Multi-Camera Tracking Helmet System

Zirui. Zhang[1,2] and Jun. Cheng[1,2,3]
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shen Zhen, China
[2]The Chinese University of Hong Kong, China
[3]Guangdong Provincial Key Laboratory of Robotics and Intelligent System
Email: zirui.dream@gmail.com; jun.cheng@siat.ac.cn

***Abstract*—Object tracking in multiple cameras has been largely studied in the literature. However there are still unsolved problems, such as camera handoff, objects association, etc. In this paper we presents a novel multiple camera tracking helmet system, we stitch views from multiple cameras mounted on the helmet to one wide view in which we tracking objects, this simplifies the task of object tracking from multiple cameras. And combining with online training tracking algorithm, our tracking system demonstrates real-time and robust tracking performance.**

***Index Terms*—Multiple cameras, helmet system, stitching, tracking.**

I. Introduction

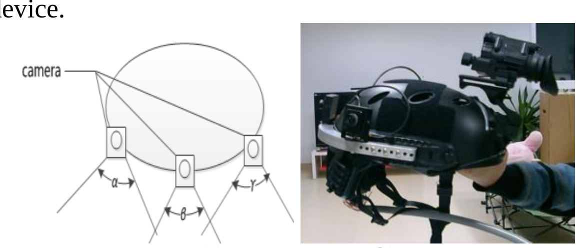
In recent years, multiple cameras tracking has attracted much attention in the computer vision community. By using multiple cameras, the camera view can be expanded largely, so more information can be obtained than using single camera. While at the same time, there are more issues need to handle, such as overlapping view, objects association (also known as consistent labeling), handoff of objects over multiple camera views. Many solutions have been addressed.

Khan [1] presents an attempt to tracking objects in multiple cameras with overlapping fields of view. Their approach is based on the computation of the so-called Edges of Field of View, i.e. the lines delimiting the field of view of each camera and, thus, defining the overlapped regions. Through a learning procedure in which a single track moves form one view to another, an automatic procedure computes these edges that are then exploited to keep consistent labels on the objects when they pass from one camera to the adjacent. It requires human feet to be visible, which may not be true in an indoor environment or in a crowd. Conversely, Lian et al. [2] try to tracking objects in multiple non-overlapping cameras, they proposes a Bayesian model to solve the consistent labeling problem across multiple non-overlapping camera views. In [3], a combined motion (position and velocity of object) and appearance model for human is built in database to help recognize objects in different camera views. Work in [4] presents a video summarization method to visualize the video object trajectories across multiple cameras in a static image for monitoring the movements of suspicious people in a building, human feature based method is used to associate objects. There are methods to associate objects observed from disjoint camera views using Markov chain Monte Carlo [5], and histogram of local feature based matching [6]. This setting is not expected to associate objects with high confidence due to the inherent ambiguity, especially in an environment with dense objects whose appearance can change significantly across views.

There are several other hybrid approaches have been introduced by researchers. In [7], Huang developed a tracking system consisting of several single camera tracking clients and information fusing server, and then use a TCP/IP network to exchange information between tracking clients. In [8], a laser scanner is used to assist the tracking task in multiple cameras, and in [9], a Multi-GPU based approach is introduced. These solutions can solve some issues of multiple cameras tracking in an extent, but they have the disadvantage that they are either too complex or high cost. In this paper we introduce a multiple cameras tracking helmet system, which is simple and low cost.

II. Overe View of Our Approach

To cover a wide area, we use multiple cameras in our helmet system. The cameras are fixed on the helmet (the helmet is not fixed). Each Camera gets view of a certain range. The adjacent camera pair has overlap in their field of views. We stitch the views from different cameras to produce a wide view and then tracking objects in the stitched view, which has the advantage that we don't need to handle those complex issues, such as object handout between multiple cameras and consistent labeling as we only need to track objects in the final one enlarged view. This simplifies our later object tracking task. Image (a) in Fig. 1 shows the configuration of the cameras on the helmet, and image (b) is our helmet device.



(a) (b)

Figure 1. Configuration of the cameras

Manuscript received February 10, 2013; revised April 16, 2013.

doi: 10.12720/joig.1.2.76-79

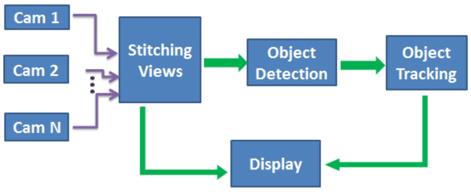


Figure 2. Block diagram of proposed system

And then, in our System, we apply the On-line Boosting algorithm which is robust to complex environment to track objects. This makes the tracking system stable even when the helmet device the cameras mounted on move slightly (at a certain degree). Finally, the panorama scene and tracking result will be shown on the micro-display mounted on the front of the helmet. The overall block diagram of our multi-cameras tracking system is show in Fig. 2.

III. STICHING VIEWS

In our system, as the cameras are relatively fixed, we can calibrate the cameras to find the focal length and the homographies between cameras. Using these pre-calibrated camera parameters, we can stitching image views from multiple cameras very fast. Generally, there are two steps to stitch views from different cameras in our system. We warp the image views to a surface by geometric transformation, and then blend them to produce the final panorama.

A. Geometric Transformation

Assuming that the camera rotates about its optical center, the group of transformations the images may undergo is a special group of homographies. We parameterize each camera by rotation vector $\theta=[\theta_1,\theta_2,\theta_3]$ and focal length f, this gives pairwise homographies

$$\tilde{x}_i = H_{ij}\tilde{x}_j \tag{1}$$

$$x_i = \frac{h_{00}x_j + h_{01}y_j + h_{02}}{h_{20}x_j + h_{21}y_j + h_{22}} \tag{2}$$

$$y_i = \frac{h_{10}x_j + h_{11}y_j + h_{12}}{h_{20}x_j + h_{21}y_j + h_{22}} \tag{3}$$

H_{ij} is a general 3 x 3 homography matrix and $\tilde{x}_i$ and $\tilde{x}_j$ are the homogeneous image positions positions. h_{ij} is the element of the homogrphy matrix.

$$H_{ij} = K_i R_i R_j^T K_j^{-1} \tag{4}$$

and K is called the intrinsic calibration matrix, defined by

$$K_i = \begin{bmatrix} f_i & 0 & c_x \\ 0 & f_i & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

The pair (c_x, c_y) in the last column of K is the optical center. R is a 3 by 3 rotation matrix.

B. Compositing

Once we have registered all of the camera view with respect to each other, we need to decide how to produce the final stitched image. The stitched images can be arranged using one of various map projections, such as flat projection, rectilinear projection, cylindrical projection, spherical projection, etc. And we also should select a final reference image. A natural approach is to select one of the images of the camera views as the reference and then warp all of the other images into the reference coordinate system. For compositing larger panoramas, the usual choice is to use a cylindrical (shown in Fig. 3 [10]) or spherical projection. From image coordinates (x, y), the projected coordinates on the cylinder (x', y') are given by

$$x' = s\theta = s\tan^{-1}\frac{x}{f} \tag{6}$$

$$y' = sh = s\tan^{-1}\frac{y}{\sqrt{x^2+f^2}} \tag{7}$$

where s is the radius of the cylinder. Similarly, we can map world coordinates into 2D spherical coordinates (θ, ϕ) using

$$x' = s\theta = s\tan^{-1}\frac{x}{f} \tag{8}$$

$$y' = s\phi = s\tan^{-1}\frac{y}{\sqrt{x^2+f^2}} \tag{9}$$

And for the selection of the reference image, a reasonable choice is the one view that is geometrically most central.

C. Blending

After mapping the source pixels onto the final composite surface, the next step is to blend theme to produce a panorama. The simplest way to create a final composite is to simply take an average value at each pixel

$$C(x) = \sum_k w_k(x)\tilde{I}_k(x) / \sum_k w_k(x) \tag{10}$$

where $\tilde{I}_k(x)$ are the warped images and $w_k(x)$ is 1 at valid pixels and 0 elsewhere. A better approach is to weight image proportional to its distance from the edge. This can be done by computing a distance map, and this is often called feathering blending. There are other approaches, such as center-weight, minimum likelihood and multiband blending [11] which can create better blended image while at a higher time cost. And to make the panorama looks ‘beautiful’, some other works should be done, such as exposure compensation, deghosting and so on.

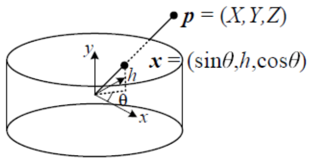


Figure 3. Cylindrical warping

IV. TRACKING

Our tracking system is designed to tracking objects in complex environment, so in this part, we utilize the On-line Boosting tracking approach [12], which considers the tracking problem as a binary classification problem between object and background, and was demonstrated good performance on object tracking in complex environment for its online learning capability. The basis is an on-line AdaBoost algorithm which allows updating features of the classifier during tracking. We use three different types of features for generating weak hypotheses, haar-like features, histogram of oriented gradients [13] and a simple version of local binary patterns, the combination of these three features makes the tracking robust in complex environment .By using integral images and integral histograms as data structures, the computation of all feature types can be done very efficiently. This allows to tracking objects in real-time.

V. EXPERIMENTS AND DISCUSSION

In this part, we perform experiments demonstrating our multiple cameras tracking helmet system. We mounted cameras on the helmet (with a 3.0 GHZ processor and 1 GB RAM), and grabbed image views of resolution of 640 by 480 in our experiments.

A. *Stitching Views*

In Fig. 4, image (a), (b), (c) are the 3 views form three cameras separately. We mapped the 3 image views onto a plane surface, and then blended them to produce the final stitched Image (d).

For 3 images of resolution of 640 by 480 taken from the cameras, mapping the images on a plane surface and applying with feathering blending algorithm, the stitching process costs about 40ms, which meets the real time requirement.

B. *Tracking*

In our experiments, we tried to track human in our stitched wide camera view. First, we used histograms of oriented gradients feature for human object detection, which is the state of the art algorithm for human detection. And then a linear SVM classifier was applied to classify the human objects. After getting the initial positions of the human objects, we then applied On-line Boosting algorithm to track the human objects.

Image (a) and image (b) In Fig. 5 were corresponding views from the left camera and the right camera, and image (d) shows the tracking result in the stitched view when the object moved across the views.

Image (a) to image (f) in Fig. 6 shows the tracking result in different illumination environment, and the appearance of the object changes during the tracking process. From the tracking result, the tracking results demonstrate robustness of our tracking system.

Image (a), (b) and (c) in Fig. 7 shows views from 3 cameras separately, from the image views in image (d) we can see that our approach enlarged our horizontal view. And then, we track human objects in the enlarged view.

(a) Left View (b) Middle View (c) Right View

(d) Stitched View

Figure 4.

(a) Left View (b) Right View

(c) Tracking Result

Figure 5.

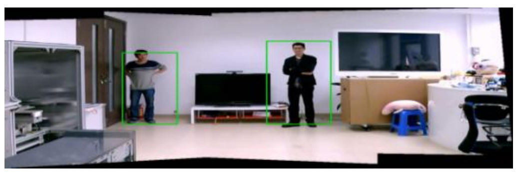
(a) (b)

(c) (d)

(e) (f)

Figure 6.

(a) Left View (b) Middle View (c) Right View



(d) Tracking objects in stitched view

Figure 7.

VI. CONCLUSION

In this paper we introduced a multiple camera tracking helmet system, our system differs from most tracking system in that we stitch views from multiple cameras, and then apply On-line Boosting algorithm to track objects in a wide view other than tracking objects in different cameras separately. As we pre-calibrate the cameras, the stitching process is fast, and we only need to tracking objects in one view (the stitched result), which simplify the task of object tracking from multiple cameras. And benefit from the robustness of the On-line Boosting tracking algorithm, the tracking process performs well in complex environments.

VII. FUTURE WORK

As aberration of different cameras exists, the image views form different cameras may look different in color which leads the final stitching result looks like not that "natural". We may do some exposure compensation works for the camera images in the future. And to achieve real time stitching, we used simple feather blending algorithm in our experiment, which was very fast but the final images may looks blur for the project error of the computing of the homographies between the cameras, as suggested by MB [11] et all, we may try to use Multi-Band Blending algorithm to get more clear stitching result and still keep the real time property at the same time. And Another problem we need to consider in our future work is that how to solve the parallax problem when the objects are too close to the cameras, which may lead blur of the stitched image.

ACKNOWLEDGMENT

Special thanks to CAS and Locality Cooperation Projects (ZNGZ-2011-012), Guangdong-HongKong Technology Cooperation Funding (2011A09- 1200001), Guangdong Innovative Research Team Program (No.201001D0104648280), Shenzhen Technology Project (ZD200904300074A).

REFERENCES

[1] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 25, pp. 1355-1360, 2003.

[2] G. Lian, J. Lai, and W.-S. Zheng, "Spatial–temporal consistent labeling of tracked pedestrians across non-overlapping camera views," *Pattern Recognition*, vol. 44, pp. 1121-1136, 2011.

[3] A. Ilyas, M. Scuturici, and S. Miguet, "A combined motion and appearance model for human tracking in multiple cameras environment," in *Proc. 6th International Conference on Emerging Technologies*, 2010, pp. 198-203.

[4] S. Liu and S. Lai, "Schematic visualization of object trajectories across multiple cameras for indoor surveillances," in *Proc. Fifth International Conference on Image and Graphics*, 2009, pp. 406-411.

[5] H. Kim, J. Romberg, and W. Wolf, "Multi-camera tracking on a graph using markov chain monte carlo," in *Proc. Third ACM/IEEE International Conference on Distributed Smart Cameras*, 2009, pp. 1-8.

[6] L. F. Teixeira and L. Corte-Real, "Video object matching across multiple independent views using local descriptors and adaptive learning," *Pattern Recognition Letters*, vol. 30, pp. 157-167, 2009.

[7] D. T. Lin and K. Y. Huang, "Collaborative pedestrian tracking with multiple cameras: Data fusion and visualization," in *Proc. International Joint Conference on Neural Networks*, 2010, pp. 1-8.

[8] H. Noguchi, T. Mori, T. Matsumoto, M. Shimosaka, and T. Sato, "Multiple-person tracking by multiple cameras and laser range scanners in indoor environments," *Journal of Robotics and Mechatronics*, vol. 22, pp. 221, 2010.

[9] S. Orts-Escolano, J. Garcia-Rodriguez, V. Morell-Gimenez, J. Azorin-Lopez, and J. Garcia-Chamizo, "Multi-GPU based camera network system keeps privacy using growing neural gas," in *Proc. International Joint Conference on Neural Networks*, 2012, pp. 1-8.

[10] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, pp. 1-104, 2006.

[11] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, pp. 59-73, Aug 2007.

[12] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *Proc. BMVC*, 2006, pp. 47-56.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886-893.

Zirui Zhang received the bachelor degree in computer science from the Chang Chun University of Science and Technology in 2010. He is currently a graduate student of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, majoring in computer science. His research interests include pattern recognition, computer vision, and machine learning.

Jun Cheng received the B.Eng., B.Fin., and M.Eng. degrees from the University of Science and Technology of China, Beijing, in 1999 and 2002 respectively, and the Ph.D. degree from the Chinese University of Hong Kong in 2006. Currently he is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, as a Professor and Director of the Laboratory for Human Machine Control. His Research interest include computer vision, pattern recognition, human machine interface, robotics