

# Video Synthesis System for Monitoring Closed Sections

<sup>1</sup>Taehyeong Kim\*, <sup>2</sup>Bum-Jin Park

<sup>1</sup>Senior Researcher, Korea Institute of Construction Technology, Korea

<sup>2</sup>Senior Researcher, Korea Institute of Construction Technology, Korea

<sup>1</sup>[tommykim@kict.re.kr](mailto:tommykim@kict.re.kr), <sup>2</sup>[park\\_bumjin@kict.re.kr](mailto:park_bumjin@kict.re.kr)

## ABSTRACT

This study is aimed at developing view synthesis system for closed sections such as Tunnel, Bridge etc. to monitor and track the vehicle movement. For this purpose, video storage and synthesis systems were developed and demonstrated, making use of the video taken and stored in 4 web cameras. Consequently original technology has been secured but for commercialization, application to the sites with system supplementation is more than important. What noteworthy is, this system is highly evaluated by the potential users including the monitoring managers of regional construction and management administration.

**Keywords:** *Video synthesis, Object tracking, Closed section, Monitoring*

## 1. INTRODUCTION

The need of advanced monitoring is increasingly growing to secure the facility safety and prevent secondary fatal accident through safe management of traffic accidents that may occur at large scale of SOC (Social Overhead Capital) facility such as a long tunnel or bridge, which is considered closed section. However, current management of SOC facility is heavily dependent on CCTV (Closed Circuit Television) screen which directly monitors the specific spot only. For such a reason, a camera is linked to a single monitor and thus the observer have to keep checking the moving object through a multiple number of monitor and to be aware of the real location displayed on monitor, which might become the negative factor for observer to take necessary action in timely manner. Thus, there is the need of video synthesis system in order to provide a continuous screen for a manager by composing the territories of CCTVs which monitor a spot and support a quick decision for a situation.

Therefore, in this study, real-time view synthesis system for closed section is developed to monitor and track the vehicle movement. Also, the developed system is demonstrated on an experiment based on 4 web cameras.

## 2. LITERATURE REVIEW

The research of video synthesis is divided into several areas such as image quality improvement, improvement of the accuracy of geographic coordinates, creation of stereo 3D data, feature improvement, improvement of image classification, change detection, restoration of the missing observations, and security-control, according to the purpose of research. This study belongs to security-control field and this area of research has been studied since the late 2000s with the concept establishment. Through the research of video synthesis, a variety of algorithms has been developed to reduce the computational complexity of the image and to increase the quality of the image. A set of procedures is needed to compose images. Thus, we summarize related works into

three parts, image stitching, image blending, and automatic extraction of image seams.

### 2.1 Image Stitching

The methods of creating panoramic images through composing captured images of the same object can be classified into two methods such as direct-based and feature-based. However, direct-based method is now rarely used due to the problem of large computational complexity and performance, and most of the research advances were made to the feature-based method.

A typical example of feature-based methods is SIFT (Scale Invariant Feature Transform) and SURF (Speed Up Robust Features). SIFT is to extract size and rotation invariant feature (Lowe, 1999 [1]). SIFT is often implemented through 4 processes such as Scale-Space extrema detection, Keypoint localization, Orientation assignment, Keypoint descriptor. SURF was developed to improve the speed by reducing the computational complexity compared to SIFT (Bay et. al, 2008 [2]). For this purpose, the detector and the descriptor can be utilized to reduce the dimension and integral image is used. The processing speed of SURF is several times faster than that of SIFT, but the accuracy of SURF drops a bit. Due to the dependence on the background, SURF may or may not have better performance than SIFT according to the situation. However, SURF is evaluated as more advanced technique than SIFT for image synthesis. SURF technique is classified into 4 processes such as interest point detection, image pyramid, interest point descriptor, and matching.

### 2.2 Image Blending

Image blending is to naturally connect boundary using the values around the boundary. There exists a variety of blending methods. The simplest one among blending methods is Alpha blending (Porter and Duff, 1984 [3]). Alpha blending is a simple algorithm with the advantage of fast speed. In addition to the RGB, one more is used for 3-D graphic, which is called as Alpha channel having the information of transparency of the color. Alpha blending mixes the background color and the color of object painted on the background using the value of alpha

<http://www.cisjournal.org>

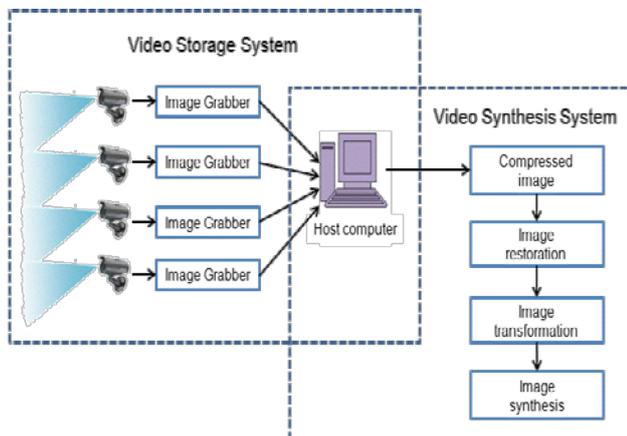
channel. If the object is fully transparent, the background color will be painted. Conversely, if the object is not completely transparent, just the object will be painted on the body. This alpha blending has good performance, but, in practice, the reliability of the image information around boundary becomes low due to a blurred image.

### 2.3 Automatic Extraction of Image Seams

The synthesized images may have the difference between the brightness values. The automatic extraction techniques were developed to adjust brightness values naturally through finding image seams for these synthetic images. There is assumption that the discontinuity and the discrepancies at the part of the seams can be minimized by selecting the part where the difference between brightness values of two images around boundary and overlap areas is minimized as seams. A typical example of the methods for automatic extraction of image seams is MAGDS (Minimum Absolute Gray Difference Sum) and Canny edge detection. MAGDS algorithm selects the part where the difference of brightness values is minimized as seams (Milgram, 1975 [4]). Canny edge detection algorithm detects the part where the gradient is maximized in smoothed image through Gaussian filter as seams (Canny, 1986 [5]).

## 3. REAL-TIME VIDEO SYNTHESIS SYSTEM

The real-time video synthesis system for the closed section is divided into two parts, video storage system and video synthesis system as shown in Fig 1.



**Fig 1:** The structure of real-time video synthesis system

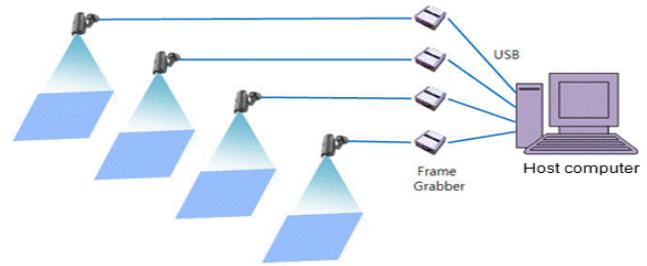
### 3.1 Video Storage System

#### 3.1.1 Overview

We made the test environment to validate the basic algorithm to develop the aimed system and performed test works repeatedly for verification of the basic algorithm. Since the real image changes every time, it is difficult to test the algorithm using this image. So, in order to overcome this problem, we made video storage system to test the algorithm using the stored images after storing images.

#### 3.1.2 Composition of the System

The video storage system is to store images from the camera into video, which consists of camera, image grabber, and video storage program. The camera is a device that receives images, the image grabber is to convert the camera's video image into the signal that can be processed by the computer, and video storage program is to make video through compressing of input signal of video from the computer. Fig 2 shows the structure of the implemented video storage system. One computer is connected to 4 image grabbers and each image grabber is connected to a camera.



**Fig 2:** The structure of video storage system

Fig 3 shows 4 cameras and 4 image grabbers that were used for the video storage system.



**Fig 3:** Cameras and image grabbers

In this study, analog cameras were used for the system because digital camera has high performance, but it is not easy to change the focal length of the lenses in digital camera. Also, the analog camera uses NTSC (National Television System Committee) signal.

Generally image grabber is classified into PCI type, USB type, PCMCIA type, etc., depending on the way to attached to a computer. We used USB types of image grabber which uses USB 2.0 standard.

The amount of video data which is transferred to a computer is very large, so it is not efficient to store video data without any process. Thus, these video data are typically saved using compression algorithm such as mpeg. In this study, mpeg4 format was used for compression, which has excellent compression ratio and image quality. The compressed video was stored in movie file format such as AVI which is the most widely used one.

### 3.2 Video Synthesis System

#### 3.2.1 Overview

The video synthesis system is to create one large image through synthesis of images from several cameras which are installed at a variety of locations.

Fig 4 shows the separate images (top) from 4 cameras and the synthesized image (bottom). We tested the system on a model road for indoor test, and after each camera shot different area, the images from cameras were synthesized. For individual images, it is not easy to understand the location of an image from the whole facility. Especially, in the case of tracking a specific object on the image, it is difficult to perform tracking because of the difficulty of finding the relationship between images. On the other hand, in the case of the synthesis image, it is easy for the user to view the whole image and to intuitively recognize it.

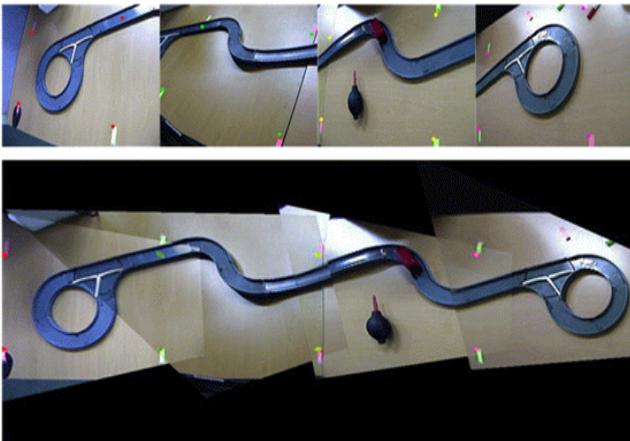


Fig 4: Individual images (top) and the synthesized image (bottom)

If we shoot the images after installing cameras by a certain distance, it is easy to synthesize the shot images. As shown in Fig 5, if we install cameras in the way where the observed areas by cameras can be overlapped, we can easily monitor the facility and analyze the consecutive screen video through image synthesis.

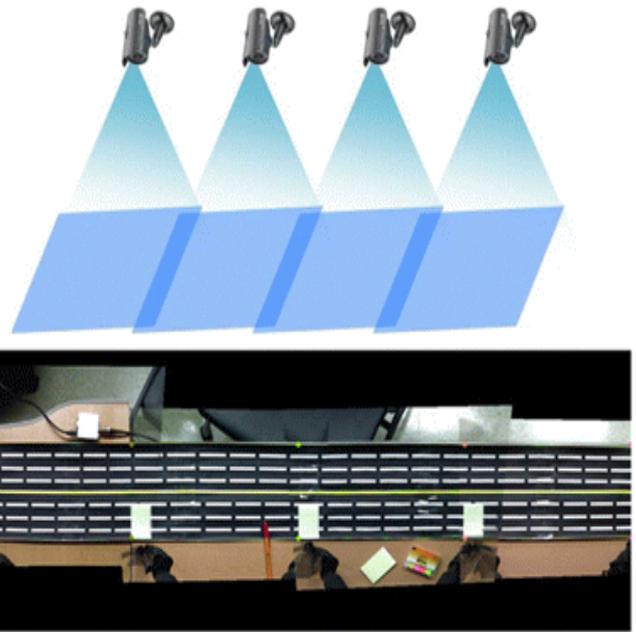


Fig 5: Image synthesis for the horizontal movements of cameras

#### 3.2.2 Image Conversion and Synthesis

If we assume that coordinate of 3-D space is on the same flat surface, we need the point on 3-D space and 4 numbers of the corresponding points to calculate P transformation matrix. It is assumed that we know the coordinates of at least 4 points on camera image, which can be markers that are defined in advance or well-known land marks. If we know the coordinates of 4 points, it is possible to calculate transformation matrix from these points and to convert all points on camera image into the points of plain coordinates on 3-D space using the matrix, as shown in Fig 6 and 7.

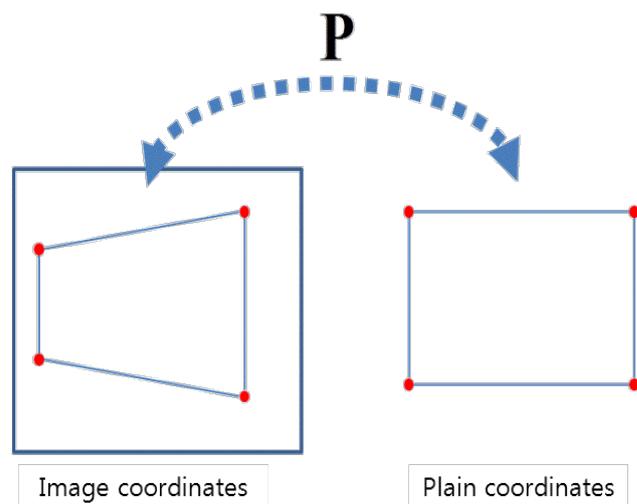
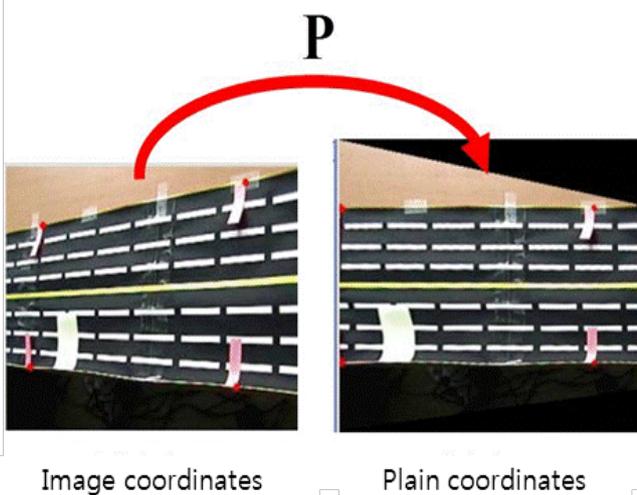


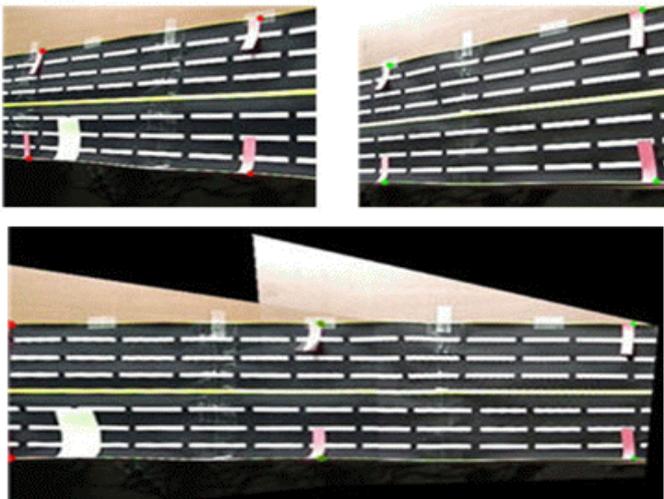
Fig 6: Matrix calculation from coordinates

<http://www.cisjournal.org>



**Fig 7:** Image transformation using P matrix

Therefore, in the case of using two cameras, we can get one image after synthesizing two areas through the calculation of P transformation matrix for each image, as shown in Fig 8.



**Fig 8:** Synthesis of transformed images

Even if the number of cameras is more than 2, we can get a synthesized image from several camera images through the calculation of P transformation matrix for each image and the definition of the coordinates on 3-D space.

### 3.2.3 Test

The proposed video synthesis system was tested using the model road and installed cameras as shown in Fig 9.



**Fig 9:** Test environment

In image synthesis, as the length of road which is shot by a camera increases, a synthesized image is much distorted. Because it is hard to quantitatively analyze this problem, we performed test to find the optimum ratio for successful image synthesis with change of the ratio of the camera height to the length of the composition area (1:1, 5:4, 5:8, and 3:8).

As test results, it is shown that as the camera is tilted horizontally, the length of road which can be shot by the camera increases, but the synthesized image is distorted and has poor quality. Even though the synthesized image has a little distortion at the ratio of the camera height to the length of the composition area, 3:8, it can be monitored qualitatively. Based on these results, we placed a model track and model cars similar to the actual road conditions and performed image synthesis, as shown in Fig 10.



(a) Test environment



(b) Individual images



(c) Synthesized image

**Fig 10:** Indoor test environment and test result

In the real road environment, because it is not easy to perform the test by installing cameras by a variety of angles and location, we performed the test using tripods and cameras. As shown in Fig 9 (b), each image seems to be distorted due to the installation location of the cameras. In the case of the practical application of the system, it is

<http://www.cisjournal.org>

hard to understand real road condition with monitoring images because the camera and monitors are installed regardless of the relationship. While, in the case of the synthesized image, it is advantageous to intuitively recognize the whole condition because one image is provided through transformation and synthesis of individual images.

#### 4. CONCLUSION

This study is aimed at developing view synthesis system for closed sections such as Tunnel, Bridge etc. to monitor and track the vehicle movement. For this purpose, video storage and synthesis systems were developed and demonstrated, making use of the video taken and stored in 4 web cameras. Also, we performed test to find the optimum ratio for successful image synthesis with change of the ratio of the camera height to the length of the composition area

Of course if high-quality blending method is used, there needs more calculation times. In this study, the way to use the average value of two images is applied to the overlapped areas to reduce the calculation times. To get the good quality of image, we need to study and apply high-quality blending method such as the way to use weights of length from each image.

For image synthesis, the selection of physical position of the camera for considering the image distortion can be one of the research areas.

#### ACKNOWLEDGEMENTS

This research was supported by a grant from a Strategic Research Project (Development of Real-time Traffic Tracking Technology Based on View Synthesis) funded by the Korea Institute of Construction Technology.

#### REFERENCES

- [1] D. G. Lowe, Object recognition from local scale invariant features, international conference on computer vision (1999), 1150-1157.
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, Speeded-up robust features (SURF), CUIV, Vol.110, No.3 (2008), 346-359.
- [3] T. Porter and T. Duff, Compositing Digital Images, Computer Graphics 18 (3) (1984), 253-259.
- [4] D. L. Milgram, Computer methods for creating photo mosaics, IEEE Transaction on Computer, Vol.C-24, No.11 (1975), 1113-1119.
- [5] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6) (1986), 679-698.

#### AUTHOR PROFILES

Taehyeong Kim received the degree in transportation engineering at the University of Maryland in the U.S. Currently; he is a senior researcher at Korea Institute of Construction Technology. His research interest covers optimization, par transit, logistics, and simulation.

Bum-Jin Park received the degree in transportation engineering at the Yonsei University in Korea. Currently, he is a senior researcher at Korea Institute of Construction Technology. His research interest covers intelligent transportation systems, traffic flow, and information technology.