and view interpolation. Among many possible applications, we focus on the virtual view synthesis in this paper. In the following subsection, we describe how this camera environment is used to the view synthesis problem.

3.2 Virtual View Synthesis

Virtual view synthesis is a technique to generate a virtual view by using existing views. In the classical stereo matching area, this can be done by estimating a disparity between two viewpoint images. However, this approach assumes a narrow baseline, i.e. two cameras are placed very close to each other. Therefore, the stereo matching based approach is not applicable to our wide-baseline situation.

In our work, we perform background and foreground interpolation separately. Since background interpolation is rather simple, we only focus on foreground interpolation in this paper. In order to interpolate foreground, i.e. tennis player, a shape of player at each view needs to be extracted. Thus the player region is first roughly found by object tracking and then the accurate shape is extracted by segmentation. Here, note that player tracking does not need to be done for every camera. Since we have the MIMO Y channel, each camera can receive the tracking result through the relay. Then, player tracking at the other cameras is simply done by projecting the available tracking result.

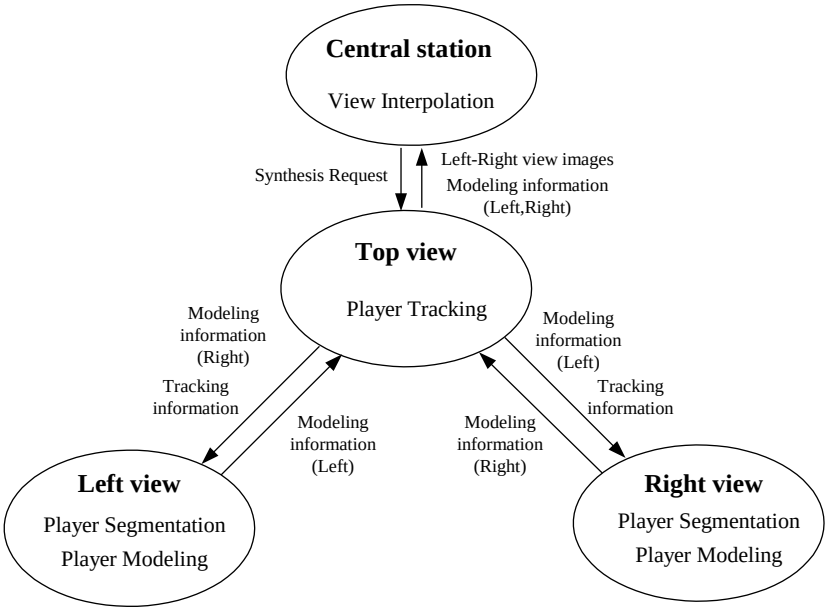

Figure 3. System architecture for view synthesis.

Figure 3 shows architecture of the proposed view synthesis system. Here, top-view camera is chosen to track a player since the player is always visible in the top-view. If top-view is not available, two or more cameras may need to track the player and exchange tracking information. After tracking is done by employing kernel based tracking (KBT) [8], the tracking information is transmitted to the other two cameras through the MIMO Y channel. In each side-view camera, the region containing the player can be obtained by projecting the top-view player region onto the side-view images. However, an accurate shape of the player cannot be obtained by this simple projection. Thus we implemented a level-set segmentation algorithm to extract a precise object shape [5]. Note that the two adopted algorithms, KBT and level-set segmentation, are widely used and efficient solution for tracking and segmentation, respectively. Detailed explanation of them can be found in many literatures [5].

Now we have two object shapes or silhouettes of the same player at the two different viewpoints. Due to a large distance between cameras, two objects have large partial occlusions. Therefore, the object to object interpolation method cannot provide accurate interpolation results. To this end, we adopt a human body modeling to segment the object [6][9]. In [6], human is modeled by connected multiple ellipses. Since tennis player holds a racket, two ellipses for the racket are additionally attached as shown in Figure 4. Then, in order to fit this model to the segmented player, the Gaussian mixture model (GMM) is used and the expectation maximization algorithm (EM) [7] is adopted to estimate the GMM parameters [6]. Since this modeling is a main part of the proposed view synthesis system, detailed description is given below.

After segmentation, pixel locations inside the player are determined. Let x be a pixel location inside the player and $p(x \mid \theta)$ be the probability distribution of x

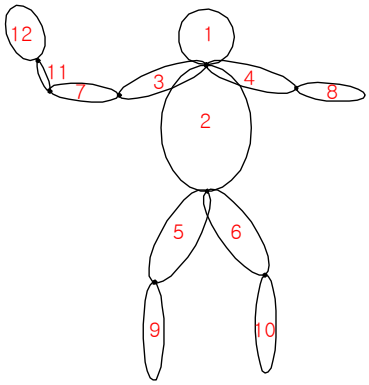

Figure 4. Ellipse model for tennis player

given parameter set $\theta = \{(\mu_i, \Sigma_i, \omega_i)\}$. Here, μ_i, Σ_i, and ω_i are the mean vector, covariance matrices, and the weight of *i*th Gaussian cluster, respectively. In GMM, $p(x \mid \theta)$ is represented by

$$p(x \mid \theta) = \sum_{i=1}^{N} w_i G_i(x; \mu_i, \Sigma_i), \tag{3}$$

where *N* is the number of clusters, $\sum_{i=1}^{N} w_i = 1$, and $G_i(\cdot)$ is an individual Gaussian probability density function.

Then, the EM algorithm finds the GMM parameters iteratively and provides the maximum likelihood estimate. However, the GMM-EM method is quite sensitive to the initial guess of $\theta^0 = \{(\mu^0, \Sigma^0, \omega^0)\}$. When the information exchange between cameras is impossible, a reasonable choice is to use the estimated parameters from the previous frame. However, since we have the MIMO Y channel camera network, the estimated parameters of the neighboring camera can be also exploited. Note that player segmentation itself is a difficult task and often inaccurate results are obtained. In such a case, if a more precise modeling is performed in the other view, the previously estimated parameters from the other view can be used as initial parameter for the current camera.

The obtained ellipses by the GMM-EM are usually disconnected. Thus the separated ellipses are forced to be connected by using the player body model in Figure 4. This compulsory connection usually improves the fitting accuracy because the GMM-EM process does not consider the human body model. Finally, in order to further enhance the fitting accuracy, the Levenberg-Marquardt (LM) algorithm is additionally employed [6]. The objective function used in the LM algorithm is composed of the sums of the distance between ellipse and silhouette contour. By minimizing this objective function, we can make the ellipses closely cover the contour of the player.

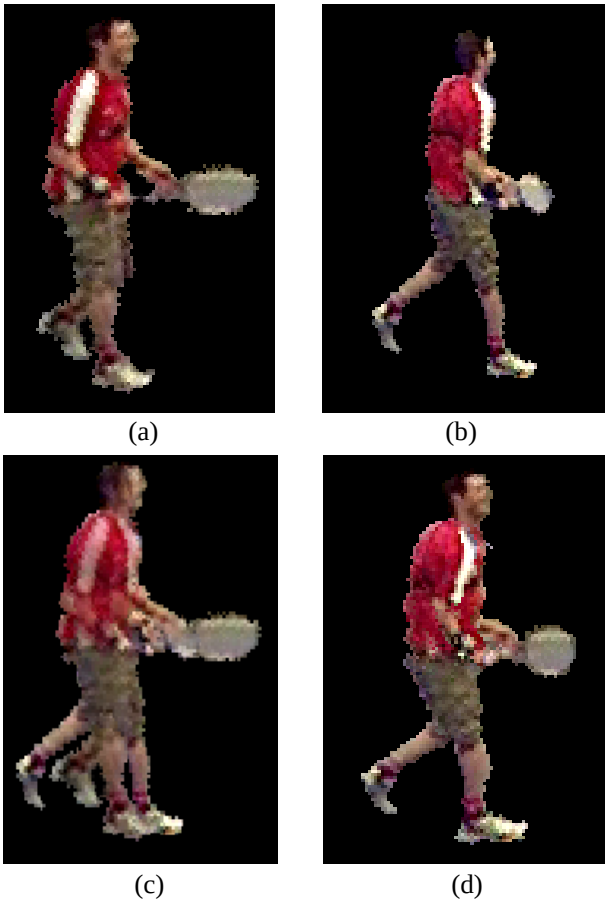

Figure 5. Player interpolation results. (a)interpolation result by transforming the image from left-view camera; (b)interpolation result by transforming the image from right-view camera; (c) average image of (a) and (b); (d) proposed segmentation based interpolation result.

When each camera finishes the player segmentation and modeling, the modeling result is transmitted to the relay, i.e. top-view camera. Then the relay sends the modeling results and the corresponding two images to a central station. The central station interpolates two side-view images using the modeling information. Since the central station is controlled by user, any arbitrary viewpoint can be generated once modeling information is given. The transformation matrix for each ellipse pair is found by [10] and the interpolation is performed independently for each pair. The interpolated segments are similarly connected by using the player body model the final interpolated foreground is achieved by combining two transformed images. The proposed method is fully image-based technique and does not require 3D human body modeling, training information, or any kinetic prior information. In the experiment section, we will demonstrate the advantage of the proposed view synthesis method.

4 Experimental Results

In this paper, we proposed a usability of the MIMO Y channel to the sports activity analysis scenario. Since we are currently constructing a network environment, the current experiment is done offline, where the test images and camera calibration data are available in [11]. After establishing our proposed camera network, we expect that we can realize our virtual view synthesis system in real time.

Among the many techniques in the proposed method, such as the KBT, level-set segmentation, ellipse fitting, and model-based image interpolation, only the final interpolation results are included. Figures 5 (a) and (b) show the interpolation results using the images from one view and another, respectively. Here, the whole body is interpolated without player body modeling. As can be seen, the whole body is simply transformed, and the player at the intermediate position is not effectively represented. Thus, the average of two results shown in Figure 5 (c) has strong artifacts. The experimental result in Figure 5(d) demonstrates that the proposed model based interpolation technique strongly outperforms the whole body interpolation.

Our view synthesis system currently does not perform well when the player is visible only at the one view. We plan to extend our work to produce reasonable visual quality results in such a case. In addition, we believe that more elaborate methods for occlusion handling can improve the interpolation performance.

5 Conclusion

In this paper, we presented an interesting application of the MIMO Y channel. By establishing a camera network using the MIMO Y channel model, each camera can exchange information and produce a high quality virtual-view image. In our camera network, a top view camera serves as a relay and two side-view cameras exchange player modeling information through the relay. A central station receives the final modeling information from the relay and reconstructs the virtual-view image using the image interpolation technique. Experimental results demonstrated that the proposed camera network can effectively support a view synthesis function and it is expected that several additional interesting applications can be devised by using the proposed camera network.

Acknowledgments. This research was supported by Seoul Future Contents Convergence (SFCC) Cluster established by Seoul R&BD Program (10570) and by Mid-career Researcher Program through National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0000449) and by a research grant from KOSEF (No. 2009-0083945) and by the International Research & Development Program of the NRF grant funded by the MEST (grant no. K20901001401-09B1300-03410).

References

1. Lee, N., Lim, J. and Chun, J.: 2010. Degrees of Freedom of the MIMO Y channel: Signal Space Alignment for Network Coding. IEEE Trans. Information Theory. vol. 56, no. 7, pp. 3332-3342 (2010)
2. Foschini, G. and Gans, M.:1998. On limits of wireless communications in a fading environment when using multiple antennas. Wireless Personal Communication. vol. 6, pp. 311-335 (1998)
3. Poppe. R.: Vision-based human motion analysis: An overview. Comput. Vis. Image Und. pp. 4-18 (2007)
4. Wang, L., Hu, W., and Tan, T.: Recent developments in human motion analysis. Pattern Recogn. pp. 585-601 (2003)
5. Shi, Y. and Karl, W. C.: Real-time tracking using level sets. In Proceedings of International Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, pp. 34-41 (2005)
6. Xu, R. Y. D. and Kemp, M.: 2010. Fitting multiple connected ellipses to an image silhouette hierarchically. IEEE Trans. Image Process. vol. 19, no. 7, pp. 1673-1682 (2010)
7. Bilmes, J. A.: A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Berkley, CA, USA: International Computer Science Institute (1997)
8. Comaniciu, D., Ramesh, V., and Meer, P.: Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. vol. 25, no. 5, pp. 564-577 (2003)
9. Rosin, R. L.: A note on the least squares fitting of ellipses. Pattern Recogn. Lett. pp. 799-808 (1993)
10. Alexa, M., Cohen-Or, D., and Levin, D.: As-rigid-as possible shape interpolator. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques ,New Orleans, LA, USA pp.157-164 (2000)., July 23-28,
11. O' Conaire, C., Kelly, P., Connaghan, D., and O'Connor, N. E.: TennisSense: A Platform for Extracting Semantic Information from Multi-camera Tennis Data. DSP 2009 - 16th International Conference on Digital Signal Processing, Santorini, Greece, pp. 1062-1067 (2009)

Seung-Won Jung received the B.S. degree in Electronics Engineering from Korea University in 2005. He is now a Ph.D. candidate in the Department of Electronics Engineering at Korea University. His research interests are image and video compression.

Byungju Lee received the B.S. degree in Radio and Communication Engineering from Korea University in 2008. He is now a Ph.D. candidate in the School of Information and Communication at Korea University. His research interests are wireless communication and statistical signal processing.

Byonghyo Shim received the B.S. and M.S. degree in Control and Instrumentation Engineering (currently Electrical Eng.) from Seoul National University in 1995 and 1997, respectively, and the M.S. degree in Mathematics and the Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2004 and 2005, respectively. From 1997 and 2000, he was with the Department of Electronics Engineering at the Korean Air Force Academy as an Officer (First Lieutenant) and an Academic Full-time instructor. He also had a short time research position in DSP group of LG Electronics and DSP R&D Center, Texas Instruments Incorporated, Dallas, TX, in 1997 and 2004, respectively. From 2005 to 2007, he was with the Qualcomm Inc., San Diego, CA as a Senior/Staff Engineer working on CDMA systems with the emphasis on the next generation UMTS receiver design. In September 2007, he joined the School of Information and Communication, Korea University, Seoul, Korea, as an Assistant Professor. His research interests include wireless communication, statistical signal processing, estimation and detection, applied linear algebra, and information theory. Dr. Shim was the recipient of the 2005 M.E. Van Valkenburg Research Award from the Electrical and Computer Engineering Department of the University of Illinois. He is a senior member of IEEE and member of Sigma Xi and Tau Beta Pi.

Sung-Jea Ko received the Ph.D. degree in 1988 and the M.S. degree in 1986, both in Electrical and Computer Engineering, from State University of New York at Buffalo, and the BS. degree in Electronic Engineering at Korea University in 1980. In 1992, he joined the Department of Electronic Engineering at Korea University where he is currently a Professor. From 1988 to 1992, he was an Assistant Professor of the Department of Electrical and Computer Engineering at the University of Michigan-Dearborn. He has published over 150 international journal articles. He also holds over 40 patents on video signal processing and multimedia communications.

He is currently a Senior Member in the IEEE, a Fellow in the IET and a Korean representative of IEEE Consumer Electronics society. He has been the Special Sessions chair for the IEEE Asia Pacific Conference on Circuits and Systems (1996). He has served as an Associate Editor for Journal of the Institute of Electronics Engineers of Korea (IEEK) (1996), Journal of Broadcast Engineering (1996 - 1999), the Journal of the Korean Institute of Communication Sciences (KICS) (1997 - 2000), and Journal of Selected Topics in Signal Processing (2009~). He has been a division editor of Journal of Communications and

Networks (JCN) (1998 - 2000). He is the 1999 Recipient of the LG Research Award given to the Outstanding Information and Communication Researcher. He received the Hae-Dong best paper award from the IEEK (1997) and the best paper award from the IEEE Asia Pacific Conference on Circuits and Systems (1996), and the research excellence award from Korea University (2004).