A proposed image stitching method for web-based panoramic virtual reality for Hoseo Cyber Museum

Irfan Khan[1] **and Hong Song Soo**[1*]

[1]**Dept. of Computer Engineering, Hoseo University**

호서 사이버 박물관: 웹기반의 파노라마 비디오 가상현실에 대한 효율적인 이미지 스티칭 알고리즘

아르판 칸[1], 홍성수[1*]

[1]호서대학교 컴퓨터공학과

Abstract   It is always a dream to recreate the experience of a particular place, the Panorama Virtual Reality has been interpreted as a kind of technology to create virtual environments and the ability to maneuver angle for and select the path of view in a dynamic scene. In this paper we examined an efficient method for Image registration and stitching of captured imaged. Two approaches are studied in this paper. First, dynamic programming is used to spot the ideal key points, match these points to merge adjacent images together, later image blending is used for smooth color transitions. In second approach, FAST and SURF detection are used to find distinct features in the images and nearest neighbor algorithm is used to match corresponding features, estimate homography with matched key points using RANSAC. The paper also covers the automatically choosing (recognizing, comparing) images to stitching method.

요 약   파노라마 가상현실이란 특정 장소의 경험을 재현하는 방식으로, 현실 세계의 장소에 직접 가보지 않고 가상 현실 속의 사물이나 정보를 보다 쉽고 빠르게 탐색하고 습득 할 수 있다. 본 논문에서는, 우리는 이상적인 키 포인트를 탐지하는 동적 프로그래밍을 사용하여 함께 이 지점과 인접한 이미지를 병합하고, 부드러운 색상 전환을 위해 이미지를 혼합하는데 사용된다. FAST와 SURF 탐지는 이미지의 확실한 특징을 찾는데 사용되고, 가장 가까운 이웃 알고리즘은 해당되는 특징을 일치시키는데 사용되며, RANSAC을 사용하여 일치하는 키 포인트를 homography로 판단한다. 이러한 방법으로 이미지를 자동 선택하여 스티칭하는 방법을 사용한다

Key Words : Panoramic VR, Image Stitching, Image Registration, SURF Descriptor

1. Introduction

A Next Generation Network (NGN) is a packet-based network able to provide services including telecommunication Services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It offers unrestricted access by users to different service providers. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users. Cyber Shell Museum is defined as the use of audio video, and other telecommunications and electronic information processing technologies for the transmission of information and data relevant to the shells. Cyber Shell Museum is designed with the aim of educating both

scholarly and public about their culture while leaving out the high costs of having to travel to particular destinations. The NGN Cyber Shell Museum Prototype System is built around the internet Protocol, offering converged, fixed and mobile real-time multimedia services. It allows users to communicate through voice, video and text and panoramic virtual reality tour. It also allows users to share files. Users can easily add and remove calling parties, media streams and switch between sessions, devices and fixed or mobile connection, start a new chat session, upgrade the session to a voice or video call, or add a new participant.
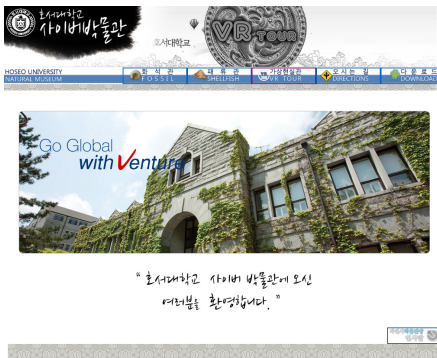
[Fig. 1] Example of Hoseo Cyber Museum Panoramic VR

1.1 Virtual Reality and Panorama VR

"Virtual Reality (VR) is a technology which allows a user to interact with a computer-simulated environment, be it a real or imagined one." Burdea and Coiffet (1994) pointed out that Virtual Reality is a high-end users' interface including immersion, interaction and imagination[10]. Xiao and Othman et al (2002) mentioned that Panorama VR can be defined as "telepresences" and showed that Panorama based VR is an extension of computer graphics and virtual reality. Also revealed that that Panorama VR is an effective communication tool for visualizing information and website representation makes it prefect for museum exhibitions.

1.2 Panning

Panorama VR that allows users to stand in a virtual place to look around and tilt up and down is the basic feature. Users drag with the mouse across to look left and right (Pan), and up and down (Tilt) across the vista to
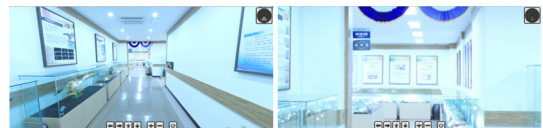
have a panoramic view as shown in Fig.2
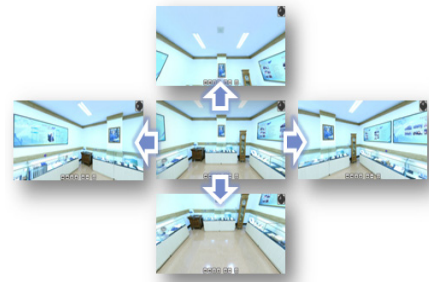
[Fig. 2] The Panorama VR shows the right, left, up and down of Hoseo Cyber Museum

1.3 Zooming In and Out

The zoom in and out function of Panorama VR is to provide a close look at an object as show in Fig.3

[Fig. 3] The zoom in and out effect

1.4 Hot Spot

The Hotspot navigation system is the common method used to explore the virtual environment. The hotspot navigation style of panoramic VR is to jump from one panorama to another.

[Fig. 4] Red Dot in the picture is a hotspot to link to jump between scenes.

2. Related Work

The most widely used detector, Harris corner detector [1], proposed in 1988. Even larger variety of feature descriptors has been proposed, like Gaussian derivatives [2], steerable filters [3], phase-based local features [4]. Various algorithms has been proposed for homography

estimation. in this paper is restricted to algorithms for point correspondences, and then we describe an optimized, automated and reliable method for both images joining and blending and also a technique for multiple panoramas.

3. SURF DESCRIPTOR

SURF (Speeded Up Robust Feature) is a robust local feature detector, first presented by Herbert Bay et al. in 2006, that can be used in computer vision tasks like object recognition or 3D reconstruction. It is partly inspired by the SIFT descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images.

It uses an integer approximation to the determinant of Hessian blob detector, which can be computed extremely quickly with an integral image (3 integer operations). For features, it uses the sum of the Haar wavelet response around the point of interest. Again, these can be computed with the aid of the integral image[5]

The proposed SURF descriptor constructs a square with the selected positions and extract the SURF descriptor from it.

3.1 Object Recognition

New features are also tested for reorganizing objects in a museum. The database consists of 155 images of 80 objects. The test images were taken under various conditions, including extreme lighting changes, objects in reflecting glass cabinets, viewpoint changes, zoom, diff erent camera qualities, etc. For object to recognition from the DB, following steps are taken. Test images are compared to all pertaining images by matching their interest points. The object with the highest number of matches with test image is picked as the recognised object.


[Fig. 5] An example image from the reference set (left) and the test set (right). Note the difference in viewpoint and colours.

4. Mathematical basic and Image Stitching Algorithm Overview

Find a set of distinctive key-points, Extract and normalize the region content, Define a region around each keypoint, Compute a local descriptor from the normalized region, Match local descriptors

Step1: Detect key points (e.g.,SIFT, FAST, SURF)

Step2: Match key points (e.g., 1st/2ndNN<thresh)

Step3: Estimate homography with four matched key points (using RANSAC)

Step4: Project onto a surface and blend

For projection we have used planer surface
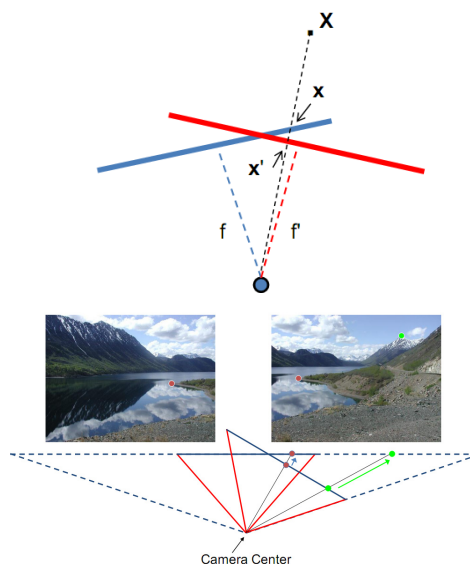
[Fig. 6] Planner Mapping
a. For blue image: pixels are already on the planar surface
b. For red image: map to first image plane

4.1 Homography (Direct Linear Transformation)

A homography is an invertible transformation from a projective space to itself that maps straight lines to straight lines.

$$\mathbf{x'} = \mathbf{Hx} \qquad \mathbf{x'} = \begin{bmatrix} w'u' \\ w'v' \\ w' \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$$

$$\begin{bmatrix} -u & -v & -1 & 0 & 0 & 0 & uu' & vu' & u' \\ 0 & 0 & 0 & -u & -v & -1 & uv' & vv' & v' \end{bmatrix} \mathbf{h} = \mathbf{0} \qquad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix}$$

$$\begin{bmatrix} -u_1 & -v_1 & -1 & 0 & 0 & 0 & u_1 u_1' & v_1 u_1' & u_1' \\ 0 & 0 & 0 & -u_1 & -v_1 & -1 & u_1 v_1' & v_1 v_1' & v_1' \\ & & & & \vdots & & & & \\ 0 & 0 & 0 & -u_n & -v_n & -1 & u_n v_n' & v_n v_n' & v_n' \end{bmatrix} \mathbf{h} = \mathbf{0} \Rightarrow \mathbf{Ah} = \mathbf{0}$$

- Apply SVD: $\mathbf{UDV}^T = \mathbf{A}$
- $\mathbf{h} = \mathbf{V}_{\text{smallest}}$ (column of $\mathbf{V}$ corr. to smallest singular value)

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_9 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$$

Matlab
```
[U, S, V] = svd(A);
h = V(:, end);
```

4.2 Normalized DLT

Step1. Normalize coordinates for each image
   a) Translate for zero mean
   b) Scale so that u and v are $\sim=1$ on average

$$\tilde{\mathbf{x}} = \mathbf{Tx} \qquad \tilde{\mathbf{x}}' = \mathbf{T'x'}$$

Step2. Compute using DLT in normalized coordinates

Step3. Unnormalize: $\mathbf{H} = \mathbf{T'}^{-1}\tilde{\mathbf{H}}\mathbf{T} \qquad \mathbf{x}_i' = \mathbf{Hx}_i$

4.3 RANSAC Algorithm

Step1: Choose number of samples N

Step2: Choose 4 random potential matches

Step3: Compute $\mathbf{H}$ using normalized DLT

Step4: Project points from $\mathbf{x}$ to $\mathbf{x'}$ for each
   $\mathbf{x}_i' = \mathbf{Hx}_i$ potentially matching pair:

Step5: Count points with projected distance < t
   (e.g. t = 3 pixels)

Step6: Repeat steps 2-5 N times choose $\mathbf{H}$ with most inliers

For probability p of no outliers:

$$N = \log(1 - p) / \log(1 - (1 - \epsilon)^s)$$

- N, number of samples
- s, size of sample set
- ϵ, proportion of outliers

e.g. for $p = 0.95$

Sample size	Proportion of outliers ϵ						
s	5%	10%	20%	25%	30%	40%	50%
2	2	2	3	4	5	7	11
3	2	3	5	6	8	13	23
4	2	3	6	8	11	22	47
5	3	4	8	12	17	38	95
6	3	4	10	16	24	63	191
7	3	5	13	21	35	106	382
8	3	6	17	29	51	177	766

5. Automatic Image Stitching Homography Estimation with RANSAC

The RANdom SAmple Consensus (RANSAC) algorithm proposed by Fischler and Bolles. RANSAC is a resampling technique that generates candidate solutions by using the minimum number observations (data points) required to estimate the underlying model parameters. RANSAC uses the smallest set possible and proceeds to enlarge this set with consistent data points.
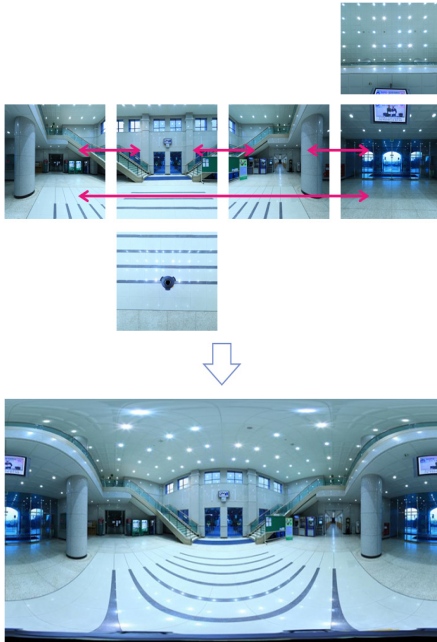
5.1 Automatically choosing images to stitch and Recognizing Panoramas

Input: N images

Step1: Extract points, descriptors from all images.

Step2: Find K-nearest neighbors for each point (K=4).

Step3: For each image
   a) Select M candidate matching images by counting matched keypoints (m=6)
   b) Solve homography Hijfor each matched image
   c) Decide if match is valid (ni>8+0.3nf)

Step4: Find connected components

Step5: For each connected component
   a) Perform bundle adjustment to solve for rotation ($\Theta1$, $\Theta2$, $\Theta3$) and focal length f of all cameras
   b) Project to a surface (plane, cylinder, or sphere)
   c) Render with multiband blending

Step6. Blending
   Gain compensation: minimize intensity difference of overlapping pixels

a) Pixels near center of image get more weight

b) Multiband blending to prevent blurring

Output: Panoramic Image

The analysis of two adjacent cameras is displayed in figure7. Because horizontal lines of adjacent mesh are matching so direct overlapping from 0 to 8% is an reasonable range of displacement as shown in figure 7(a,b,c). 9% overlapping shows diverging as shown in figure 7(e), the result gives evidence that the proposed stitching method is potentially workable. Red points are positioned on the mesh for identifying the divergence areas as shown in figure 7(d,e).

(a) 0 % overlapping

(b) 5% overlapping

(c) 8% overlapping

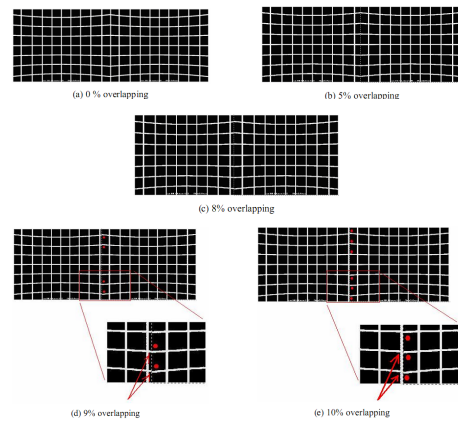(d) 9% overlapping

(e) 10% overlapping

6. Image stitching and overlap displacement analysis

Developing panorama VR using the multi-camera system is inevitable to face the issues of image alignment. To find out the amount of alignment a connecting simulated mesh is used to analyse the camera positioning and stitching process as shown in figure 7, surrounding the virtual multi-camera system to capture images for the purpose of analysing the displacement of alignment. The parameters setting of the virtual cameras are simulating the real digital camera. There are ten equally divided sections on the mesh can be seen in the lens of each camera for displacement analysis.
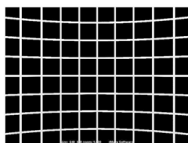
[Fig. 7] View of the virtual mesh from camera lens

7. Conclusions and future work

In this paper, we have discussed the method for generating panoramic images, the stitching of the acquired images by detecting, matching and computing homography and we hope to provide a better understanding of the different stages involved in the generation of panoramic images. The proposed method provides a cost effective and flexible alternative to acquire panoramic images using a panoramic camera. These panoramic images can be used for 360 degree interactive panoramic virtual reality movies, or multi-node panoramic movies.

AR apps and QR codes are getting plenty of media attention. The Next step is to study and develop the QR code and augmented reality system for museum related exhibits so people can access them anywhere they want and learn about them in friendly environment with touch of augmented reality.

References

[1] 10. Harris, C., Stephens, M.: A combined corner and edge detector. (1988)

[2] Florack, L.M.J., Haar Romeny, B.M.t., Koenderink, J.J., Viergever, M.A.: General intensity transformations and differential invariants (1994).

[3] Freeman, W.T., Adelson, E.H. The design and use of steerable filters (1991).

[4] Carneiro, Jepson, Multi-scale phase-based local features (2003)

[5] Wikipedia: http://en.wikipedia.org/wiki/SURF

[6] R. Hartley and A. Zisserman. Multiple View Geomerty in Computer Vision. Cambridge University Press, second edition, 2003.

[7] R. Hartley. In defense of the eight-point algorithm, June 1997.

[8] Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: CVPR (2). (2004).

[9] Mikolajczyk, K.Schmid: A performance evaluation of local descriptors. PAMI (2005)

[10] Brudea and Coiffet (1994) Virtuality Reality Techonology, Volume1.

[11] Konstantinos G. Derpanis, "Overview of the RANSAC Algorithm" May 13, 2010.

Irfan Khan   [Regular member]

- 2005년 12월 : Hoseo University, Asan, Korea, Computer Engineering
- 2012년 12월 : Hoseo University, Asan, Korea, Computer Engineering

<Research Interests>
His research interest includes mobile anbd wireless multimedia application, coding.

Sung–Soo Hong   [Regular member]

- Dec. 1983 : Kwangwoon University, Computer Engineering
- Dec. 1990 : Kwangwoon University, Computer Engineering
- Mar. 1984 ~ current : until present: Professor at Hoseo University.

<Research Interests>
His main interest of research includes computer vision, algorithm, data structure and mobile engineering