

# Automatic Generation of 3D Models from UAV-captured Image Data for Immersive VR Applications

Nei-Hao Perng<sup>1</sup>, Bo-Sheng Bai<sup>1</sup>, Po-Han Chen<sup>1\*</sup>, Jen-Yu Han<sup>1</sup>,

Ming-Yi Jiang<sup>2</sup>, Jyun-Hao Huang<sup>2</sup>, Cheng-Wei Su<sup>2</sup>, and Po-Yuan Chen<sup>2</sup>

<sup>1</sup>National Taiwan University

<sup>2</sup>Ministry of Transportation and Communications, Taiwan

\*Corresponding author

E-mail: [pohanchen@ntu.edu.tw](mailto:pohanchen@ntu.edu.tw), [jyhan@ntu.edu.tw](mailto:jyhan@ntu.edu.tw)

## Abstract –

In this paper, the way to automatically convert UAV-captured image data to 3D models as well as the subsequent immersive virtual reality (VR) applications will be presented. It is expected that the proposed development would help identify the defects on the inspected bridge, and assist with the training of junior bridge inspectors.

## Keywords –

UAV; VR; 3D-Model; Automatic

## 1 Introduction

There are some twenty-nine thousand bridges in Taiwan, managed respectively by Taiwan Railways Administration, Taiwan Area National Freeway Bureau, Directorate General of Highways, and local governments. According to the current bridge maintenance guidelines in Taiwan, bridge inspection is mainly carried out through visual inspection, supplemented with equipment inspection. Bridge inspectors would usually approach the bridge structure for visual inspection. However, for high-rise or cross-river bridges, it would be hard to carry out visual inspection due to the difficulty in accessing the bridge structure.

In view of this, the Institute of Transportation under the Ministry of Transportation and Communications (MOTC) in Taiwan conducted a research project on “Enhancing the effectiveness of bridge inspection equipment” in 2014, and successfully carried out bridge inspection using a camera-mounted unmanned aerial vehicle (UAV). In this paper, the way to automatically convert UAV-captured image data to 3D models as well as the subsequent immersive virtual reality (VR) applications will be presented.

## 2 Research Limitations

In our research, we build up a 3D model bases on the photos took by UAV with Photogrammetry. Eq 1-1 is the formula about Collinearity, the fundamental principle of Photogrammetry (Fig.1 is the Schematic diagram).

$$\begin{bmatrix} x_p^L - x_0^L \\ y_p^L - y_0^L \\ -f^L \end{bmatrix} = R^L \begin{bmatrix} X_0^L - X_p \\ Y_0^L - Y_p \\ Z_0^L - Z_p \end{bmatrix}, \begin{bmatrix} x_p^R - x_0^R \\ y_p^R - y_0^R \\ -f^R \end{bmatrix} = R^R \begin{bmatrix} X_0^R - X_p \\ Y_0^R - Y_p \\ Z_0^R - Z_p \end{bmatrix} \quad (1-1)$$

$(x_0^L, y_0^L, f^L)$  &  $(x_0^R, y_0^R, f^R)$  = parameters of camera (could get from manufacturer)

$(X_0^L, Y_0^L, Z_0^L)$  &  $(X_0^R, Y_0^R, Z_0^R)$  = camera position (while shooting)

$R^L$  &  $R^R$  = shooting attitude

$(x_p^L, y_p^L)$  &  $(x_p^R, y_p^R)$  = values of Point P in the “photos”

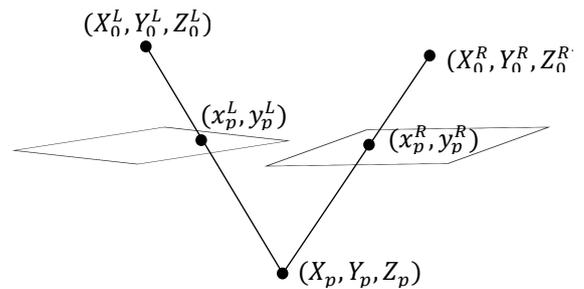


Fig 1. Schematic Diagram about Photogrammetry

Limited by matching technology nowadays, the photos we used for research must conform to following conditions to make sure the model quality is good enough:

1. The object should be shown in multiple different photos, and overlap over certain percentage.
2. The conjugate features between images should be recognized by computer.
3. The area with same color tone will lead to bad outputs, because of the lack of Eigenvalues.

4. Moving object is not included in research scope.

Besides, 3D-model is only able to show a rough outline without high definition texture, so we need to involve panorama as an assistance mode to provide user immersive experience. (Since the panoramas are shot at discrete points, we cannot move smoothly in panorama mode)

### 3 Literatures Review

#### 3.1 UAV (Unmanned Aerial Vehicle)

There are lots of software in the market could automatically analyze 2D pictures, then build up a 3D-model. In our case, we fly UAV to take photos about the object, and input to those software to get the model we need.

##### 3.1.1 Agisoft PhotoScan

Agisoft PhotoScan (commonly known as PhotoScan) is a professional tool for a photogrammetry pipeline. It is a stand-alone software product that performs photogrammetric processing of digital images and generates 3D spatial data to be used in GIS applications, cultural heritage documentation, and visual effects production as well as for indirect measurements of objects of various scales. The software is developed by Agisoft LLC located in St. Petersburg in Russia. It is widely used by archaeologists, and many UAV companies are also using it.

Agisoft PhotoScan is an off-the-shelf photogrammetric kit that allows for photography (both metric and non-metric, aerial and close range) to be...

- ✓ calibrated using a physical camera model;
- ✓ georeferenced utilizing ground control points or camera locations;
- ✓ block adjusted through aerial triangulation including the use of tie points;
- ✓ employed to create TIN/DEM products of the overlap areas using autocorrelation algorithms;
- ✓ orthorectified using a TIN and the calculated exterior orientations; and
- ✓ used to carry out area and volume measurements.

Agisoft PhotoScan is capable of processing up to tens of thousands of photos yielding result products characterized by high degree of accuracy in both the horizontal and vertical dimensions.

##### 3.1.2 Pix4Dmapper

Pix4Dmapper software offers up to centimeter-grade 3D precision, features a fully automatic workflow and

can be used by anyone as it is very intuitive and easy to use. You can assess, edit and improve projects directly in the software using the rayCloud and Mosaic Editor and seamlessly import results into any professional GIS, CAD and traditional photogrammetry software package.

Pix4Dmapper's goal is to provide customers with cuttingedge technology and groundbreaking tools combining the latest innovations of Computer Vision and Photogrammetry. Orthomosaics, DSMs and Point clouds generated with Pix4Dmapper achieve surveygrade accuracy from any aerial or oblique imagery, ranging from lightweight compact cameras used on UAVs (including GoPro, multispectral cameras and Fisheye lens) to SLR's and large photogrammetric sensors carried by aircraft. New algorithms allow for even more precise and stunning results and help customers to get the most out of their UAV imagery.

#### 3.2 VR (Virtual Reality)

Virtual reality (VR) typically refers to computer technologies that use software to generate realistic images, sounds and other sensations that replicate a real environment (or create an imaginary setting), and simulate a user's physical presence in this environment, by enabling the user to interact with this space and any objects depicted therein using specialized display screens or projectors and other devices.

VR has been defined as "a realistic and immersive simulation of a three-dimensional environment, created using interactive software and hardware, and experienced or controlled by movement of the body" or as an "immersive, interactive experience generated by a computer".

A person using virtual reality equipment is typically able to "look around" the artificial world, move about in it and interact with features or items that are depicted on a screen or in goggles. Virtual realities artificially create sensory experiences, which can include sight, touch, hearing, and, less commonly, smell. Most 2016-era virtual realities are displayed either on a computer monitor, a projector screen, or with a virtual reality headset (also called head-mounted display or HMD). HMDs typically take the form of head-mounted goggles with a screen in front of the eyes. Some simulations include additional sensory information and provide sounds through speakers or headphones. Virtual Reality actually brings the user into the digital world by cutting off outside stimuli. In this way user is solely focusing on the digital content.

Some advanced haptic systems in the 2010s now include tactile information, generally known as force

feedback in medical, video gaming and military training applications. Some VR systems used in video games can transmit vibrations and other sensations to the user via the game controller. Virtual reality also refers to remote communication environments which provide a virtual presence of users with through telepresence and telexistence or the use of a virtual artifact (VA), either through the use of standard input devices such as a keyboard and mouse, or through multimodal devices such as a wired glove or omnidirectional treadmills.

The immersive environment can be similar to the real world in order to create a lifelike experience—for example, in simulations for pilot or combat training, which depict realistic images and sounds of the world, where the normal laws of physics apply (e.g., in flight simulators), or it can differ significantly from reality, such as in VR video games that take place in fantasy settings, where gamers can use fictional magic and telekinesis powers.

#### 4 System Structure

Because our research objects are bridges, there are more adverse factors must be faced when we are trying to get a 3D model from 2D images by software (such as Pix4Dmapper or Agisoft PhotoScan) automatically. For example, the deck is in the same tone, the plants around are in the same tone, the water in the river is moving, the hollow structure is complicated, so and on. Therefore, the 3D model is only able to give a rough outline. It's not detailed enough for safety checking or other operations.



Fig 2. Auto-build-up Model from UAV Pictures



Fig 3. Panorama about Under-bridge

The panorama technology could merge multiple pictures into an “image ball”, which allows user to locate on the center and look around, such as Google StreetView or Youtube360. Thanks to the minimization of camera, there are lots of products in the market can take panoramas with a single click (Ricoh Theta S, Kodak sp360, etc.).

Since the both two mentioned above fit part of the definition about virtual reality, we include them all in our system. When we want to see a topography (Fig 2), we use 3D model; when we need to check some details, we use panorama (Fig 3). The system structure is as follows:

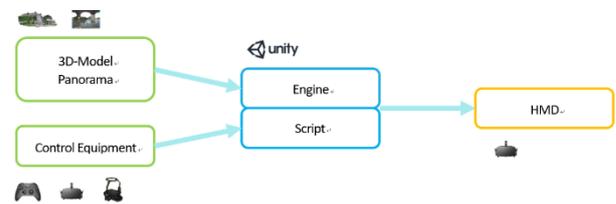


Fig 4. System Structure

#### 5 Data Analysis

The research result is shown as Fig 5. The upper right is VR goggle, and the lower right are the respective images with two eyes. The VR goggle includes a triaxial accelerometer, so it could catch the attitude of user and make computer output the corresponding scene.

Fig 6 is showing how user controls the system. As you can see in the lower right, the user moves virtual camera through the model with joystick, and looks around by turning his head.



Fig 5. The View in VR Goggle

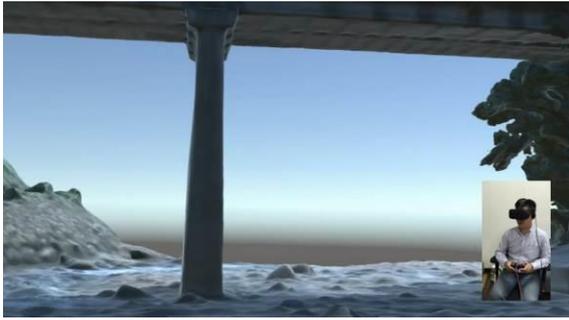


Fig 6. VR Model & Controller

Users could move to the checking point in model-mode, then switch to panorama-mode to verify testing items. Fig 7 is the difference between these two modes.

The left side, the model, allows more latitude for moving around; and the right side, the panorama, has a better texture to help user understanding reality problems (such as fractures or defects).



Fig 7. Model-Mode vs Panorama-Mode

## 6 Conclusions & Future Works

The system we developed could make user feel like in the objective environment. So that, they can learn the bridge structure and other neighboring information through model-mode, and check details with panorama-mode. There are some future works as follows:

- ✓ Improve the quality of auto-build-up model. Maybe trying to change the data type or add some artificial adjustments.
- ✓ Connect a walking pad such as Virtuix Omni to the control system, make user feel more immersive.
- ✓ Eliminate the limitations in the panorama-mode, such as discrete moving and space confusion.

## References

- [1] Tsai, Yi-Ting. A paper. The Study on DSM Generated from Point Clouds by UAV Image, Taiwan, 2012
- [2] Lapiere. A, "3D City Geographic Information System – A Major Step towards Sustainable Infrastructure". A Bentley White Paper, August, 2015.
- [3] Maher. K, "ContextCapture: Bentley's new take on digital reality", Graphic Speak, 2015.
- [4] Substitution of 123D Catch. On-line: <http://www.makezine.com.tw/make2599131456/diy123d-catch>, Accessed: 20/05/2016
- [5] VisualSFM. On-line: <http://blog.csdn.net/u012989207/article/details/45032469>, Accessed: 20/05/2016
- [6] Introduction of 3D Image Data Showing. On-line: <http://content.teldap.tw/index/blog/?p=1666>, Accessed: 20/05/2016
- [7] Microsoft. Photo tourism: Exploring photo collections in 3D. On-line: <http://phototour.cs.washington.edu/>, Accessed: 20/05/2016
- [8] 3D, M. Digital Elevation Model (DEM). On-line: <http://www.mosaic3d.com/en/products-service/dem/>, Accessed: 20/05/2016
- [9] iWitnessPRO-Agilis V3 has released. On-line: <http://www.iwitnessphoto.com/support/AgilisV3>, Accessed: 22/12/2013
- [10] Zhu. L, Bentley. Promote VR Mode. On-line: [https://communities.bentley.com/communities/other\\_communities/chinafirst/b/blog/archive/2015/03/03/bentley-acute3d](https://communities.bentley.com/communities/other_communities/chinafirst/b/blog/archive/2015/03/03/bentley-acute3d), Accessed: 20/05/2016
- [11] Acute 3D. On-line: <https://www.acute3d.com/>, Accessed: 20/05/2016
- [12] Turn photos into 3D models automatically with ContextCapture. On-line: <https://www.acute3d.com/contextcapture/>, Accessed: 20/05/2016
- [13] LeaderTek. On-line: <http://www.leadertek.com.tw/>, Accessed: 20/05/2016
- [14] GeoForce. On-line: <http://www.geoforce.com.tw/#1=0>, Accessed: 20/05/2016
- [15] Pix4D. On-line: <https://www.pix4d.com/>, Accessed: 20/05/2016