ASIAGRAPH 2008 The Intermediate View Synthesis System For Soccer Broadcasts

Songkran Jarusirisawad, Kunihiko Hayashi, Hideo Saito (Keio Univ.), Naho Inamoto (SGI Japan Ltd.), Tetsuya Kawamoto (Chukyo Television Broadcasting Co., Ltd.), Naoki Kubokawa, Toru Fujiwara (Nippon Television Network Corporation) songkran@ozawa.ics.keio.ac.jp, hayashi@ozawa.ics.keio.ac.jp

Abstract: In recent years, computer vision technologies are intensively applied for generating novel visual presentation of visual media for variety practical use including TV broadcast. In this paper, we will present a new technique for generating freeviewpoint video of soccer games that aim practical use with broadcasting quality. The key-point for the practical broadcast use is not automatic procedure for generating the free-viewpoint videos with some error frames, but efficient procedure for generating the free-viewpoint videos without any errors in frame. For achieving such efficient procedure, we propose an interactive user-interface for non-erroneous segmentation of the player from input soccer video. The segmentation of players is very important for generating high quality free-viewpoint videos, but it is no easy task. Therefore, our interactive segmentation with graph-cut algorithm can significantly contribute the efficient generation of free-viewpoint videos. In this paper, we will also demonstrate how the developed interface system works by showing the generated free-viewpoint videos.



1. Introduction

In traditional broadcasting system, viewer can see a scene only from the capturing position. New viewpoint video can break this limit by allowing viewer to see the scene from their desired viewpoint. By generating new viewpoint video in post-production of broadcasting system, virtual playbacks of the key events can be viewed from many angles. This technology will enhance the viewer's experience and it is a breakthrough of traditional broadcasting system.

Ideally, automatic view generating system is desired. However, some error frames in new view images usually occur due to the errors in player segmentation. Segmentation of player in a soccer scene is not trivial due to low resolution of the players in the input video and uncontrolled environment. Only a little miss segmentation may lead to obvious artifact in output new viewpoint video. In broadcasting where image quality is uncompromisable, perfect visual result is required. Manual operations for correct some error frames in automatic process are necessary.

In this paper, we will present a new technique for generating free-viewpoint video of soccer games target for practical use with broadcasting quality. The key-point for the practical broadcast use is not automatic procedure for generating the free-viewpoint videos with some error frames, but efficient procedure for generating the free-viewpoint videos without any errors in frame.

For achieving such efficient procedure, we propose an interactive user-interface for non-erroneous segmentation of the player from input soccer video. The segmentation of players is very important for generating high quality free-viewpoint videos, but it is no easy task. Therefore, our interactive segmentation with graph-cut algorithm [Boykov01] can significantly contribute the efficient

generation of free-viewpoint video. In this paper, we will also demonstrate how the developed interface system works by showing the generated free-viewpoint videos.

1.1. Related Works

Pioneering research in new viewpoint image synthesis of a dynamic scene is Virtualized Reality [Kanade95]. In that research, 51 cameras are placed around hemispherical dome called 3D Room to transcribe a scene. 3D structure of a moving human is then reconstruct for rendering from new view.

Many methods for improving quality of free viewpoint image have been proposed. Carranza recover human motion by fitting a human shaped model to multiple view silhouette input images for accurate shape recovery of the human body[Carranza03]. Starck optimizes a surface mesh using stereo and silhouette data to generate high accuracy virtual view image [Starck03]. Saito et al. propose appearancebased method [Sait099], which combines advantage from Image Based Rendering and Model Based Rendering.



In the mentioned researches, they proposed techniques mainly for using with studio environment. In outdoor environment, it is more difficult to generate high quality new viewpoint video due to uncontrolled light condition, natural background and less resolution of object (in case that a scene is large).

For the new viewpoint of outdoor sport scene, Kanade et al. have developed the eyevision system that was demonstrated at the Superbowl XXXV. Inamoto et al. have demonstrated a new viewpoint image system for soccer scene using image morphing [Inamoto04]. Billboards representation of soccer players have been used in new viewpoint system as well [Koyama03] [Hayashi06]. However these systems are limited either in the quality and some also require special equipment.

2. Capturing Environment

Figure 1 is the capturing environment and cameras position. We use 4 cameras to capture the soccer scene. The input is captured from the qualify match on December 2nd 2006 of 85th nationwide high school soccer championships Aichi prefecture between Chukyo high school of Chukyo University v.s. Tohou high school. Camera used for capturing the scene is Sony HDR-HC3 with video resolution 1440x1080 RGB 24 bits color. This capturing is supported by Chukyo Television Broadcasting and Nihon Television Broadcasting Company. Image and Video copyright is owned by Chukyo television.

3. New View Rendering Method

Our system use the view interpolation method proposed by Inamoto et.al [Inamoto03] and add interactive graph-cut functionality for segmenting soccer players without any errors for broadcasting quality. In this section, we will firstly explain Inamoto's method and the case where artifact might occur and need some manual correction. In section 4, we will explain our interactive system for generating new viewpoint image which can produce broadcasting quality new viewpoint image by resolving the case where artifact occurs in fully automatic system.

Figure 2 shows the overview of Inamoto's new viewpoint image generation method. From figure 2, soccer scene is divided into three types of area which are moving area (players and ball area), field area and audience area. From input images, fundamental matrices and homography of field and background regions between cameras are needed to render new viewpoint images. 2D-2D corresponding points for such estimation are selected from natural features manually.

Field and background regions are approximated as planar. Homography matrices are used for warping these regions to new viewpoint image.

Moving objects are firstly segmented from static area (ground and background) using ordinary background subtraction. This step sometimes produces error of segmentation in some frames and result to visual artifact in output new viewpoint image. Corresponding silhouette is



Figure 2 Overview of Inamoto's Method



Figure 3 Making Corresponding Silhouettes

made automatically by the following method. From the assumption that the feet of soccer players normally attach to the ground, correspondence of silhouette in another camera can be found using homography of the soccer field to warp foot position (assume to be the most left bottom pixel) to the another camera and find the closest foot position to match silhouette. Figure 3 shows the explained matching algorithm.

From corresponding silhouettes, dense pixels correspondence are found by project a pair of epipolar line onto both reference views. There will be four points, a_1 , b_1 , a_2 , and b_2 , that intersect the edge of corresponding silhouettes as figure 4. This line correspondence will be used for linear interpolation on a new view image. By

projecting enough pairs of epipolar line and make linear interpolation, new viewpoint image can be generated by this method.



Figure 4 Making Dense Correspondences

4. Proposed System

Using the method described in section 3, new viewpoint image can be generated automatically. However using conventional background subtraction causes bad silhouette segmentation result. Moreover, in some cases there could be miss corresponding silhouette which cause noticeable artifact in the rendered new viewpoint image.

We have developed Graphical User Interface (GUI) for solving these problems by having two functionalities. The first one is interactive foreground segmentation using graphcut. The second one is silhouette division and silhouette corresponding.

In the following section we will describe both functionalities in detail. Figure 5 shows our GUI for generating new viewpoint image.



Figure 5 Our Graphical User Interface

4.1 Interactive Soccer Players Segmentation Using Graph-Cut

Because automatic segmentation sometimes results in wrong segmentation, we develop interactive system for soccer players segmentation using graph-cut optimization method proposed in [Boykov01]. The user marks certain pixels as "object" or "background" to provide hard constraints for segmentation. The marked hard constrain pixels are guarantee to be segmented as "object" or "background" as user specify.

These hard constraints provide clues on what the user intends to segment. The rest of the image is segmented automatically by computing a global optimum among all segmentations satisfying the hard constraints. The cost function is defined in terms of boundary and region properties of the segments. These properties can be viewed as soft constraints for segmentation.

Once user specify hard constrain, graph-cut can find global segmentation result that gives globally optimal solution based on given hard constrain. If user found the result still unsatisfied, user can redefine hard constrain to get better result as many times as needed. Re-computation time is also fast enough to give interactive feeling to the user and convenient to use.



Figure 6 Soccer Players Segmentation Using Graph-Cut (1 Input Frame)

Figure 6 show snapshots of soccer player segmentation using graph-cut (one input frame). In figure 6 the green line is user-specified area of foreground (soccer players) and the red line is user-specified area of background. We can see that user need only a few interactions to complete segmentation task.

In case of input video that consists of consecutive frames rather than only single image, graph-cut segmentation is also suitable to apply in this case. User only needs to specify hard constrain in some key-frame, then this hard constrain will take effect to the previous and the next frame as well.

Figure 7 shows the example segmentation result of consecutive five frames. Hard constrains are only specified in three frames. The result shows a satisfy segmentation.

4.2 Interactive Silhouette Labelling and Corresponding

Silhouette labelling and finding corresponding silhouette in another camera could be done automatically using described method in section 3. By the way, problems of miss corresponding can occur in the following situations.

- Many players are segmented to the same silhouette

- Warped foot accidentally closer to another silhouette (of not the same person) in another camera

In the first case, user of the system can correct by manually dividing silhouette that contains multiple players into each player. In the second case, user can correct by redefining corresponding silhouette in the other camera. Using our GUI, these tasks can be done easily.

Figure 8 shows input images from two cameras. Figure 9 is the segmentation and labeling result using automatic method describe in section 3. Figure 10 is the segmentation and labeling result using our GUI. The same silhouette color means two silhouettes are corresponding to each other.

5. Experimental Result

In this section, we firstly compare the new viewpoint video generated using fully automatic method described in section 3 with the one from our interactive method. Figure 11 is new viewpoint images generated from segmentation and labeling result as in Figure 9 while Figure 12 shows new viewpoint images generated from Figure 10 (our system).

Figure 13 shows the result new viewpoint images from our proposed system. The weight ratios between two reference cameras are written under each figure. We can see that output rendered image has almost the same quality as input image itself. We cannot distinguish between rendered view and the actual input image so it is suitable for using in television broadcasting.

6. Conclusions

In this paper, we have presented the developed system for generating new viewpoint images targeted for practical use in broadcasting. Because quality of output new viewpoint images is the most important criteria, manual operation is needed when automatic process fails to achieve perfect result. Our system use graph-cut for interactive soccer player segmentation[Boykov01]. This could reduce tedious time user need to do manual segmentation. The result new viewpoint also shows promising quality of output images which is good enough and suitable for broadcasting.



Figure 8 Input Images from Camera3 and Camera4



(b) camera4

Figure 9 Segmentation and Corresponding Result Using Method Described in Section 3



(a) camera3



(a) camera3

(b) camera4

Figure 10 Correct Segmentation and Corresponding Result Using Our GUI

References

- Boykov,Y., Jolly,M.P., Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images, *ICCV2001*, vol. I, pp. 105-112
- Carranza, J., Theobalt, C., Magnor, M. Seidel, H.-P. Free viewpoint video of human actors. In *Proceedings of SIGGRAPH'03*, pages 569–577.
- Hayashi,K. Saito,H., Synthesizing free-viewpoint images from multiple view videos in soccer stadium. In *International Conference on Computer Graphics, Imaging and Visualisation,* 0:220–225, 2006.
- Inamoto, N., Saito, H., 2003. Fly-through viewpoint video system for multi-view soccer movie using viewpoint interpolation. In *Visual Communications and Image Processing (VCIP2003)*. Vol.5150, 122. Jul2003
- Inamoto, N., Saito, H., Arbitrary viewpoint observation for soccer match video. *European Conference on Visual Media Production*, pages 21–30, 2004

- Kanade, T., Rander, P. W., Narayanan, P.J. Virtualized reality: concepts and early results. In *IEEE Workshop on Representation of Visual Scenes*, pages 69–76, 1995.
- Koyama, T., Kitahara, I., Ohta, Y.. Live mixed-reality 3d video in soccer stadium. *International Symposium on Mixed and Augmented Reality*, pages 178–186, 2003.
- Saito, H., Kanade, T., Shape reconstruction in projective grid space from large number of images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, June 1999.
- Starck, J., Hilton, A., Towards a 3D virtual studio for human appearance capture. In *IMA International Conference on Vision*, *Video and Graphics (VVG)*, pages 17–24, 2003.



(a) User Specified Hard Constrain on Some Frames



(b) Segmentation Result Using Graph-Cut

Time

Figure 7 Consecutive Five Frames Which Are Segmented Using Graph-Cut



(a) Weight 2:8



(b) Weight 4:6



(c) Weight 6:4



Figure 11 New Viewpoint Image between Camera3 and Camera4 Generate from Figure 9



(a) Weight 2:8(b) Weight 4:6(c) Weight 6:4(d) Weight 8:2Figure 12 New Viewpoint Image between Camera3 and Camera4 Generated from Figure 10



(a) Camera 1

(b) Weight 2:8

(c) Weight 4:6

(d) Weight 6:4



(e) Weight 8:2



(f) Camera 2



(g) Weight 2:8



(h) Weight 4:6



(i) Weight 6:4



(j) Weight 8:2



(k) Camera 3



(l) Weight 8:2



(m) Weight 6:4

(n) Weight 4:6

(o) Weight 2:8

(p) Camera 4

