

Integration of 3D Geospatial Information for Radar Viewshed Analysis of Urban Warfare Simulation

Ming-Da Tsai^{1*} and Peng-Wen Lai²

¹*Environmental Information and Engineering Dept., Chung Cheng Institute of Technology, National Defense University*

²*Institute of Electrical and Control Engineering, National Chiao Tung University*

ABSTRACT

The booming economy and the advance of society lead to densely populated settlements of towns and cities in Taiwan, and some of these towns and cities have been developed into very important centers of local politics, economy, culture and transportation. The occupation or defense of these urban settings has profound impact to the overall warfare in times of war. Modern buildings are structurally fortified and strengthened. High-rise buildings are next to one another. These buildings can be integrated into a strong defense system if troops and fire power are arranged properly. Therefore, the “urban warfare” has become one of very important patterns of warfare in the defense of Taiwan and Penghu. Geospatial information of terrains and radar viewshed analysis were combined for this study, with the simulation of the 3D building barrier and ground radar field of vision, to simulate the battlefield environment of urban operations.

Keywords: Urban Warfare, Geospatial Information, Radar Viewshed Analysis, Simulation

整合 3 維空間資訊於城鎮作戰模擬之雷達視域分析

蔡明達^{1*} 賴鵬文²

¹國防大學理工學院環境資訊及工程學系

²國立交通大學電機學院

摘 要

隨著社會經濟蓬勃發展，臺灣本島城鎮密集且多為政治、經濟、文化及交通的中心，其攻佔或固守，對整體作戰有長遠之影響；現代化建物強固堅韌、高樓鱗比節次，如善加運用適切兵力、火力可成為堅強之防禦體系，故「城鎮戰」已成為未來臺澎防衛作戰中重要作戰型態。本研究結合地形地物之空間資訊與雷達視域分析，藉由仿真的 3D 建物屏障以及地面雷達的視域，來模擬城鎮作戰之戰場環境。

關鍵字: 城鎮戰、空間資訊、雷達視域分析、模擬

I. INTRODUCTION

The western part of Taiwan is mostly flat and densely populated (as shown in Fig.1). Decades of development has resulted in significant urbanization as we see today. This indicates that the “urban warfare” has a very important part in the study of land-based warfare. The US military has learned from the numerous combats in urban settings ever since the 1940s that the urban warfare features a number of characteristics unlike other types of warfare, including complicated preparation works for battlefield intelligence, difficulties in battle planning, complexity of command and control, limited room for maneuver of troops in a mobile manner, serious concern of fire power demonstration, and challenging supplies and maintenance. In an urban setting, roads and streets crisscross among high and low buildings. The topographic features that are complicated and easy to shelter and hide in are often a headache to commanders at all levels when it comes to having access to enemy intelligence [1].

Military exercise is a method and means of training that enables commanders and their staff to be familiar with a battlefield environment. However, limitations, such as the overall external environment, public opinions, training grounds, safety of exercise, possible loss of weapons and equipment, and huge training budgets, have prevented live exercises where large scale of forces are involved and which are necessary to maintain the quality and quantity of military training[2]. In response to the dilemma of “three highs and one difficulty” (that is, “high value,” “high risk,” “high energy consumption” and “difficulty in sourcing for a proper training ground”) in training, the Armed Force of Taiwan is currently seeking out ways to replace “live exercise” with “computerized wargame” in order to work around the limitation of training facilities. The use of computer simulation technology in exercise and training of troops of all levels, analysis and planning of deployment of force in battlefield, as well as the development of national strategies has become a trend of military technology development around the world. One that is most widely used in the world is “computerized wargame system” [3] in combination with the use of results derived from battle simulation

analysis.



Fig.1 Most of the population in Taiwan is concentrated in the western part of the island where the terrains are flat (the red areas shown in the picture are buildings)

The analytic models that are current deployed in the Armed Force of Taiwan are the “Integrated Theater Engagement Model (ITEM)” and the “Extended Air Defense Simulation (EADSIM).” However, the influence of buildings on combat effectiveness is not taken into consideration for either of these two models. It is impossible to have a thorough characterization and analysis on the information of urban space where it is complicated in nature and easy to shelter and hide in.

1.1 Integrated Theater Engagement Model (ITEM)

ITEM is a high level model with low solution that allows simulation of engagement in land-based environment, naval settings (on the surface, under water and amphibious), air space and missiles. It is an analysis model at the integrated battle level that is focused on joint combat. It provides comparison of various types of (joint) combat action solutions, combat schedule and combat effectiveness analysis. As a result, ITEM is used mostly for high level federations with low solution. The primary types of spatial data are land and sea (or water). The analysis on terrains and topography is rather limited. The simulation is based on the mathematic models of ultimate federated engagements, and digital topographic information has no influence, as shown in Fig.2.

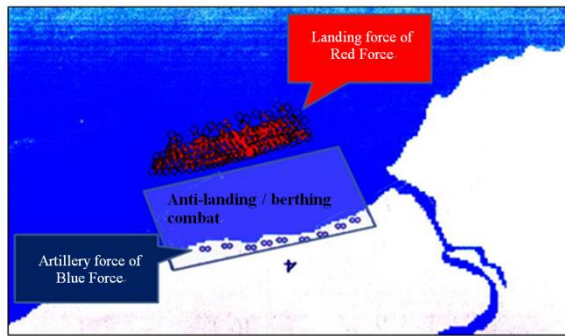


Fig.2 A scheme of computerized wargame using ITEM

1.2 Extended Air Defense SIMulation (EADSIM)

EADSIM is a mission level model with high solution. Commonly used for combat study and analysis in relation to air defense, space warfare and missile defense, EADSIM is capable of simulating “single platform engagement,” “multiple force engagement” and “theater level” warfare. It is equipped to provide overall combat planning, training and analysis for combatants. It consists of several sub-models, including air warfare, space warfare, electronic warfare, command / control / communication / intelligence and missile defense [4].

When it comes to the processing and analysis of battlefield environment, EADSIM is able to read ADRG (ARC Digital Raster Graphics) in NIMA (National Imagery and Mapping Agency) format, CADRG (Compressed ARC Digital Raster Graphics), CIB (Controlled Image Base), image files and topographic data such as Digital Terrain Elevation Data (DTED). These data allow for the sideview profile analysis in addition to search and calculation of location and distance in general. As shown in the following figures, Fig.3 (a) is an example of how the display of Extended Air Defense Simulation looks like. The model provides tools for sideview profile analysis, and the result is shown in Fig.3 (b). In addition, this model allows the analysis of effective detection range of sensors. EADSIM considers the shielding influence of ups and downs in topographic elevation on sensors, as shown in Fig.4. In Fig. 4(b), the white lines are detection lines at the elevation of 3000m, whereas the cyan lines are those at the elevation

of 500m. The cyan lines are more susceptible to terrains since they are closer to ground. As for white lines, the blind zone of radar is larger. For weapon platforms, the settings of movement must take into consideration the influence of terrains. However for topographic features (such as buildings, electric towers, chimneys and trees), the topographic feature simulation is not included due to the limits in the calculation of digital elevations in high resolution. It is unable to satisfy the need for the environment analysis of a battlefield space as densely urbanized as the island of Taiwan.

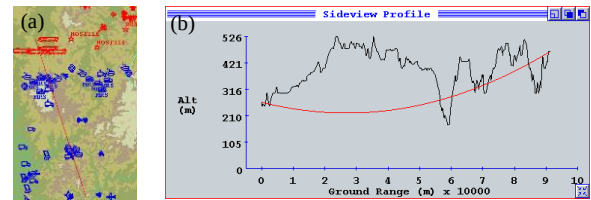


Fig.3 (a) It is allowed to use the sideview profiling tool in the EADSIM window and perform analysis along the red line; (b) display of sideview profiling data

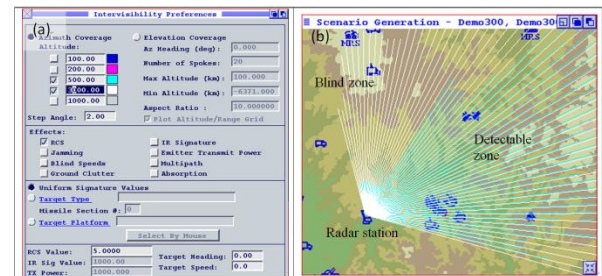


Fig.4 (a) This is the setting window for the reference of sensors' effective detection range; (b) this shows the effective detection range of a sensor, where the white lines and cyan lines represent the detection lines at the elevations of 3000m and 500m, respectively

At present, most of countries around the world are facing a number of difficulties, including military staff downsizing, shrinking size of battle exercises and training grounds, and training expenses too expensive to afford. However, the maturity of various types of modeling and simulation systems facilitates the use of computer simulation in the aid of battle training of military forces, analysis and planning of battlefield force deployment and the strategic goals of country at the moment. The spatial

information is put together and life-like battlefield environment is established to improve the operations of modeling and simulation. It is now a trend in most of the countries in terms of military technologies.

1.3 3-dimensional modeling

The 3-dimensional modeling is a process in which a 3D digital city is divided into 5 levels of detail (LODs) [5] using the City Geography Markup Language (CityGML) that was developed by the Open Geospatial Consortium (OGC), as shown in Fig.5. The LODs are listed as follows and shown in Fig. 5 below: (1) LOD0: this level covers only the topography and overlapping of different layers of theme and provides 2.5D display; (2) LOD1: at this level, buildings that look like toy building blocks (without roof types) are added on topography. This level is applicable to regional or city scale; (3) LOD2: roof structure types and wall finishing textures are added at this level, which is at the scale of architectural models; (4) LOD3: this level consists of models of attachments on roof and walls at the scale of architectural models; (5) LOD4: this includes the models of building interior, which provides an architectural model with walkable details.

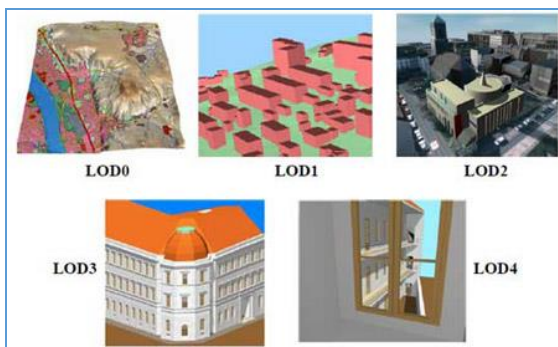


Fig.5 The 5 levels of detail (LODs) in CityGML [5]

1.4 Introduction to STK program

Developed by AGI (Analytical Graphics, Inc.), the STK (System Tool Kit) program (as shown in Fig.6) is a state of the art program for simulation and analysis for space and national defense missions. It analyzes the relative position and orientation of a vehicle on land, at sea, in the sky and in the vast space. It is capable of simulating and calculating communications,

remote sensing and radar payloads, and integrating the entire scenario in 4D visual animation [6].

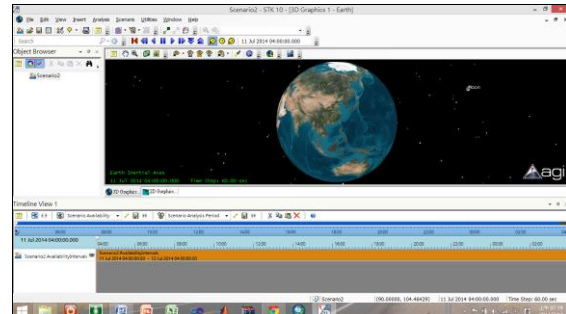


Fig.6 Interface of the simulation and analysis program, STK

Some examples of the common applications of STK are: automatic warning of satellite detection and photographing, automatic interception of ballistic missiles, 4D flight simulator in a war game, and anti-radar / missile jamming. It is equipped to interface with geographic information system (GIS) and combine with the spatial data of the atmospheric environment for various capabilities of simulation and analysis. For spatial data, it is compatible with ESRI (Arc GIS), Google Earth and Microsoft Virtual Earth, and CADRG, JPEG2000, CIB, GeoTIFF and NITF for image. For elevation data, it reads DTED and DEM. For atmospheric environment data, STK provides the atmosphere model. For the analysis of rain attenuation estimation, it provides the rain attenuation model developed by Crane, R. K. (1978) and the rain attenuation model of International Telecommunication Union (ITU-R P838 1992). STK is capable of analyzing and studying 3D terrain data and often used in satellite projects, space science, research and development of military equipment, space intelligence, exercise and training of military forces, wargame, and combat simulation and analysis.

II. Data Collection and Processing

The Yuan Shu Lin area of Daxi, Taiyuan was selected as the experiment area, as shown in Fig.7 below. Topographic surveying and plotting and 3D modeling of buildings on the ground were performed on this area using various surveying techniques in order to establish the

spatial information of this area for the simulation and testing in STK.



Fig.7 Aerial photo of Yuan Shu Lin, Daxi, Taoyuan, which is the selected experiment area

2.1 3D modeling of the experiment area

The acquisition of Geospatial data in the experiment area and the process of 3D building modeling are provided in Fig.8. The first step started by measuring the terrains in a large area using aerial (or remote) sensing techniques with the aid of the precision positioning data obtained from the global positioning system (GPS). Thus, the topography of the experiment area and its coordinates were determined. Next, the 3D topographic model was built to gain the access the elevation data of terrains and geographic features. Finally, 3D modeling was performed on the buildings.

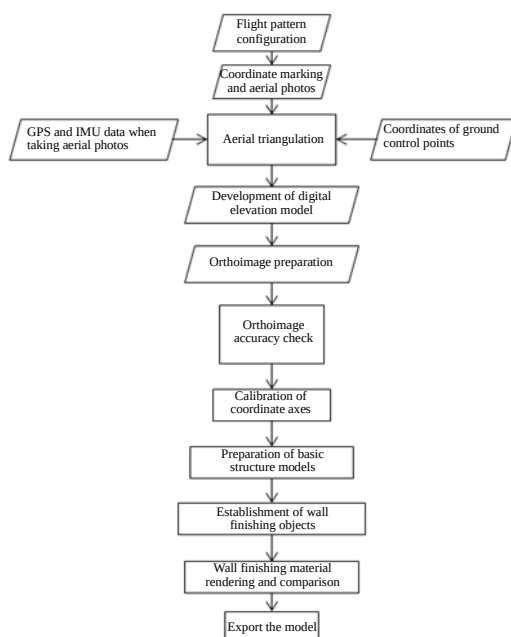


Fig.8 Acquisition of Geospatial data in the experiment area and the process of 3D building modeling

2.2 Surveying of the experiment area

1. Flight pattern configuration for unmanned aerial vehicle

The Yuan Shu Lin area at Daxi, Taoyuan was selected as the experiment area. The flight pattern configuration program, eMotion2, was used to develop a flight pattern for the aerial photography. Two flight patterns were developed for the unmanned aerial vehicle used. The first was with Chung Cheng Institute of Technology in it (as shown in Fig.9 (a)). The aerial photo flight was taken in July 2013 at an elevation of 200m above ground. The other flight pattern included Daxi. The flight was taken in October 2013 at an elevation of 130m above ground (as shown in Fig.9 (b)). The forward overlap of the aerial photos taken in both the flights was 70% and the lateral overlap was 60%. The size of single frame was 4,608 pixels x 3,456 pixels for all photos taken.

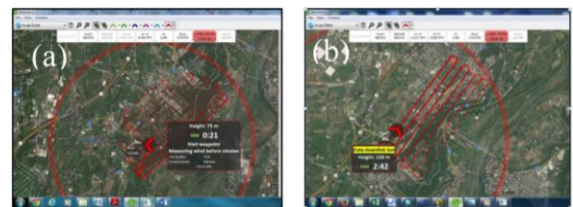


Fig.9 Flight patterns for (a) Chung Cheng Institute of Technology and (b) Daxi

(1) Surveying of ground control point coordinates and elevation data

The data processing came after the aerial photo flights of UAV were completed. The ground control points were added and aerotriangulation procedure was performed. The ground control point coordinates were measured using e-GPS. 3 points and 6 points were selected in the surveying areas as the control points. The surveyings were taken every second and more than 60 surveyings were received for both flights. The “electronic global satellite real-time kinematic positioning system (the e-GPS real-time kinematic positioning system or just e-GPS) was developed by National Land Surveying and Mapping Center, Ministry of the Interior in 2004 using the technique called “Virtual Base

Station Real-Time Kinematic” (VBS-RTK) [4]. The working principle of Unmanned Aerial Vehicle (network-based GPS real-time positioning system) is that the RTK station is located on a ground control point to receive signals from satellite positioning base stations established throughout Taiwan and evaluate the positioning errors in the areas covered by these base stations. With the observation data from the nearest physical base station, the RTK is considered a virtual base station with which to identify its coordinates and elevation, and finally, the orthometric height is obtained through the direct leveling in elevation (National Land Surveying and Mapping Center, Ministry of the Interior, 2007)

As the aerotriangulation procedure was finished, the outer orientation parameter of every UAV photo and orthoimage mosaic of the surveying areas were decoded and the orthoimages with geographic positioning coordinates went through the planar precision test. The test points were determined using e-GPS, and the root mean square error (RMSE) was calculated. The orthoimage tested was imported into the UAV 3D image platform developed, which allows the comparison between images in two windows or image overlapping analysis in a single window. Also available is the simple surveying and analysis functions.

(2) Setting up ground control points

The purpose to set up ground control points was to allow the interconnection between aerial image data and the coordinate data in real space. In the old days, navigation marks had to be set up in the surveying area when performing aerial photo surveying as the ground control points to improve accuracy of aerial photo surveying. However, thanks to the global positioning system (GPS) on board of UAV, the coordinates of aerial images are easily obtained. In order to determine the influence of number of ground control points on the positioning accuracy, three different numbers of ground control points were selected for both the surveying areas (as shown in Tab.1). One of the reasons was to prevent errors resulted from different ground control points selected. For 6 ground control points, there had to be the positions of 3 ground control points in that area.

Tab.1 Number of control points

	CCIT	BM3
No. of ground control points (GCP)	No ground control point	
	3 control points	
	6 control points	

(3) Preparation of UAV orthoimages

The image data from the aerial photos taken by UAV were imported into Postflight Terra 3D-EB along with the coordinates and elevations of ground control points. The projection coordinate of 121 degree zone in Taiwan Datum 1997 was selected for calculation of 3 different numbers of ground control points described in Table 1. Hence the point cloud of images and individual orthoimages were obtained. Then, the orthoimage of the entire area is achieved with matching and color rendering (as shown in Fig. 10 (a) and (b)).

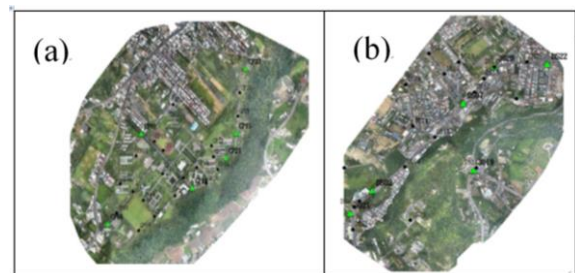


Fig.10 Plot maps for UAV control points and check points at (a) Chung Cheng Institute of Technology and (b) Daxi (: control point; : check point)

2. UAV orthoimages

The images from UAV aerial photos were added with ground control points for AT adjustment and the orthoimages with geographic positioning were overlapped on Google Map (as shown in Fig.11 (a) (b)). It is easy to tell the difference in accuracy between the UAV orthoimage and Google Map with naked eyes, and the color tone of UAV orthoimage is closer to the real terrains.



Fig.11 Othroimages of (a) Chung Cheng Institute of Technology and (b) Daxi

Institute of Technology and (b) Daxi overlapped on Google Map

3. Accuracy check

The root mean square error (RMSE) is commonly used for position accuracy in accuracy check. The root mean square error is a type of standard deviation in a wider sense. It works by calculating the square root of the mean square error (MSE) of the different at point i in either the x or y direction (ΔX_i , ΔY_i) on the coordinate where the check point on the image checked and the coordinate of the check point on another image data with greater accuracy. It works for checking multiple different physical surveyings, such as the root mean square errors of a check point group (RMSE_x, RMSE_y) and planar root mean square error (RMSE_r), in order to have a clear picture of the less accurate indicator between the two sets of coordinates:

$$RMSE_x = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{X_{data,i} - X_{check,i}}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \Delta X_i^2}$$

$$RMSE_y = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{Y_{data,i} - Y_{check,i}}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \Delta Y_i^2}$$

$$RMSE_r = \sqrt{RMSE_x^2 + RMSE_y^2}$$

where,

ΔX_i is the difference of check point i in the X direction coordinate;

ΔY_i is the difference of check point i in the Y direction coordinate;

n is the total number of check points;

$X_{data,i}$ and $Y_{data,i}$ are the coordinates of check point i in the map image to be checked; and

$X_{check,i}$ and $Y_{check,i}$ are the reference coordinates of check point i .

7 check points were selected for this study. Surveyings were made with e-GPS at one surveying every second. More than 60 surveyings were made for each of the points. Another comparison was made between the orthoimages prepared with and without ground control points in the calculation of AT adjustment. The root mean square errors (RMSEs) of the coordinates of check points in the image and the ones from the actual surveyings are shown in Tab.2 to 4. The relationship between the number of GCPs and RMSE is provided in Fig. 12.

When there was no ground control point,

the positioning was accurate to 4m for CCIT and 2m for Daxi. When three ground control points were added, the positioning was accurate to 0.7m for CCIT and 0.2m for Daxi. When six ground control points were added, the accuracy improved to 0.3m for CCIT and 0.1m for Daxi.

Tab.2 RMSEs of check points (without ground control point), in m

Area	X dir.	Y dir.	Planar
CCIT	±2.84278	±2.78318	±3.97838
Daxi	±0.49741	±1.81958	±1.88634

Tab.3 RMSEs of check points (with three ground control points), in m

Area	X dir.	Y dir.	Planar
CCIT	±0.63055	±0.24234	±0.67552
Daxi	±0.12474	±0.10295	±0.16174

Tab.4 RMSEs of check points (with six ground control points), in m

Area	X dir.	Y dir.	Planar
CCIT	±0.26352	±0.11600	±0.28792
Daxi	±0.08387	±0.04103	±0.09337

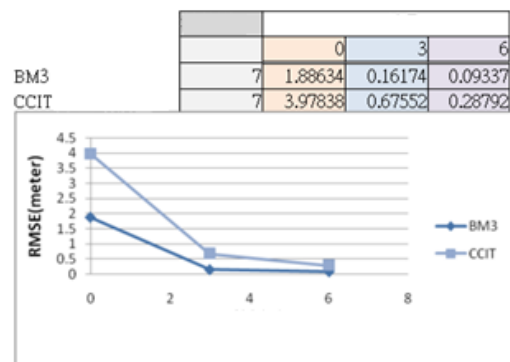


Fig.12 Number of GCPs vs. RMSE for CCIT and Daxi

2.3 Development of 3D building models

1. Calibration of coordinate axes

When it comes to the 3D building model construction, the modeling criteria of LOD3 was adopted for the establishment of building models in the experiment area. The first step to build a

model was to recalibrate the coordinate axes. The Google Map functions included in Sketch Up were introduced to convert the satellite images of buildings into background image. The corners of building were used to define new coordinate axes. There was a very visible difference in establishing the contour of a building before calibration (as shown in Fig. 13(a)) and after calibration (as shown in Fig. 13 (b)) for particularly buildings with a simpler shape or structure (similar to, for example, a rectangle). It was much easier to refine the details after recalibrating the coordinate axes. However, if dealing with a building that is complex (a building with an irregular shape such as the Armament Building), the orthoimages of UAV helped characterize a more accurate frame.

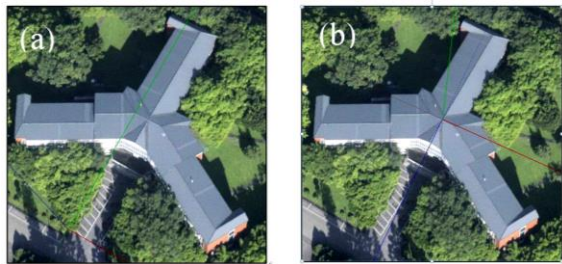


Fig.13 A building (a) before calibration and (b) after calibration

2. Preparation of basic structure models

The orthoimages produced by UAV were used to characterize the contour of building and define the length and width of building. Then, LiDAR, or Light Detection and Ranging, was introduced for its advantages of rapid observation, high accuracy and the ability to observe terrains and objects on the ground at the same time. The use of data from LiDAR observation allowed the development of multiple digital topographic models. In general, these models were Digital Elevation Modeling (DEM), Digital Terrain Modeling (DTM), Digital Surface Modeling (DSM) and Digital Building Model (DBM). By subtracting DEM from DSM, the elevation of building was obtained, and then the basic structure model of building was achieved (as shown in Fig.14) [7,8]

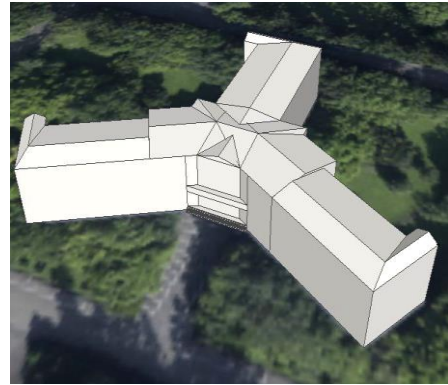


Fig.14 Basic structure model of building

3. Establishment of wall surface objects

The photos of various walls were taken on site. These photos were taken to cover all four corners of the walls (as shown in Fig. 15) in order to identify the positions of objects on the wall, such as eaves, windows, stairs, and so on. Once the positions were identified, the wall objects were plotted to finish the object refinement of building details.



Fig.15 Position of objects

4. Wall finishing material rendering

There are always unwanted objects when taking the photos of walls, such as people, cars, trees and anything that would obscure the look of building (as shown in Fig.16). If it is impossible to avoid objects that cover part of the wall, the Sketch Up has the “Set as the only texture” function for help. Since the walls to be taken photos of were covered mostly with ceramic tiles, which are the same texture over and over again, the only thing to do was to identify the repetitive unit texture (as shown in Fig.17) and the rest was to use this texture to cover the objects that obscured the wall and piece together the entire surface of wall. As shown in Fig.18, (a) Before the obscuring objects were removed and (b) after they were removed.



Fig.16 Example of objects that obscured a building

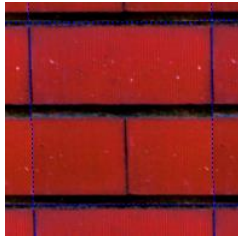


Fig.17 Unit texture of wall surface

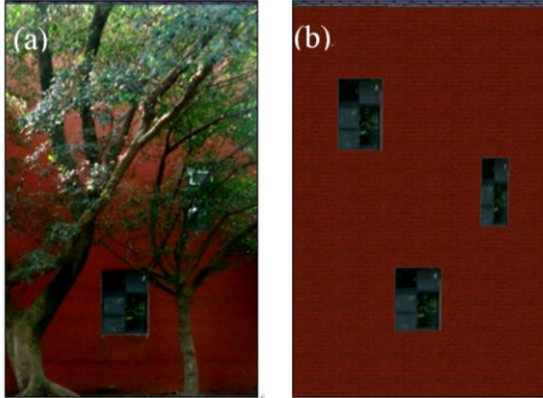


Fig.18 (a) Before the obscuring objects were removed and (b) after they were removed

5. Exporting the models

Once developed, the building models were exported in .dae format to allow STK and other programs to load these various refined models at LOD3. When importing these models into the interface of STK or other programs, it will be easier to import these data. The result of models is shown in Fig.19.



Fig.19 Result of models

III. Simulation Scenario Design and 4D Simulation Data Processing

3.1 Loading 3-dimensional geospatial data

The Yuan Shu Lin area of the township of Daxi was selected as the experiment area. The surveying was made possible by Geo-spatial information technology and the 3D map of the selected area was developed. The result was the topographic map of the area and 3D models of buildings in this area. The STK program used for this study is capable of reading spatial information in various formats, including: JPEG2000, BMP, PNG, TIF, Shape File, DTED (*.dtd, *.dtd0, *.dtd1, *.dtd2, *.dtd3, *.dtd4, *.dtd5, *.dtd6), pdtt, pdttx, kml, and kmz. For 3D building models, it is capable of reading formats such as mdl and dae.

For the elevation data of the terrains, high resolution digital elevation model (DEM) was obtained through LiDAR or aerial photos. The data were later converted into DTED format. The 3D models of building were developed using the refined modeling specifications at the LOD3 level. The prepared building objects were exported into files in mdl and dae formats, which were imported into STK. The procedure to import spatial data is described as follows: firstly the coordinate positions of the experiment area were established and the image files from aerial (satellite) photos of that area were imported. Then, the DTED elevation data were imported. Finally, the 3D building objects were imported. At this moment, the building objects were placed and positioned by hand based on the locations shown on the aerial (satellite) photos, as shown in Fig.20.



Fig.20 Importing the 3D building objects developed into system

3.2 Scenario design

The design of scenario, as shown in Fig.21, was tested to see whether topographic features (or buildings) have any impact on the detection viewshed when radar is detecting an object. An air defense radar station was established among the buildings in the experiment area. A general helicopter was added to hover around the radar station at 100 nautical miles per hour. This setup was to determine whether the radar viewshed was compromised due to the shielding effect of buildings. Fig.21 shows the detection range of the radar station. For the purpose of this study, the radius of the radar's detection range was defined at 3km. The range and distance of the yellow dome shown in the figure indicates the viewshed in which the radar is effective in detection. As shown in Fig.22, the rugged base of the yellow dome indicates the impacts created by the shielding of topographic features and buildings. It is clear in Fig.23 that a radar blind spot was created at the area behind the building shown, which is the result of compromised detection range of radar due to the shielding impact of building.



Fig.21 The radar station was designed among several buildings to test the impact of buildings on the radar's ability to detect

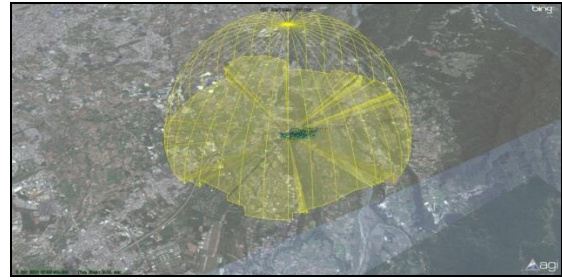


Fig.22 The detection range of air defense radar

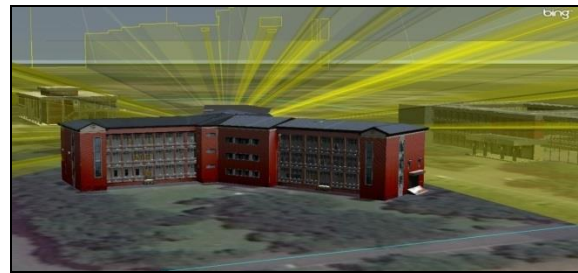


Fig.23 The radar detection blind spot (blind zone) was created behind the building in the figure due to the shielding effect of building

IV. Result Display and Analysis

In Fig.24, the window on the top is a 3D simulation scenario. For this study, the radar station was located between 4 buildings. A helicopter took off from the lawn next to the radar station and hovered around the station. The time series was shown in the window at the bottom. The red section of line meant that the helicopter was detected by the radar in this particular period of time (4D simulation was achieved by adding the element of time). When the helicopter was picked up by the radar station, a yellow line of detection connected the radar station and the hovering helicopter as shown in the window on the top. The yellow area was the detectable zones of the radar, whereas the rest was the radar's blind zones.

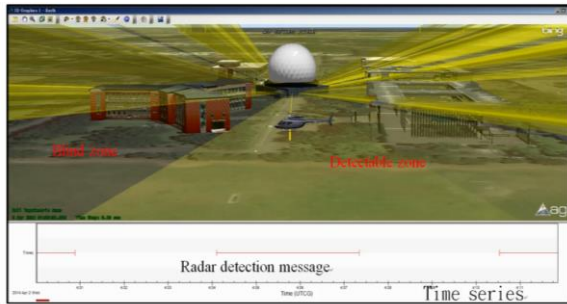


Fig.24 Description of simulation display

The helicopter took off at minute 4 and flew out of the detection range of the radar at minute 4.01 and then back in the detection range of the radar at minute 4.04. It hovered around the radar station to test the influence of the buildings on the ground on the detection ability of radar. When the radar detected the helicopter, a detection line of radar was generated in the 4D simulation scenario. If not, the detection line disappeared, as shown in Fig.25. The helicopter hovered for 12 minutes. Fig.26 showed the trajectory of the helicopter at minute 11, which indicates that the helicopter was within the detection range of the radar and therefore detected by the radar.

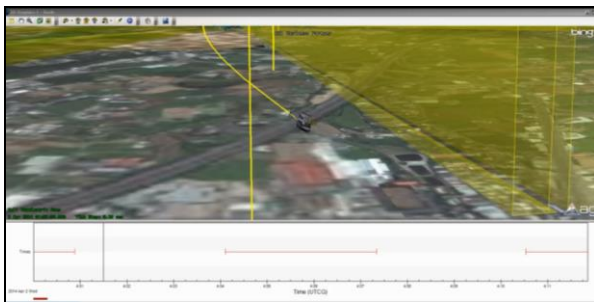


Fig.25 The detection line between the radar and the helicopter disappeared when the helicopter flew out of the detection range of radar

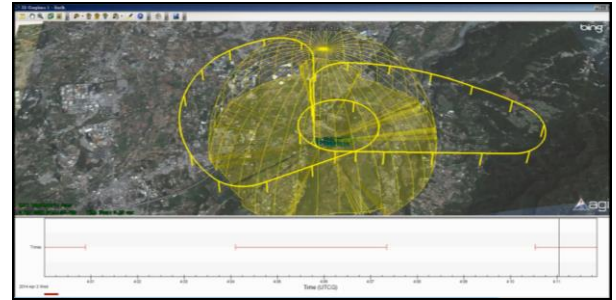


Fig.26 The flight trajectory of helicopter hovering around the radar station

V. Conclusions and Suggestions

5.1 Conclusions

Sun Tzu, the author of the Art of War, once said: “The natural formation of the country is the soldier's best ally.” Similarly, western military scientists said: “map is the eyes and ears of a military force.” This is just to show how important a clear map is for troops in a battle. Everything from the development of combat plan to the basics of command and control depends on it. Commanders and their staff acquire the information of terrain intelligence in maps for coordination and communications in federated combat maneuvers and space coordination of tri-force jointed warfare [9]. As a result, the decision on national defense and security, command and control in war time and everything military are built upon the correct knowledge and judgment about the space in which combats take place. The military exercise and training with the aid of computerized war gaming and modeling and simulation analysis is one of the best methods and means to train commanders and their staff to familiarize themselves with the environment of possible battlefield.

At present, the Armed Force of Taiwan is still working on integrating spatial information (topographic features in particular) models into tools used for simulations of urban combat in high resolution, including model analysis program currently employed in the Armed Force, such as Extended Air Defense Simulation (EADSIM), Integrated Theater Engagement

Model (ITEM) and the Synthetic Theater Operation Research Model (STORM) that will soon to be introduced. It is found in the results of this study that the detection zone of radar is susceptible to terrains and topographic features and therefore blind spots or blind zones are created for the radar. A helicopter is very likely to be detected when it flies into the detection range of radar. However, the shielding and covering of buildings help the helicopter stay out of the radar detection. As a result, a realistic 3-dimensional battlefield was created for this study. With the careful setup of the simulation scenario (i.e. the detection range of radar and simulated flight patterns of helicopter), it is much easier to develop the knowledge about the environment of a battlefield in real life.

Thanks to the advance in the development of geo-spatial information technology, the surveying of terrains and topographic features was improved more effectively by employing unmanned aerial vehicle (UAV) in combination with spatial surveying instruments such as remote sensing (RS), global positioning system (GPS) and inertia navigation system (or IMU), and it is now much quicker and more accurate to acquire spatial information on the surface of Earth. The surveying was performed in three different conditions for this study. For the assessment and analysis of accuracy, the positioning was accurate to 2 to 4m with no ground control point. If three ground control points were added, the positioning was accurate to 1m or less; and finally, if six ground control points were added, the coordinate positioning was more accurate to 0.5m or less. These levels of accuracy were enough to satisfy the need for information of space at the time of simulation. In terms of timeliness, this study showed that it took approximately a work day to finish the entire process from the onset of aerial photography to the acquisition of the orthoimages needed with ground control points. However, it took longer to establish the models of topographic features as the building models developed for this study were refined ones at LOD3 level. It took 5 work days in average to build the model of a building. This was exactly

the reason why no more than 8 building objects were loaded for the simulation test described in the study because of the inherent limits in the performance of the computers used (as shown in Fig.27).



Fig.27 8 building objects were imported in the area enclosed by the red lines

5.2 Future development and suggestions

The knowledge about and attention to the environment of a battlefield was limited only on terrains in the past, and topographic features were usually neglected. If the terrains and topographic features are displayed in 3D in the future in military intelligence gathering and establishment or in exercises, training events or war games, it will be able to improve the perception of commanders or staff about the “battlefield space.” The addition of time in exercise or training simulation makes the 4D simulations possible. For example, as shown in Fig.28, the coastal areas of Tamsui are riddled with ocean view buildings, which are possible hiding places for attack or reconnaissance helicopters to prevent enemy detection. In Fig.28, the green parts are detectable zones for the radars of red force, while the red zones are the blind zones of radar. In Fig.29, helicopters may hide in the radar blind zones created by the buildings from enemy radars. Apart from this, such simulation can be used for prevention of potential flight accidents. For example, April 3 2007, an Army UH-1H crashed into the transmission tower of a radio transmission station of the Police Radio in the Zhongliao Mountains at Qishan, Kaohsiung. The helicopter went down with all 8 soldiers and officers killed. The major reason for this disastrous accident was that the flight crew was unfamiliar with the

local terrains and topographic features. It is a future possibility to deploy UAVs to scan the topography for spatial information surveyings and simulation and analysis of disaster rescue routes. It can come in handy when the Armed Force is trying to develop a rescue plan and perform effectiveness simulation of air defense position deployment using 3D military map intelligence.

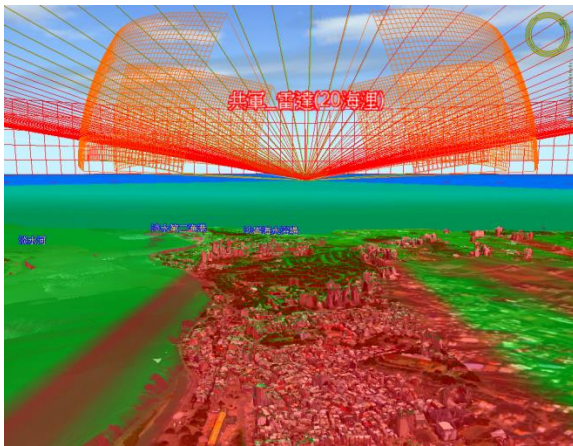


Fig.28 The green parts were the detectable zones for the radar of red force, whereas the red zones were the blind zones of radar



Fig.29 Topographic features provide hiding places for helicopter

The building models developed by this study were based on the specifications of refined modeling at LOD3. Therefore, the files of these objects were so big that the computer performance was unable to meet the requirement. If a model of a city is to be built in this way, it could be a challenge for computer wargame or modeling and simulation analysis. Hence, it is

suggested to develop large scope of 3D models with LOD1 or LOD2 specifications (as shown in Fig.30) in order to reduce the size of 3D object files and improve the performance of simulation systems.



Fig.30 Specifications of 3D town at LOD2

Acknowledgement

The study group would like to appreciate AGI for providing the STK program for trial and their assistance, and Dataforce for providing technical supports that were vital to make this study possible.

References

- [1]Guang-Jiang Pang, "The Tactics of The Urban Warfare in US. Army and Its Enlightenment to Us", Infantry Magazine Quarterly, No. 219, 2005. (in Chinese)
- [2]Cun-Zhou Lai, "The Study on the Troops Training with War-game Simulation - Taking JCATS as an Example", Army Bimonthly, Vol. 49, No.530, pp.4-22, 2013. (in Chinese)
- [3]Chuan-Wei Chiang, "The Study of Terrain Files for Joint Conflict and Tactical Simulation", Master Thesis, Chung Cheng Institute of Technology, National Defense University, Taoyuan, Taiwan, ROC, 2012. (in Chinese)
- [4]The Modeling and Simulation Office, Department of Integrated Assessments of MOD., "2008 EADSIM Analyze Training Handouts", Taipei, ROC., 2008. (in Chinese)
- [5]Gröger, G., T.H. Kolbe, A. Czerwinski and C.

- Nagel, 2008. OpenGIS City Geography Markup Language (CityGML) Encoding Standard, OGC-08-007r1, v. 1.0.0, Open Geospatial Consortium Inc., URL: <http://www.opengeospatial.org/standards/citygml> (accessed on Jan. 10 2010).
- [6][http://www.agi.com/products/stk/\(2014/07/18\)](http://www.agi.com/products/stk/(2014/07/18))).
- [7]Chia-Ming Liu, "Feature Extraction from LIDAR PointCloud Data", Master Thesis, National Cheng Kung University, Tainan, Taiwan, ROC, 2005. (in Chinese)
- [8]Clément Mallet, and Frédéric Bretar, "Full-waveform topographic lidar: State - of - the-art" , ISPRS Journal of Photogrammetry and Remote Sensing 64, pp.1-16, 2009 .
- [9]Ming-Da Tsai and Chia-Chyang Chang, "Impact of Digital Battlefield on Future Wars", Army Bimonthly, No. 470, pp. 78-79, 2004. (in Chinese)

Computationally-Efficient Image-Processing Scheme for Unilateral Parkinson's Disease Rodent Model

Chien-Jen Wang¹, Yaw-Syan Fu^{2*}, Jer-Min Tsai³, and Yin-Mou Shen⁴

¹Department of Computer and Communication, Kun Shan University

²Department of Biomedical Science and Environmental Biology, Kaohsiung Medical University

³Department of Information and Communication, Kun Shan University

⁴Department of Information Management, Kun Shan University

ABSTRACT

An automatic animal motion activity measurement system is proposed. In the proposed approach, the input image frames (acquired in real-time by a webcam or retrieved from a pre-saved video file) are processed by a combined Background Subtraction and Region Growing method in order to separate the object of interest from the background. The contours of the extracted object are determined using a chain code algorithm and are used to determine the principal axis of the object. Finally, the change in angle of the principal axis over successive frames is used to evaluate the rotation motion of the object. The feasibility of the proposed method is demonstrated by examining the apomorphine-induced asymmetric rotational motion behavior of male adult Sprague–Dawley rats with nigrostriatal lesions. The motion activity measurement system proposed in this study has the advantages of a simple low-cost setup, the potential for real-time implementation, and an accurate and reliable measurement performance. As such, it provides a useful tool for normative, pharmacological and neurophysiologic animal behavioral activity studies.

Keywords: Parkinson's disease, animal tracking, image processing, real time

具高計算效率影像處理之齧齒目動物單側帕金森氏症行動記錄系統

王建仁¹ 傅耀賢^{2*} 蔡哲民³ 沈英謀⁴

¹崑山科技大學電腦與通訊系

²高雄醫學大學生物醫學暨環境生物學系

³崑山科技大學資訊傳播系

⁴崑山科技大學資訊管理系

摘 要

動物神經性病變，例如帕金森氏症，對動物行進影響是醫學研究的重要課題。但是記錄動物行進行為，需要大量的人力且極為耗時。本論文介紹一可自動記錄動物行進軌跡之動物運動行為量測系統。系統藉由影像處理分析即時拍攝之動物行進並記錄運動軌跡，也可依預錄之影片進行分析，再以背景相減及區域成長將所追蹤動物自背景影像分離，分離之動物影像以鍊碼法描述其外型並計算形狀之主軸，最後記錄動物在連續影像中的位移軌跡及主軸角度

文稿收件日期 106.9.18;文稿修正後接受日期 107.4.20; *通訊作者

Manuscript received September 18, 2017; revised April 20, 2018; * Corresponding author

變化以供學者分析。為驗證系統可用性，研究團隊記錄以 Sprague – Dawley 鼠為例，注射阿朴嗎啡於其腦部特定位置，引發不對稱運動，以模擬其黑質紋狀體病變對行動影響；經與人力觀察比對，系統可正確追蹤及紀錄其行進及方向變化。本系統除即時自動追蹤紀錄以節省人力外，系統硬體僅包含攝影設備及電腦，建置成本低廉。本系統可提供研究運動行為受神經生理病變及藥物影響學者一組有用的工具。

關鍵詞：帕金森氏症，動物追蹤，影像處理，即時處理

I . INTRODUCTION

Parkinson's disease (PD) is a progressive disorder [1,2] resulting mainly from a de-generation of the dopaminergic neurons of the nigrostriatal tract of the brain, and characterized by motor impairments such as akinesia, bradykinesia, postural instability, and tremors [3]. 6-Hydroxydopamine (6-OHDA) is a specific neurotoxin for the dopaminergic pathways [4], and is transported into the cells of the nigrostriatal tract using the same membrane transporters as those used for catecholamine. The toxic effects of 6-OHDA have been attributed to the formation of various oxidants and reactive oxygen radicals [5], which cause lipid peroxidation, protein damage, and amino acid modification. On the other hand, when injuries occur on the nervous system, such as stroke [6], bleeding [7], brain trauma [8], subarachnoid hemorrhage [9], spinal cord injury [10], or neurodegeneration [11], most phenotypes of these nervous injuries may impair motor ability of the patients or modeling animals. In the experimental animal models, the behavioral testing or motor activity is applied for evaluating the degree of brain injury or recovery status, especially in brain injuries.

In response to apomorphine administration, rats with unilateral 6-OHDA induced lesions of the nigrostriatal dopaminergic pathway circle to the contralateral side of the lesions. This circular movement model is widely used in pharmacological and neuro-physiologic studies of Parkinson's disease to test and evaluate the striatal dopaminergic activity or dopaminergic neuron survival in substantia nigra. In such studies, the motion activity (rotational behavior) of the rats is examined using either observational methods or automated methods. Observational methods involve the direct open-field monitoring of the rats in their home cage [12-14], and have the advantage that the qualitative

behavioral patterns of the rats can be observed directly. However, observation methods are time consuming and labor intensive. Accordingly, automated methods are commonly preferred.

Automated animal motion activity measuring systems can be broadly classified as either hardware-based or software-based, depending on the particular method used to collect and analyze the data. Hardware-based systems have the advantage of de-skilling the system design. However, they are rather expensive. As a result, various automatic measuring systems have been developed [15-18] and brought to market (e.g., Videomex-V, Auto-track, Columbus Instruments). Most software-based automated animal behavior measuring systems are based on image-processing techniques, such as object segmentation, contour tracking, and so on [19,20].

In general, physical objects can be categorized as either rigid or non-rigid. For rigid objects, the contours do not readily change (e.g., a vehicle or a ball), and hence their images are more easily processed using automated methods. The literature contains various proposals for vehicle segmentation systems [21] or insect-image recognition [22] based on a rigid assumption. In addition, Chang et al. [23] proposed an intelligent data fusion system for vehicle collision warning based on a combined vision / GPS sensing approach. However, the segmentation of non-rigid objects, such as humans and animals, poses a significant challenge to automated image-processing methods since the contours of such objects typically change over a sequence of frames. Thus, the segmentation of non-rigid objects generally requires a more complex procedure based on the use of level sets [24,25], boundary-based models [26], or active contour-based models [27-30].

As described above, the asymmetric circling rodent model is widely used to test the

effects of pharmacological compounds on striatal dopaminergic activity. One of the earliest uses of automatic measuring methods to assess the effects of neural damage in mice was that reported by [15], in which both the area of the circles circumscribed by the rat and the number of rotational circles walked were observed. However, the method is based on the path cross of the gravity center, which differs from that of the actual rat rotation behavior. Hence significant errors are produced, e.g. ~10% in the case of slow rat rotation. Furthermore, the proposed system requires special hardware and thus has a high cost. In performing behavioral studies based on the asymmetric circling rodent model, the direction of rotation is of interest in addition to the diameter and number of circumscribed circles. However, the literature presently contains very few automated tracing systems capable of providing such an insight.

Accordingly, the present study proposes an automatic system capable not only of recording and analyzing the asymmetric rotational motion activity of a rat, but also determining the direction of rotation. The proposed system has many practical advantages, including a low-cost setup, the potential for real-time implementation, and a reliable and repeatable measurement performance. As a result, it provides a useful tool for a variety of normative, pharmacological and neurophysiologic animal behavioral activity studies.

II. MATERIALS AND METHOD

2.1 Animals

Male adult Sprague-Dawley (SD) rats, aged 8 weeks and weighing 250-300 g, were purchased from the National Laboratory Animal Center, Taiwan. The rats were housed in an experimental animal room with a 12:12 h light-dark cycle and a temperature of 22±1°C. The rats were fed a standard pellet diet and water ad libitum. The study was conducted according to the guidelines of the National Science Council of the Republic of China, Taiwan, and was approved by the Animal Care and Use Committee of Kaohsiung Medical University.

2.2 Surgery

The experiments were conducted between 7:00 AM and 7:00 PM (i.e., during the light cycle). The rats were anaesthetised with chloral hydrate (400 mg/kg, i.p., Sigma) and then placed in a stereotaxic apparatus (Stoelting Instruments, IL, USA). 4 microliter 6-OHDA.Br (1 µg/µl, Sigma) or vehicle (sterile normal saline with 0.2% ascorbic acid) was injected into the left medial forebrain bundle (coordinates as A: -2.0 mm, L: 2.0 mm, V: -8.4 mm) at a rate of 1.0 µl/min for 4 min. The injection was unilateral, with the contralateral structures serving as controls. Following injection, the injection cannula was left in place for 4 min before removal.

4 weeks after surgery, the rats were tested with systemic apomorphine (0.75 mg/kg in sterile normal saline with 0.2% ascorbic acid, i.p.; Sigma, USA) to evaluate the behavioral effect of the nigrostriatal lesion. Specifically, 10 minutes after apomorphine administration (i.p.), the rotational behavior of the rats was evaluated by placing the animals in a black circular cage and then monitoring their movement using the proposed automatic animal motion activity measurement system.

III. SYSTEM DESIGN AND IMPLEMENTATION

3.1 System Architecture and Implementation

Figure 1 shows the basic architecture of the proposed animal motion activity measurement system. As shown, the system comprises four main components, namely Object Search, Object Segmentation, Contour Extraction, and Axis Computation. The input video stream is first processed by the Object Search procedure to identify the region of the input image containing the object of interest (i.e., the rat, in the present case). The Object Segmentation procedure, consisting of a Background Subtraction method [31-33] and a Region Growing method [34,35], is then applied to segment the object from the background.

A variety of foreground and background segmentation algorithms have been proposed in the literature, such as background subtraction, color clustering [36], image-based distance and area measurement [37], etc. Among them, background subtraction is widely adopted due to

its computation efficiency.

The object contour is then extracted using a chain code algorithm [38]. Finally, the principal axis of the object is computed from the extracted contour and used to evaluate the change in rotational position of the object from one frame to the next. The details of each component of the proposed motion measurement system are described in the following sub-sections.

3.2 Object Search

Traditional Background Subtraction methods operate at the pixel level over the entire input image, and tend to include significant noise in the extracted results. Consequently, post-processing using some form of filtering technique such as a Median Filter, Average Filter or Morphology Filter [39,40] is generally required. However, these post-processing operation sometimes blur the contours of the segmented object, and hence degrade the performance of downstream processes. Accordingly, in the present study, the input image is first processed using an Object Search algorithm to identify the region of the image containing the object of interest, and the segmentation process is then performed only in this more restricted region of the image. The Object Search algorithm is as follows:

(1) A cross-shape mask ('+') with an assigned scale is set. In practice, the scale of the mask should be carefully assigned since while a larger value results in a more robust noise removal performance, it may also cause some candidate objects of interest to be filtered out. In this study, both vertical scale(VS) and horizontal scale(HS) of the mask are assumed to be 11. In addition, this mask operates from left to right and top to bottom every 11 pixels apart.

(2) The region that can contain the whole mask is assumed as the candidate object. The center location of the mask is marked as O_i , which will serve as the seed point of region growing in the later process. The remaining regions are assumed as noise. Considering the illustrative example shown in Fig. 2, the green regions indicate objects or noises. The pseudo code of the algorithm is as follows:

```

N=1
FOR I=7 to Height step VS
{
  FOR J=7 to Width step HS
  {
    Flag=true;
    FOR m=0 to VS-1 step 1
    {
      If (b (J, I+m) !=1)
      { Flag=false; break m loop}
    }

    If(Flag==true)
    {
      FOR n=0 to HS-1 step 1
      {
        If (b (J+n, I) !=1)
        {Flag=false; break n loop}
      }
       $O_N(x, y) = (J, I)$ 
      N=N+1
    }
  }
}

```

A bitmap $b(x, y)$ consisting of "1" (seed points) and "0" (background) is constructed.

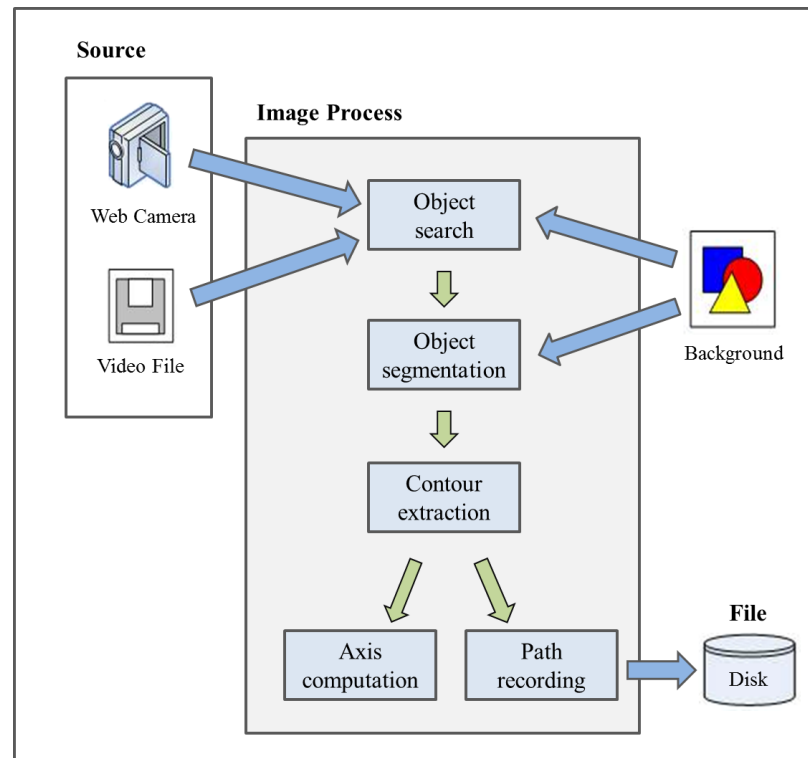


Fig.1. System architecture

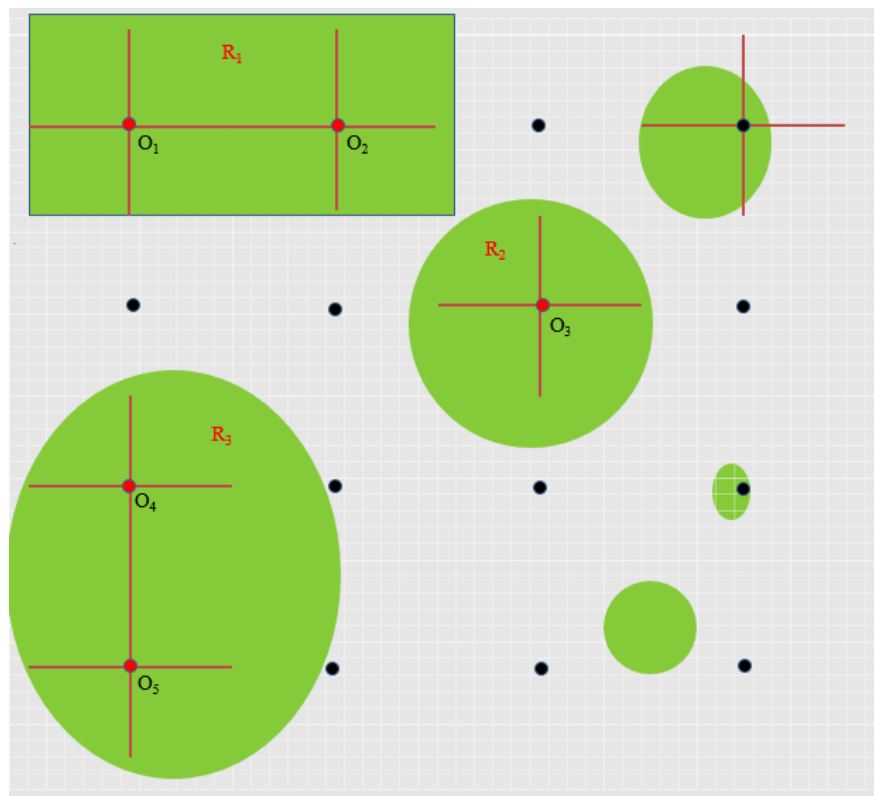


Fig.2. Object search example (Note that green regions indicate objects or noises. The red points will serve as the seeds of region growing.)

3.3 Object Segmentation

Having identified the candidate seeds, we perform region-growing process from the first seed. During growing, if other seeds are

covered, then those seeds belong to the same object and those seeds won't perform region-growing any more. For example, in Fig. 2, when the region is growing from seed O_1 , O_2 will be covered during the growing process since these two seeds belong to the same region, R1.

New region-growing is performed from the remaining seeds. This procedure continues until all the seeds are performed or covered by some region. These resulting regions are identified as $R \{R_1, R_2, R_3\}$. In this study, only one mouse is tested in the experiment, the largest region is chosen to be the object of interest.

3.4 Contour Extraction

Once the object of interest has been segmented, the contour of the segmented object is then extracted using the chain code algorithm. Once the contour is extracted, all of the pixels within the contour are set to "1" in order to fill any intra-object gaps, as shown in Fig. 3.

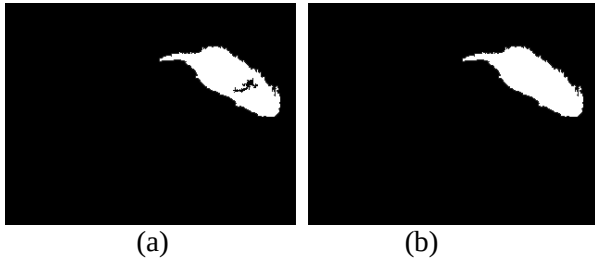


Fig. 3. Filling of intra-object gaps: (a) Segmented image, (b) After contour extraction

3.5 Axis Computation

The binarized image obtained from the chain code algorithm (denoted by the function $f(x, y)$) is processed to calculate the axis of the extracted object. The change in the axis position between successive frames is then used to detect the rotational movement of the object. The axis is calculated using Eq. (1) based on the relative distribution of the pixels within the object. The change in the axis angle, θ , is then calculated using Eq. (2). Note that the coordinates of the object mass center (x_c, y_c) are derived from Eq. (2) in accordance with Eq. (3)

$$M_{p,q} = \sum_x \sum_y x_p y_q f(x, y) \quad (1)$$

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2M_{1,1}}{M_{2,0} - M_{0,2}} \right) \quad (2)$$

$$(x_c, y_c) = \left(\frac{M_{1,0}}{M_{0,0}}, \frac{M_{0,1}}{M_{0,0}} \right) \quad (3)$$

3.6 Additional Functions

In addition to measuring the animal motion activity, the proposed system also facilitates the real-time saving to disc of the webcam video file to provide a permanent record of the original laboratory data and to enable subsequent double-checking of the analysis results by alternative methods if desired. In addition, the calculated coordinates of the object mass center are also saved (together with a timing stamp) in order to support further data processing and analysis, e.g., the movement speed and travel distance of the object, the travel track of the object, and so on.

IV. RESULT AND DISCUSSIONS

To demonstrate the computational efficiency of the proposed system, the image-processing algorithm was implemented on a basic PC with limited processing capabilities (i.e., P-4 2.8 GHz with 1 G RAM). The input images were captured using a commercial webcam connected through a USB 2.0 interface. The resolution of the input images was set as 320×240 with 24-bit RGB color. The captured images were processed at a rate of 20 frames per second with a 50% CPU utility. The rats were administered an intracranial injection of 6-OHDA to induce unilateral PD and were then placed (10 minutes later) in the round cage (diameter of 40 cm). Note that a round cage was deliberately chosen in preference to a quadrate cage in order to enable a freer motion of the rats. Nine PD-induced rats were used as observations targets; with each rat repeating the test protocol 5 or 6 times.

As described in Section 3.2, the image-processing algorithm proposed in this study performs an initial Object Search procedure to simplify the subsequent segmentation process. The resulting performance improvement was evaluated by

comparing the segmentation time for three objects of different sizes with that obtained using a traditional Background Subtraction method based on the full image. The segmentation times of the two systems are

shown in Table 1. It is seen that the proposed method reduces the segmentation time by up to 42%. In other words, the potential of the proposed algorithm for real-time implementation is confirmed.

Table 1 Segmentation times of traditional method and proposed method for objects of different sizes

Number of border pixels	Segmentation time for Background Subtraction method	Segmentation time for proposed method	Time saving
203	79 ms	46 ms	42%
321	78 ms	53 ms	32%
1023	79 ms	63 ms	20%

Table 2 Effect of Median Filter in improving counting performance of proposed system

Manual count	System count without Median Filter	System count with Median Filter	Count error without Median Filter	Count error with Median Filter
50	37	49	26 %	2 %
50	41	50	18 %	0 %
50	40	50	20 %	0 %

As described in Section 3.5, the image-processing algorithm proposed in this study evaluates the rotational behavior of the rat by computing the change in the principal axle angle of the extracted object over successive frames. Thus, compared to existing methods, which count the number of rotation circles by reconstructing the entire travel track, the present study determines the number of rotation circles directly from the computed change in the principal axle angle. Notably, by observing the direction in which the principal axle angle changes, the system is also able to determine the direction of the rotational movement of the rat.

In performing rotational motion, the swing of the rat tail may degrade the accuracy of the axis computation process and thus impair the ability of the system to count the number of rotational circles. Accordingly, in implementing the algorithm, a Median Filter is used to adjust the object image, thereby decreasing tail-swing interference and promoting the system accuracy as a result. Table 2 compares the number of rotational circles counted by the system with and without the Median Filter, respectively, with that observed manually. The results confirm that the filter significantly improves the reliability of the proposed system.

Table 3 Experimental results for apomorphine-induced rotational motion of PD rats

Test number	Manual count	System count	FPS	System accuracy
1	100	99	21	99 %
2	100	99	22	99 %
3	100	99	23	99 %
4	100	99	22	99 %
5	100	100	21	100 %
6	100	99	21	99 %
7	100	97	21	97 %

8	100	98	20	98 %
9	100	97	19	97 %
10	100	99	20	99 %
11	100	99	22	99 %
12	100	98	22	98 %
13	100	99	22	99 %
14	100	99	21	99 %
15	100	100	22	100 %
Average			21.26	98.7 %

As described earlier, the experiments were deliberately performed using a black cage in order to increase the contrast between the background and the SD rats. However, as the experiments continue, the cage tends to be spoiled by urine and feces. Consequently, the contrast diminishes and the likelihood of object axis computation errors increases. Accordingly, in performing the segmentation process, a high threshold value of $T=100$ was deliberately chosen in order to filter out such experimental noise.

Table 3 summarizes the experimental results obtained for the apomorphine-induced rotational motion of the SD rats (Table 3). As shown, the average accuracy of the proposed system is over 98%, as evaluated 15 practical tests. Consequently, the reliability and stability of the proposed automated system is confirmed.

Various systems have been proposed (and commercialized) for the automated measurement of animal motor activity. Generally speaking, these systems rely on the interruption of an infrared beam or the triggering of a touch-sensitive sensor as the animal performs motion [41-44]. However, these systems have a high price (typically greater than USD 20,000) due to their hardware requirements. Moreover, the behavioral mode which can be observed or measured by such systems is limited by their hardware design. By contrast, the system proposed in this study has a cost of less than USD 3,000 and provides the ability to examine a greater range of motional behaviors since the behavioral data are computed from an inspection of the pixel changes between successive images rather than by means of some form of optical or mechanical device. Furthermore, in an alpha test, it was found that the resolution of the proposed

system (more than 320x240) was sufficient to segment even small animals such as ants and crickets from a monotonous background.

The literature contains several proposals for automatic animal motion analysis systems based on tracking the spatial coordinates of multiple markers affixed to the animals' bodies [45]. However, webcam-based marker-free systems such as that proposed in this study have the ability to monitor a greater range of real-time movements of the object animals and to gain improved quantitative measurements of the overall animal movements under a variety of conditions [46,47].

In response to amphetamine administration, rats with unilateral 6-OHDA induced lesions of the nigrostriatal dopaminergic pathway circle to the ipsilateral lesioned side. By contrast, the rats circle to the contralateral side of the lesion when administrated with apomorphine [48]. In addition, the experimental results (Table 3) have shown that the count performance of the proposed system is accurate to within 1.5% of that obtained manually. Finally, the system is based on the RGB color space and therefore avoids the need for color space transformation. In addition, the system utilizes an Object Search algorithm to reduce the computational complexity of the object segmentation process, and applies a chain code algorithm to detect the object contour such that the rotational change of the object can be determined. Consequently, the system has a lower computational complexity than existing methods, and thus has significant potential for real-time implementation.

V. CONCLUSIONS

This study has presented an automated

animal motion activity measurement system based on a webcam and a computationally efficient image processing algorithm. The feasibility of the proposed system has been demonstrated experimentally by measuring the apomorphine-induced asymmetric rotational motion behavior of rats with nigrostriatal lesions. Notably, compared to existing automated methods, the proposed analysis system enables the direction of rotation to be reliably determined. The system has minimal hardware requirements, a high measurement accuracy and reliability, and the potential for real-time implementation. Furthermore, the hardware setup can be easily scaled to accommodate animals of a larger size. Consequently, the proposed system provides an ideal solution for a wide variety of normative, pharmacological and neurophysiologic animal behavioral activity studies.

REFERENCES

- [1] Beitz, J. M., "Parkinson's disease: a review," *Frontiers in Bioscience*, Vol. 6, pp. 65-74, 2014.
- [2] Fritsch, T., Smyth, K. A., Wallendal, M. S., Hyde, T., Leo, G., and Geldmacher, S., "Parkinson disease: research update and clinical management," *Southern Medical Journal*, Vol. 105, pp. 650-656, 2012.
- [3] Deumens, R., Blokland, A., and Prickaerts J., "Modeling Parkinson's disease in rats : an evaluation of 6-OHDA lesions of the nigrostriatal pathway," *Experimental Neurology*, Vol. 175, pp. 303-317, 2002.
- [4] Cohen, G. and Heikkila, R. E., "The generation of hydrogen peroxide, superoxide radical, and hydroxyl radical by 6-hydroxydopamine, dialuric acid, and related cytotoxic agents," *Journal Of Biological Chemistry*, Vol. 249, pp. 2447-2452, 1974.
- [5] Cohen, G., "Oxy-radical toxicity in catecholamine neurons," *Neurotoxicology*, Vol. 5, No. 1, pp. 77-82, 1983.
- [6] Shin D. H., Kim, G. H., Lee, J. S., Joo, I. S., Suh-Kim, H., Kim, S. S., Hong, J. M., "Comparison of MSC-Neurogenin1 administration modality in MCAO rat model," *Translational Neuroscience*, Vol. 7, No. 1, pp. 164-172, 2016.
- [7] Yang, Y., Zhang, M., Kang, X., Jiang, C., Zhang, H., Wang, P., and Li, J., "Thrombin-induced microglial activation impairs hippocampal neurogenesis and spatial memory ability in mice," *Behavioral and Brain Functions.*, Vol. 11, No. 1, 30, 2015.
- [8] Cline, M. M., Yumul, J. C., Hysa, L., Murra, D., Garwin, G. G., Cook, D. G., Ladiges, W. C., Minoshima, S., and Cross, D. J., "Novel application of a Radial Water Tread maze can distinguish cognitive deficits in mice with traumatic brain injury," *Brain Research*, Vol. 1657, pp. 140-147, 2017.
- [9] Zibly, Z., Fein, L., Sharma, M., and Assaf, Y., Wohl, A., Harnof, S., "A novel swine model of subarachnoid hemorrhage-induced cerebral vasospasm," *Neurology India*, Vol. 65, No. 5, pp. 1035-1042, 2017.
- [10] Manohar, A., Foffani, G., Ganzer, P. D., Bethea, J. R., and Moxon, K. A., "Cortex-dependent recovery of unassisted hindlimb locomotion after complete spinal cord injury in adult rats," *Elife*, Vol. 6, pii: e23532, 2017.
- [11] Kühn, J., Haumesser, J. K., Beck, M. H., Altschüler, J., Kühn, A. A., Nikulin, V. V., and van Riesen, C., "Differential effects of levodopa and apomorphine on neuronal population oscillations in the cortico-basal ganglia loop circuit in vivo in experimental parkinsonism," *Experimental Neurology*, Vol. 298, pp. 122-133, 2017.
- [12] Ivinskis, A., "The reliability of behavioural measures obtained in the open field," *Australian Journal of Psychology*, Vol. 22, pp. 175-183, 1968.
- [13] Denenberg, V. H., "Open-field behavior in the rat: What does it mean?" *Annals of the New York Academy of Sciences*, Vol. 159, pp. 852-859, 1969.
- [14] Walsh, R. N. and Cummins, R. A., "The open-field test: a critical review," *Psychological Bulletin*, Vol. 83, pp. 482-504, 1976.
- [15] Torello, M. W., Czekajewski, J., Potter, E. A., Kober, K. J., Fung, Y. K., "An Automated Method for Measurement of Circling Behavior in the Mouse," *Pharmacology Biochemistry & Behavior*, Vol. 19, pp. 13-17, 1983.

- [16] Young, M. S., Li, Y. C., Lin, M. T., "A modularized infrared light matrix system with high resolution for measuring animal behaviors," *Physiology & Behavior*, Vol. 53, pp. 545-551, 1993.
- [17] Kao, S. D., Shaw, F. Z., Young, M. S., Jan, G. J., "A new automated method for detection and recording of animal moving path," *Journal of Neuroscience Methods*, Vol. 63, pp. 205-209, 1995.
- [18] Fitch, T., Adams, B., Chaney, S., and Gerlai, R., "Force transducer-based movement detection in fear conditioning in mice: a comparative analysis," *Hippocampus*, Vol. 12, pp. 4-17, 2002.
- [19] Dai, H. and Carey, R. J., "A new method to quantify behavioral attention to a stimulus object in a modified open-field," *Journal of Neuroscience Methods*, Vol. 53, pp. 29-34, 1993.
- [20] Kato, S., Nakagawa and T., Ohkawa, M., "A computer image processing system for quantification of zebrafish behavior," *Journal of Neuroscience Methods*, Vol. 134, pp. 1-7, 2003.
- [21] Wang, J. M., Chung, Y. C., Chang, C. L., and Chen, S. W., "Shadow detection and removal for traffic images". In *Networking, Sensing and Control*, IEEE International Conference on, Vol. 1, pp. 649-654, 2004.
- [22] Shen, Y. M., Wang, C. J., Tsai, J. M., and Fu, Y. S. "Insect-image recognition system based on intensity gradient feature," *Optical Engineering*, Doi: 10.1117/1.OE.54.4.043103, Vol. 54, No. 4, 10 pages, 2015.
- [23] Chang, B., R., Tsai, H. F., and Young, C. P., "Intelligent data fusion system for predicting vehicle collision warning using vision/GPS sensing," *Expert Systems with Applications*, Vol. 37 No. 3, pp. 2439-2450, 2010.
- [24] Wren, C. R., Azarbajani, A., Darrell, T., and Pentland, A. P., "Pfindex: Real-time tracking of the human body," *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 19, No. 7, pp. 780-785, 1997.
- [25] Sun, X., Yao, H., Zhang, S., and Li, D., "Non-Rigid Object Contour Tracking via a Novel Supervised Level Set Model," *IEEE Transactions on Image Processing*, Vol. 24 No. 11, pp. 3386-3399, 2015.
- [26] Tillet, R. D., Onyango, C. M., and Marchant, J. A., "Using model-based image processing to track animal movements," *Computers and Electronics in Agriculture*, Vol. 17, pp. 249-261, 1997.
- [27] Yilmaz, A., Li, X., and Shah, M., "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 26, No. 11, pp. 1531-1536, 2004.
- [28] Paragios, N. and Deriche, R. "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 22, No. 3, pp. 266-280, 2000.
- [29] Araki, S., Matsuoka, T., Yokoya, N., and Takemura, H., "Real-time tracking of multiple moving object contours in a moving camera image sequence," *IEICE TRANSACTIONS on Information and Systems*, Vol. 83, No. 7, pp. 1583-1591, 2000.
- [30] Kass, M. and Witkin, A., "Terzopoulos D. Snake: Active contour model," *International Journal of Computer Vision*, Vol. 1, No. 4, pp. 321-331, 1988.
- [31] Kehtarnavaz N. and Rajkotwala F., "Real-Time Vision-based Detection of Waiting Pedestrians", *Real-Time Imaging*, pp. 433-440, 1997.
- [32] Yong, H., Meng, D., Zuo, W., and Zhang, L., "Robust Online Matrix Factorization for Dynamic Background Subtraction," *IEEE transactions on pattern analysis and machine intelligence*, 2017.
- [33] Arya, G., Singh, M., and Gupta, M., "Human-Computer Interaction based on Real-time Motion Gesture Recognition," *Human-Computer Interaction*, Vol. 4, No.3, 2016.
- [34] Yu, Y. W. and Wang, J. H. "Image segmentation based on region growing and edge detection," In *Systems, Man, and Cybernetics*, 1999. *IEEE SMC International Conference on*, Vol. 6, pp. 798-803, 1999.
- [35] Pavlidis, T. and Horowitz S., "Segmentation of plane curves," *IEEE transactions on Computers*, Vol. 100, No. 8, pp. 860-870, 1974.
- [36] Hong, A. X., Chen, G., Li, J. L., Chi, Z. R., and Zhang, D., "A flower image retrieval

- method based on ROI feature,” *Journal of Zhejiang University-Science A*, Vol. 5, No.7, pp. 764-772, 2004.
- [37] Lu, M. C., Wang, W. Y., and Chu, C. Y., “Image-based distance and area measuring systems,” *IEEE Sensors Journal*, Vol. 6, No. 2, pp. 495-503, 2006.
- [38] Shahab, W., Al-Otun, H., and Al-Ghoul, F., “A modified 2D chain code algorithm for object segmentation and contour tracing,” *Int. Arab J. Inf. Technol.*, Vol. 6, No. 3, pp. 250-257, 2009.
- [39] Jin, X. C., Ong, S. H., and Jayasooriah, J., “A domain operator for binary morphological processing,” *IEEE transactions on image processing*, a publication of the IEEE Signal Processing Society, Vol. 4, No. 7, pp. 1042-1046, 1994.
- [40] Heijmans, H. J. and Ronse, C., “The algebraic basis of mathematical morphology I. Dilations and erosions,” *Computer Vision, Graphics, and Image Processing*, Vol. 50, No. 3, pp. 245-295, 1990.
- [41] Irwin, I., DeLanney, L. E., Forno, L. S., Finnegan, K. T., Di Monte, D. A., and Langston, J. W., “The evolution of nigrostriatal neurochemical changes in the MPTP-treated squirrel monkey,” *Brain Research*, Vol. 531, pp. 242-52, 1990.
- [42] Pearce, R. K. B., Jackson, M., Smith, L., Jenner, P., and Marsden, C. D., “Chronic L- DOPA administration induces dyskinesias in the 1- methyl- 4- phenyl- 1, 2, 3, 6- tetrahydropyridine- treated common marmoset, (*Callithrix jacchus*),” *Movement disorders*, Vol. 10, No.6, pp. 731-740, 1995.
- [43] Smith, R. D., Zhang, Z., Kurlan, R., McDermott, M., and Gash, D. M., “Developing a stable bilateral model of parkinsonism in rhesus monkeys,” *Neuroscience*, Vol. 52, pp. 7-16, 1993.
- [44] Bushnell, P. J., Moser, V. C., MacPhail, R. C., Oshiro, W. M., Derr-Yellin, E. C., Phillips, P. M., and Kodavanti, P. R., “Neurobehavioral assessments of rats perinatally exposed to a commercial mixture of polychlorinated biphenyls,” *Toxicological Sciences*, Vol. 68, No. 1, pp. 109-120, 2002.
- [45] Pan, W. H., Lee, C. R., and Lim, L. H., “A new video path analyzer to monitor travel distance, rearing, and stereotypic movement of rats,” *Journal of neuroscience methods*, Vol. 70, No. 1, pp. 39-43, 1996.
- [46] Hashimoto, T., Izawa, Y., Yokoyama, H., Kato, T., and Moriizumi, T., “A new video/computer method to measure the amount of overall movement in experimental animals, two-dimensional object-difference method),” *Journal of neuroscience methods*, Vol. 91, No. 1, pp. 115-122, 1999.
- [47] Togasaki, D. M., Hsu, A., Samant, M., Farzan, B., DeLanney, L. E., Langston, J. W., and Quik, M., “The Webcam system: a simple, automated, computer-based video system for quantitative measurement of movement in nonhuman primates,” *Journal of neuroscience methods*, Vol. 145, No. 1, pp. 59-166, 2005.
- [48] Ahmad, M., Yousuf, S., Khan, M. B., Hoda, M. N., Ahmad, A. S., Ansari, M. A., Ishrat, T., Agrawal, K. A., and Islam, F., “Attenuation by Nardostachys jatamansi of 6-hydroxydopamine-induced parkinsonism in rats, behavioral, neurochemical, and immunohistochemical studies,” *Pharmacol. Pharmacology Biochemistry and Behavior*, Vol. 83, pp. 150-160, 2006.

The Response and Failure of Local Sharp-cut 6061-T6 Aluminum Alloy Tubes with Different Diameter-to-thickness Ratios Submitted to Cyclic Bending

Kuo-Long Lee^{1*}, Wei-Yi Huang², Wen-Fung Pan³

¹ Department of Innovative Design and Entrepreneurship Management, Far East University

^{2,3} Department of Engineering Science, National Cheng Kung University

ABSTRACT

In this paper, the response and failure of local sharp-cut 6061-T6 aluminum alloy tubes with different diameter-to-thickness ratios submitted to cyclic bending were studied. Three different diameter-to-thickness ratios of 16.5, 31.0 and 60.0 were considered. The experimental moment-curvature relationship rapidly became a steady loop from the beginning of the first bending cycle. Moreover, the cut depth had almost no influence on the moment-curvature relationship. As for the ovalization-curvature relationship, when the number of cycles increased, it exhibited an increasing and ratcheting manner. It was seen that the greater the depth of the cut, the more asymmetrical ovalization-curvature relationship and the greater the increase of the ovalization. Furthermore, for a certain diameter-to-thickness ratio, five unparallel straight lines corresponding to five different cut depths were found for the controlled curvature-number of cycles required to produce failure relationship on a log-log scale. Finally, a theoretical model was proposed in this study for simulating aforementioned relationship. It was found that the experimental and analytical data agreed well.

Key words: Local Sharp-cut, 6061-T6 Aluminum Alloy Tubes, Diameter-to-thickness Ratio, Cyclic Bending, Response, Failure.

不同外徑/壁厚比局部切痕 6061-T6 鋁合金圓管在循環彎曲 負載下行為與損壞之研究

李國龍^{1*} 黃文義² 潘文峰³

¹ 遠東科技大學創新設計與創業管理系

^{2,3} 國立成功大學工程科學系

摘 要

本文主要是研究不同外徑/壁厚比的局部尖銳切痕 6061-T6 鋁合金圓管在循環彎曲負載下的行為及損壞，而所考慮的三種不同外徑/壁厚比包含有：16.5、31.0 及 60.0。從實驗的彎矩-曲率曲線中顯示，彎矩-曲率關係從第一圈開始即快速呈現一穩定的迴圈，且切痕深度對彎矩-曲率關係幾乎沒有影響。至於橢圓化-曲率關係則隨著循環圈數的增加而呈現棘齒狀的成長，且切痕深度

文稿收件日期 106.9.18；文稿修正後接受日期 107.4.20；*通訊作者

Manuscript received September 18, 2017；revised April 20, 2018；*Corresponding author

越深時，橢圓化-曲率關係就越不對稱，橢圓化增加就越大。此外，當考慮個別外徑/壁厚比時，五種不同切痕深度在雙對數座標的控制曲率-循環至損壞圈數關係呈現出五條不平行的直線。最後，本研究提出一個理論模式來描述控制曲率-循環至損壞圈數關係，在與實驗結果比較後發現，理論能夠合理描述實驗結果。

關鍵詞：局部尖銳切痕、6061-T6 鋁合金管、外徑/壁厚比、循環彎曲、反應、損壞。

I. INTRODUCTION

Circular tubes are often used in mechanical or structural components, for example earthquake-prone-area structures, drilling oil platform, subsea pipelines, and heat exchangers in power plants or nuclear reactors. However, these tube components are frequently submitted to bending load, which brings about the ovalization of tube's cross section. And reverse and continuous cyclic bending leads to gradual increase in ovalization. The ovalization phenomenon leads to the degradation of the tube's rigidity. The circular tube failures (buckling or fracture) when the ovalization reaches a crucial value. Therefore, it is of importance to completely understand the response during the cyclic bending process and failure at the final stage of circular tubes submitted to cyclic bending in many industrial applications.

In 1985, Prof. Kyriakides and his research team designed a bending machine and began a series of experimental and theoretical studies on tubes submitted to monotonic or cyclic bending with or without external or internal pressure. Shaw and Kyriakides [1] studied the elastoplastic behavior of tubes under cyclic bending. The moment-curvature and ovalization-curvature relationships were theoretically simulated using the principle of virtual work. Kyriakides and Shaw [2] then extended the aforementioned research to evaluate the buckling failure of tubes under cyclic bending. They suggested an empirical form to describe the controlled curvature and the number of cycles required to produce a buckling relationship. Meanwhile, Corona and Kyriakides [3] experimentally investigated the behavior of tubes subjected to cyclic bending with some external pressure. They discovered that the accumulation of ovalization was accelerated by applying some external pressure. Vaze and Corona [4]

employed a similar bending machine to examine the deterioration and collapse of square steel tubes submitted to cyclic bending. They discovered that the tubes deteriorated because of the growth of periodic deflections in the flanges. Moreover, Corona and Kyriakides [5] studied the tube's buckling under bending with some external pressure. In their work, 304 stainless steel tubes exhibited the angle buckling oriented at 20-45° to the direction of the bending moment. Similarly, Corona *et al.* [6] investigated the anisotropy of the plastic deformation for tubes submitted to bending. The material anisotropy behavior at the pre-buckling, post-buckling, and bifurcation stages was simulated by flow and deformation theories. Limam *et al.* [7] experimentally and theoretically discussed the inelastic collapse of tubes subjected to bending with some internal pressure. They employed the shell model in the finite element method to evaluate tube wrinkling and localization. Limam *et al.* [8] also examined the response and collapse of local-dented tubes undertaking pure bending with some internal pressure. The dent processing, tube pressurization, and tube bending to collapse were described through the finite element method. Bechle and Kyriakides [9] experimentally investigated the localization of NiTi tubes submitted to bending. In addition, the influence of the texture-driven and material asymmetry on the tube structure was studied. Jiang *et al.* [10] studied the pseudoelastic response of NiTi tubes subjected to bending. A recently developed constitutive model for the pseudoelastic behavior was implemented in a finite element analysis used to simulate the tube bending experiments.

Other researchers have also published numerous related studies. Yuan and Mirmiran [11] experimentally and theoretically investigated the static buckling of fiber-reinforced plastic tubes filled with concrete subjected to bending. Elchalakani *et al.* [12]

experimentally tested on grade C350 steel tubes with different diameter-to-thickness ratios (D_o/t ratios) submitted to pure bending. They proposed two theoretical models to simulate the experimental results. Meanwhile, Jiao and Zhao [13] experimentally investigated the bending behavior of very high-strength steel tubes and proposed a plastic slenderness limit for the material they tested. Moreover, Houliara and Karamanos [14] investigated the elastic response of long-pressurized, thin-walled tubes at the buckling and post-buckling stages undertaken in-phase bending. Elchalakani and Zhao [15] focused on concrete-filled and cold-formed steel tubes subjected to monotonic and cyclic bending with variable amplitudes. Yazdani and Nayebi [16] investigated the response and damage of thin-walled tubes submitted to cyclic bending with a steady internal pressure. Fan *et al.* [17] presented an analytical study on the critical dynamic buckling load of cylindrical shells under a uniform external pressure. Elchalakani *et al.* [18] determined new ductile slenderness limits of CFT structures during plastic design using measured strains in plastic bending tests. Shamass *et al.* [19] investigated the elastoplastic buckling of thin circular shells subjected to a non-proportional loading.

In 1998, Pan *et al.* [20] designed and set up a new measurement apparatus. The apparatus was used with a cyclic bending machine to study various types of tube under different cyclic bending conditions. For examples: Pan and Fan [21] studied the effect of the prior curvature-rate at the preloading stage on subsequent creep or relaxation behavior, Pan and Her [22] investigated the response and stability of 304 stainless steel tubes that were subjected to cyclic bending with different curvature rates, Lee *et al.* [23] studied the influence of the D_o/t ratio on the response and stability of circular tubes that were subjected to symmetrical cyclic bending, Lee *et al.* [24] experimentally explored the effect of the D_o/t ratio and curvature-rate on the response and stability of circular tubes subjected to cyclic bending, Chang *et al.* [25] studied the mean moment effect on circular, thin-walled tubes under cyclic bending, and Chang and Pan [26] discussed the buckling life estimation of

circular tubes subjected to cyclic bending.

In 2010, the research team of Prof. Pan began to experimentally and theoretically investigate the response and the collapse of sharp-notched circular tubes submitted to cyclic bending. Lee *et al.* [27] experimentally studied the relationship between the variation of ovalization and the number of bending cycles for sharp-notched circular tubes subjected to cyclic bending. Lee [28] investigated the response of sharp-notched SUS304 stainless steel tubes under cyclic bending, and found the asymmetrical, ratcheting, and increasing trends of ovalization-curvature relationships. Hung *et al.* [29] then later experimental investigated on the mechanical response and buckling failure of SUS304 stainless steel tubes with five different sharp-notched depths subjected to cyclic bending with three different curvature rates. In addition, Lee *et al.* [30] evaluated the viscoplastic response and buckling of sharp-notched SUS304 stainless steel circular tubes undertaking cyclic bending. They observed that the cyclic controlled curvature and the number of bending cycles required to produce buckling relationships on a log-log scale revealed parallel lines for every notch depths for a certain curvature rate. Lee *et al.* [31] studied the response of sharp-notched circular tubes under pure bending creep. They found that the creep curvature and ovalization increase with time for pure bending creep and higher held moment leads to the higher creep curvature and ovalization of the tube's cross-section. However, the sharp notch for the aforementioned investigations was a circumferential sharp notch as shown in Fig. 1.

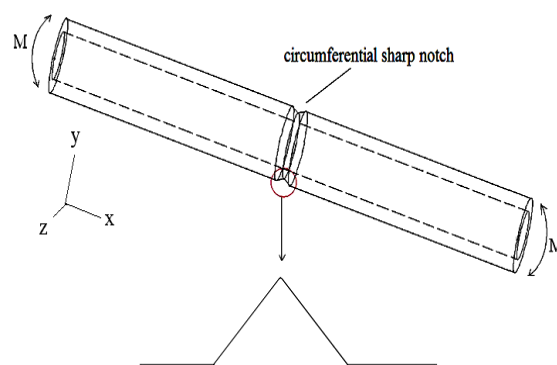


Fig. 1 A schematic drawing of the circumferential sharp-notched tube

The investigation of the influence of the D_o/t ratio on the response and failure was first investigated by Lee *et al.* [23]. The 304 stainless steel smooth tubes with four different D_o/t ratios of 30, 40, 50 and 60 were tested under cyclic bending. Later, Chang *et al.* [32] studied the buckling failure of 301 stainless steel smooth tubes with different D_o/t ratios subjected to cyclic bending. In 2012, Lee *et al.* [33] started to investigate the influence of the D_o/t ratio on the response and failure of sharp-notched 304 stainless tubes under cyclic bending. The type of the sharp notch was in Fig. 1 and three different D_o/t ratios were 16.25, 21.20 and 24.46. They discover that five almost parallel straight lines corresponding to five different notch depths for each D_o/t ratio were observed from the controlled curvature and number of bending cycles required to produce failure relationship on a log-log scale. In addition, the slopes for the aforementioned relationship for three different D_o/t ratios were almost the same.

In this paper, local sharp-cut 6061-T6 aluminum alloy tubes with different D_o/t ratios of 16.5, 31.0 and 60.0 submitted to cyclic bending were experimentally studied. Related experimental tests were conducted by using the tube-bending machine and curvature-ovalization measurement apparatus. The quantities of bending moment, curvature and ovalization were measured by sensors on testing facilities. Additionally, the number of bending cycles required to produce failure was also recorded.

II. EXPERIMENTS

A tube-bending machine and a curvature-ovalization measurement apparatus were employed to conduct the cyclic bending test on local sharp-cut circular tubes with different D_o/t ratios. The details on the experimental device, materials, specimens, and test procedures were presented in the sections that follow.

Experimental Device

Fig. 2 schematically shows the experiments executed by a specially built tube-bending machine. This facility was set up to

conduct monotonic, reverse, and cyclic bending tests. A detailed explanation of the experimental facility can be found in many papers (Pan *et al.* [20], Pan and Fan [21], Pan and Her [22], Lee *et al.* [23]). Pan *et al.* [20] designed a new light-weight apparatus to measure the curvature and the ovalization of the tube cross section as shown in Fig. 3. Two side-inclinometers in the apparatus were used to detect the tube's angle variation during cyclic bending. The tube curvature can be determined by a simple calculation according to the angle changes. An extended version of the calculation can be found in the work by Pan *et al.* [20]. In addition, the tube ovalization can be measured in the center part of the apparatus that included a magnetic detector and a magnetic block.

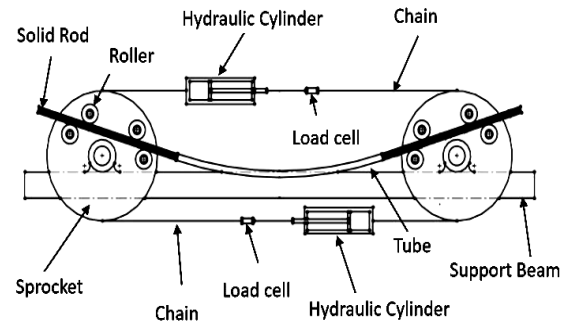


Fig. 2 A schematic drawing of the tube bending machine

Material and Specimens

6061-T6 aluminum alloy tubes were used for the experimental testing. Table 1 depicts the chemical composition (weight %) of the tested material. The mechanical properties were 0.2% offset yield stress (σ_o) = 166 MPa, ultimate stress (σ_u) = 258 MPa, and percent elongation = 23 %.

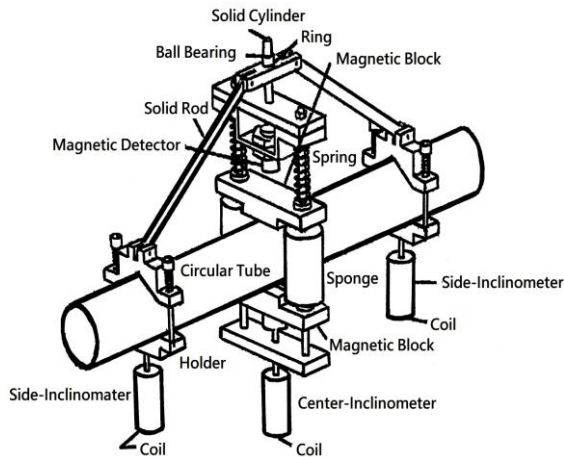


Fig. 3 A schematic drawing of the curvature-ovalization measurement apparatus

Table 1 Chemical composition of 6061-T6 aluminum alloy (weight %)

Chemical Composition	Al	Mg	Si	Cu	Ti	Fe
Proportion (%)	97.40	0.916	0.733	0.293	0.268	0.256
Chemical Composition	Mn	Zn	Cr	Ni	Pb	Sn
Proportion (%)	0.132	0.098	0.068	0.006	0.005	<0.001

The raw, unnotched 6061-T6 aluminum alloy circular tubes had a length L_o of 800 mm, an outside diameter D_o of 35.0 mm and a wall thickness t of 3.0 mm. The raw tubes were machined on the outside surface to obtain the desired D_o/t ratios of 16.5, 31.0 and 60.0 as shown in Fig. 4. However, the inner radiuses of all tested tubes were intact (29.0 mm).

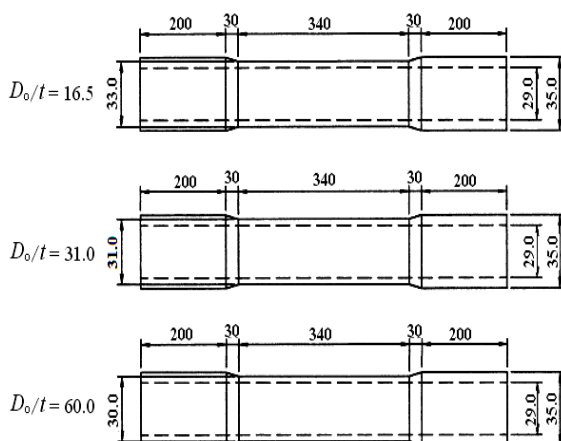


Fig. 4 A schematic drawing of tube's dimensions for D_o/t ratios of 16.5, 31.0 and 60.0

Next, tubes with a certain D_o/t ratio were

processed on the outside surface again to obtain the desired shape and depth of the cut. Fig. 5 shows a schematic drawing of the local sharp-cut circular tube, where the cut depth is denoted as a and the cut length is denoted as L_a . The relationship between a and L_a was determined by Lee *et al.* [34]. In this study, five different depth-to-thickness (a/t) ratios were considered: 0.0, 0.15, 0.3, 0.45, and 0.6. Note that $a/t = 0.0$ represents a tube with a smooth surface. It can be seen from Fig. 4 that the values of t for D_o/t ratios of 16.5, 31.0 and 60.0 are 2.0, 1.0 and 0.5 mm, respectively. Therefore, the magnitudes of a for $D_o/t = 16.5$, 31.0 and 60.0 are shown in Table 2. Fig. 6 shows a picture a local sharp-cut 6061-T6 aluminum alloy tube with $D_o/t = 16.5$ and $a = 0.0, 0.3, 0.6, 0.9$ and 1.2 mm. In addition, according to our measurement of a before the test, the error was within 10%.

Table 2 Magnitudes of a for $D_o/t = 16.5, 31.0$ and 60.0 (unit: mm)

D_o/t	a	a	a	a	a
16.5	0.0	0.3	0.6	0.9	1.2
31.0	0.0	0.15	0.3	0.45	0.6
60.0	0.0	0.075	0.15	0.225	0.3

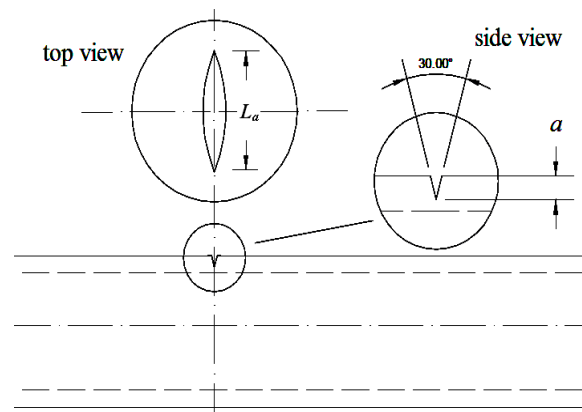


Fig. 5 A schematic drawing of the local sharp-cut circular tube



Fig. 6 A picture a local sharp-cut 6061-T6 aluminum alloy tube with $D_o/t = 16.5$ and $a = 0.0, 0.3, 0.6, 0.9$ and 1.2 mm

Test Procedures

The test involved a curvature-controlled cyclic bending. The controlled-curvature ranges were from ± 0.2 to ± 1.1 m^{-1} , and the curvature rate of the cyclic bending test was 0.035 $\text{m}^{-1}\text{s}^{-1}$. The magnitude of the bending moment was measured by two load cells mounted to the bending device. The magnitude of the curvature and ovalization of the tube cross-section were controlled and measured by the curvature-ovalization measurement apparatus. In addition, the number of cycles required to produce failure was recorded. The timing of failure was the magnitude of the bending moment dropped 20%.

III. EXPERIMENT RESULTS AND DISCUSSION

Mechanical Behavior

Fig. 7 shows a typical set of experimentally determined cyclic moment (M) - curvature (κ) curves for a local sharp-cut 6061-T6 aluminum alloy circular tube, with $D_o/t = 16.5$ and $a = 1.2$ mm, subjected to cyclic bending. The tubes were cycled between $\kappa = \pm 0.3$ m^{-1} . It was observed that the M - κ relationship was linear for a small curvature. However, it became nonlinear for a large curvature. In addition, the M - κ response was seen to be characterized by a nearly closed and steady hysteresis loop from the first bending cycle. Since the sharp cut is small and local, the cut depth has almost no influence on the M - κ

curve. Therefore, the M - κ curves for $D_o/t = 16.5$ and different a are not shown in this paper.

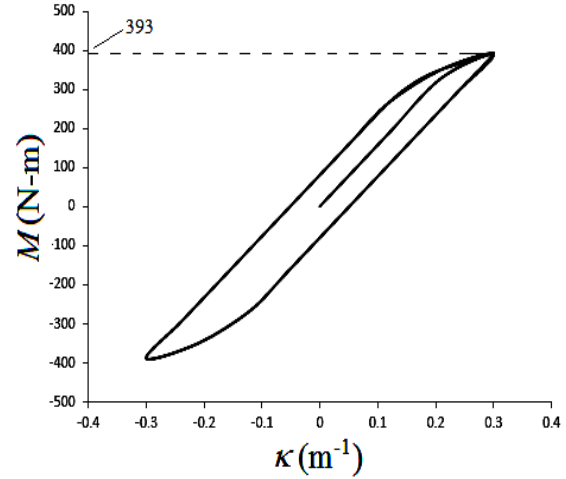


Fig. 7 Experimental moment (M) - curvature (κ) curve for a local sharp-cut 6061-T6 aluminum alloy tube with $D_o/t = 16.5$ and $a = 1.2$ mm under cyclic bending

Figs. 8 and 9 demonstrate typical sets of experimentally determined cyclic moment (M) - curvature (κ) curves for local sharp-cut 6061-T6 aluminum alloy circular tubes, with $D_o/t = 31.0$, $a = 0.6$ mm and $D_o/t = 60.0$, $a = 0.3$ mm, respectively, subjected to cyclic bending. The tubes were cycled between $\kappa = \pm 0.3$ m^{-1} . Similar phenomenon of the M - κ curves with Fig. 7 was found. Because the thicknesses for $D_o/t = 31.0$ and 60.0 are smaller, so the bending moments at the $\kappa = + 0.3$ m^{-1} are smaller. Again, the sharp cut is small and local, the cut depth has almost no influence on the M - κ curve. Therefore, the M - κ curves for $D_o/t = 31.0$ and 60.0 and different a are not shown in this paper.

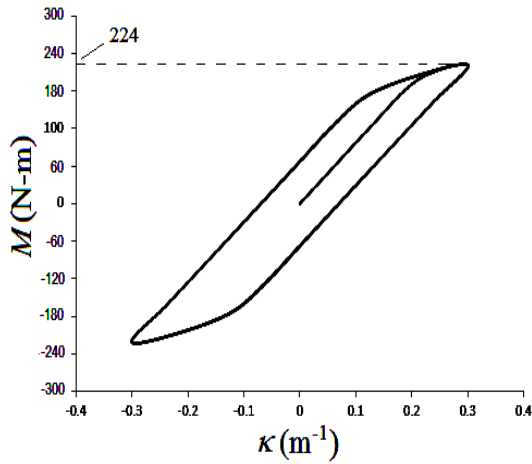


Fig. 8 Experimental moment (M) - curvature (κ) curve for a local sharp-cut 6061-T6 aluminum alloy tube with $D_o/t = 31.0$ and $a = 0.6$ mm under cyclic bending

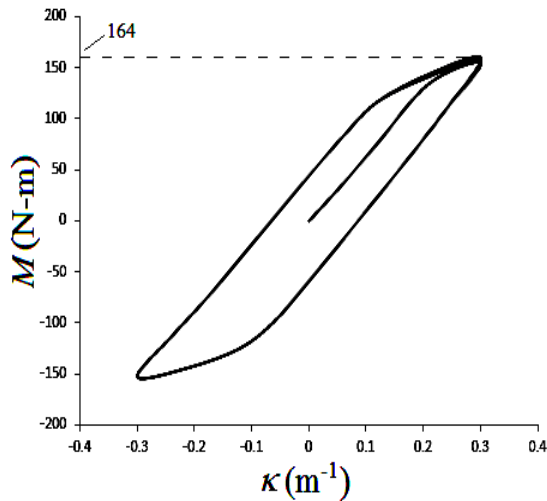


Fig. 9 Experimental moment (M) - curvature (κ) curve for a local sharp-cut 6061-T6 aluminum alloy tube with $D_o/t = 60.0$ and $a = 0.3$ mm under cyclic bending

Figs. 10(a)-(e) depict the experimental relationships between the ovalization ($\Delta D_o/D_o$) and curvature (κ) for local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t = 16.5$ and $a = 0.0, 0.3, 0.6, 0.9$, and 1.2 mm, respectively, under cyclic bending. The quantity of ΔD_o is the change in the outer diameter. The $\Delta D_o/D_o$ - κ relationships exhibited a ratcheting and an increasing trend with the number of bending cycles. A larger a led to a more asymmetrical appearance of the $\Delta D_o/D_o$ - κ relationship. Moreover, a larger a of the cut tubes caused a larger ovalization.

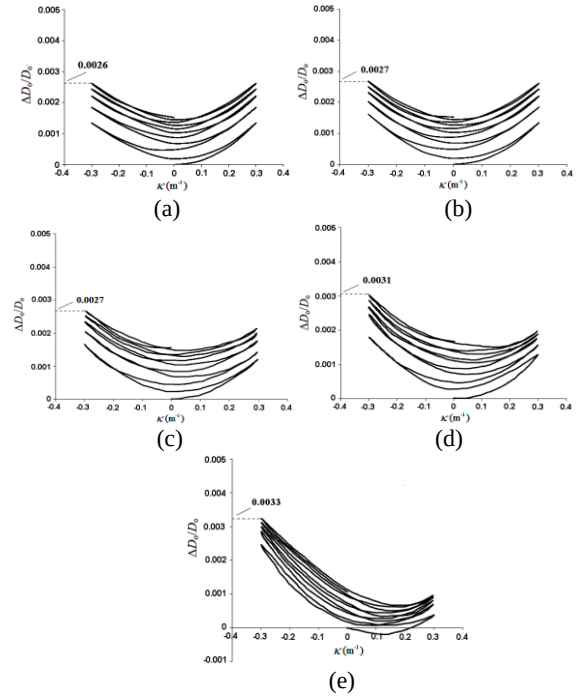
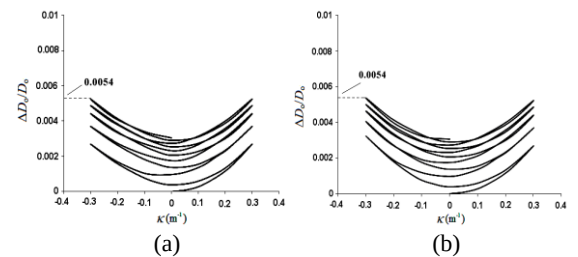


Fig. 10 Experimental ovalization ($\Delta D_o/D_o$) - curvature (κ) curves for local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t = 16.5$ and $a =$ (a) 0.0, (b) 0.3, (c) 0.6, (d) 0.9 and (e) 1.2 mm under cyclic bending

Figs. 11(a)-(e) depict the experimental relationships between the ovalization ($\Delta D_o/D_o$) and the curvature (κ) for local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t = 31.0$ and $a = 0.0, 0.15, 0.3, 0.45$, and 0.6 mm respectively, under cyclic bending. Figs. 12(a)-(e) demonstrate the experimental relationships between the ovalization ($\Delta D_o/D_o$) and the curvature (κ) for local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t = 60.0$ and $a = 0.0, 0.075, 0.15, 0.225$, and 0.3 mm, respectively, under cyclic bending. Similar phenomena with local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t = 16.5$ were found. In addition, a larger D_o/t ratio has a smaller wall thickness, thus, the ovalization increases faster.



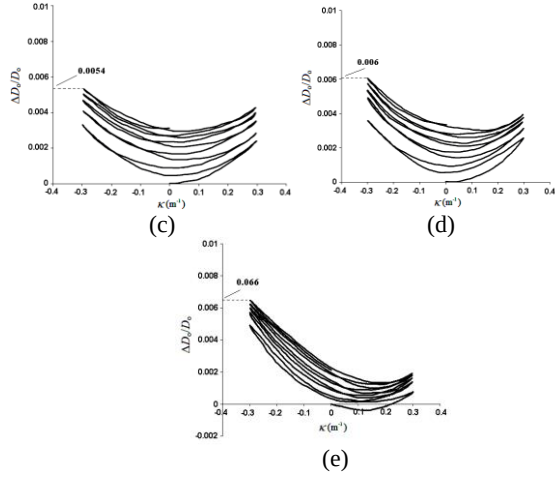


Fig. 11 Experimental ovalization ($\Delta D_0/D_0$) - curvature (κ) curves for local sharp-cut 6061-T6 aluminum alloy tubes with $D_0/t = 31.0$ and $a =$ (a) 0.0, (b) 0.15, (c) 0.3, (d) 0.45 and (e) 0.6 mm under cyclic bending

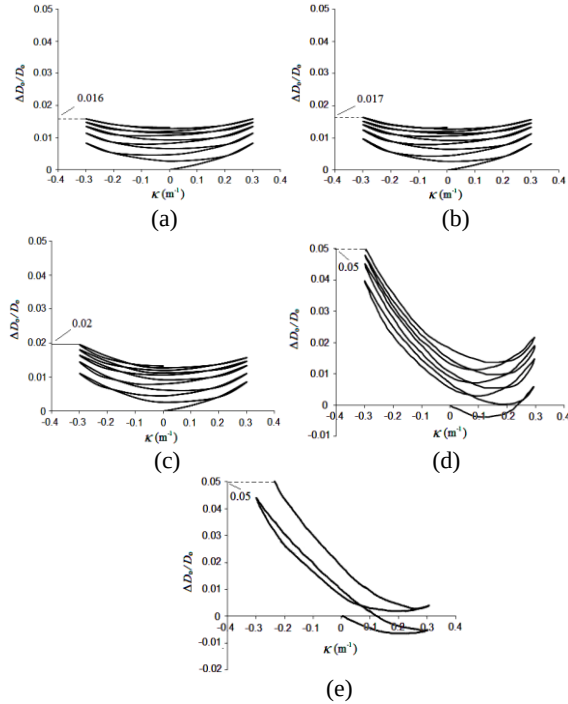


Fig. 12 Experimental ovalization ($\Delta D_0/D_0$) - curvature (κ) curves for local sharp-cut 6061-T6 aluminum alloy tubes with $D_0/t = 60.0$ and $a =$ (a) 0.0, (b) 0.075, (c) 0.15, (d) 0.225 and (e) 0.3 mm under cyclic bending

Failure

Figs. 13(a)-(c) present the experimental data of the cyclic controlled curvature (κ_c/κ_0) versus the number of bending cycles required to produce failure (N_f) for local-cut 6061-T6

aluminum alloy tubes with $D_0/t = 16.5$, 31.0 and 60.0, respectively, under cyclic bending. The controlled curvature was normalized by $\kappa_0 = t/D_0^2$ [2]. For a certain κ_c/κ_0 and a , tubes with a larger D_0/t ratios led to a lower N_f . In addition, for a certain D_0/t ratio and κ_c/κ_0 , tubes with a larger a led to a lower N_f .

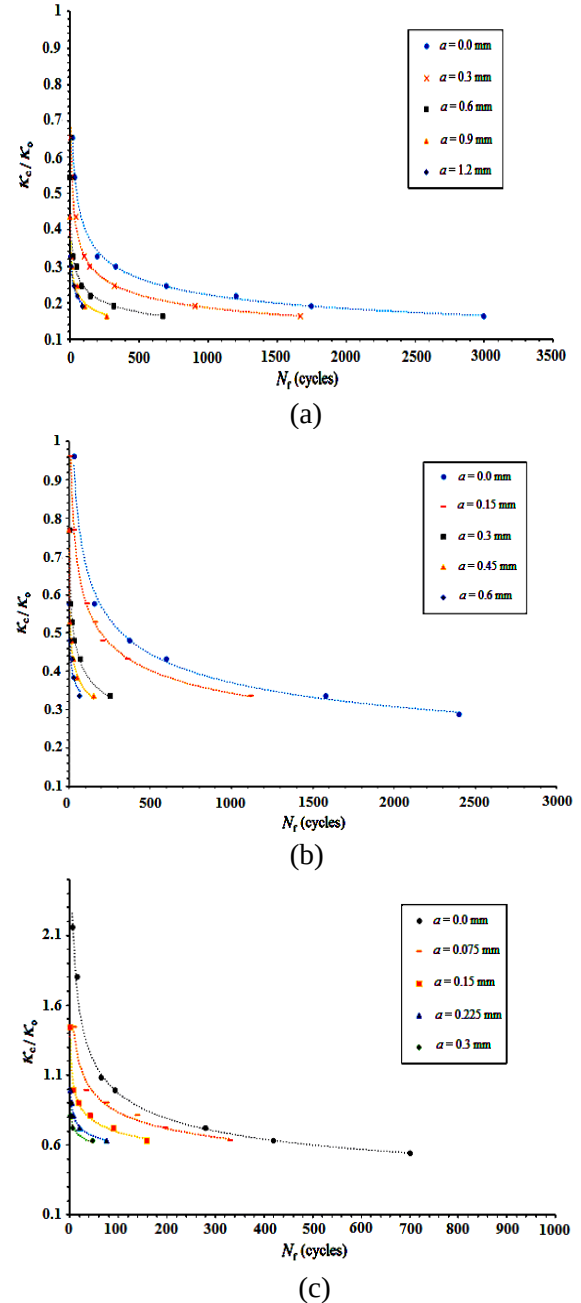
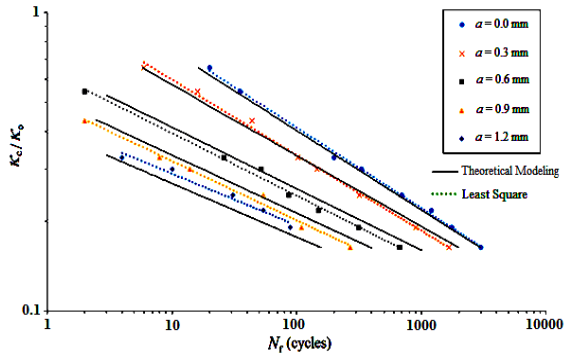
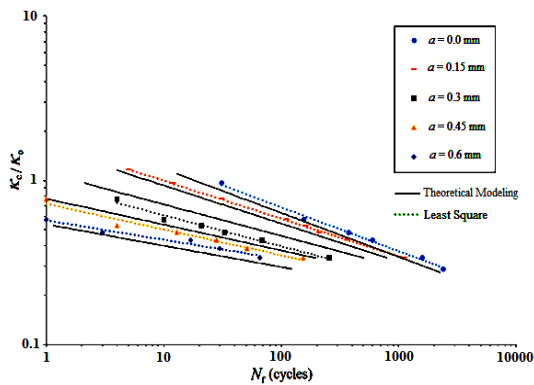


Fig. 13 Experimental controlled curvature (κ_c/κ_0) - number of bending cycles required to produce failure (N_f) curves for local sharp-cut 6061-T6 aluminum alloy tubes with $D_0/t =$ (a) 16.5, (b) 31.0 and (c) 60.0 under cyclic bending

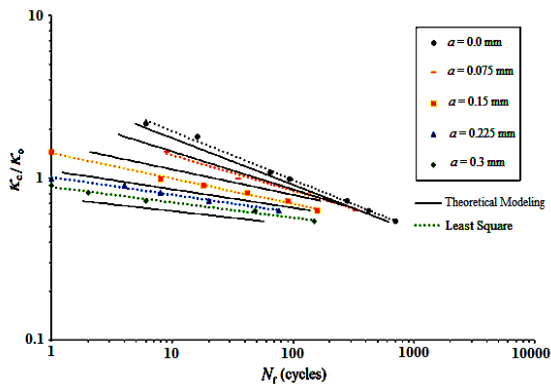
Subsequently, the experimental data in Figs. 13(a)-(c) were plotted on a log-log scale, and five straight dot lines were observed in Figs. 14(a)-(c). Note that the dot lines were determined by the least square method. It can be seen that five unparallel dot lines corresponding to five different a for any D_o/t ratio.



(a)



(b)



(c)

Fig. 14 Experimental and modeled controlled curvature (κ_c/κ_0) - number of bending cycles required to produce failure (N_f) curves for local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t =$ (a) 16.5, (b) 31.0 and (c) 60.0 under cyclic bending on a log-log scale

In 1987, Kyriakides and Shaw [2] proposed an empirical formulation to describe the relationship between κ_c/κ_0 and N_f to be:

$$\kappa_c/\kappa_0 = C (N_f)^{-\alpha} \quad (1)$$

or

$$\log \kappa_c/\kappa_0 = \log C - \alpha \log N_f \quad (2)$$

where C and α are the material parameters, which are related to the material and D_o/t ratios. The parameter C is the amount of κ_c/κ_0 when $N_f = 1$, and the parameter α is the slope of the log-log plot. The formulation has been successfully simulated the κ_c/κ_0 - N_f relationships for 6061-T6 aluminum alloy and 1018 carbon steel smooth tubes subjected to cyclic bending [2]. In this study, the liner relationships between $\ln C$ and a/t and $\ln \alpha$ and a/t were found in Figs. 15(a)-(c) and Fig. 16(a)-(c). Therefore, we write

$$\ln C = C_0 - \beta(a/t) \quad (3)$$

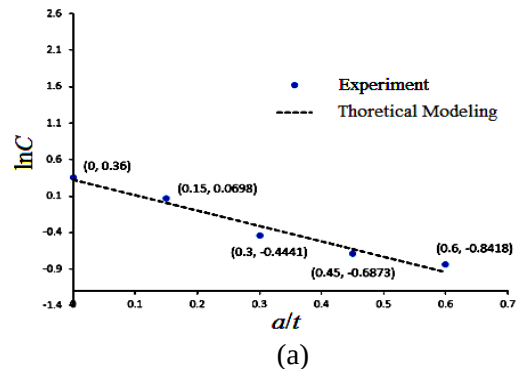
and

$$\ln \alpha = \alpha_0 - \gamma(a/t) \quad (4)$$

where C_0 , β , α_0 and γ are material parameters. Due to the linear relationship, the quantities of C_0 and β were the intercepts and slopes in Figs. 15(a)-15(c) for $D_o/t = 16.5$, 31.0 and 60.0, respectively, and quantities of α_0 and γ were the intercepts and slopes in Figs. 16(a)-16(c) for $D_o/t = 16.5$, 31.0 and 60.0, respectively. Table 3 show the magnitudes of C_0 , β , α_0 and γ for $D_o/t = 16.5$, 31.0 and 60.0.

Table 3 Magnitudes of C_0 , β , α_0 and γ for $D_o/t = 16.5$, 31.0 and 60.0

D_o/t	C_0	β	α_0	γ
16.5	0.356	2.083	-1.321	0.717
31.0	0.849	2.667	-1.332	1.167
60.0	1.258	2.495	-1.227	2.167



(a)

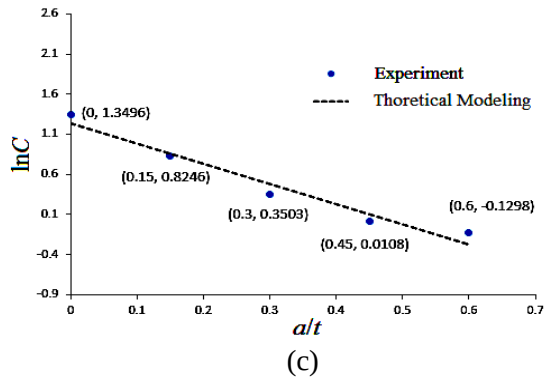
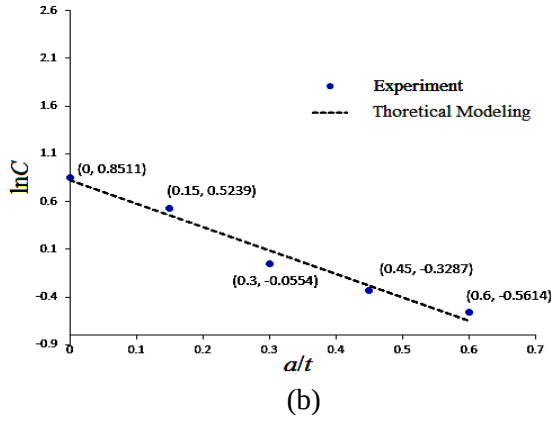


Fig. 15 The relationship between $\ln C$ and a/t for $D_o/t =$ (a) 16.5, (b) 31.0 and (c) 60

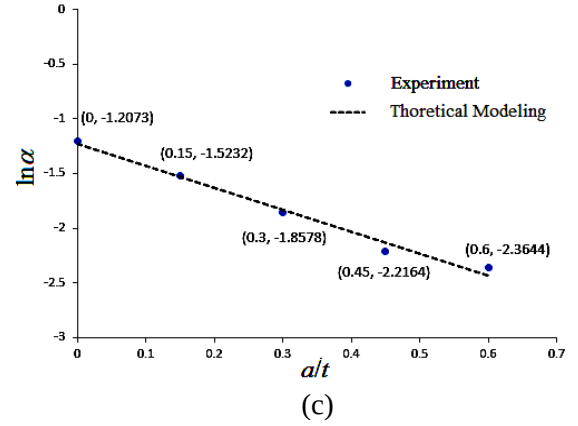
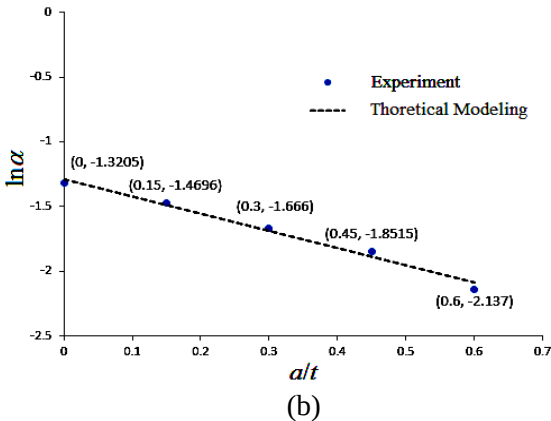
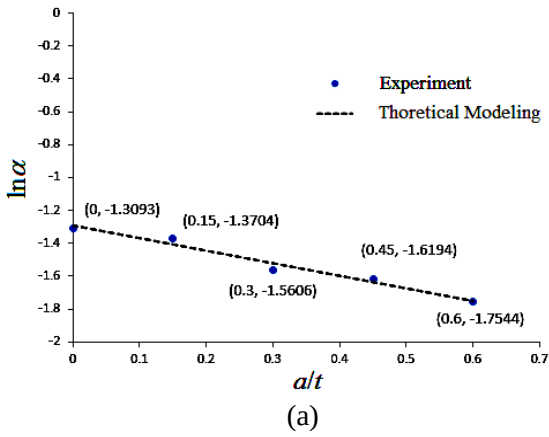


Fig. 16 The relationship between $\ln \alpha$ and a/t for $D_o/t =$ (a) 16.5, (b) 31.0 and (c) 60.0

Next, we assumed that the material parameters C_o , β , α_o and γ were related to the D_o/t ratios. In this study, we tried to find the linear relationships. After a lot of attempts, the linear relationships for C_o and D_o/t ratios, β and D_o/t ratios, α_o and D_o/t ratios, γ and D_o/t ratios were obtained in Figs. 17(a)-(d), respectively. Thus, the linear relationships were written as:

$$C_o^2 = a_1(D_o/t) + a_2 \quad (5)$$

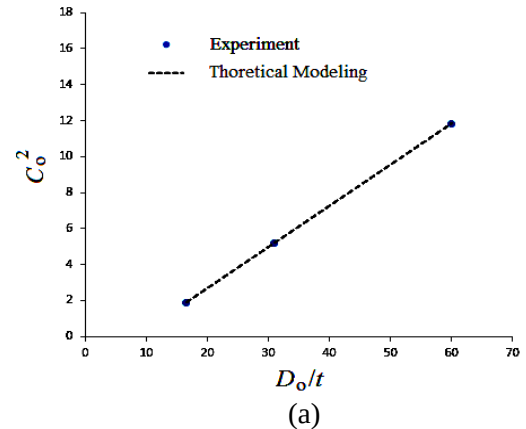
$$\beta^2(D_o/t) = b_1(D_o/t) + b_2 \quad (6)$$

$$\alpha_o^2(D_o/t) = c_1(D_o/t) + c_2 \quad (7)$$

and

$$1/\gamma = d_1/(D_o/t) + d_2 \quad (8)$$

where a_1 , a_2 , b_1 , b_2 , c_1 , c_2 , d_1 and d_2 are material constants. The magnitudes of a_1 and a_2 were the slope and intercept in Fig 17(a), respectively, amounts of b_1 and b_2 were the slope and intercept in Fig 17(b), respectively, values of c_1 and c_2 were the slope and intercept in Fig 17(c), respectively, and magnitudes of d_1 and d_2 were



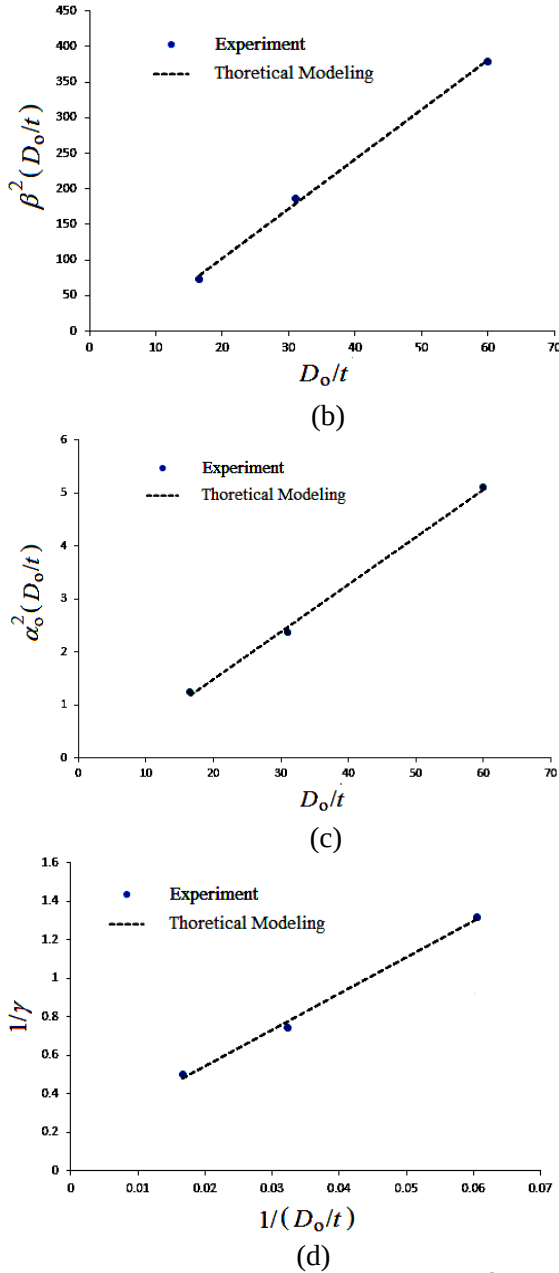


Fig. 17 (a) The relationship between C_0^2 and D_o/t , (b) the relationship between $\beta^2(D_o/t)$ and D_o/t , (c) the relationship between $\alpha_o^2(D_o/t)$ and D_o/t , and (d) the relationship between $1/\gamma$ and $1/(D_o/t)$

the slope and intercept in Fig 17(d), respectively. These material constants were determined to be 0.228, -1.781, 7.011, -40.661, 0.089, -0.191, 0.021 and -0.062, respectively. In addition, Eqs. (5)-(8) are applicable in $16.5 \leq D_o/t \leq 60.0$.

Finally, Eqs. (2)-(8) and the material parameters had been obtained were used to simulate the experimental data in Figs. 14(a)-(c).

The modeled results of the relationship between the controlled curvature (κ_c/κ_o) and number of bending cycles required to produce failure (N_f) curves for local sharp-cut 6061-T6 aluminum alloy tubes with $D_o/t = 16.5, 31.0$ and 60.0 , respectively, under cyclic bending on a log-log scale are depict in Figs. 14(a)-(c) in solid lines. Good agreement between the experimental and simulated results has been achieved.

V. CONCLUSIONS

This study investigated the response and failure of the local sharp-cut 6061-T6 aluminum alloy tubes with different D_o/t ratios submitted to cyclic bending. Some important conclusions are sorted as follows according to the experimental and simulated results:

- (1) The experimental $M-\kappa$ relationship for the local sharp-cut 6061-T6 aluminum alloy tubes with any a or D_o/t ratio displayed a closed hysteresis loop from the first bending cycle. Since the sharp cut was small and local, the cut depth had almost no influence on the $M-\kappa$ curve for a certain D_o/t ratio.
- (2) The experimental $\Delta D_o/D_o-\kappa$ relationship for local sharp-cut 6061-T6 aluminum alloy tubes with any D_o/t ratio or a revealed an increasing and ratcheting trend with the number of bending cycles. The $\Delta D_o/D_o-\kappa$ relationships were symmetrical for $a = 0.0$ mm, but asymmetrical for $a \neq 0.0$ mm. In addition, the tubes with a larger D_o/t or a led to more asymmetrical trend and a larger ovalization.
- (3) The empirical formulation of Eq. (2) proposed by Kyriakides and Shaw [2] was modified to simulate the κ/κ_o-N_f relationship for the local sharp-cut 6061-T6 aluminum alloy tubes with different D_o/t ratios submitted to cyclic bending. According to the experimental data, the forms of the material parameters, C and α , were proposed in Eqs. (3) and (4), respectively. In addition, The forms of the material parameters, C_o , β , α_o and γ , were proposed in Eqs. (5)-(8), respectively. The results simulated by Eqs. (2)-(8) were in good agreement with the experimental findings (Figs. 14(a)-(c)).

V. REFERENCES

- [1] Shaw, P. K. and Kyriakides, S., "Inelastic Analysis of Thin-walled Tubes under Cyclic Bending," *Int. J. Solids Struct.*, Vol. 21, No. 11, pp. 1073-1110, 1985.
- [2] Kyriakides, S. and Shaw, P. K., "Inelastic Buckling of Tubes under Cyclic Loads," *ASME J. Pre. Ves. Tech.*, Vol. 109, No. 2, pp. 169-178, 1987.
- [3] Corona, E. and Kyriakides, S., "An Experimental Investigation of the Degradation and Buckling of Circular Tubes under Cyclic Bending and External Pressure," *Thin-Walled Struct.*, Vol. 12, Is. 3, pp. 229-263, 1991.
- [4] Vaze, S. and Corona, E., "Degradation and Collapse of Square Tubes under Cyclic Bending," *Thin-Walled Struct.*, Vol. 31, No. 4, pp. 325-341, 1998.
- [5] Corona, E. and Kyriakides, S., "Asymmetric Collapse Modes of Pipes under Combined Bending and Pressure," *J. Eng. Mech.*, Vol. 126, No. 12, pp. 1232-1239, 2000.
- [6] Corona, E., Lee, L. H. and Kyriakides, S., "Yield Anisotropic Effects on Buckling of Circular Tubes under Bending," *Int. J. Solids Struct.*, Vol. 43, No. 22, pp. 7099-7118, 2006.
- [7] Limam, A., Lee, L. H., Corana, E. and Kyriakides, S., "Inelastic Wrinkling and Collapse of Tubes under Combined Bending and Internal Pressure," *Int. J. Mech. Sci.*, Vol. 52, Is. 5, pp. 637-647, 2010.
- [8] Limam, A., Lee, L. H. and Kyriakides, S., "On the Collapse of Dented Tubes under Combined Bending and Internal Pressure," *Int. J. Mech. Sci.*, Vol. 55, Is. 1, pp. 1-12, 2012.
- [9] Bechle, N. J. and Kyriakides, S., "Localization of NiTi Tubes under Bending," *Int. J. Solids Struct.*, Vol. 51, No. 5, pp. 967-980, 2014.
- [10] Jiang, D., Kyriakides, S., Bechle, N. J. and Landis, C. M., "Bending of Pseudo Elastic NiTi Tubes," *Int. J. Solids Struct.*, Vol. 124, pp. 192-214, 2017.
- [11] Yuan, W. and Mirmiran, A., "Buckling Analysis of Concrete-filled FRP Tubes," *Int. J. Struct. Stab. Dyn.*, Vol. 1, No. 3, pp. 367-383, 2001.
- [12] Elchalakani, M., Zhao, X. L. and Grzebieta, R. H., "Plastic Mechanism Analysis of Circular Tubes under Pure Bending," *Int. J. Mech. Sci.*, Vol. 44, No. 6, 1117-1143, 2002.
- [13] Jiao, H. and Zhao, X. L., "Section Slenderness Limits of Very High Strength Circular Steel Tubes in Bending," *Thin-Walled Struct.*, Vol. 42, No. 9, pp. 1257-1271, 2004.
- [14] Houliara, S. and Karamanos, S. A., "Buckling and Post-buckling of Long Pressurized Elastic Thin-walled Tubes under In-plane Bending," *Int. J. Nonlinear Mech.*, Vol. 41, No. 4, pp. 491- 511, 2006.
- [15] Elchalakani, M., Zhao, X. L. and Grzebieta, R. H., "Variable Amplitude Cyclic Pure Bending Tests to Determine Fully Ductile Section Slenderness Limits for Cold-formed CHS," *Eng. Struct.*, Vol. 28, No. 9, pp. 1223-1235, 2006.
- [16] Yazdani, H. and Nayebi, A., "Continuum Damage Mechanics Analysis of Thin-walled Tube under Cyclic Bending and Internal Constant Pressure," *Int. J. Appl. Mech.*, Vol. 5, No. 4, 1350038 [20 pages], 2013.
- [17] Fan, H. G., Chen, Z. P., Feng, W. Z., Zhou, F. and Cao, G. W., "Dynamic Buckling of Cylindrical Shells with Arbitrary Axisymmetric Thickness Vibration under Time Dependent External Pressure," *Int. J. Struct. Stab. Dyn.*, Vol. 15, No. 3, 1450053 [21 pages], 2015.
- [18] Elchalakani, M., Karrech, A. and Hassanein, M. F. and Yang, B., "Plastic and Yield Slenderness Limits for Circular Concrete Filled Tubes Subjected to Static Pure Bending," *Thin-Walled Struct.*, Vol. 109, pp. 50-64, 2016.
- [19] Shamass, R., Alfano, G. and Guarracino, F., "On Elastoplastic Buckling Analysis of Cylinders under Nonproportional Loading by Differential Quadrature Method," *Int. J. Struct. Stab. Dyn.*, Vol. 17, No. 7, 1750072 [40 pages], 2017.
- [20] Pan, W. F., Wang, T. R. and Hsu, C. M., "A Curvature-Ovalization Measurement Apparatus for Circular Tubes under Cyclic Bending," *Exp. Mech.*, Vol. 38, No. 2, pp.

- 99-102, 1998.
- [21] Pan, W. F. and Fan, C. H., "An Experimental Study on the Effect of Curvature-rate at Preloading Stage on Subsequent Creep or Relaxation of Thin-walled Tubes under Pure Bending," *JSME Int. J., Ser. A*, Vol. 41, No. 4, pp. 525-531, 1998.
- [22] Pan, W. F. and Her, Y. S., "Viscoplastic Collapse of Thin-walled Tubes under Cyclic Bending," *ASME J. Eng. Mat. Tech.*, Vol. 120, No. 4, pp. 287-290, 1998.
- [23] Lee, K. L., Pan, W. F. and Kuo, J. N., "The Influence of the Diameter-to-thickness Ratio on the Stability of Circular Tubes under Cyclic Bending," *Int. J. Solids Struct.*, Vol. 38, No. 14, pp. 2401-2413, 2001.
- [24] Lee, K. L., Pan, W. F. and Hsu, C. M., "Experimental and Theoretical Evaluations of the Effect between Diameter-to-thickness Ratio and Curvature-rate on the Stability of Circular Tubes under Cyclic Bending," *JSME Int. J., Ser. A*, Vol. 47, No. 2, pp. 212-222, 2004.
- [25] Chang, K. H., Pan, W. F. and Lee, K. L., "Mean Moment Effect of Thin-walled Tubes under Cyclic Bending," *Struct. Eng. Mech.*, Vol. 28, No. 5, pp. 495-514, 2008.
- [26] Chang, K. H. and Pan, W. F., "Buckling Life Estimation of Circular Tubes under Cyclic Bending," *Int. J. Solids Struct.*, Vol. 46, No. 2, pp. 254-270, 2009.
- [27] Lee, K. L., Hung, C. Y. and Pan, W. F., "Variation of Ovalization for Sharp-notched Circular Tubes under Cyclic Bending," *J. Mech.*, Vol. 26, No. 3, pp. 403-411, 2010.
- [28] Lee, K. L., "Mechanical Behavior and Buckling Failure of Sharp-notched Circular Tubes under Cyclic Bending," *Struct. Eng. Mech.*, Vol. 34, No. 3, pp. 367-376, 2010.
- [29] 洪兆宇、李國龍和潘文峰，"不同尖銳凹槽深度 304 不銹鋼管在循環彎曲負載下黏塑性行為之實驗分析"，*技術學刊*，第 26 卷，第 4 期，235-242 頁，2011。
- [30] Lee, K. L., Hsu, C. M. and Pan, W. F., "Viscoplastic Collapse of Sharp-notched Circular Tubes under Cyclic Bending," *Acta Mech. Solida Sinica*, Vol. 22, No. 6, pp. 629-641, 2013.
- [31] Lee, K. L., Hsu, C. M. and Pan, W. F., "Response of Sharp-notched Circular Tubes under Bending Creep and Relaxation," *Mech. Eng. J.*, Vol. 1, No. 2, pp. 1-14, 2014.
- [32] Chang, K. H., Lee, K. L. and Pan, W. F., "Buckling Failure of 310 Stainless Steel Tubes with Different Diameter-to-thickness Ratios under Cyclic Bending," *Steel Comp. Struct.*, Vol. 10, No. 3, pp. 224-245, 2010.
- [33] Lee, K. L., Hsu, C. M. and Pan, W. F., "The Influence of the Diameter-to-thickness Ratio on the Response and Collapse of Sharp-notched Circular Tubes under Cyclic Bending," *J. Mech.*, Vol. 28, No. 3, pp. 461-468, 2012.
- [34] Lee, K. L., Chern, R. T. and Pan, W. F., "The Response of 6061-T6 Aluminum Alloy Tubes with Different Chop Depths under Cyclic Bending," *J. Chung Cheng Inst. Tech.*, Vol. 45, No. 1, pp. 19-29, 2016.

Lee et al.

The Response and Failure of Local Sharp-cut Circular Tubes with Different Diameter-to-thickness Ratios Submitted to Cyclic Bending

國軍接艦訓練成效之研究

蒲澤春^{1*} 盧文民² 徐祥禎¹

¹義守大學工業管理學系

²國防大學財務管理學系

摘 要

本文主要目的為探討艦船移轉任務官、士、兵參加美軍流路訓練與基地在職訓練之滿意度與學習成效之差異分析。實證結果發現基地在職訓練滿意度與學習成效均優於流路訓練，尤其在場所與設備、成果獎賞、生活津貼等項目，則是顯示基地在職訓練顯著高於流路訓練。此外，職務不同對訓練之滿意度與學習成效皆呈現顯著差異。此結果建議管理單位未來在規劃接艦訓練課程時，可以適度調整兩類訓練流路訓練時數，著重於裝備實際操作與維護，以提升訓練滿意度與學習成效；在遴選受訓人員時，除了必要專業人員外亦可遴選無接艦經驗人員，以擴大訓練成效，廣儲人才。

關鍵詞：艦船移轉，流路訓練，基地在職訓練，訓練滿意度，學習成效

Effectiveness of Republic of China (Taiwan) Navy's Training for Naval Vessel Transfer

Tse-Chun Pu ^{1*}, Wen-Min Lu ², and Hsiang-Chen Hsu ²

¹ Department of Industrial Management, I-SHOU University

² Department of Financial Management, National Defense University

ABSTRACT

This paper investigated and analyzed differences in the satisfaction and effectiveness of training commissioned officers, noncommissioned officers, and sailors who received pipeline training and on-the-job training (OJT) provided by active duty and retired military from the United States. The empirical results revealed that the satisfaction and effectiveness of OJT training were higher than those of pipeline training. Factors such as location and equipment, recognition, and rewards, and per diem allowances, make OJT training significantly superior to pipeline training in terms of effectiveness, and satisfaction of the trainees. In addition, there was significant difference in training satisfaction and effectiveness in specific jobs. The findings can be used to develop recommendations for supervisory staff to adjust the training hours for both types of training when planning future naval vessel transfers. Emphasis should be placed on practical operation and equipment maintenance to enhance training effectiveness and satisfaction. Management can use OJT training for inexperienced personnel in addition to experienced professionals to enlarge the pool of candidates for training, and improve overall training effectiveness.

Keywords: Naval vessel transfer, Pipeline training, OJT, Training satisfaction, Learning effectiveness

文稿收件日期 106.8.16; 文稿修正後接受日期 106.11.10; *通訊作者

Manuscript received August 16, 2017; revised November 10, 2017; * Corresponding author

一、前言

循軍事售予軍艦是國家對國家的實質軍事協助的一種行為，故國軍官兵赴國外接艦執行任務，責任重大，代表著國家軍人之能力，亦包括軍紀、管理與軍事專業(裝備之操作與保養、故障隔離與排除及戰術等相關之訓練)，故接艦工作執行之良窳，代表著國軍之教育、訓練之成效，更甚而可評估出其戰力之水準，故本研究從接艦訓練進行深入探討。

過去十餘年來，人力資源逐漸變成企業價值的核心理論 (Kim and Ployhart, 2014; Fulmer and Ployhart, 2014; Ployhart and Hale, 2014; Nyberg et al., 2014)。人力資本、策略、人力資源管理、創新及資本是組織成長的五個關鍵因素，此外組織如果投資特定的人力資本，會強化及建構組織的核心競爭力 (Demir et al., 2017)。而訓練是組織提升人力資源素質的重要方法 (Chatzimouratidis et al., 2012)，許多組織都希望提升的員工素質以強化其競爭能力，進而增進組織的經營績效 (Wright et al., 1994)。對員工個人而言，訓練後可以提高他的工作滿足感和信心，並且改善其工作績效，因此對組織而言，透過有效的員工訓練可增加組織的競爭力，並激勵員工來達成組織的營運目標 (Baldwin and Ford, 1988; Curry et al., 1994)。Dolezalek (2005) 企業或組織每年通常都要支出數十億美元以上的經費來從事員工的教育訓練，員工教育訓練費用的支出，已成為企業組織的一項成本負擔。美國訓練發展學會 (American Society of Training and Development, ASTD) 在 2002 年指出，美國企業 2001 年花費在員工教育訓練的費用大約是 156 億美元，此外，該學會在 2014 年調查 340 間美國企業，發現企業 2013 年花費在每位員工教育訓練的費用平均是 1,208 美元，而每位員工投入在教育訓練的時數平均為 31.5 小時；在國內部分，依據行政院勞動部 104 年公布之勞動統計通報指出，民國 99-103 年間勞動部辦理職業訓練結訓人數 (包含補助在職訓練) 平均每年約 115 萬人，其中補助在職訓練佔 93%，顯示政府及企業亦相當重視員工的教育訓練。

Georgenson (1982) 指出受訓者在受課程中所學，只有 10% 的能成功地移轉至工作環境。這意味著企業浪費了約 90% 的訓練費用，而 Lorge (1998) 也指出，美國麥當勞公司

估計，該公司在員工教育訓練後，大約只有 10% 的訓練成果能運用在實務工作上。因此對於訓練成效的評估，衡量訓練之後所能帶來的效益，以及是否能將所學應用在實務工作上，便成為企業組織相當重視的一項課題，也逐漸採取正式的訓練評估程序檢驗 (Noe, 2002)。Waldman and Balven (2015) 也指出訓練與發展、衡量及評估是未來組織在提升領導責任時應該深入研究及關切的重要課題。

近年我國積極推動國防自主，在海軍艦艇方面規劃以國艦國造為主，惟以往我國在考量國防財力與裝備籌購時效等因素，為快速、合理價格提升海軍戰備能力，多向美國循軍售採購艦艇，20 年來我國海軍歷經陽級、永陽級、中和級、旭海級、基隆級及獵雷艦、派里級艦等 20 餘艘美國海軍艦船移轉，有效提升我國海軍的戰力。因為美國對外軍售 (Foreign Military Sales, FMS) 移轉友邦海軍之艦船係屬國防剩餘物資 (Excessive Defense Article, EDA)，因此，美國售予我國多屬已計畫汰除熱艦 (Hot Ships) 或已除役封存冷艦 (Cold Ships)。就冷艦而言，即使艦齡僅 20 年(艦艇自除役前不執行深度維修工程，僅實施最低限度之維修工程)，但是經過除役多年閒置時的風雨摧殘，艦艇狀況多不佳，艦上裝備也經常短少，要使艦艇恢復以往作戰狀況必需選派人員赴美接收艦艇，藉由接艦訓練、廠修及艦力自修工程復原艦艇上的各項裝備及接受艦艇各項裝備的操作與保養訓練，完成啟封工程及性能提升等作業。艦艇完成啟封工程返國後，經過重點輔訓後，即可擔負起作戰任務，另於美國先行經過裝備保養、訓練、操作與維持等四階段(如圖 1)，其中裝備保養區分基本保養(O 級)、中繼維修(I 級)、廠級維修(D 級)等三個維修層級；訓練則包含基礎訓練 (Basic Training)、個人合格簽證 (Personnel Qualification standard)、值更訓練 (Watch Team Replacement Plan)、組合訓練 (Training Team Evolutions) 等四個類別；而維持則是指艦船依計畫維修保養制度 (Planned Maintenance System, PMS)，實施保養並定期進廠執行維修使各項裝備仍能具備(或超過)原設計之功能。

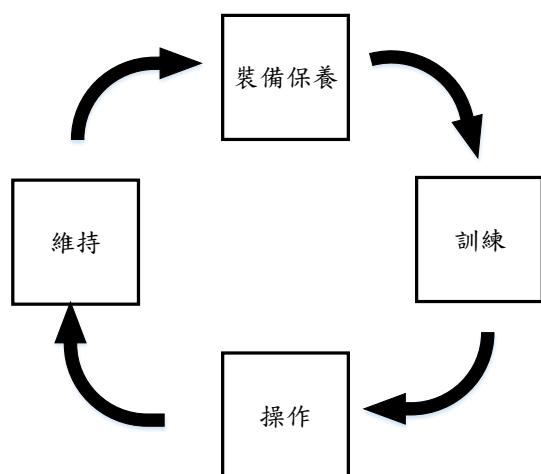


圖 1 軍艦完成戰備流程

每次接艦官兵的人數少則數十人(獵雷艦)，多則上千人(飛彈驅逐艦)，時間長達約 2~4 年，在美國期間，除了要接受訓練外，也要投入相當多的時間和人力在裝備的保養與維護上，在官兵夜以繼日的努力下，才能如期的將艦艇開回或塢運回國(獵雷艦)，納入戰鬥序列，但是國軍投入了鉅額的經費與人力後，國防部與官兵個人從訓練中所獲得的成效如何？則顯少有研究探討。因此本研究之目的有下列三個：

- (1)探討官兵赴美接艦接受不同訓練型態時對訓練滿意度與成效的差異分析。
- (2)瞭解不同職務的官兵對訓練滿意度與成效的差異分析。
- (3)瞭解人員參訓次數多寡對訓練滿意度與成效的差異分析。

二、文獻回顧

2.1 國艦換裝現況

赴美接艦視艦艇的噸位與類型，其作業的複雜度差異相當大，為了縮短人員赴美後的適應期及提升作業效率，大多會在國內先期加強語文訓練、專長訓練，甚至是到國內性質類似的現役艦艇實作。赴美後，美國對我執行接艦人員訓練的單位為美國海軍教育訓練技術移轉安全援助處(Naval Education Training Security Assistance Facility, NETSAF)，其會依照工程期間規劃出接艦人員之訓練期程，包括

美軍訓練系統中之流路班次及接艦基地之基礎理論、單機及系統之操作，保養與故障控制訓練。接艦官兵抵美後，分到美軍流路班次訓練或至接艦基地執行在職訓練，訓練類型、流程分別如圖 2、圖 3 所示。

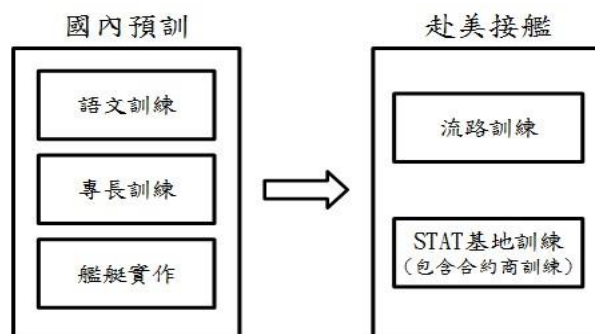


圖 2 赴美接艦訓練類型

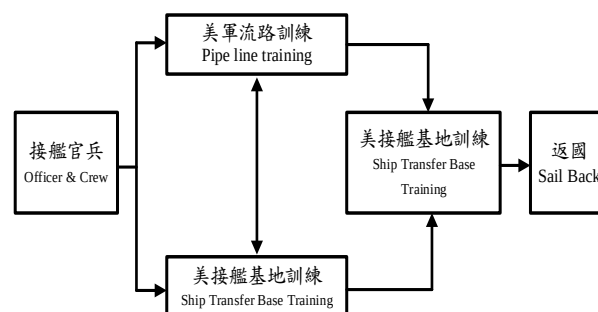


圖 3 赴美接艦訓練流程

在美國的接艦訓練課程主要區分為美軍基地流路訓練(如雷達、聲納、燃氣渦輪機輪機軍官班)、合約商基地訓練(如 GPS、導航雷達訓練)、美軍基地訓練及驗收(依專長分別對工作班授課)。整體而言，美軍基地流路訓練與基地訓練主要區分為裝備原理原則說明，及裝備實際操作練習，而艦艇實際練習裝備操作時，如有對裝備原理疑惑，授課教官仍會予以講解。因此，兩類課程並無技術等級差異，而是理論與實作的差別，流路訓練與基地訓練的差異如表 1。

表 1 美軍流路訓練與基地訓練差異

分	流路訓練	基地訓練
優點	1.較不受天候影響 2.可以快速對大量學習者傳播特定知識與技能	1.學員除對基礎理論學習外，亦可參與裝備之線路追蹤，維保及操作訓練 2.老師理論到實作親自指導 3.經由學員的互動學習與成長 4.訓練成本較低 5.受訓時數較長 6.官員可協助督導
缺點	1.依班次不同上課地點分散在美國各基地 2.以講師為中心多為單向傳播知識 3.學員通常是被動學習，較少機會參與討論 4.訓練成本較高 5.訓練時數較少	1.訓練壓力較大 2.體力負擔較重

2.2 相關訓練方式之文獻

2.2.1 訓練方式

針對不同的訓練目標，組織或團體可運用的訓練方式包含了體驗法、講授法、研討法、個案研究法、分組討論法、遊戲法等(Noe, 2002)。美國體驗教育學會 (Association for Experiential Education, AEE) 在 1995 年時提出，教育訓練依據訓練場所的不同一般可區分為室內講授訓練 (Indoor Lecture Training)，與戶外體驗訓練(Outdoor Experiential Learning)兩類：

- (1)室內講授訓練：係指在特定的教室對訓練人員，透過講師的課程講解與說明來傳授相關的知識與技巧；室內講授訓練是一種有效傳播知識與技能的方法，透過室內講授訓練課程，一方面藉由專業人員對知識與技能的傳授，能使受訓學員充分學習；另一方面，受訓學員可以在課程中主動提出問題，加深對團隊對課程中知識的認知，是比較常見的訓練方式。
- (2)戶外體驗訓練：係將訓練的地點移往實地

場所來進行，參與受訓的人員，透過直接體驗而建構知識以及獲得技能。企業經常運用團隊訓練課程來改善員工的團隊合作、團隊溝通等技能巧。Yamazaki and Kayes (2004) 研究認為外派人員的跨文化學習或適應能力與體驗學習方式有關；而 Kanters et al. (2002) 則實證研究發現，經過一天的戶外體驗訓練，能夠協助研究所學生克服求學壓力，有效降低他們的緊張、沮喪等情緒狀態。因此，戶外體驗訓練是強調「直接經驗 (Direct Experience)」，透過直接、實作的學習的過程。Molinsky (2015) 認為領導者必須讓員工有機會在不同文化環境實地練習、磨練技巧。把不同背景的員工聚集起來，一同執行長期專案，鼓勵資深員工輔導資淺員工，共同提升員工的技能。Hoover (2017) 則提出 TIPS (Teach, Illustrate, Practice, Simulate)的訓練模式，並認為透過 TIPS 訓練模式，可以使無經驗員工訓練成熟稔作業流程，而訓練時間可減少 35%，有效減少企業的訓練成本。而教學、示範兩個程序應不超過總訓練時數的 20~25%，練習、模擬兩個程序則不低於總訓練時數的 75~80%。

2.2.2 訓練成效評估

關於訓練成效的評估有相當多提出不同的評估模式，包含參加訓練人員的人格特質、在組織內表現及課程訓練的投資報酬為衡量標準等。Galvin (1983) 提出以背景、投入、過程及結果等四個構面的 CIPP(Context-Input-Process-Output) 模式來評估訓練的成效；Kirkpatrick (1979) 的反應、學習、行為、結果四個層次模式；Bushnell (1990) 提出以投入、過程及產出等構面為評估標準的 IPO(Input-Process-Output) 模式。其中 Kirkpatrick (1979)的四層次訓練評估模式最廣被接受與採用的模式(Arthur et al., 2003)。Kirkpatrick 將訓練成效評估依其評量準則區分為四個層次，層次一是反應(Reaction)，用以衡量員工對訓練的想法與感受；層次二是學習(Learning)，衡量員工從訓練中學到什麼；層次三是行為(Behavior)，衡量員工工作行為是否改變；層次四是結果(Results)，衡量訓練是否改變組織績效；上述四個層次是階層關係，每個訓練方案都應要產生這四層次的變化，缺一不可，故四層次都要衡量才能得知訓

練成效全貌(Kirkpatrick, 1998)。Leach and Liu (2003)更指出這四個評估層次，約有 78% 的企業組織僅實施第一層次的訓練評估、32% 的企業組織僅實施第二層次的學習評估、9% 的企業組織僅實施第三層次的行為改變、7% 的企業組織僅實施第四層次的組織評量，顯示這四個層次的評估複雜度和耗用成本是逐漸遞增。

三、研究設計

3.1 研究架構

由於國內外有關訓練成效衡量的研究，多著重在探討訓練能夠為組織帶來的效益或單一訓練方法的討論 (如 Baldwin and Ford, 1988；陳銘薰及王滢婷, 2006 等)，較少有針對不同訓練方法之成效評估與比較的研究 (林文政等, 2010)，更鮮少有評估接艦訓練課程訓練滿意度與學習成效之研究。本研究以 Kirkpatrick (1979) 的四層次訓練評估模式中的反應、學習等兩個層次評估來探討海軍赴美接艦訓練課程的成效與差異，並將反應層次稱為「訓練滿意度」，學習層次稱為「訓練學習成效」，研究架構如圖 4。

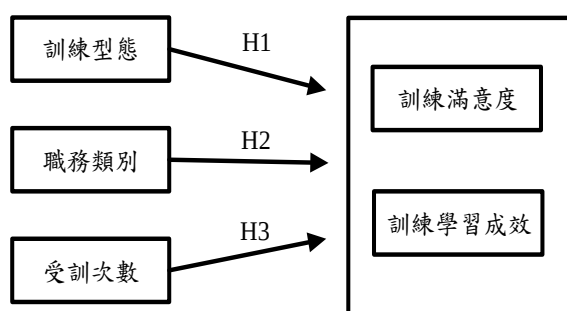


圖 4 研究架構圖

3.2 研究假設

Leone and Pinkston (2017) 認為讓員工學習進修對組織是有幫助，但是如果學習中包含了實務練習或實作課程則會有更佳的學習效果。Bushnell (1990)指出訓練的成效是藉由短期及長期的訓練結果組成，藉由教育訓練，組織短期內可使員工獲得技能與知識，長期而言

可使組織增加利潤、提升客戶滿意度及組織的生產力。林文政等 (2010) 以某一製藥廠為研究對象，運用 Kirkpatrick (1979) 的四層次訓練評估模式，探討室內講授式與戶外體驗式團隊訓練的成效差異，發現兩種訓練都有明顯的訓練成效，但是戶外體驗式團隊訓練的成效並沒有明顯的高於室內講授式訓練課程。考量接艦基地執行在職訓練是以裝備實際操作、武器系統原理及故障排除等為主，應比流路班次訓練採室內授課、以講師為中心、屬單向傳播知識，有較好的學習成效。因此，提出下列假設：

H1 訓練型態的不同對訓練績效有影響

H1-1 訓練型態的不同對訓練滿意度呈現影響

H1-2 訓練型態的不同對學習成效呈現影響

陳銘薰及王滢婷(2006)運用 Bushnell (1990) 的投入、過程及產出等三個構面的 IPO 模式，以我國 832 家中小型製造業為研究對象，評估訓練投入、訓練實施程序及訓練成效的差異，並將員工依職務階層分為「管理職務人員」與「非管理職務人員」兩類，研究發現，企業員工人數的多寡，不會影響「管理職務人員」對於訓練成效的看法，但是會影響「非管理職務人員」的看法。海軍派員赴美參加接艦訓練官兵依職務區分管理職與非管理職、依專長職務區分輪機、補給、政戰等專長，為瞭解不同職務的官兵對訓練績效的差異，提出下列假設：

H2 職務不同對訓練績效有影響

H2-1 職務不同對訓練滿意度有影響

H2-2 職務不同對學習成效有影響

王居卿 (2000) 以國內高科技產業為研究對象探討影響訓練成效的因素，發現學員訓練的期間愈長，則訓練成效愈差，且訓練滿意度會影響學員自身能力增進的程度。海軍每次選派人員赴美接艦，都會考量專長與職務需求，遴選適當人員，有些人員因為工作績優或有特殊專長，經常出國接艦。但是國家資源有限，應使有限資源發揮最大效益，因此為了解赴美接艦次數多寡對於訓練成效及滿意度的差異，提出下列假設：

H3 受訓次數愈多則訓練績效有正向影響

H3-1 受訓次數愈多則訓練滿意度有正向影響

H3-2 受訓次數愈多則學習成效有正向影響

3.3 問卷設計

發放之問卷係根據以往的文獻，採用適合且相關的衡量題項與尺度，根據國內外相關

文獻設計，並經專家學者檢視，因此量表具有相當的內容效度。問卷共區分為三個部份，第一個部份衡量官兵參加美軍流路班次訓練的訓練滿意度與學習成效，第二個部份衡量接艦基地執行在職訓練的訓練滿意度與學習成效，第三個部份則為受測者的基本資料。訓練滿意度問項參考吳秉恩 (1984) 問卷，及接艦訓練實況修改而成，計有「對訓練時段的安排是否滿意」等 9 題(含 1 題開放式問項)，依程度高低依序為「非常同意」、「同意」、「普通」、「不同意」及「極不同意」等，以李克特(Likert)五點尺度衡量，分別給予 5、4、3、2、1 分的分數；考量選派赴美接艦人員目的為培養未來海軍各專業領域管理人才，因此訓練成效問卷之問項，依據 Katz (1955) 提出主管人員應該具備專業(Technical)、人際 (Human)、理念(Conceptual)等三項技能的概念，並參考吳秉恩(1984)及接艦訓練實況修改而成，計有「與任務有關的專業技能方面」等 10 題(含 1 題開放式問項)，其中 1~3 題為專業技能、4~6 為人際技能、7~9 為理念技能，依程度高低排列依序為「進步相當多」、「進步很多」、「進步頗多」、「進步很少」及「進步相當少」等，由填答者任選其一，以李克特(Likert)五點尺度衡量，分別給予 5、4、3、2、1 分的分數分別。第三部份則為收集受測者的性別、教育程度等基本資料，共 5 題。

3.4 分析方法

- (一)描述性統計分析：以次數分配的方式分析受訪者基本變項資料，用以瞭解研究對象的特性。
- (二)信度分析：信度是指一份量表所測得分數的可靠度、一致性及穩定性，信度係數的種類與衡量方法，主要區分再測性度、複本信度、折半信度與 Cronbach α 信度。本研究採用 Cronbach α 值分析問卷的信度，並採 Nunnally and Bernstein (1994) 與 DeVellis (2003) 的觀點，認為信度測量最好大於 0.70，其內部一致性方可接受。 α 係數愈大表示各題項間的關聯性愈大，亦即一致性越高。
- (三)成對樣本 t 檢定：檢驗兩個相依樣本平均值的差異是否相等，或平均數差異值達到統計上有顯著的意義。
- (四)單因子變異數分析：檢定單因子在不同組別的平均數是否有顯著的差異。

3.5 資料蒐集

本文為國內外首次衡量赴美接艦訓練滿意度之研究，問卷發放對象為曾赴美參加過美軍流路班次訓練與接艦基地在職訓練兩類訓練之官兵，為擴大研究樣本，問卷以電子方式寄發，由於電子問卷無法精確統計實際發放數量，僅統計回收數量與有效回收率，於 2017 年 5 月共發出 196 份問卷，剔除無同時參加兩類訓練課程者，有效問卷 127 份，回收率 86.7%。

為驗證本次研究之合理樣本，運用樣本數估算公式 $n = \left(z_{\alpha/2} \sigma / e \right)^2$ ，選定準確性在信賴係數為 95% ($1 - \alpha$)，容忍誤差界限不超過 0.15 (e) 時，如母體標準差未知，可用樣本取代之，分別針對流路訓練滿意度、流路訓練學習成效、基地在職訓練滿意度及基地在職訓練學習成效估算合理之樣本(此四面向之樣本標準差如表 3)，所需樣本數分別為 70, 96, 86 和 112。因此，本研究的抽樣樣本符合統計學中抽樣的要求，亦能在抽樣上有效估計母體，即具有抽樣的代表性。

四、資料分析與討論

4.1 樣本基本資料分析

表 2 敘述樣本組成分析，受訪者人口統計變項包括性別、教育程度、階級、服務年資、接艦時服務的部門、參與訓練的次數等五項。在性別方面，受訪者 126 人為男性佔 99.2%；教育程度方面受訪者中以高中(職) 51 人佔 40.2%最多、大專 26 人佔 20.5%次之，其中大專以上者 76 人佔 59.8%，顯示海軍選派赴美受訓人員的學歷，整體而言較高；服務年資以已退役 36 人佔 28.3%最多、16~20 年者 28 人佔 20%次之；軍官 36 人佔 28.5%、士官 91 人佔 71.5%；階級以士官長 39 人佔 30.7%最多、上士 29 人佔 22.8%次之，顯示海軍赴美受訓人員，多以任務、專長考量遴選專業士官為主；接艦時任職的部門以輪機部門 47 人佔 36.2%最多、兵器部門 37 人佔 29.1%次之；接艦次數則以 1 次 90 人佔 70.9%最多、2 次 30 人佔 23.6%次之。

表 2 樣本組成分析(N=127)

變項	次數	項目	次數
性 男	126(99.2)	士兵	9(7.1)
別 女	1(0.8)	下士	5(3.9)
變項	次數	項目	次數
高中(職)	51(40.2)	中士	9(7.1)
教 大專	26(20.5)	上士	29(22.8)
育 大學	23(18.1)	士官	39(30.7)
程 碩士或指參	20(15.7)	少尉	-
度 博士或戰院	7(5.5)	中尉	1(0.8)
1 年以下	1(0.8)	上尉	2(1.6)
1~5 年	15(11.8)	少校	5(3.9)
服 6~10 年	6(4.7)	中校	18(14.2)
務 11~15 年	15(11.8)	上校	8(6.3)
年 16~20 年	28(22.0)	少將	2(1.6)
資 21~24 年	17(13.4)	支隊	2(1.6)
25 年以上	9(7.1)	作戰	31(24.4)
已退役	36(28.3)	艦 兵器	37(29.1)
1 次	90(70.9)	時 輪機	47(37)
接 2 次	30(23.6)	的 補給	7(5.5)
艦 3 次	5(3.9)	部 政戰	-
次 4 次	1(0.8)	門 教訓	3(2.4)
數 5 次	1(0.8)	廠訓	-

4.2 問卷之信度分析

回收問卷之 Cronbach α 係數如表 3，其中美軍流路班次訓練的訓練滿意度為 0.93、訓練學習成效為 0.93；接艦基地執行在職訓練的訓練滿意度為 0.92、訓練學習成效為 0.91。

表 3 變數 Cronbach α 值(N=127)

變數	α 值	平均值	標準差
流路訓練滿意度	0.93	3.87	0.64
流路訓練訓練學習成效	0.93	3.83	0.75
基地在職訓練滿意度	0.92	3.96	0.71
基地在職訓練學習成效	.91	.03	.81

Nunnally and Bernstein (1994)與 Devellis (2003) 認為信度測量最好大於 0.70，其內部一致性方可接受，故問卷的內部一致性為可接受範圍。此外，問卷中流路訓練滿意度問項平均值為 3.87，標準差為 0.64，屬於中等滿意度；而流路訓練學習成效問項平均值為 3.83，標準差為 0.75；基地在職訓練滿意度問項平均值為 3.96，標準差為 0.71，屬於中等滿意度，基地在職訓練學習成效問項平均值為 4.03，標準差為 0.81。

4.3 訓練滿意度比較分析

運用成對樣本 t 檢定檢定不同的訓練型態在滿意度上的差異，以下區分整體訓練滿意度、個別項目訓練滿意度來探討：

整體而言，在訓練滿意度方面，美軍基地在職訓練滿意度的平均數(M= 3.96, SD= 0.71)，高於流路訓練滿意度的平均數(M= 3.87, SD= 0.64)，且達顯著水準 ($p < 0.001$)，如表 4。顯示，受訪者在同時接受流路訓練、基地在職訓練後，對於基地在職訓練的滿意度高於流路訓練。因此，本研究 H1-1 訓練型態的不同對訓練滿意度呈現影響，假說成立。

表 4 整體訓練滿意度成對樣本 t 檢定

項目	平均數(標準差)		t 值
	流路訓練	基地訓練	
滿意度	3.87(0.64)	3.96 (0.71)	-3.29***

*** $p < 0.001$

為了解受訪者在不同訓練模式下個別項目平均數的差異，本研究再以成對樣本 t 檢定檢驗(如表 5)，發現在訓練時段、課程編排、課程內容、訓練師資等部分，流路訓練滿意度的平均數均高於基地在職訓練滿意度，其中訓練師資項目達顯著水準，顯示受訪者普遍認為流路訓練所檢派的教官或授課人員、授課環境、課程規劃等均較基地在職訓練為佳，應該是流路訓練授課內容多屬原理、原則教授，我海軍派訓人員多已有裝備的基礎知識，甚至是具備操作經驗，較易吸收授課內容。但是在場所與設備、成效獎勵、生活津貼等項目，則是顯示基地在職訓練顯著高於流路訓練，應是因為基地在職訓練的教學地點多是接收艦艇或附近教室，支隊部較能支援教學資源及管理。

表 5 個別項目訓練滿意度成對樣本 t 檢定

項目	平均數(標準差)		t 值
	流路訓練	基地訓練	
訓練時段	4.21(0.74)	4.16(0.67)	1.09
項目	平均數(標準差)		t 值
	流路訓練	基地訓練	
訓練時數	3.99(0.97)	3.99(0.81)	0
課程編排	4.13(0.71)	4.07(0.75)	1.24
課程內容	4.09(0.76)	4.06(0.82)	0.53
訓練師資	4.17(0.71)	4.04(0.74)	2.96**
場所設備	3.51(0.70)	4.00(0.83)	-7.50***
成果獎賞	3.36(0.83)	3.79(0.95)	-5.74***
生活津貼	3.47(1.16)	3.61(1.15)	-2.33*

*p<0.05, **p<0.01, ***p<0.001

4.4 學習成效比較分析

對於不同的訓練型態在學習成效上的差異，區分整體學習成效、個別項目學習成效來探討。整體而言，在學習成效方面，美軍基地在職訓練學習成效的平均數(M= 4.03, SD= 0.81)，顯著高於流路訓練學習成效的平均數(M=3.83, SD= 0.75)， $t(126)=-6.05$ ， $p<0.001$ (如表 6)。

表 6 整體學習成效成對樣本 t 檢定

項目	平均數(標準差)		t 值
	流路訓練	基地訓練	
學習成效	3.83(0.75)	4.03(0.81)	-6.05***

***p<0.001

顯示，受訪者在同時接受流路訓練、基地在職訓練後，認為基地在職訓練的學習成效高於流路訓練。因此，本研究之 H1-2 訓練型態的不同對學習成效呈現影響，假說成立。

為了解受訪者在不同訓練模式下個別項目訓練學習成效平均數的差異，以成對樣本 t 檢定檢驗(如表 7)，發現在工作操作效率、對工作目標認知與合作態度、整合分析能力、應用推廣能力、領導統御等項目，基地在職訓練的學習成效均高於流路訓練，且整合分析能力、領導統御等項目達顯著水準。顯示受訪者普遍認為經過基地在職訓練，能於返國後在任

務執行時，在增加業務的整合分析及領導統御能力方面，均較流路訓練為佳。

表 7 個別項目訓練滿意度成對樣本 t 檢定

項目	平均數(標準差)		t 值
	流路訓練	基地訓練	
專業技能	4.16(0.90)	4.13(0.85)	0.69
工作效率	4.06(0.89)	4.09(0.84)	-0.46
事務處理	4.02(0.97)	4.00(0.93)	0.29
協調能力	3.98(0.96)	3.98(0.92)	0
目標認知	4.01(0.88)	4.02(0.89)	-0.15
專業知識	4.21(0.90)	4.13(0.85)	1.52
整合分析	3.40(0.68)	4.06(0.86)	-11.84**
應用推廣	3.91(0.96)	3.93(0.90)	-0.48
領導統御	2.77(0.65)	3.91(1.02)	-16.41**

*p<0.05, **p<0.01, ***p<0.001

4.5 職務差異比較分析

在赴美參加接艦訓練的受訪官兵中，依照職務與專長區分支隊部、作戰部門、兵器部門、輪機部門、補給部門、教訓小組等部門，因為美軍的接艦訓練會針對艦艇的類型與我國的使用需求開設各專長課程，因此本研究運用單因子變異數分析，探討不同職務官兵在參加美軍接艦訓練時，是否會因為職務差異已對相關課程、裝備操作程序熟稔，而對美軍的流路訓練課程及基地在職訓練課程滿意度與學習成效產生差異。檢驗發現，不同的職務對訓練滿意度有顯著差異 ($p<0.05$)；不同職務對學習成效有顯著差異($p<0.05$)。因此，H2-1 職務不同對訓練滿意度有影響及 H2-2 職務不同對學習成效呈現影響，假說均成立(如表 8、9)。

表 8 職務不同訓練滿意度差異

變異來源	SS	df	MS	F
職務類別	5.35	5	1.07	2.69*
變異來源	SS	df	MS	F
誤差	48.19	121	0.39	
總和	53.54	126		

*p<0.05

表 9 職務不同訓練學習成效差異

變異來源	SS	df	MS	F
------	----	----	----	---

職務類別	6.41	5	1.28	2.33*
誤差	66.48	121	0.55	
總和	72.89	126		

*p<0.05

進一步檢視不同職務類別之訓練滿意度與學習成效（如表 10），其指出作戰部門、兵器部門與補給部門皆呈現偏低現象（皆低於 4 分），由於接艦訓練是一個整體性之評估，自艦務、兵器、作戰、輪機，補給等職務都會影響到接艦訓練之成效，建議未來培訓課程要特別留意此三類職務培訓人員的課程設計、職務分工、情緒管理等，並追蹤訓練之成效回饋，以有效掌握學習成效。

表 10 職務類別對訓練滿意度與學習成效差異

職務類別	平均數	
	訓練滿意度	學習成效
支隊長等	4.03	4.33
作戰部門	3.66	3.77
兵器部門	3.89	3.94
輪機部門	4.12	4.08
補給部門	3.61	3.25
教訓小組	4.33	4.52

註：訓練滿意度為流路訓練與基地在職訓練兩類課程訓練滿意度之平均值；訓練學習成效為流路訓練與基地在職訓練兩類課程學習成效之平均值

4.6 訓練次數比較分析

為探討參加赴美參加接艦訓練的官兵，是否會因為參訓次數多寡，影響訓練滿意度與訓練學習成效，將赴美參訓次數區分 1 次(90 人)、2 次(30 人)及 3 次以上(7 人)等三個類別。檢驗發現，參訓次數在對訓練滿意度沒有顯著差異；不同的參訓次數對學習成效也沒有顯著差異。顯示受訪者並沒有因為參訓次數較多而對美軍的訓練滿意與學習成效有差異。檢驗結果與王居卿 (2000)學員訓練時間愈長，則訓練成效愈差結果相似。雖然國軍在積極落實國艦國造的政策下，未來艦艇將以國造為優先考量，循軍售管道籌獲，仍為一重要且可遇不可求機會，未來仍可能向美軍購入軍艦，而海軍未來在選派人員出國時，除必要專業人員外，可考量遴選無赴美接艦經驗人員，擴大訓練成效。

因此，假設 H3-1 受訓次數愈多則訓練滿意度有正向影響及假設 H3-2 受訓次數愈多則

訓練成效有正向影響，假說均不成立（如表 11、12），假設驗證結果整理如表 13。

表 11 訓練次數訓練滿意度差異

變異來源	SS	df	MS	F	p
受訓次數	1.50	2	0.75	1.79	0.17
誤差	52.04	124	0.42		
總和	53.54	126			

表 12 訓練次數學習成效差異

變異來源	SS	df	MS	F	p
受訓次數	1.15	2	0.58	0.99	0.37
誤差	71.73	124	0.58		
總和	72.89	126			

表 13 假設驗證結果

假說	結果
H1-1 訓練型態的不同對訓練滿意度呈現影響	成立
H1-2 訓練型態的不同對學習成效呈現影響	成立
H2-1 職務不同對訓練滿意度有影響	成立
H2-2 職務不同對學習成效呈現影響	成立
H3-1 受訓次數愈多則訓練滿意度有正向影響	不成立
H3-2 受訓次數愈多則學習成效有正向影響	不成立

五、結論與管理意涵

人是組織中最重要的資產，「為用而訓、計畫培養」一直是國軍人才培育及教育訓練的宗旨，在國防資源有限的情況下，提高訓練的成效也是備受矚目的課題。向美國採購除役之軍艦，必須掌握機會，以較低價格、極短時間，並且可快速提升我國海軍防衛戰力，此為美國對台灣關係法較實務作為之一，惟其中除武器、裝備等硬體設施外，軟體亦為重要之項目，訓練即為其中必須配合在接艦前需完成之最重要之軟體工作。

美國海軍訓練課程設計，係因其艦艇官兵操縱著困難度較高之作戰艦艇，其訓練要求非常嚴格與審慎，端賴於健全的制度下逐級督管及測考單位亦於公平、公正、公開基礎上實施鑑測，相信此為美軍能維持高水準戰力之憑藉。要使一艘軍艦完成戰備訓練，能順利執行其所賦予之任務並且長時維持於高戰備水準，除硬體再精良，亦必須依賴良好的訓練，方能將其發揮至最大之戰力。

本研究以有效管理之方式，對接艦人員訓練方式之成效，予以合理客觀之分析、意見調查、驗證何種方式最為有效，以達到專案管理之設定目標。針對曾赴美參加過美軍流路訓練與基地在職訓練兩類訓練班次的官兵，探討訓練的滿意度與學習成效。實證發現，受訪者在同時接受流路訓練與基地在職訓練後，對於基地在職訓練的滿意度高於流路訓練，其中以訓練師資、場所與設備、成效獎勵及生活津貼等項目達顯著水準，故赴美接艦訓練型態的不同，會影響訓練滿意度；此外，受訪者在同時接受流路訓練與基地在職訓練後，基地在職訓練的學習成效亦高於流路訓練，因此赴美接艦訓練型態的不同，也會影響訓練成效。整體而言，受訪的官兵對於赴美接艦基地在職訓練的滿意度與學習成效均較流路訓練為佳。

在赴美參加接艦訓練的受訪官兵中，依照單位職務與專長區分支隊部、教訓、廠訓小組等，艦上亦分為作戰、兵器、輪機、補給等部門，發現不同職務官兵在參加美軍接艦訓練時，對美軍的流路訓練課程及基地在職訓練課程滿意度與學習成效產生差異。其次探討參加赴美參加接艦訓練的官兵，是否會因為參訓次數多寡，影響訓練滿意度與學習成效。實證發現不同的參訓次數對於訓練滿意度與學習成效沒有顯著差異。

在我國海軍前輩奠定良好的基礎上，一直精進人員訓練、裝備保養、戰備演訓工作，期達成護衛海疆的任務，而訓練並非一蹴可幾，它必須在有組織、有系統，循序漸進式且長時期的教育培養，勤訓苦練，方能奏效。但訓練、保養均為相當繁雜的工作，必須反覆不斷的訓練、檢討、再訓練，方能於時間內，精確的發揮其功能。本研究發現接艦的官兵曾接受過美軍接艦流路訓練與基地在職訓練後，多認為基地在職訓練的滿意度與學習成效明顯高於流路訓練。此外，多次赴美接艦的官兵，並沒有因為參訓次數較多，而認為有較佳的學習成效。國防資源有限，必須要兼顧預算使用效益與訓練成效，因此，管理單位未來在規劃接艦訓練課程時，可以適度調整訓練流路與訓練時數，甚至減併流路班次訓練，著重於接艦基地之基礎理論教學及裝備實際操作與維護，以提升訓練滿意度與學習成效。此外，Kirkpatrick (1988)指出反應、學習、行為、結果四個層次訓練評估模式是階層關係，每個訓練方案都應要產生這四層次的變化，四個層次

都要衡量才能得知訓練成效全貌。本研究因時間點限制，無法同時衡量同一批接艦官兵在接艦前及參加接艦訓練課程後行為、能力的改變及對於所屬組織所產生的結果效益，未來可長期追蹤，以完整運用 Kirkpatrick (1979)的四個層次訓練評估模式，了解訓練成效的全貌。

本案主要探討，中華民國海軍至美海軍循環軍售管道接收美國除役之艦船，此為高額度之國防預算投資，其管理工作雖錯綜複雜，惟無外乎於品質、期程及預算，而影響這些最大的就是”人的訓練”，故本篇強調有關之人員訓練之效果評估，俾利後續執行相關工作時參考依據，因為訓練成效對接艦工作執行是否順利佔了決定性之比重，因為他們負責自修工程，裝備保養、操作、值更、故障隔離與排除，進而達到高級戰力之水準，故接艦人員之訓練不可不加以重視！

誌謝

感謝參與問卷調查的官、士、兵同袍，由於您們回饋寶貴的接艦經驗，使得本研究能實際反應出接艦期間，以何種方法訓練接艦官兵最符合經濟效益，此舉有助於未來國軍赴外接艦訓練之品質效能提升，進而快速提升國軍戰力，在此致上崇高的謝意。

參考文獻

- [1] Arthur, W. Jr., Bennett, W. Jr., Edens, P. S., and Bell, S. T., “Effectiveness of Training in Organizations: A Meta-Analysis of Design and Evaluation Features,” *Journal of Applied Psychology*, Vol. 88, No. 2, pp. 234–245, 2003.
- [2] Baldwin, T. T. and Ford, J. K., “Transfer of training: A review and directions for future research,” *Personnel Psychology*, Vol. 41, pp. 63–105, 1988.
- [3] Bushnell, S. D., “Input Process output: A model for evaluating training,” *Training and Development Journal*, Vol. 44, pp. 41–43, 1990.
- [4] Chatzimouratidis, A., Theotokas, I., and Lagoudis, I. N., “Decision support systems for human resource training and development,” *The International Journal of Human Resource Management*, Vol. 23, pp. 662–693, 2012.
- [5] Curry, D., Dobbins, G., and Ladd, R.,

- “Transfer of training and adult learning (TOTAL),” *Journal of Continuing Social Work Education*, Vol. 6, No. 1, pp. 8–14, 1994.
- [6] Demir, R., Wennberg, K., and McKelvie, A., “The Strategic Management of High-Growth Firms: A Review and Theoretical Conceptualization,” *Long Range Planning*, Vol. 50, pp. 431–456, 2017.
- [7] Demir, R., Wennberg, K., and McKelvie, A., “Advancing the Human Capital Perspective on Value Creation by joining Capabilities and Governance Approaches,” *Academy of Management Perspectives*, Vol. 29, No. 3, pp. 296–308, 2017.
- [8] DeVellis, R. F., “Scale Development Theory and Applications”, London: SAGE, 2003.
- [9] Dolezalek, H., “2005 industry report,” *Training*, Vol. 42, No. 12, pp.14–25, 2005.
- [10] Fulmer, I. S. and Ployhart, R. E., “Our most important asset: a multidisciplinary/multilevel review of human capital valuation for research and practice,” *Journal of Management*, Vol. 40, No. 1, pp. 161–192, 2014.
- [11] Galvin, J. C., “What Can Trainers Learn From Educators About Evaluating Management Training? ,” *Training and Development Journal*, Vol. 37, No. 8, pp. 52–57, 1983.
- [12] Georgenson, D. L., “The Problem of Transfer Calls for Partnership,” *Training & Development Journal*, Vol. 36, No. 10, pp. 75–78, 1982.
- [13] Hoover, J. B., “TIPS” for Training,” *Training*, Vol. 54, No. 2, pp. 12-13, 2017.
- [14] Kanters, M. A., Bristol, D. G., and Attarian, A., “The Effects of Outdoor Experiential Training on Perceptions of College Stress,” *Journal of Experiential Education*, Vol. 25, No. 2, pp. 257–267, 2002.
- [15] Katz, R. L., “Skill of an effective Administrator,” *Harvard Business Review*, Vol. 33, pp. 33–42, 1955.
- [16] Kim, Y., and Ployhart, R., “The effects of staffing and training on firm productivity and profit growth before, during, and after the great recession,” *Journal of Applied Psychology*, Vol. 99, No. 3, pp. 361–389, 2014.
- [17] Kirkpatrick, D. L., “Techniques for Evaluating Training Programs,” *Training and Development Journal*, Vol. 33, No. 6, pp. 78–92, 1979.
- [18] Kirkpatrick, D. L., “Evaluating training programs: The four levels,” San Francisco: Berrett-Koehler Publishers, 1998
- [19]
- [20] Leach, M. P., and Liu, A. H., “Investigating Interrelationships among Sales Training Evaluation Methods,” *Journal of Personal Selling and Sales Management*, Vol. 23, No. 4, pp. 327–339, 2003.
- [21] Leone, P. and Pinkston, R., “Practice Makes Perfect,” *Talent Development*, Vol. 71, No. 5, pp. 62–63, 2017
- [22] Lorge, S., “Getting into their heads,” *Sales and Marketing Management*, Vol. 150, No. 2, pp. 58–67, 1998.
- [23] Molinsky, A., “The Mistake Most Managers Make with Cross-Cultural Training,” *Harvard Business Review*, January, 2015.
- [24] Noe, R. A., “Employee Training and Development,” New York: Irwin/McGraw-Hill, 2002.
- [25] Nunnally, J. C., and Bernstein, I. H., “Psychometric Theory,” New York: McGraw-Hill, 1994.
- [26] Nyberg, A. J., Moliterno, T. P., Hale, D. Jr. and Lepak, D. P., “Resource-based perspectives on unit-level human capital: a review and integration,” *Journal of Management*, Vol. 40, No. pp. 316–346, 2014.
- [27] Ployhart, R. E., and Hale, D., “The fascinating psychological microfoundations of strategy and competitive advantage,” *Annual Review of Organizational Psychology and Organizational Behavior*, Vol. 1, pp. 145–172, 2014.
- [28] Waldman, D. V., and Balven, R. M., “Responsible leadership: theoretical issues and research directions,” *The Academy of Management Perspectives*, Vol. 3015, No. 1, pp. 19–29, 2015.
- [29] Wright, P. M., McMahan, G. C., & McWilliams, A., “Human resources and sustained competitive advantage: A resource-based perspective,” *International Journal of Human Resource Management*, Vol. 5, pp. 301–326, 1994.
- [30] Yamazaki, Y. and Kayes, D. C., “An Experiential Approach to Cross-Culture Learning: A Review and Integration of Competencies for Successful Expatriate Adaptation,” *Academy of Management*

Learning & Education, Vol. 3, No. 4, pp. 362-379, 2004.

- [31] 王居卿，“影響訓練成效相關因素模式之實證研究：認知及多變量的觀點”，臺大管理論叢，第 10 卷，第 2 期，第 135-166 頁，2000。
- [32] 吳秉恩，1984，管理才能發展方案實施與成效關係之研究，國立政治大學企業管理研究所博士論文。
- [33] 林文政、王湧水、許智翔，“不同訓練方法成效評估之研究：一項室內講授式與戶外體驗式團隊訓練的比較”，管理與系統，第 17 卷第 2 期，第 229-254 頁，2010。
- [34] 陳銘薰及王澄婷，“訓練投入、訓練實施程序、訓練成效評估模式之探討”，人力資源管理學報，第 6 卷，第 1 期，第 75-99 頁，2006。

最小一乘法 WiFi 匹配定位之研究

林老生^{1*} 林子添¹

¹國立政治大學地政學系

摘 要

由於 WiFi 定位常使用的 KNN 演算法，無法對訊號資料有進一步的分析，找出不良的訊號觀測量。因此，本文嘗試利用具有穩健性的最小一乘法處理 WiFi 匹配定位(簡稱 LAD)，以獲得較精確的 WiFi 定位成果。實驗資料，包含模擬資料、室內及室外資料等 WiFi 資料。首先，以實驗資料測試 LAD 精度，並與最小二乘法 WiFi 匹配定位(簡稱 LS)及 KNN 定位精度比較。然後，再將粗差加入實驗資料，測試 LAD 與 LS 的穩健性及偵錯能力。根據實驗成果顯示：(1)在定位精度方面，通常 LAD 優於 LS，例如 LS 與 LAD 的室內定位資料，分別為 2.41 與 2.20 公尺。(2)在穩健性方面，於 RSSI 建模階段，LAD 有很好的穩健性，LS 則無；於定位解算階段，當粗差數量小於或等於 2 時，LAD 有好的穩健性，LS 則無。(3)在偵錯能力方面，於 RSSI 建模階段，LAD 有很好的偵錯能力，可以從參考點的 RSSI 殘差判斷粗差所在；於定位解算階段，當粗差數量小於或等於 2 時，LAD 有好的偵錯能力。

關鍵詞：WiFi 匹配定位、最小一乘法、最小二乘法、穩健性

Study on WiFi Matching Positioning Using Least Absolute Deviation

Lao-Sheng Lin^{1*}, Tzu-Tian Lin¹

¹The Department of Land Economics, National Chengchi University

ABSTRACT

Since the KNN algorithm of common WiFi positioning can't find the outliers in the WiFi signal. Hence, this research tries to use the robustness of Least Absolute Deviation to deal with WiFi matching positioning (short name, LAD) to overcome the sensibility of WiFi signal strength, expecting to make the result of WiFi positioning more reliable. Test data sets include simulated data, indoor data and outdoor data. At first, test data sets are used to study the accuracies of LAD, Least Squares WiFi matching positioning (short name, LS) and KNN. Then, the outliers are added to the test data sets. The robustness and the outlier detection ability of LAD and LS are tested accordingly. According to the test results, there are some findings. (1) Usually LAD has better position accuracies than LS. For example, the indoor positioning accuracies of LS and LAD are 2.41 and 2.20 m respectively. (2) LAD's robustness is better than LS. (3) LAD's outlier detection ability is better than LS. The outlier can be detected by the largest RSSI residual of reference points or check points.

Keywords: WiFi Matching Positioning, Least Absolute Deviation, Least Squares, Robustness

一、前言

訊號紋法(Finger Printing)為 WiFi 定位的方法之一,雖然對於環境有較強的適應性,然而訊號紋法中所使用 KNN (K- Nearest Neighbor)演算法僅能利用現有資料進行計算,並無法對訊號資料有進一步的分析,找出不良的訊號觀測量;因此,李楨、譚先科(2015)[1]提出最小二乘法(Least Squares, LS)WiFi 匹配定位。首先,透過 WiFi 訊號強度以及已知的參考點坐標,建立訊號強度與地面點平面坐標之數學模型;其次,利用導航者接受的 WiFi 訊號及已建好的數學模型推估導航者之(未知)位置;在建立數學模型或者利用數學模型推估未知點位置時,皆可以產生觀測量殘差等資訊,供後續成果分析。若改以具有穩健性的最小一乘法(Least Absolute Deviation, LAD)進行演算,則可進一步分析是否有不良的訊號在觀測量之中,也可以產生較穩健的成果。所謂穩健性,代表 WiFi 演算法不易受含粗差觀測量影響到定位精度[2]。

然而,在李楨、譚先科(2015)之研究中,僅使用模擬資料進行最小二乘法 WiFi 匹配定位解算,且文中並未對於訊號強度與地面點平面坐標模型之解釋能力進行檢定。再者, WiFi 訊號強度容易受到環境因素影響,例如降雨、建物遮蔽、人群擾動等因素,皆會使訊號強度降低,若以受影響的訊號強度進行匹配定位,將使定位成果與真實位置產生偏移。若改以最小一乘法進行匹配定位演算,是否能克服上述訊號干擾問題?

因此,本文研究目的:(1)首先,以模擬資料來對李楨、譚先科(2015)所提的 WiFi 訊號強度與參考點平面坐標之模型進行統計檢定;(2)其次,再以模擬資料及室內、室外真實資料,測試最小一乘法用於 WiFi 匹配定位的精度、穩健性及偵錯能力。

二、理論基礎

2.1 WiFi 定位研究

2.1.1 室內定位

Bahl et al. (2000)[3]設定 70 個參考點,並於實驗中發現使用者的方位,會顯著的影

響 WiFi 訊號(以後簡稱訊號)的強度高達正負 5dBm,在某一方向下,使用者的身體將成為阻擋訊號的障礙,使行動裝置的天線可能與 AP(Access Point)發射源產生非視線連結。因此,在建立訊號強度資料庫時,在訊號強度的紀錄上會新增方向(Direction)欄位,選擇最強訊號方向來代表參考點訊號資訊,並用 KNN 法選取 3~5 個最鄰近點進行平均,使得定位精度獲得提升。在實驗區域的平均位置誤差可在 2 至 3 公尺之內,文章中並表明參考點以及最鄰近點數量會影響定位的精度。

余宗鴻(2015)[4]在 20x18 平方公尺的室內環境,使用 KNN 演算法進行訊號紋法於室內定位解算,平均定位誤差在 2 至 3 公尺左右。

莊豐錨(2015)[5]以 1.5 倍標準差為標準,過濾離群訊號,若收到的訊號強度大於或小於 1.5 倍的標準差,則將該筆訊號剔除,並且利用使用者定位歷史紀錄,推估使用者未來的位置,藉此減少樣式比對次數及定位運算量,降低行動裝置的運算負荷,並得到較精確的定位結果,該研究以 2 公尺為間距建立參考網格,並將訊號強度低於 90dBm 的訊號剔除,實驗結果表示該方法可以將定位精度控制在 3 公尺左右,並且最大誤差不超過 9 公尺。

2.1.2 室外定位

劉科宏(2008)[6]結合 WiFi 與 GPS 兩種定位方法,觀測在不同氣候條件下(晴天、陰天、雨天)GPS 的定位成果,並以網格大小 1 公尺的格網,建立無線電地圖(Radio Map)進行 WiFi 輔助 GPS 定位,利用可調變的權重機制隨環境變化進行調整,修正 GPS 與 WiFi 誤差,計算行動裝置位置。研究結果顯示,在三種氣候條件下,單純使用 KNN 演算法進行 WiFi 室外定位之平均誤差為 18.9 公尺。

崔文、陳昭男(2011)[7]以智慧型手機為平台,透過手機內建的 GPS、WiFi 接收晶片、加速度計與電子羅盤,提出混合式的定位系統,在 WiFi 部份透過粒子濾波器推估 WiFi 定位結果,並利用基於信心度的混合讓整體平面誤差能收斂在 20 公尺以內,其中 WiFi 定位平均誤差為 18.37 公尺。

郭清智(2011)[8]結合 WiFi 定位與 GPS

技術於室外導航，並建立混合定位流程，針對 GPS 衛星顆數小於 4 顆或者 HDOP 值大於 4.5 的狀態引入 WiFi 定位，設定平均誤差與權重成反比進行加權平均。研究結果顯示，在 160x110 平方公尺的室外環境，參考點間距約為 3~5 公尺，單純使用 KNN 演算法進行 WiFi 室外定位之平均誤差約為 20.0 公尺。

由室內與室外 WiFi 定位的文獻可知，在室內的環境下，WiFi 定位精度皆可控制在 3 公尺之內；而在室外的環境下，WiFi 定位精度則約在 20 公尺。故可以室內定位精度 3 公尺與室外定位精度 20 公尺為標準，檢驗 WiFi 匹配定位精度之成果是否能比一般的 WiFi 定位成果優良？

2.2 最小二乘法 WiFi 匹配定位

李楨、譚先科(2015)[1]以最小二乘法進行 WiFi 匹配定位，透過模擬資料的設定，在網格大小 3 公尺的狀況下，定位精度可在 1 公尺左右，並與訊號紋法比較，發現訊號紋法精度受限於格網的大小，精度僅有 3 公尺。

最小二乘法的目標函數(S)為所有殘差(V)平方和為最小，如(1)式所示。

$$S = \sum_{i=1}^m V_i^2 \Rightarrow \min \quad (1)$$

最小二乘法 WiFi 匹配定位主要分為 RSSI 建模階段與定位解算階段兩個流程，說明如下：

2.2.1 RSSI 建模階段

若有參考點的坐標與各 AP 之訊號強度值(Received Signal Strength Indicator, RSSI)等資訊，可以利用參考點坐標及參考點所接收的 AP 訊號強度(RSSI)透過多項式擬合的方式，建立參考點坐標與訊號強度之間的函數關係，兩者的函數關係若以(2)式表達： r_{i-j} 表示第 j 個參考點所收到的第 $i(i=1, \dots, m)$ 個 AP 的訊號強度，單位為 dBm； (x_j, y_j) 表示參考點 j 的平面坐標， $j(j=1, \dots, n)$ 代表參考點的編號，單位為公尺； a_i, b_i, c_i 則為第 i 個 AP 的待求解模型參數。若有 i 個 AP， j 個參考點，則能將(2)式寫成矩陣型式如(3)式及(4)式所

示，以最小二乘法透過(5)式求解模型參數[1]；式中， A 、 X 、 L 及 V ，分別代表係數矩陣、參數向量、觀測量向量及殘差(剩餘誤差)向量。

$$r_{i-j} = a_i x_j + b_i y_j + c_i x_j y_j \quad (2)$$

$$A = \begin{bmatrix} x_1 & y_1 & x_1 y_1 \\ \vdots & \vdots & \vdots \\ x_j & y_j & x_j y_j \end{bmatrix}, X = \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix}$$

$$L = \begin{bmatrix} r_{i-1} \\ \vdots \\ r_{i-j} \end{bmatrix}, V = \begin{bmatrix} v_{i-1} \\ \vdots \\ v_{i-j} \end{bmatrix} \quad (3)$$

$$AX = L + V \quad (4)$$

$$X = (A^T A)^{-1} (A^T L) \quad (5)$$

為了降低資料的運算量，透過粗匹配的方法減少參與計算的參考點資料，並且提升定位精度。利用歐基里德距離的計算，找出最鄰近參考點作為粗匹配的代表點，並以該參考點為中心各向四周擴展 1 個參考點，其可能情形依粗匹配點的位置而有不同數量的相鄰參考點。相鄰參考點數量的數量有三種可能，分別為 4、6、9 個參考點，即代表(3)式的觀測量有 j 的最大值有 4、6 與 9 等三種可能。如圖 1 所示，每一方格皆代表 1 個參考點，而藍色實心方格代表粗匹配的參考點位置，由該參考點為中心各向四周擴展 1 個參考點，即有 4、6、9 個參考點數量的可能性。4 點的情況分佈於角點，6 點情況分佈於邊緣非角點上，而 9 點情況則為剩下的點分佈情形。依照粗匹配的位置向四周擴散一個格點，即會有 4、6、9 個參考點來進行模型參數的解算，並求得 a_i, b_i, c_i 等轉換參數。

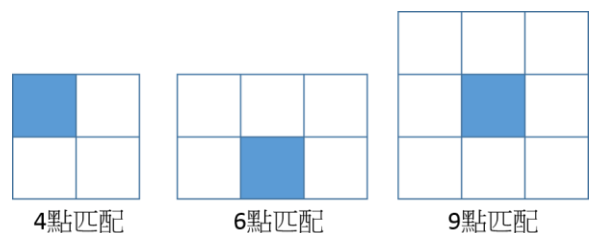


圖 1. 粗匹配網格點可能位置示意圖(改繪自李

楨、譚先科，2015)。

2.2.2 定位解算階段

在求得所有 AP 的模型參數 (a_i, b_i, c_i) 後，即可利用在未知點 (x, y) 所蒐集到的 i 個 AP 訊號及每個 AP 透過前述步驟所建立的數學模型寫下(6)式[1]：

$$\begin{bmatrix} r_1 \\ \vdots \\ r_i \end{bmatrix} = \begin{bmatrix} a_1 x + b_1 y + c_1 xy \\ \vdots \\ a_i x + b_i y + c_i xy \end{bmatrix} \quad (6)$$

(x, y) 為未知點平面坐標，單位為公尺； (a_i, b_i, c_i) 為已知的第 $i(i=1, \dots, m)$ 個 AP 的模型參數(已在 RSSI 建模階段求得)，無單位； r_i 為未知點所收到的第 $i(i=1, \dots, m)$ 個 AP 的訊號強度(RSSI)，單位為 dBm，每個 AP 可列出一條觀測方程式，若有 2 個以上的 AP 即可使用已知的 AP 模型參數 (a_i, b_i, c_i) 與未知點 AP 訊號 r_i 解算 x 、 y 兩未知數。(6)式的觀測方程式，需要求解的對象為未知點坐標，因此方程式為非線性，將方程式線性化後寫成矩陣型式如(7)式與(8)式所示，初值 (x_0, y_0) 由粗匹配的點所給定，透過(9)式求解改正數 $(\Delta x, \Delta y)$ ，並如(10)式所示更新後的初值，迭代計算至改正數 $(\Delta x, \Delta y)$ 皆小於 0.001m 為止[1]，最終更新的初值 (x_0, y_0) ，即為未知點的坐標值(註：(8)式的 F_i ，代表未知平面點坐標與訊號強度之間的函數關係，如(6)式第 1 列， $F_1 = r_1 = a_1 x + b_1 y + c_1 xy$ 。

$$JX = K + V \quad (7)$$

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \vdots & \vdots \\ \frac{\partial F_i}{\partial x} & \frac{\partial F_i}{\partial y} \end{bmatrix} = \begin{bmatrix} a_1 + c_1 y_0 & b_1 + c_1 x_0 \\ \vdots & \vdots \\ a_i + c_i y_0 & b_i + c_i x_0 \end{bmatrix}$$

$$K = \begin{bmatrix} r_1 - a_1 x_0 - b_1 y_0 - c_1 x_0 y_0 \\ \vdots \\ r_i - a_i x_0 - b_i y_0 - c_i x_0 y_0 \end{bmatrix} \quad (8)$$

$$X = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, V = \begin{bmatrix} v_1 \\ \vdots \\ v_i \end{bmatrix}$$

$$X = (J^T J)^{-1} (J^T K)$$

(9)

$$x_0 = x_0 + \Delta x, y_0 = y_0 + \Delta y \quad (10)$$

2.3 最小一乘法原理

假設有 n 個未知數， m 個觀測方程式，且 $m > n$ ，今將觀測方程式以矩陣型態表示將如(11)式與(12)式所示， A 為設計矩陣，擺放各觀測方程式的參數係數， X 為未知參數向量， L 為觀測量向量， V 為殘差向量。

$$AX = L + V \quad (11)$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

$$L = \begin{bmatrix} L_1 \\ \vdots \\ L_m \end{bmatrix}, V = \begin{bmatrix} V_1 \\ \vdots \\ V_m \end{bmatrix} \quad (12)$$

最小一乘法的目標函數(S)即為殘差 V 的絕對值和最小，如(13)式所示，在面對觀測量含粗差的狀況下，則能降低粗差影響得到符合實際的模型[9,10]。

$$S = \sum_{i=1}^m |V_i| \Rightarrow \min \quad (13)$$

2.3.1 解決最小一乘法目標函數含有絕對值的方法

由於最小一乘法的目標函數具有絕對值不可微分的問題，因此在求解最小一乘法問題時，改由線性規劃方式來求解，然而因線性規劃中可行解需要皆為正值，若要應用在測量領域中，則無法應付負實數解的情形[9-12]。

為了解決此矛盾，將待求參數 x_i 表示為兩個非負實數 x_i^+ 及 x_i^- 。 x_i^+ 代表待求參數為正值時之非負實數， x_i^- 代表待求參數為負值時之非負實數。若 x_i 值為正，則 x_i^- 為 0；若 x_i 值為負，則 x_i^+ 為 0，因此，至少有一個值為 0，如(14)式所示[9-12]。

$$x_i \in R; \quad x_i^+ = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$x_i^- = \begin{cases} 0, & \text{if } x_i \geq 0 \\ -x_i, & \text{otherwise} \end{cases}; \quad x_i^+, x_i^- \geq 0$$

(14)

2.3.2 將觀測方程式化為線性規劃模型

將觀測方程式表示為 $AX=L+V$ 形式後，即可將最小一乘法的目標函數與觀測方程式轉化為線性規劃的模式，詳細內容如下所述 [9-12]：

最小一乘法之函數模型與目標函數：

$$\begin{cases} AX = L+V & \text{函數模型} \\ [P|V] = \min & \text{目標函數} \end{cases} \quad (15)$$

(15)式中的 P ，代表權矩陣，本研究在後續相關 WiFi 定位演算上皆假設為等權的情形，即 P 為單位矩陣(I)。線性規劃之約束條件及目標函數：

$$\begin{cases} Ax = b & \text{約束條件} \\ c^T x = \min & \text{目標函數} \end{cases}, x \geq 0$$

(16)

因線性規劃中，未知數(待求參數)必須為正值 $x_i \geq 0$ 。而最小一乘法之函數模型 $AX=L+V$ 中待求之參數(X)及殘差值(V)均未知正負。因此，透過(17)式，令 X 與 V 為兩個非負實數之差， $|V|$ 為兩個非負實數之和：

$$\begin{aligned} X &= X^+ - X^- \\ V &= V^+ - V^- \\ |V| &= V^+ + V^- \\ X^+, X^-, V^+, V^- &\geq 0 \end{aligned}$$

(17)

將最小一乘法函數模型 $AX=L+V$ 轉換為線性規劃約束條件 $Ax=b$ 模式，如(18)至(19)式，其中 P 為單位矩陣(I)：

$$AX = L+V \rightarrow AX - V = L$$

(18)

$$\begin{bmatrix} A & -A & -I & I \end{bmatrix} \begin{bmatrix} X^+ \\ X^- \\ V^+ \\ V^- \end{bmatrix} = [L]$$

(19)

接著，將最小一乘法之目標函數 $[P|V] = \min$ ，表示為線性規劃之目標函數 $c^T x = \min$ 形式，如(20)式：

$$c^T x = [P|V] = P^T \begin{bmatrix} V^+ & V^- \end{bmatrix} = \min$$

(20)

其中， c^T 與 x 分別為：

$$c^T = [0 \quad 0 \quad P \quad P], x = \begin{bmatrix} X^+ \\ X^- \\ V^+ \\ V^- \end{bmatrix}$$

(21)

在透過線性規劃求解後，則可求得 X^+, X^-, V^+, V^- ，若欲求得 X 與 V 則透過(17)式還原即可。

2.4 最小一乘法 WiFi 匹配定位

2.4.1 RSSI 建模階段

若將(4)式寫成線性規劃的約束條件將如(22)至(24)式所示，目標函數則如(25)式。透過(24)式約束條件與(25)式的目標函數，以線性規劃求解模型參數的最優化解，其中 I 為 $j \times j$ 大小的單位矩陣，亦即為等權 P 矩陣，並藉由 MATLAB 的 `linprog` 函數進行計算，最後依照(17)式將 x^+ 與 x^- 相減即可得到 AP 模型參數(a_i, b, c_i)，而將 v^+ 與 v^- 可得到 RSSI 建模階段的 RSSI 殘差值[13]。

$$AX - V = L$$

(22)

$$A = \begin{bmatrix} x_1 & y_1 & x_1 y_1 \\ \vdots & \vdots & \vdots \\ x_j & y_j & x_j y_j \end{bmatrix}, X = \begin{bmatrix} x^+ \\ x^- \\ v^+ \\ v^- \end{bmatrix}, L = \begin{bmatrix} r_{APi-1} \\ \vdots \\ r_{APi-j} \end{bmatrix}$$

$$x^+ = \begin{bmatrix} a^+ \\ b^+ \\ c^+ \end{bmatrix}, x^- = \begin{bmatrix} a^- \\ b^- \\ c^- \end{bmatrix}, v^+ = \begin{bmatrix} v_1^+ \\ \vdots \\ v_j^+ \end{bmatrix}, v^- = \begin{bmatrix} v_1^- \\ \vdots \\ v_j^- \end{bmatrix}$$

(23)

$$\begin{bmatrix} A & -A & -I & I \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \\ v^+ \\ v^- \end{bmatrix} = L$$

(24)

$$\begin{bmatrix} 0 & 0 & -I & I \end{bmatrix} \begin{bmatrix} x^+ \\ x^- \\ v^+ \\ v^- \end{bmatrix} = \min$$

(25)

例如，有 4 個參考點的 RSSI 觀測值要解算某一個 AP 的模型參數(a, b, c)，有關 A 矩陣、L 矩陣的內容如(26)式所示：

$$A = \begin{bmatrix} 10.04 & 10.00 & 100.41 \\ 9.99 & 12.0 & 119.91 \\ 12.01 & 9.98 & 119.84 \\ 11.99 & 12.00 & 143.89 \end{bmatrix}, L = \begin{bmatrix} -67.27 \\ -65.88 \\ -66.88 \\ -64.86 \end{bmatrix}$$

(26)

透過(24)式與(25)式與 A、L 矩陣，確定約束條件與目標函數之內容，並以 MATLAB 的 linprog 函數進行線性規劃解算，即可求得 x^+ 、 x^- 、 v^+ 、 v^- ，若以模型參數 x 為例，依照(17)式即可求得該 AP 的模型參數，如(27)式所示：

$$x^+ = \begin{bmatrix} 0 \\ 0 \\ 0.58 \end{bmatrix}, x^- = \begin{bmatrix} 6.40 \\ 5.91 \\ 0 \end{bmatrix}$$

$$x = x^+ - x^- = \begin{bmatrix} -6.40 \\ -5.91 \\ 0.58 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

(27)

2.4.2 定位解算階段

將(6)式方程式線性化後寫成線性規劃的約束條件如(28)至(30)式所示，其中 I 為 $i \times i$

大小的單位矩陣，目標函數則如(31)式，透過(30)式約束條件與(31)式目標函數以 MATLAB 的 linprog 函數求解坐標改正數 ($\Delta x, \Delta y$) 的最優化解，並依照(17)式將 X^+ 與 X^- 相減即可得到坐標改正數 ($\Delta x, \Delta y$)，而將 V^+ 與 V^- 相減即可得到定位解算階段的 RSSI 殘差值，使用(10)式更新初值 (x_0, y_0)，迭代計算至改正數小於 0.001m 為止，最終更新的初值，即為未知點的坐標值[13]。

$$JX - V = K$$

(28)

$$X^+ = \begin{bmatrix} \Delta x^+ \\ \Delta y^+ \end{bmatrix}, X^- = \begin{bmatrix} \Delta x^- \\ \Delta y^- \end{bmatrix},$$

$$V^+ = \begin{bmatrix} v_1^+ \\ \vdots \\ v_i^+ \end{bmatrix}, V^- = \begin{bmatrix} v_1^- \\ \vdots \\ v_i^- \end{bmatrix}, X = \begin{bmatrix} X^+ \\ X^- \\ V^+ \\ V^- \end{bmatrix}$$

(29)

$$J = \begin{bmatrix} a_1 + c_1 y_0 & b_1 + c_1 x_0 \\ \vdots & \vdots \\ a_i + c_i y_0 & b_i + c_i x_0 \end{bmatrix}$$

$$K = \begin{bmatrix} r_1 - a_1 x_0 - b_1 y_0 - c_1 x_0 y_0 \\ \vdots \\ r_i - a_i x_0 - b_i y_0 - c_i x_0 y_0 \end{bmatrix}$$

$$\begin{bmatrix} J & -J & -I & I \end{bmatrix} \begin{bmatrix} X^+ \\ X^- \\ V^+ \\ V^- \end{bmatrix} = K$$

(30)

$$\begin{bmatrix} 0 & 0 & -I & I \end{bmatrix} \begin{bmatrix} X^+ \\ X^- \\ V^+ \\ V^- \end{bmatrix} = \min$$

(31)

例如，有 4 組 AP 的模型參數已知，要解算某檢核點坐標的改正數，J 矩陣、K 矩陣的內容如(32)式所示：

$$J = \begin{bmatrix} -2.04 & 0.85 \\ 1.13 & -1.53 \\ 0.18 & -0.09 \\ -0.07 & 0.54 \end{bmatrix}, L = \begin{bmatrix} -17.79 \\ -19.08 \\ 3.52 \\ 6.49 \end{bmatrix}$$

(32)

透過(30)式與(31)式與 A、L 矩陣，確定約束條件與目標函數之內容，並以 MATLAB

的 linprog 函數進行線性規劃解算即可求得 X^+ 、 X^- 、 V^+ 、 V^- ，依照(17)式即可求得檢核點坐標改正數 X ，如(33)式所示：

$$X^+ = \begin{bmatrix} 0.83 \\ 1.08 \end{bmatrix}, X^- = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$X = X^+ - X^- = \begin{bmatrix} 0.83 \\ 1.08 \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (33)$$

三、實驗資料與方法

3.1 實驗資料

實驗資料，包含模擬資料、室內資料及室外資料；參考點與檢核點的長寬間距皆為 2 公尺。WiFi 訊號部分，模擬資料由模擬所得；室內資料及室外資料，以 Vistumbler 軟體蒐集東西南北每方向各 20 秒，然後取其平均值。

3.1.1 模擬資料

假設參考點與檢核點的坐標是由 GPS-RTK 所測得，平面 X、Y 方向的中誤差皆為 $\pm 0.02\text{m}$ [14]，並以 2 公尺間距設立正方形網格建立參考點與檢核點；在 WiFi 訊號強度的部分，則是參考 Li et al.(2011)[15]的訊號衰減模型如(34)式所示；其中， $P(d)$ 代表接收機所接收的訊號功率，單位為分貝毫瓦(dBm)； d 為參考點與訊號發射源之間的距離，單位為公尺。

$$P(d) = -45 - 10 \times 2 \times \log(d) \quad (34)$$

模擬資料含 36 個參考點、25 個檢核點以及 4 個訊 AP 號源，如圖 2 所示。其中，參考點資料(含點位平面坐標、接收相關 AP 的 RSSI 等)，用於 WiFi 匹配定位之 RSSI 建模階段，求取相關 AP 的模型參數；檢核點資料，則用於定位解算階段，以分析定位精度。

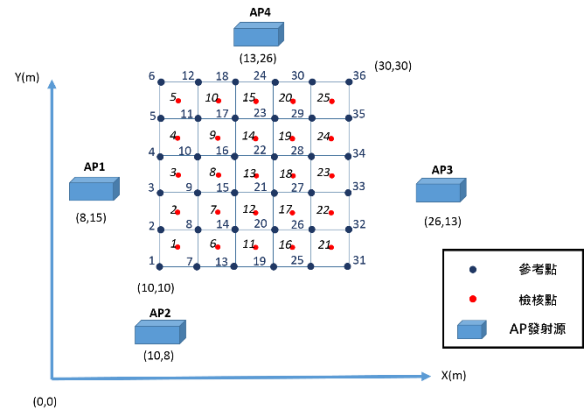


圖 2. 模擬資料之參考點、檢核點以及 AP 訊號源略圖。

3.1.2 室內資料

以政治大學綜合院館 6 樓之廊道為室內實驗區，其像片圖如圖 3 所示。室內資料含 22 個參考點與 10 個檢核點，如圖 4 所示。實驗區的大小為 20×2 平方公尺，參考點與檢核點的坐標以全測站測得，WiFi 訊號資料蒐集日期為 2017 年 3 月 21 日夜間，坐標系統為自行建立的平面坐標系。

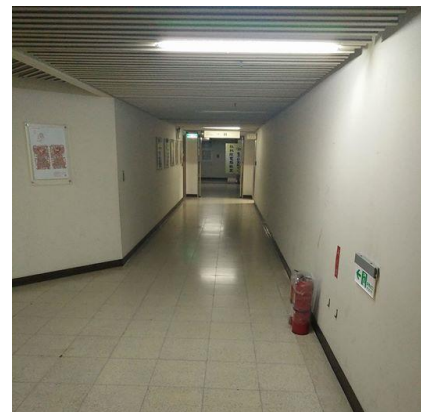


圖 3. 室內實驗區像片圖。

3.1.3 室外資料

選取國立政治大學山下校區噴水池廣場做為室外實驗區域，其像片圖如圖 5 所示。室外資料，如圖 6 所示，其大小為 20×8 平方公尺，參考點與檢核點數量分別為 55 點及 40 點，坐標以全測站測得；WiFi 資料蒐集日期為 2016 年 10 月 30 日白天，坐標系統為自行建立的平面坐標系。

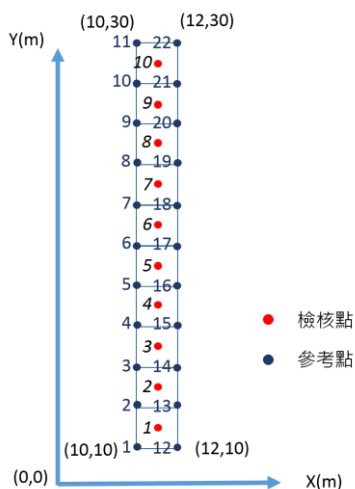


圖 4. 室內資料之參考點與檢核點略圖。



圖 5. 室外實驗區像片圖。

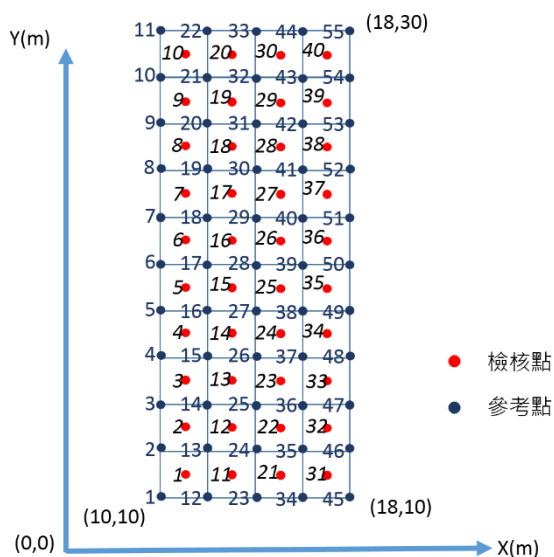


圖 6. 室外資料之參考點與檢核點略圖。

3.2 實驗方法

3.2.1 模擬資料實驗方法

- 檢定 WiFi 訊號強度與參考點平面坐標之函數關係：在無粗差之情形，使用 AP1 的訊號強度值與 36 點參考點之資料來進行相關函數關係統計檢定。
- 最小一乘法 WiFi 匹配定位之穩健性與偵錯能力測試：利用模擬資料，透過下列 3 種粗差模擬情形來測試最小一乘法 WiFi 匹配定位之穩健性與偵錯能力，並與最小二乘法 WiFi 匹配定位成果進行比較。粗差模擬的 3 種情境：
 - 無粗差之狀況(僅有偶然誤差)。
 - 某些參考點接收之 AP 訊號含粗差：在 RSSI 建模階段，分別於不同參考點接收的某一 AP 訊號加入粗差，亦即針對 RSSI 建模階段 $AX = L + V$ 的 L 矩陣加入訊號強度粗差。
 - 某些檢核點接收 AP 訊號含粗差：在定位解算階段，分別於不同檢核點接收的某一 AP 訊號加入粗差，亦即針對定位解算階段 $JX = K + V$ 的 K 矩陣加入訊號強度粗差。

3.2.2 室內與室外資料實驗方法

分別以室內資料與室外資料，測試最小一乘法 WiFi 匹配定位之穩健性與偵錯能力，並與最小二乘法 WiFi 匹配定位與 KNN 演算法的定位精度比較。

3.3 定位精度檢核

下列實驗之定位精度檢核，採用檢核點的平面坐標、位置均方根誤差(Root Mean Square Error, RMSE)作為評估指標，有關均方根誤差之定義如下[8,9,13]：

$$\Delta X_i = X_{WiFi,i} - X_{check,i} \quad (35)$$

$$\Delta Y_i = Y_{WiFi,i} - Y_{check,i} \quad (36)$$

$$\Delta P_i = \sqrt{\Delta X_i^2 + \Delta Y_i^2} \quad (37)$$

$$RMSE_x = \sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n \Delta X_i^2} \quad (38)$$

$$RMSE_y = \sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^n \Delta Y_i^2}$$

(39)

$$RMSE_p = \sigma_p = \sqrt{\sigma_x^2 + \sigma_y^2}$$

(40)

式中, $i=1, 2, \dots, n$, n 為檢核點數量。 $X_{WiFi,i}$ 、 $Y_{WiFi,i}$ 代表檢核點 i 經 WiFi 定位得到的 X 、 Y 坐標值。 $X_{check,i}$ 、 $Y_{check,i}$ 為檢核點 i 已知的平面 X 、 Y 坐標值。 ΔX_i 、 ΔY_i 分別為檢核點 i 在 X 、 Y 方向坐標值的較差。 ΔP_i 為檢核點 i 的平面誤差。 σ_x 、 σ_y 分別為檢核點 X 、 Y 方向較差之均方根誤差, σ_p 為檢核點位置誤差之均方根誤差。

四、實驗成果與討論

4.1 檢定 WiFi 訊號強度與參考點平面坐標之函數關係

最小二乘法 WiFi 匹配定位所使用的模型如(2)式, 由李楨、譚先科(2015)提出。該模型代表 WiFi 訊號強度 r_{ij} 與參考點平面坐標 (x_j, y_j) 的非線性函數關係; 然而, 該文獻並未對於訊號強度與地面點平面坐標模型之解釋能力進行檢定。

因此, 使用模擬資料含 36 個參考點與 AP1, 來對最小二乘法 WiFi 匹配定位 RSSI 建模階段之模型進行統計檢定。AP1 與 36 個參考點之間的關係可參考圖 2。透過 R^2 與 F 統計檢定對模型進行測試。由 R^2 來評估坐標值 (x, y) 對於訊號強度值 (γ) 的配合度高低; 另外透過 F 統計檢定, 在 $\alpha=0.05$ 的信心水準下, 驗證模型是否具有解釋能力? 若 F 檢定結果足以拒絕虛無假設 H_0 , 即自變數「坐標值 (x, y) 」對應變數「訊號強度 (γ) 」具有顯著的聯合解釋能力[13]。

此處的 F 檢定, 乃是對於回歸模型解釋能力的檢定; 而測量領域常用的 F 檢定, 是對於 2 個或者 2 個以上的樣本變異數差異性檢定[16]。兩者所假設的虛無假設(H_0)與對立假設(H_1)內容不同。對於回歸模型解釋能力的 F 檢定, 是期望能夠拒絕 H_0 , 代表模型具有解釋能力。而樣本變異數差異性的 F 檢定則是希望能夠接受 H_0 , 代表樣本之間的無顯著的差異性。

採取粗匹配之方法, 則模型參數在求解時, 依據粗匹配的位置, 將有 4、6、9 個觀測量等 3 種可能, 利用前述模擬實驗區選取 4、6、9 個參考點進行模型參數推估, 選取粗匹配資料位置圖如圖 7 所示。

模型粗匹配的統計檢定結果如表 1 所示, 除了在 4 點匹配區的 F 檢定無法通過外, 6 點與 9 點匹配之 F 檢定皆通過; 顯示若採用 6 點與 9 點匹配, 「坐標值 (x, y) 」對「訊號強度 (γ) 」具有顯著的聯合解釋能力。另外, 就 R^2 而言, 三種粗匹配的 R^2 皆大於 0.60, 顯示坐標值 (x, y) 對於訊號強度值 (γ) 的配合度高。

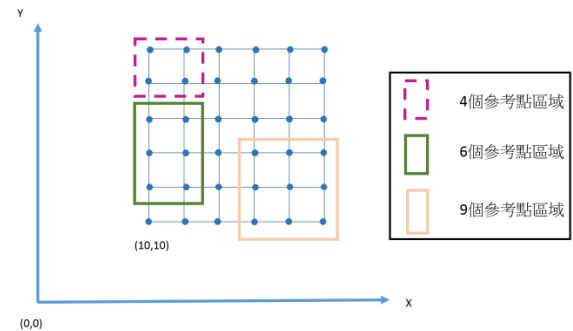


圖 7. 選取粗匹配資料位置圖。

表 1. 粗匹配之統計檢定結果

粗匹配建模	4 點匹配	6 點匹配	9 點匹配
判定係數 R^2	0.98	0.87	0.64
F 檢定統計量	29.12	13.73	11.62
臨界值 F	215.71	9.28	4.76
檢定結果	接受 H_0 (無顯著聯合解釋能力)	拒絕 H_0 (有顯著聯合解釋能力)	拒絕 H_0 (有顯著聯合解釋能力)

4.2 最小一乘法 WiFi 匹配定位之精度

為測試最小一乘法 WiFi 匹配定位之精度, 透過模擬、室內與室外資料, 採用粗匹配來進行最小一乘法 WiFi 匹配定位(以後簡稱 LAD)、最小二乘法 WiFi 匹配定位(以後簡稱 LS)與 KNN 演算法 WiFi 定位(以後簡稱 LS)。所謂 KNN 演算法, 為 WiFi 訊號紋法之一,

是選取 K 個最臨近點來進行加權平均，計算出待測點(如檢核點)的位置；本研究採用 K 值為 3 的設定，即代表選取 3 個最臨近點來進行加權平均的計算，加權平均的方法為歐基里德距離倒數[13]。

4.2.1 模擬資料

在無粗差狀況下，採用 LS 或 LAD 後，模擬實驗區中 25 個檢核點的 σ_x 、 σ_y 、 σ_p 成果，如表 2 所示。可由表中成果發現在無粗差的狀況下，LS 的定位精度比 LAD 稍佳。

經由變異數分析 F 檢定後，LAD 與 LS 之 σ_p 值雖然不同，但在顯著水準 $\alpha=0.05$ 的情況下，沒有排斥兩者定位精度是相同的假設。因此，在無粗差的情形下，使用 LAD 與 LS 的精度並無太大的差異。

表 2. 模擬資料，在無粗差狀況下，使用 LAD 與 LS，25 個檢核點的精度統計表

定位方法	σ_x (m)	σ_y (m)	σ_p (m)
LAD	0.23	0.38	0.44
LS	0.12	0.26	0.28

4.2.2 室內資料

利用室內資料，10 個檢核點之 LAD、LS 及 KNN 之定位結果，如表 3 所示。各檢核點採用 LS、LAD、KNN 之平面誤差 Δ_p ，如圖 8 所示。從表 3 與圖 8 結果可見：(1)室內 WiFi 定位精度都在 3.00 公尺以內。(2)就定位精度而言，LAD 最好，其次是 LS，最後是 KNN。

表 3. 室內資料，10 個檢核點之 KNN、LS 及 LAD 定位結果

定位方法	KNN	LS	LAD
σ_p (m)	2.69	2.41	2.20

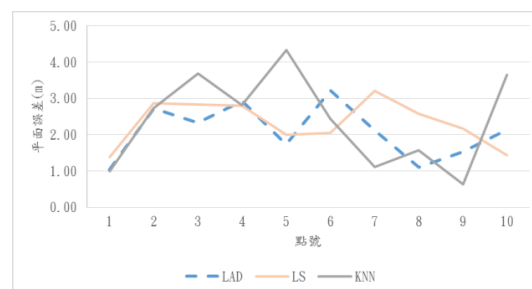


圖 8. 室內資料，各檢核點採用 LAD、LS 及 KNN 之平面誤差。

4.2.3 室外資料

40 個檢核點的 LS、LAD 及 KNN 定位成果，如表 4 所示。各檢核點採用 KNN、LAD 及 LS 之平面誤差，如圖 9 所示。從表 4 與圖 9 結果可見：(1)室外 WiFi 定位精度都在 9.00 公尺以內。(2)就定位精度而言，KNN 最好，其次是 LAD，最後是 LS。

如圖 10 所示，為室外資料，LS、LAD、KNN 平面誤差累積分佈函數圖。在 50% 的機率下 KNN 的誤差可在 5.37 公尺內，LAD 在 8.31 公尺內，LS 則在 8.55 公尺內。

表 4. 室外資料，採用 LS、LAD、KNN 後 40 個檢核點的定位成果

	LS	LAD	KNN
σ_p (m)	8.59	8.46	7.40
最小平面誤差(m)	2.93	2.68	0.37
最大平面誤差(m)	13.63	13.08	12.44

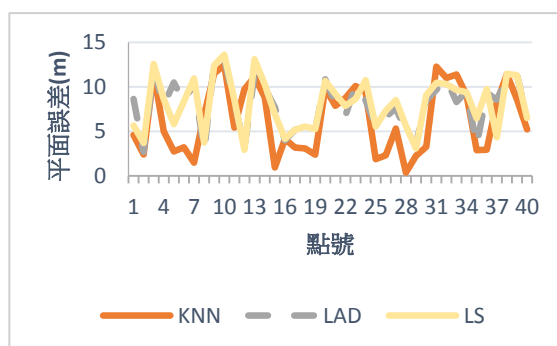


圖 9. 室外資料，各檢核點採用 KNN、LAD 及 LS 之平面誤差。

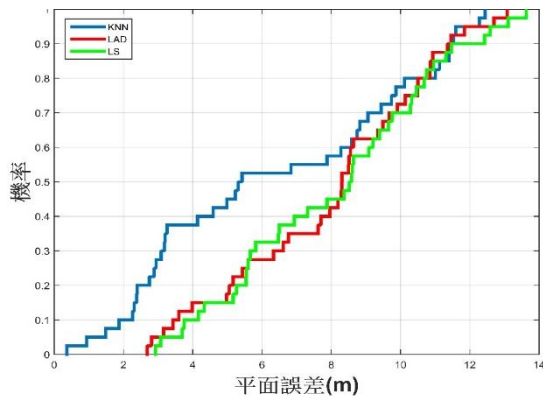


圖 10. 室外資料，LS、LAD、KNN 平面誤差累積分佈函數圖。

4.2.4 小結

根據上述 LAD 精度之實驗結果顯示:(1) 在無粗差的狀況下，模擬資料的 LAD 與 LS 精度相當；室內或室外資料，LAD 的定位精度皆優於 LS。(2)LAD 若與 KNN 相較，在室內資料 LAD 的定位精度優於 KNN；然而，在室外資料 KNN 的定位精度優於 LAD。可能原因，由於在室外環境各參考點所蒐集到的 AP 訊號強度、與 AP 的距離等跟模擬資料環境不相同；故在 RSSI 模型推估時，無法產生良好的 AP 模型參數。此外，AP 的模型參數可能尚有其他因子未考量。(3)LAD 的室內、室外定位精度分別為 2.20、8.46 公尺。

4.3 最小一乘法 WiFi 匹配定位之穩健性

本節主要在分別 LAD、LS 的 RSSI 建模階段及定位解算階段的 RSSI 觀測量，故意加入一定大小的粗差，然後，觀察 LAD、LS 的定位精度是否會受到粗差的影響？換言之，由此測試 LAD 或 LS 具有穩健性。

4.3.1 參考點接收 AP 訊號含有粗差之狀況

4.3.1.1 模擬資料

依序於不同參考點接收的不同訊號源 (AP) 加入粗差進行計算；換言之，對 RSSI 建模階段 $AX = L + V$ 的 L 向量加入訊號強度粗差，以測試 LAD 的穩健性。由於模擬資料的 AP 有 4 個，分別於不同參考點接收的不同 AP 訊號加入粗差，假定每個 AP 的粗差只存在一個參考點上，故粗差數最多為 4 個，加入的

粗差量為 -8dBm。

假設粗差量為 -8dBm 的主要原因，依據 (34) 式訊號衰減模型進行誤差傳播公式的推導，可推求各參考點與檢核點的訊號強度的中誤差值，模擬實驗區最大與最小訊號中誤差值的絕對值，分別為 0.54 與 0.09 dBm。若將 3 倍中誤差之值視為粗差，則代表所加入的訊號強度值大於 1.62dBm ($=0.54 \times 3$)，即可視為粗差；另外，由於本文於真實環境中，以人體作為障礙物，阻擋接收機接收 AP 訊號，所造成的訊號衰減量為 8dBm，大於模擬資料 3 倍中誤差的標準，符合視為粗差的條件，故以 -8dBm 為粗差的加入量，粗差分佈以及加入順序如表 5 所示。

表 6 所示，為參考點接收之 AP 訊號含數個粗差時，使用 LS 與 LAD，25 個檢核點精度比較表。參考點接收的 AP 訊號強度含有粗差狀況，使用 LS 與 LAD 之檢核點 σ_p 比較圖，如圖 11 所示。

從表 6 與圖 11 結果可見：(1) LAD 的定位精度，不會隨著粗差的增加而有明顯的下降，大致維持在 0.56 公尺左右。(2) LS 的定位精度，則隨著粗差的增加而有明顯的下降，從 0.28 公尺下降至 0.87 公尺。

經由變異數分析 F 檢定後，在第 4 個參考點接收 AP 訊號粗差加入時，LAD 與 LS，在顯著水準 $\alpha=0.05$ 的情況下，拒絕兩者定位精度是相同的假設。因此，證實在參考點含有接收 AP 訊號粗差的情形下，使用 LAD 與 LS 的定位精度是有差異性的。

表 5. 模擬參考點對接收 AP 訊號粗差之分佈及加入順序

加入順序	參考點號	AP 訊號源	粗差
1	2	2	-8dBm
2	35	3	-8dBm
3	18	4	-8dBm
4	10	1	-8dBm

表 6. 模擬資料，參考點接收之 AP 訊號含數個粗差時，使用 LS 與 LAD，25 個檢核點精度比較(單位：公尺)

	LS			LAD		
粗差數量	σ_x	σ_y	σ_p	σ_x	σ_y	σ_p

1	0.21	0.64	0.67	0.30	0.46	0.55
2	0.28	0.64	0.70	0.30	0.48	0.56
3	0.38	0.73	0.82	0.30	0.44	0.53
4	0.48	0.72	0.87	0.32	0.44	0.55

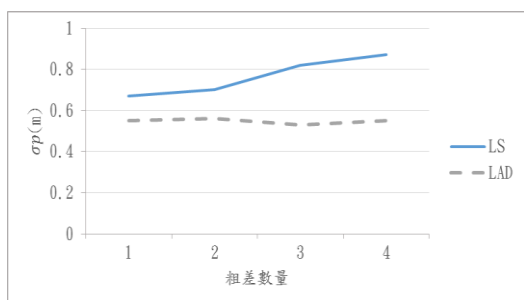


圖 11. 模擬資料，參考點接收的 AP 訊號強度含有粗差狀況，使用 LS 與 LAD 之檢核點 σ_p 比較圖。

4.3.1.2 室內資料

表 7 所示，為參考點 20 接收的 AP6 加入 -8dBm 前後，檢核點 9 之平面誤差比較表。從表 7 結果顯示：(1) 利用 LAD，檢核點 9 之平面誤差，不會隨著粗差的加入而有明顯的下降，皆維持在 1.53 公尺左右。(2) 利用 LS，檢核點 9 之平面誤差，則會隨著粗差的加入而有明顯的下降，從 2.17 公尺下降至 2.30 公尺。

請注意，在真實 WiFi 資料裡，WiFi 匹配定位，不需要知道各 AP 的位置，只需要知道各 AP 的 MAC(Media Access Control)位址(Address)及其訊號強度(RSSI)。

表 7. 室內資料，參考點 20 接收的 AP6 加入 -8dBm 前後，檢核點 9 之平面誤差比較 (單位：公尺)

	LS	LAD
粗差加入前	2.17	1.53
粗差加入後	2.30	1.53

4.3.1.3 室外資料

表 8 所示，為參考點 8 接收的 AP1 加入 -8dBm 前後，檢核點 40 之平面誤差(m)比較表。從表 8 結果顯示：(1) 利用 LAD，檢核點 40 之平面誤差，不會隨著粗差的加入而有明顯的下降，皆維持在 6.33 公尺左右。(2) 利

用 LS，檢核點 40 之平面誤差，則會隨著粗差的加入而有明顯的變大，從 6.50 公尺上升至 6.61 公尺。

表 8. 室外資料，參考點 8 接收的 AP1 加入 -8dBm 前後，檢核點 40 之平面誤差比較 (單位：公尺)

	LS	LAD
粗差加入前	6.50	6.33
粗差加入後	6.61	6.33

4.3.1.4 小結

根據上述在參考點接收 AP 訊號(在 RSSI 建模階段)含有粗差的穩健性實驗結果顯示：(1) 不管是模擬資料、室內資料、或室外資料，LAD 皆有很好的穩健性，LS 則無。(2) 使用模擬資料，即使在含有 4 個粗差的情況下，LAD 仍然具有很好的穩健性。

4.3.2 檢核點接收 AP 訊號含有粗差之狀況

4.3.2.1 模擬資料

依序於不同檢核點的不同訊號源加入 -8dBm 的粗差進行計算；即針對定位解算階段 $JX = K + V$ 的 K 矩陣加入訊號強度粗差，測試最小一乘法的穩健性，粗差分佈以及加入順序如表 9 所示。

檢核點接收之 AP 訊號強度含粗差狀況，使用 LS 與 LAD 之精度比較，如表 10 所示。圖 12 所示，為檢核點接收的 AP 訊號強度含有粗差狀況，使用 LS 與 LAD 之 σ_p 比較圖。

從表 10 與圖 12 的結果顯示：(1) 檢核點接收 AP 訊號含有粗差之狀況，LAD 的定位精度皆優於 LS 的定位精度。(2) 當粗差數量小於或等於 2 時，LAD 是穩健的。

表 9. 模擬檢核點對接收 AP 訊號粗差之分佈及加入順序

加入順序	檢核點號	AP 訊號源	粗差
1	2	2	-8dBm
2	10	3	-8dBm
3	24	4	-8dBm
4	18	1	-8dBm

表 10. 模擬資料，檢核點接收之 AP 訊號強度含粗差狀況，使用 LS 與 LAD，25 個檢核點之精度比較(單位：公尺)

粗差數量	LS			LAD		
	σ_x	σ_y	σ_p	σ_x	σ_y	σ_p
1	0.42	0.47	0.63	0.34	0.40	0.53
2	0.45	0.83	0.95	0.39	0.41	0.56
3	0.46	1.39	1.47	0.40	1.15	1.22
4	1.05	1.60	1.92	1.10	1.49	1.85

4.3.2.2 室內資料

表 11 所示，為檢核點 1 接收的 AP5 加入 -8dBm 粗差前後之平面誤差比較。從表 11 的結果顯示：(1) LAD 的定位精度優於 LS 的定位精度(請注意：本實驗結果來自單一檢核點，沒有多餘觀測，因此，無法進行差異顯著性統計檢定)。(2)當粗差數量等於 1 時，LAD 是穩健的。

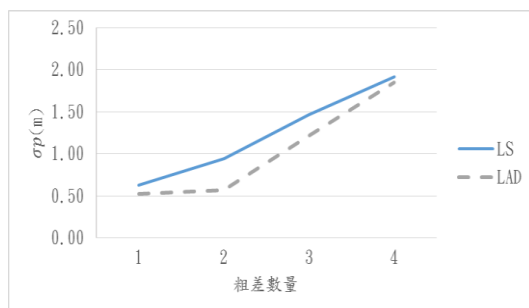


圖 12. 模擬資料，檢核點接收的 AP 訊號強度含有粗差狀況，使用 LS 與 LAD 之 σ_p 比較圖。

表 11. 室內資料，檢核點 1 接收的 AP5 加入 -8dBm 粗差前後之平面誤差比較(單位：公尺)

	LS	LAD
粗差加入前	1.39	1.04
粗差加入後	2.07	1.04

4.3.2.3 室外資料

表 12 所示，為檢核點 4 接收的 AP5 加入 -8dBm 粗差前後之平面誤差比較。從表 12 的結果顯示：(1) LAD 的定位精度優於 LS 的定

位精度(請注意：本實驗結果來自單一檢核點，沒有多餘觀測，因此，無法進行差異顯著性統計檢定)。(2)當粗差數量等於 1 時，LAD 是穩健的。

表 12. 室外資料，檢核點 4 接收的 AP5 加入 -8dBm 粗差前後之平面誤差比較(單位：公尺)

	LS	LAD
粗差加入前	6.50	6.33
粗差加入後	6.61	6.33

4.3.2.4 小結

根據上述在檢核點接收 AP 訊號(在定位解算階段)含有粗差的穩健性實驗結果顯示：(1)當模擬資料的粗差數量小於或等於 2、室內資料及室外資料的粗差數量等於 1 時，LAD 皆有很好的穩健性。然而，當粗差數量大於 2 時，在模擬資料的 LAD 則無穩健性；可能原因為定位解算階段的 K 向量，其為 RSSI 值、AP 的模型參數、待測點的坐標近似值等之函數，其較 RSSI 建模階段的 L 向量(其指包含 RSSI 值)的複雜度高。(2)根據林老生與林怡君(2014)[10]，利用最小一乘法實施地籍坐標轉換的研究結果顯示，若用 24 個參考點(可列 48 個觀測方程式)，當粗差數量小於 21 時，最小一乘法有穩健性，成功率約 44%。因此，LAD 的穩健性與粗差的數量有關，概略估計若粗差數量大於觀測方程式的 50% 時，LAD 的穩健性則可能失效。(3)不管是模擬資料、室內資料、或室外資料，LAD 的定位精度皆優於 LS 定位精度。

4.4 最小一乘法 WiFi 匹配定位之偵錯能力

4.4.1 參考點接收 AP 訊號含有粗差之狀況

4.4.1.1 模擬資料

由於粗匹配之緣故，導致並非所有的參考點皆會參與模型參數的計算。故接下來將針對受粗差影響的檢核點、檢核點在計算中所使用的參考點、受粗差影響的 AP、LS 與 LAD 的 RSSI 殘差進行說明。

以參考點 18 的 AP4(參考圖 2)加入-8 dBm 粗差為例來進行說明。圖 13 至圖 16，分別代表檢核點 5、10、14 及 15 利用 LS 及 LAD 定位時相關參考點之 RSSI 殘差分析圖(註:所有殘差已經過絕對值處理)。

由圖 13 至圖 16 的成果顯示:(1) 若使用 LAD 定位，在 4 個檢核點上，參考點 18 的殘差值皆為所有殘差值的最大項，且與粗差量較為接近。因此，以 LAD 法，可以根據參考點的最大 RSSI 殘差值，找出粗差所在。(2) 以 LS 法，無法完全根據參考點的最大 RSSI 殘差值，找出粗差所在。

4.4.1.2 室內資料

若沒有粗差存在時，各參考點的 RSSI 殘差值並不會有特別突出的現象。以檢核點 9 為例，在其定位解算階段，會用到 AP6 的模型參數及 RSSI；在 RSSI 建模階段，以 LAD 求解 AP6 的模型參數時，會用到 9、20、10、21、8 及 19 等 6 個參考點的 RSSI 資料。若無粗差時，上述 6 個參考點之 RSSI 殘差並不會有特別突出的，如圖 17 所示。

若在參考點 20 接收的 AP6 之 RSSI 值加入-8dBm 的粗差，則可發現在 RSSI 建模階段中，參考點 20 的 RSSI 殘差為最大值，且其殘差值接近粗差的加入量-8dBm，如圖 18 所示。

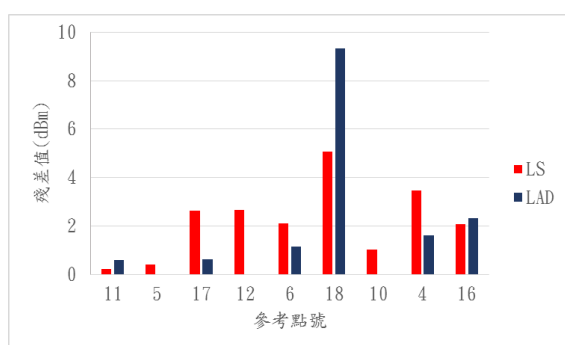


圖 13. 模擬資料，檢核點 5 相關參考點之 RSSI 殘差分析圖。

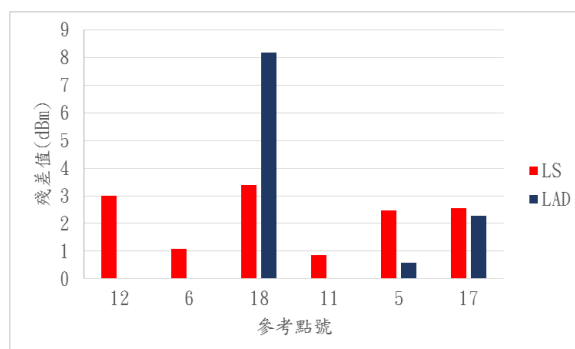


圖 14. 模擬資料，檢核點 10 相關參考點之 RSSI 殘差分析圖。

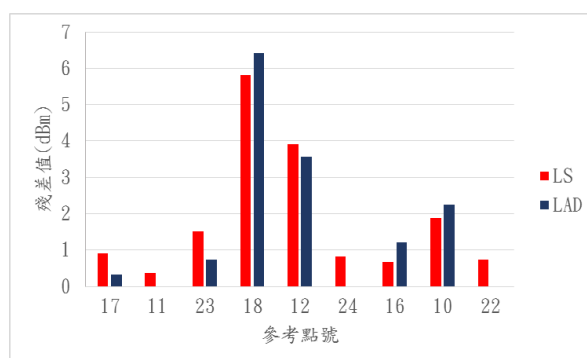


圖 15. 模擬資料，檢核點 14 相關參考點之 RSSI 殘差分析圖。

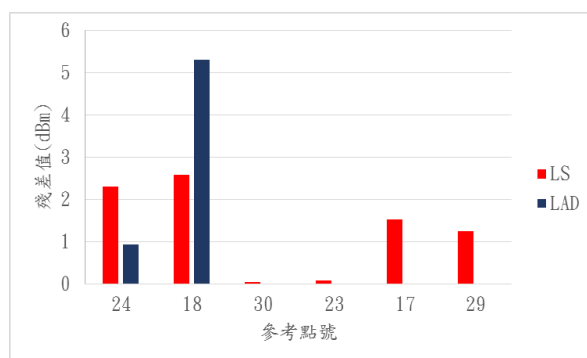


圖 16. 模擬資料，檢核點 15 相關參考點之 RSSI 殘差分析圖。

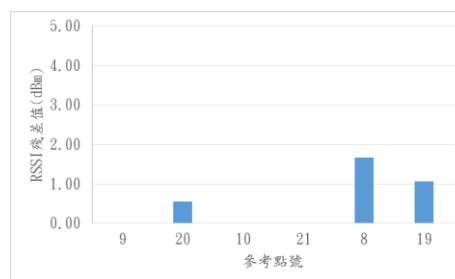


圖 17. 室內資料，無粗差時，以 LAD 求解 AP6 模型參數的 6 個參考點 RSSI 殘差圖。

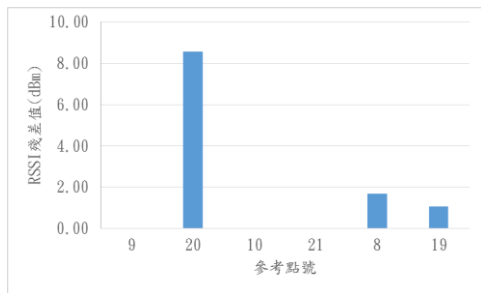


圖 18. 室內資料，在 AP6 加入-8dBm 粗差後，以 LAD 求解 AP6 模型參數的 6 個參考點 RSSI 殘差圖。

4.4.1.3 室外資料

以檢核點 40 為例，在 RSSI 建模階段，以 LAD 求解 AP1 的模型參數時，會用到 9、20、10、21、8 及 19 等 6 個參考點的 RSSI 資料。若無粗差時，上述 6 個參考點之 RSSI 殘差不會有特別突出的，如圖 19 所示。

若在真實資料參考點 8 接收的 AP1 之 RSSI 值加入-8dBm 的粗差，則可發現在 RSSI 建模階段中，參考點 8 的 RSSI 殘差為最大值，且其殘差值接近粗差的加入量-8dBm，如圖 20 所示。

4.4.1.4 小結

根據上述 LAD 之 RSSI 建模階段偵錯能力的實驗結果顯示:(1)不管是模擬資料、室內資料、或室外資料，LAD 皆有很好的偵錯能力。(2)LAD 可以根據參考點最大的 RSSI 殘差值判斷粗差的位置。

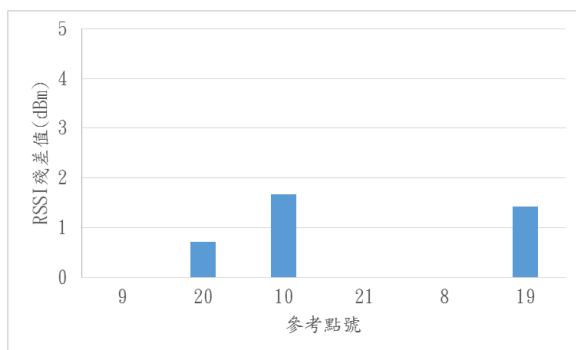


圖 19. 室外資料，無粗差時，以 LAD 求解 AP1 模型參數的 6 個參考點 RSSI 殘差圖。

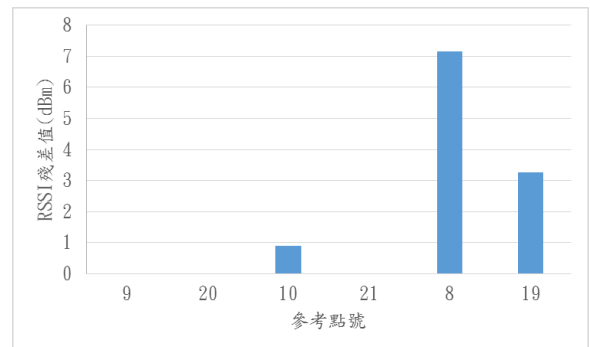


圖 20. 室外資料，在 AP1 加入-8dBm 粗差後，以 LAD 求解 AP1 模型參數的 6 個參考點 RSSI 殘差圖。

4.4.2 檢核點接收 AP 訊號含有粗差之狀況

4.4.2.1 模擬資料

表 13 所示，為 LS、LAD 於定位解算階段加入-8 dBm 粗差後相關之 RSSI 殘差分析表，黃底的部分表示在該檢核點加入粗差的 AP 位置。圖 21 及 22，分別代表檢核點 2、18 接收相關 AP 之 RSSI 殘差圖。

由表 13 及圖 21 至圖 22 的成果顯示:在定位解算階段，若使用 LAD 定位，當粗差數量小於或等於 2 時，可由各 AP 的殘差最大值來判斷模擬資料粗差的位置；此特性與前述穩健性的實驗結果相似。

4.4.2.2 室內資料

在未加入粗差前，檢核點 1 在 LAD 定位解算階段，接收的各 AP 之 RSSI 殘差值，如圖 23 所示。從圖 23 可見，檢核點 1 在 LAD 定位解算坐標時，共用 AP1~AP9 等 9 個 AP 之模型參數及相關 RSSI 值；其中，AP5 的 RSSI 殘差值為 5.5dBm。

表 13. 模擬資料，LS、LAD 於定位解算階段加入-8 dBm 粗差後相關之 RSSI 殘差分析表 (單位：dBm)

	檢核點號	AP1	AP2	AP3	AP4
LS	2	1.48	-3.39	-2.12	-1.32
	10	0.31	-2.01	-2.08	2.46
	24	0.09	-2.77	-1.83	0.35
	18	-3.14	3.82	-1.44	-0.83

LAD	2	0	5.23	1.76	0
	10	0	0	4.14	-2.11
	24	2.67	-4.4	0	0
	18	0	2.72	1.97	0

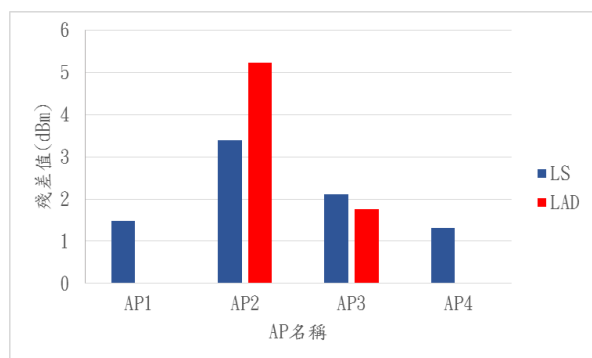


圖 21. 模擬資料，檢核點 2 接收相關 AP 之 RSSI 殘差圖。

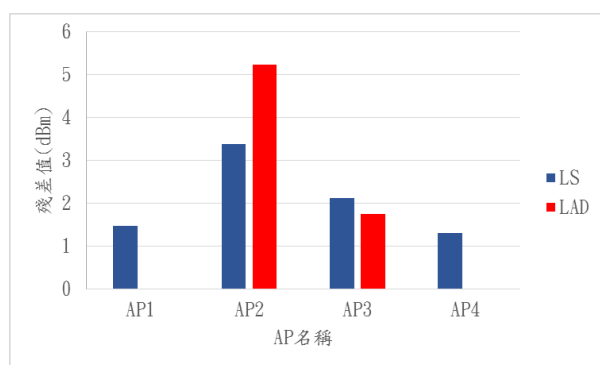


圖 22. 模擬資料，檢核點 18 接收相關 AP 之 RSSI 殘差圖。

若在 AP5 加入 -8dBm 的粗差，檢核點 1 在 LAD 定位解算階段，接收的各 AP 之 RSSI 殘差值，如圖 24 所示。從圖 24 可見，在 LAD 定位解算階段，AP5 的 RSSI 殘差值 13.5dBm 為 9 個殘差值中的最大值。因此，LAD 可由定位解算階段 AP 的 RSSI 殘差最大值來判斷粗差的位置。

4.4.2.3 室外資料

在未加入粗差前，檢核點 4 在 LAD 定位解算階段，接收的各 AP 之 RSSI 殘差值，如圖 25 所示。檢核點 4 在 LAD 定位解算坐標時，共用 AP1~AP7 等 7 個 AP 之模型參數及相關 RSSI 值。

若在檢核點 4 接收的 AP5 加入 -8dBm 的粗差，檢核點 4 在 LAD 定位解算階段，接收

的各 AP 之 RSSI 殘差值，如圖 26 所示。從圖 26 可見，在 LAD 定位解算階段，AP5 的 RSSI 殘差值 6.0dBm 為 7 個 RSSI 殘差值中的最大值。因此，可由 LAD 定位解算階段 AP 的 RSSI 殘差最大值來判斷粗差的位置。

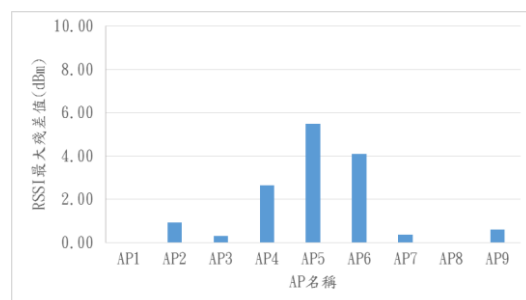


圖 23. 室內資料，未加入粗差前，檢核點 1 在定位解算階段，接收的各 AP 之 RSSI 殘差值。

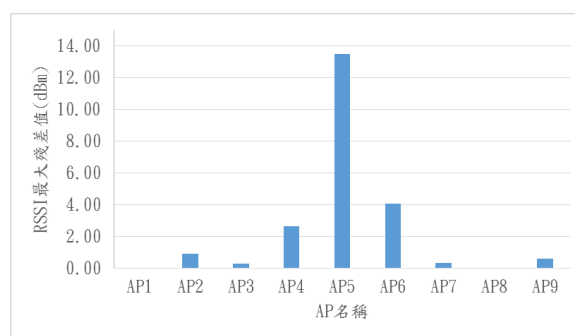


圖 24. 室內資料，在 AP5 加入 -8dBm 後，檢核點 1 在定位解算階段，接收的各 AP 之 RSSI 殘差值。

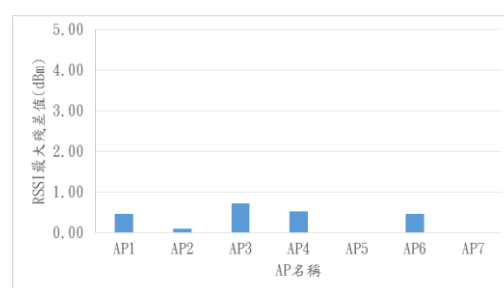


圖 25. 室外資料，未加入粗差前，檢核點 4 在定位解算階段，接收的各 AP 之 RSSI 殘差值。

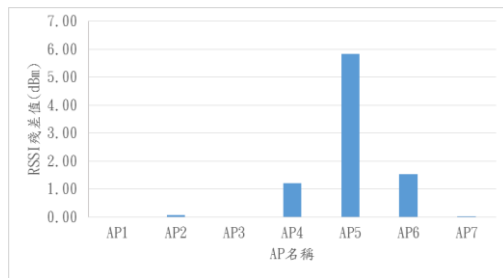


圖 26. 室外資料，在 AP5 加入 -8dBm 後，檢核點 4 在定位解算階段，接收的各 AP 之 RSSI 殘差值。

4.4.2.4 小結

根據上述 LAD 定位解算階段偵錯能力的實驗結果顯示：(1) 模擬資料，當粗差數量小於或等於 2 時，LAD 可由各 AP 的 RSSI 殘差最大值來判斷粗差的位置。然而，當粗差數量大於 2 時，LAD 則無法偵錯；可能原因為定位解算階段的 K 向量，其為 RSSI 值、AP 的模型參數、待測點的坐標近似值等之函數，其較 RSSI 建模階段的 L 向量(其指包含 RSSI 值)的複雜度高。此結果與 LAD 在定位階段的穩健性實驗結果相似。(2) 室內及室外資料，當粗差數量等於 1 時，LAD 可由各 AP 的 RSSI 殘差最大值來判斷粗差的位置。

五、結論

本文利用三個實驗區(模擬資料、室內資料及室外資料)的 WiFi 資料，探討最小一乘法 WiFi 匹配定位(簡稱 LAD)的模型檢定、精度、穩健性及偵錯能力等議題。

根據實驗結果顯示：(1)透過 F 、 R^2 等統計檢定，驗證點位的坐標值 (x, y) 對訊號強度 (γ) 具有顯著的聯合解釋能力；而且，坐標值 (x, y) 對於訊號強度值 (γ) 的配合度高。(2)在定位精度方面，通常 LAD 定位精度優於最小二乘法 WiFi 匹配定位(簡稱 LS)。(3)在穩健性方面，於 RSSI 建模階段，LAD 有很好的穩健性，LS 則無；於定位解算階段，當粗差數量小於或等於 2 時，LAD 有好的穩健性，LS 則無。(4)在偵錯能力方面，於 RSSI 建模階段，LAD 有很好的偵錯能力，可以從參考點的 RSSI 殘差判斷粗差所在；於定位解算階段，當粗差數量小於或等於 2 時，LAD 有好的偵錯能力。

從上述實驗結果顯示，LAD 不管在精度、穩健性及偵錯能力等方面皆優於 LS，然而，LAD 仍有待探討的議題，如在室外資料時 LAD 的定位精度稍遜於 KNN 等。

誌謝

本文承行政院科技部提供經費補助(計畫編號 MOST 105-2410-H-004-169)，特此誌謝。

參考文獻

- [1] 李楨、譚先科，“基於最小二乘的 WiFi 匹配導航”，測繪地理資訊，第 40 卷，第 3 期，第 60-62 頁，2015。
- [2] 陳希孺，“最小一乘線性回歸(上)”，數理統計與管理，第 43 卷，第 5 期，第 48-55 頁，1989。
- [3] Bahl, P. and Padmanabhan, V. N., “RADAR: An in-building RF-based user location and tracking system”, INFOCOM, Nineteenth Annual joint Conference of the IEEE Computer and Communications Societies. Proceedings, Vol. 2, pp. 775-784, 2000.
- [4] 莊豐錨，基於 AP 選擇和適應性樣式比對之 WiFi 室內定位策略，國立臺北科技大學電子工程系研究所碩士論文，臺北，2015。
- [5] 余宗鴻，WiFi 室內定位使用粒子群演算法，國立中央大學通訊工程學系研究所碩士論文，桃園，2015。
- [6] 劉科宏，結合 GPS 與無線網路之定位系統研究，銘傳大學資訊傳播工程學系碩士論文，桃園，2008。
- [7] 崔文、陳昭男，“用於智慧型手機上之混合式行人定位技術”，電腦與通訊，第 137 期，第 89-96 頁，2011。
- [8] 郭清智，整合 WiFi 與 GPS 技術於室外定位之研究，國立政治大學地政學系研究所碩士論文，臺北，2011。
- [9] 林怡君，利用最小一乘法在地籍坐標轉換資料偵錯之研究，國立政治大學地政學系研究所碩士論文，臺北，2013。
- [10] 林老生、林怡君，“基於最小一乘法的穩健地籍坐標轉換”，中正嶺學報，第 43 卷，第 2 期，第 199-218 頁，2014。

- [11] 方述誠，“線性優化及擴展-理論與演算法”，數學傳播，第 17 卷，第 1 期，第 1-12 頁，1993。
- [12] 方述誠、普森普拉，線性優化及擴展-理論與演算法，北京，科學出版社，1994。
- [13] 林子添，基於最小一乘法的室外 WiFi 匹配定位之研究，國立政治大學地政學系研究所碩士論文，臺北，2017。
- [14] 何維信、詹君正，“虛擬基準站即時動態定位辦理土地複丈精度之研究”，臺灣土地研究，第 13 卷，第 2 期，第 79-100 頁，2010。
- [15] Li, S. and An, X., “A WiFi-based Adjustment Algorithm for GPS Positioning on Smartphones”, Department of Hydraulic Engineering, Tsinghua University, Beijing, 2011.
- [16] Ghilani, C. D., Adjustment Computations: Spatial Data Analysis, 5th Edition, John Wiley & Sons, Inc, 2010.

手背靜脈影像邊緣特徵識別

詹閱傑¹ 莊尚仁¹ 黃煌初¹ 李仁軍^{2*}

¹國立高雄科技大學電訊工程系(所)

²海軍軍官學校電機工程系

摘 要

手背靜脈識別為一種新興生物認證技術，本研究主要目的在於提出一種強健、穩定且即時性的手背靜脈識別方法。我們將手背靜脈骨架化影像實施靜脈血管紋路骨架與邊緣相交點之特徵值擷取，於特徵比對階段提出動態模式樹加快比對速度，並配合最小距離分類器實施特徵識別。最後實驗結果經由效能評估後整體正確識別率為 99.72%與其他識別方法比較下本識別方法更能快速進行特徵比對且具備極佳識別率，因此我們提出之手背靜脈識別對於生物認證領域上具有高度實用性。

關鍵詞：手背靜脈識別，生物認證，動態模式樹，最小距離分類器

Dorsal Hand Vein Image Edge Feature Recognition

Hong-Jie Zhan¹, Shang-Jen Chuang¹, Huang-Chu Huang¹, and Jen-Chun Lee^{2*}

¹National Kaohsiung University of Science and Technology, Department of Electronic Communication Engineering

²R.O.C.Naval Academy, Dept. of Electrical Engineering

ABSTRACT

With the increasing needs in security systems, vein recognition is one of the important and reliable solutions of identity security for biometrics-based identification systems. This paper presents a novel, local feature-based vein representation method based on minutiae features from skeleton images of venous networks. These minutiae features include end points and the arc lines between the two end points as measured along the boundary of the region of interest. In addition, we propose a dynamic pattern tree to accelerate matching performance and evaluate the discriminatory power of these feature points for verifying a person's identity. In a comparison with existing verification algorithms, the proposed method achieved the highest accuracy in the lowest tested matching time. Our results demonstrate the effectiveness of the proposed method as a promising approach to vein recognition. Therefore, we propose the dorsal hand vein recognition for the field of biometric authentication with high practicability rate.

Keywords: dorsal hand vein recognition, biometrics, dynamic pattern tree, minimum distance classifier

一、前言

隨著資訊科技發展，人們對資訊服務需求越來越高，相對也帶給使用者許多安全性的隱憂與資訊安全需求，其中影響最大的莫非於使用者身份認證技術。現今實施的傳統身份認證技術(密碼、IC卡及磁卡等)隨著駭客技術不斷進步已經不再安全，更無法解決對於人本身之識別，於是近年來生物認證(Biometrics)識別技術越來越引起大家高度重視，希望由生物認證來解決資訊系統在安全上所碰到的各種問題。生物認證主要利用人體固有的生理特徵或使用人的行為特徵，透過影像處理技術等方法來鑑別個人身份。與傳統身份認證技術相比，生物認證具有更好的安全性、可靠性和有效性。

目前三種最普遍的生物特徵分別為指紋、人臉、與虹膜，而我們將這三種生物特徵進行比較後，仍發現這些識別特徵存在諸多問題。首先，指紋識別(Fingerprint recognition)存在容易進行偽造的安全隱憂，在擷取過程中需接觸擷取設備，會受到手指皮膚狀況及按壓情況影響。另外，人臉識別(Face recognition)易受光線或陰影干擾識別外，還有隨著人的年紀增長或是外型改變而變化的問題。識別率較高之虹膜識別(Iris recognition)也存在擷取設備昂貴和照射眼球時需要克服心理障礙等缺點。鑒於指紋、人臉與虹膜的諸多問題，我們發現另一種新的生物特徵識別，人類手背靜脈影像識別技術能夠解決上述三種生物特徵識別缺點。

手背靜脈紋路與指紋一樣是每個人所獨有的，且位於人體皮下組織，不易遭受擷取及竄改，因此可以擷取手背的靜脈影像特徵來進行高精度身份認證。而目前國內手背靜脈識別技術的研究不太多，因此對於手背靜脈識別演算法之研究更具有重要的意義，而本論文主要目的在於提出一個強健、穩定且即時性的手背靜脈識別方法，對於手背靜脈影像前置處理及靜脈骨架特徵擷取與比對進行深入研究。

本論文提出之手背靜脈識別流程(如圖 1 所示)總共分為註冊(Enrollment)與識別(Identification)兩大階段，在註冊階段，獲取檢

測者手背靜脈影像後，先實行影像前置處理，然後執行特徵擷取並獲取特徵向量建立模板資料庫；於識別階段，在獲得檢測者手背靜脈影像後同樣實施影像前置處理及特徵擷取，並將擷取的靜脈影像特徵向量與已註冊資料庫之特徵向量進行模板比對，得到最終識別結果。

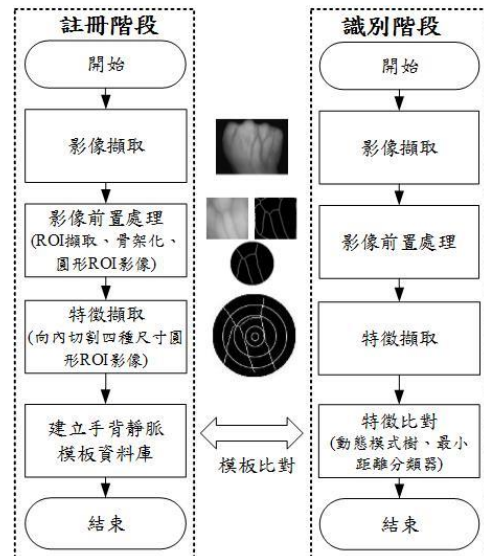


圖 1. 手背靜脈識別方法流程圖

二、文獻探討

人類手部靜脈影像的生物特徵識別技術，從 1995 年，J.M. Cross 等人對手背靜脈識別進行研究[1]，不過當時受到感測器技術限制，因此該技術並沒有進一步發展。直到 2001 年，南韓研製出較為成熟的手背靜脈身份識別產品[2]，人體手背靜脈識別技術才備受關注，成為最近幾年大家研究的方向。國內從 2004 年范國清教授發表了關於手背靜脈遠紅外影像識別相關論文[3]開始，手背靜脈身份識別技術逐漸得到學術界和企業界重視。

在手背靜脈識別技術中，該如何拍攝出高品質靜脈影像為目前靜脈影像取像裝置需要面對的問題，1998 年南韓 NEXTERN 公司使用紅外線 CCD 視頻攝影機擷取靜脈影像，其中 BK 系列為非常成熟的產品[4]。2004 年范國清教授運用遠紅外光照射人體手背[3]，並利用一台配有濾光片的 CCD 視頻攝影機擷取靜脈影像。2005 年新加坡南洋理工大學則使用 NEC 熱追蹤儀擷取手背靜脈影像。2007 年 Ahmed M. Badawi 運用 CCD 視頻攝影機拍攝

血管影像，並在 CCD 鏡頭上安裝紅外濾光片，藉此排除可見光干擾。

人體靜脈影像在進行拍攝時，受到拍攝時間、拍攝時的亮度與拍攝者皮膚厚薄度三種因素影響，因此靜脈影像像素值在灰階分布程度上有相當大的差異，造成特徵擷取與比對困難度，所以靜脈影像前置處理在靜脈識別技術中是一個重要的流程。Im 等人[4]使用一系列的影像處理技術，Gauss 低通濾波、Gauss 高通濾波、二值化處理、雙線性濾波及改進的中值濾波等去除靜脈影像雜訊，提高影像品質與突顯靜脈特徵。Im 等人採用之去除雜訊技術，也是現今大部分靜脈識別中使用的影像處理方式。

得到清晰的靜脈紋路影像後，下一個階段就進行靜脈影像特徵擷取與特徵比對。目前在靜脈識別研究中，大多擷取靜脈血管的交叉點、端點數目及相對位置或者是管曲線的曲率、長短及相對位置等特徵。特徵點演算法使用靜脈紋路骨架化影像中的特殊點作為比對，雖然可以得到精確的識別，不過須花費大量時間在於特徵點及交叉點之尋找。Ding 等人[5]運用靜脈血管端點、交叉點間的距離作為比對依據。

Lee[6]使用手背靜脈骨架特徵擷取演算法，將靜脈影像骨架化後，擷取可識別的特徵點及對應的特徵距離視為有效特徵，並提出 EP-tree 搜尋方式找尋相似的靜脈影像，不過該方法於特徵比對時需要找尋全部三組樹狀階層，需花費較多搜尋時間。另外 Lee 使用之靜脈影像於前置處理完成後，在骨架化影像邊緣仍有細節及陰影部分等雜訊，這些邊緣雜訊會導致後續無法擷取正確的靜脈影像特徵。

將特徵空間演算法用於影像識別亦能獲得較好效果，不需針對影像進行過多處理，即可達成所需目的。Lu 等人[7]提出 Eigenpalm 演算法，運用 PCA 進行掌紋影像特徵擷取。Hsu 等人[8]運用改良的 $(2D)^2$ PCA 取代傳統 PCA 進行手背靜脈特徵擷取。

影像上的二維賈伯濾波器(2D Gabor filter)為採用紋理結構進行特徵擷取代表性演算法，不論使用在空間域或頻率域皆具備良好性能，因此廣泛運用於紋理分析及識別。Lee[9]以賈伯濾波器為基，提出運用方向性編碼技術，將掌靜脈特徵編碼成一系列的位元碼，其識別效果及執行時間優於原始賈伯濾波器。Wu 等人

[10]運用高斯方向特徵編碼 (Gaussian directional binary code, GDBC)，進行靜脈影像特徵擷取。

近幾年深度學習(Deep learning)在人工智慧(Artificial intelligence, AI)的風潮帶領下，也逐漸引起高度關注。其中 Yann LeCun[11]針對影像識別提出卷積類神經網路 (Convolutional neural network, CNN) 後，對於生物認證識別產生重大突破。CNN 主要結構分為影像特徵擷取的卷積層、池化層以及完全連結神經網路兩大架構。與其它類神經網路架構相比，CNN 需考慮的參數較少，因此擁有高識別精確率。目前 CNN 廣泛應用於人臉、手寫與語音識別中。Wan 等人[12]將自行拍攝的手背靜脈影像使用 CNN 中 VGG-19 網路架構實施特徵比對亦獲得較好的識別結果。

目前各種特徵擷取演算法中都採取不同方式進行靜脈影像特徵擷取，提出之識別正確率與錯誤率證明這些特徵擷取方法能達成一定效果，但是大家依然不斷地追求生物認證識別中的正確率和錯誤率能夠趨近於完美，因此針對手背靜脈影像提出可靠且高效率的特徵識別與擷取演算法仍為手背靜脈識別研究目標。

三、手背靜脈影像前置處理

手背靜脈識別首要步驟就是檢測者手背靜脈影像擷取，並對擷取的手背靜脈影像進行 ROI(Region of interest, ROI)區域選取並使其正規化，並運用各種影像處理技術針對 ROI 影像進行影像強化、影像濾波與骨架化。最後為了避免骨架化影像邊緣雜訊，我們將方形 ROI 影像轉換為圓形 ROI 影像。完成上述影像前置處理步驟就可獲取高品質與無雜訊的手背靜脈影像並且實施特徵擷取與識別，本論文使用之手背靜脈影像前置處理方法如圖 2 所示。

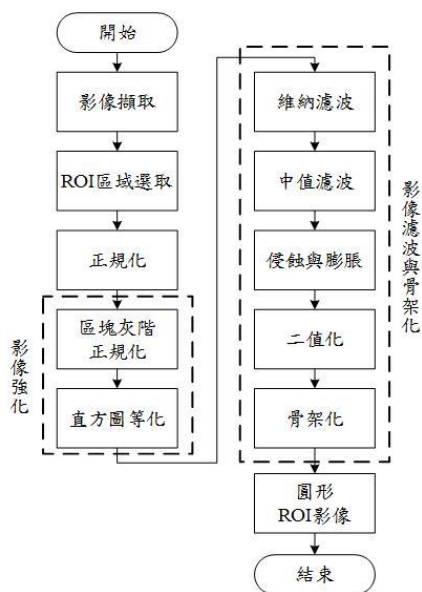


圖 2. 手背靜脈識別之影像前置處理流程

3.1 手背靜脈影像資料庫

手背靜脈血管於拍攝過程中光源占了相當大的因素，照攝靜脈血管光線均勻程度直接影響影像品質。因此光源照攝均勻度應達到70%以上，若均勻度太低影像容易有陰影或亮斑，而陰影被當作靜脈血管機率相當大，若把陰影當成靜脈血管將會嚴重影響識別準確度。手背靜脈血管紋路之成像主要利用血液中的還原血蛋白對近紅外光的吸收特性來實現。一般來說用波長在 720 nm~950 nm 的近紅外光照射手背，手背靜脈血管中的血色素相較於皮下組織可以吸收更多近紅外輻射，可以很好地呈現出靜脈血管的紋路。

使用一套可以被接受之手背靜脈影像資料庫，並用此驗證研究結果可行性及穩定性是非常重要的，因此我們使用本研究團隊於2011年開發的影像取像裝置所建立的一套手背靜脈影像資料庫[13]。其中 129 位檢測者當中，包含不同人種(白種、黃種以及黑人)，年齡範圍從 18 歲至 65 歲，性別男女皆有。共分為兩個階段拍攝手背靜脈影像，每個人每隻手每次擷取 10 張手背靜脈影像，影像擷取時間間隔約為 1-3 個月左右。靜脈影像資料庫中，不同檢測者之手背靜脈原始影像如圖 3 所示。

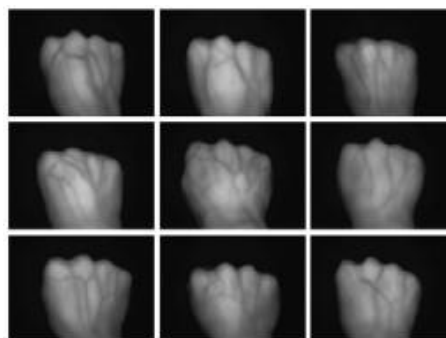


圖 3. 不同檢測者手背靜脈原始影像[13]

3.2 感興趣區域擷取及正規化

本論文手背靜脈影像資料庫於拍攝中為符合實際應用情況下只固定手背與拍攝鏡頭間的距離，並沒有規定檢測者手背擺放位置和傾斜角度。因此本識別方法為了符合實際應用以便後續執行影像前置處理時針對手背位置進行定位與去除不需識別的影像資訊，因此需要實施感興趣區域擷取與影像正規化，若 ROI 擷取位置錯誤，後續將嚴重影響識別正確率。因此針對手背靜脈影像 ROI 區域擷取及定位，我們運用 2011 年本團隊提出之 ROI 區域選取與定位方式[13]，ROI 區域定位過程(ROI 區域定位過程如圖 4 所示)，詳述如下：

1. 利用 Otsu 門檻值[14]，將手背影像二值化，擷取手背影像主體。
2. ROI 區域定位點選取，主要步驟如下：
 - (1). 計算手背重心 (Center of Gravity, COG)，如圖 4(b)。
 - (2). 運用 Inner Border Tracing Algorithm[15]邊緣偵測器擷取出影像邊緣 後擷取手背輪廓。以 COG 為基準點，將手背影像區分為上半部及下半部，如圖 4(c)。
 - (3). P_1 與 P_2 分別代表於上半部手背輪廓的極左點和極右點，將 $\overline{P_1P_2}$ 投射於 x 軸，並劃分成四等分，如圖 4(c)。在上半部手背輪廓， P_1 到 P_3 為第一子區域， P_4 到 P_2 為第四子區域。
 - (4). 計算 COG 至第一子區域及第四子區域內每一點的距離。
 - (5). P_5 與 P_6 分別為第一子區域及第四子區域距離 COG 最遠點，如圖 4(d)。 P_5 與 P_6 幾乎接近食指和小指拳眼的端點，故選用 P_5 與 P_6 為定位點，用於定

位 ROI 區域。

- 為了解決手背旋轉並將手背調整到固定角度便於定義 ROI 區域，手背靜脈影像須旋轉 θ° 如圖 4(e)。 θ 為 $\overline{P_5P_6}$ 與水平線之間的夾角，可由下式表示：

$$\theta = \tan^{-1}(Y_{P_6} - Y_{P_5}) / (X_{P_6} - X_{P_5}) \quad (1)$$

其中 (X_{P_5}, Y_{P_5}) 是 P_5 的座標， (X_{P_6}, Y_{P_6}) 是 P_6 的座標。

- 以 $\overline{P_5P_6}$ 水平線中心 P_m 為基準點，取一正方形平行 $\overline{P_5P_6}$ ，其中 C_1 、 C_2 、 C_3 、 C_4 正方形區域的四個端點，正方形邊長 $\overline{C_1C_2}$ 為 $\overline{P_5P_6}$ 長度的 3/5 倍。我們將正方形 C_1 、 C_2 、 C_3 、 C_4 定義為 ROI 區域，如圖 4(f)表示。
- 由於每位檢測者之手背大小不同，所擷取的 ROI 區域大小亦不相同。因此我們需要再進行 ROI 區域正規化處理，將手背靜脈 ROI 區域影像運用內插法正規化成相同大小。

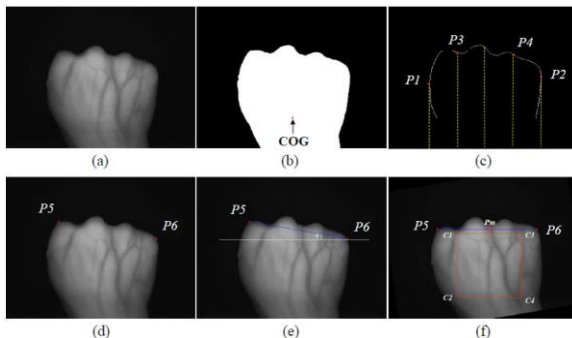


圖 4. ROI 區域定位過程[13]:(a)人體手背靜脈影像；(b)二值化影像；(c)上半部手背輪廓區域化；(d)選取定位點；(e)運用 P_5 與 P_6 點與水平線找出 θ 可將影像旋轉；(f)擷取 ROI 區域 $R_{C_1C_2C_3C_4}$ 。

3.3 ROI 影像前置處理

手背靜脈影像正規化後，每張影像均有相同大小，但仔細觀察仍存有對比度低及非均勻亮度的問題，為了解決背景亮度問題，我們運用相當成熟的影像處理技術，如影像強化(Image enhance)及銳化、二值化門檻法(Binary thresholding)、中值濾波(Median filter)及骨架化(Thinning)等處理，將手背靜脈 ROI 影像，完成靜脈影像骨架特性的擷取(ROI 影像前置處理步驟如圖 5)。

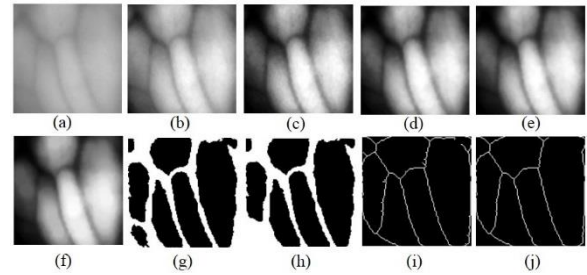


圖 5. 手背靜脈 ROI 影像前置處理(a)原始影像；(b)區塊亮度灰度正規化；(c)直方圖等化；(d)維納濾波；(e)中值濾波；(f)侵蝕膨脹；(g)二值化；(h)填充空洞；(i)影像骨架化；(j)去除毛刺。

3.4 圓形 ROI 影像

獲得乾淨且無毛刺的手靜脈骨架化影像後，我們發現部分骨架化影像其邊緣仍有細節及陰影部分等雜訊，而這些雜訊無法運用前述之影像處理演算法完整去除(如圖 6(b)與(e))。為減少影像邊緣部分無法去除的雜訊，本研究將原始方形 ROI 影像轉換為圓形 ROI 影像(如圖 6(c)及(f))去除方形骨架化影像邊緣雜訊問題。

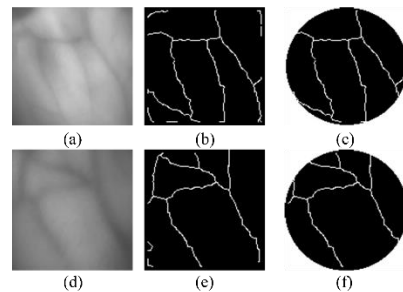


圖 6. 手背靜脈 ROI 影像前置處理:(a)(d)原始影像；(b)(e)骨架化影像；(c)(f)圓形 ROI 影像。

四、特徵擷取及比對

4.1 特徵擷取

一般來說，骨架化影像特徵以點或線的方式來表示，因此其特徵擷取主要是尋找影像間易於區分的環節，像是紋線交叉點和端點，以及它們的數量與相對位置都可以作為特徵加以擷取。有鑑於此本研究運用影像邊界點方法擷取影像特徵，不但能壓縮影像資訊量，也可以簡單進行特徵描述與識別。我們將圓形 ROI 影像以中心點為圓心，依序向內分別取直徑

96, 48, 32, 16 像素，分割為等四種不同圓形手背靜脈影像，如圖 7 所示。在每一張不同大小圓形影像上找尋圓形邊緣上與靜脈交會的邊緣點，並分別記錄每一個影像邊緣上的邊緣點數量與邊緣點間的距離，使每一種手背靜脈圓形 ROI 影像形成一組有效識別的特徵向量，本論文提出之手背靜脈影像特徵擷取演算法敘述如下[16]：

1. 將圓形 ROI 影像向內分別取直徑 96, 48, 32, 16 像素，分割為等四種不同圓形 ROI 影像(如圖 7 所示)。
2. 不同尺寸影像分別擷取影像圓周邊緣上與靜脈交會的邊緣點，以圖 7 為例，依圓形尺寸從大至小之邊緣點數量分別為 5、4、2、0。
3. 不同尺寸影像分別計算影像各圓周邊緣點彼此間的距離，並以最大邊緣距離為起始距離順時鐘方向排列，以圖 7(a)為例，該影像邊緣點為 5，其距離分別為 $[P_1, P_2, P_3, P_4, P_5]$ ；圖 7(b)影像邊緣點為 4，其距離分別為 $[Q_1, Q_2, Q_3, Q_4]$ ；圖 7(c)影像邊緣點為 2，其距離分別為 $[R_1, R_2]$ ；而圖 7(d)影像邊緣點為 0，其距離排列分別為 $[S_1]$ 。
4. 由上述邊緣點及邊緣距離資料的特徵，可將每一張靜脈 ROI 影像分別組成不同維度的特徵向量，而每張影像之特徵維度依據其邊緣點的數量訂定；以圖 7 為例，影像尺寸從大至小之邊緣點分別為 5、4、2、0，而若有 n 個邊緣點，即有 n 個邊緣距離，因此本研究以邊緣距離視為影像之特徵維度內容值，其圖 7 影像特徵維度為 1×11 (其中 $11 = 5 + 4 + 2 + 0$)，而特徵維度內每一個元素即為相對應的特徵距離值。

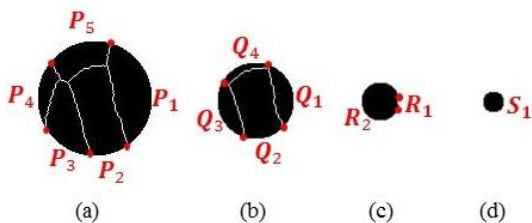


圖 7. 手背靜脈影像四個圓周邊緣上與靜脈交會的邊緣點與各點之間的距離

經由上述演算法，可以找到手背靜脈血管位於圓形 ROI 影像邊緣點數量與邊緣距離資訊，此外本演算法無須搜尋整張影像，只需針

對影像邊緣擷取其所需的特徵即可代表該影像，可大幅減少計算時效，增加系統即時性需求。

4.2 特徵比對

一般而言，針對每一手背靜脈影像進行特徵向量擷取後，即可把它儲存入資料庫內；當需要進行身分驗證時，首先透過影像擷取裝置取獲被檢測者的靜脈影像，然後進行特徵擷取以獲取特徵向量，並將該特徵向量與資料庫中的已有特徵向量逐一比對，計算其相似度，加以識別；由於特徵向量需要逐一比對且每張靜脈影像所擷取之特徵維度不一，需要龐大計算量；因此本論文提出動態模式樹(Dynamic pattern tree)演算法實施特徵比對，該演算法只需針對相同特徵點向量實施特徵比對，不但縮短計算時間且提高識別效能，更解決各影像特徵維度不一的問題，整個比對演算法的描述如下：

1. 以邊緣點為主，建立一組動態模式樹，此模式樹每一階層均代表每一尺度影像上的邊緣點數量，每張影像均分割為 4 種不同尺度(直徑分別為 96, 48, 32, 16 像素)，因有些影像分割至四個尺寸(直徑為 16 像素)時已無靜脈，故此動態模式樹有 3 至 4 個階層。
2. 每張靜脈影像可依據其邊緣點特徵，運用動態模式樹於資料庫內迅速的找到近似的靜脈影像，本研究提出的動態模式樹(如圖 8)比對敘述如下：
 - (1) 第一階層為靜脈影像直徑為 96 像素的圓周上邊緣點數量(假設有 n 個樣本，第一個樣本邊緣點數量為 P^1 ，而最後樣本邊緣點數量為 P^n)
 - (2) 第二階層為靜脈影像直徑為 48 像素的圓周上邊緣點數量(假設模式樹第二階層內屬於 P^1 類別顯示有 m 個樣本，最後樣本邊緣點數量為 Q^m ；屬於 P^2 類別顯示有 p 個樣本，最後樣本邊緣點數量為 Q^p)。
 - (3) 第三階層為靜脈影像直徑為 32 像素的圓周上邊緣點數量(模式樹第三階層內屬於 P^1 且屬於 Q^1 類別顯示有 r 個樣本，最後樣本邊緣點數量為 R^r ；

- 屬於 P^2 且亦屬於 Q^p 類別顯示有 u 個樣本，最後樣本邊緣點數量為 R^u 。
- (4) 第四階層為靜脈影像直徑為16像素的圓周上邊緣點數量(模式樹第四階層內屬於 P^1 且屬於 Q^1 及 R^r 類別顯示有 e 個樣本，最後樣本邊緣點數量為 S^e ；屬於 P^2 且亦屬於 Q^p 及 R^1 類別顯示有 g 個樣本，最後樣本邊緣點數量為 S^g)。

- (5) 一張經前置處理及特徵擷取後之待檢測靜脈影像，本研究依據邊緣點運用動態模式樹搜尋後即可找尋出所對應的相同維度特徵向量 K (如圖8中 $K_1 \sim K_{16}$)，其中 K 非唯一。關於特徵向量 K 之產生由公式(2)表示：

$$k = [P^1, \dots, P^n, Q^1, \dots, Q^m, R^1, \dots, R^r, S^1, \dots, S^e] \quad (2)$$

在此， P 、 Q 、 R 和 S 分別為動態模式樹中的第一、第二、第三和第四階層(四種尺寸圓形 ROI 靜脈影像)。例如， P^n 在半徑為64像素的圓形 ROI 靜脈影像之邊界具有 n 個特徵點。

3. 由於運用動態模式樹找出的近似影像並為唯一，故須再以特徵距離為主的特徵向量實施分類，經過動態模式樹比對後找出的近似影像均有相同的距離特徵向量維度，因此採用最小距離分類器(Minimum distance classifier, MDC)實施分類。

$$m = \arg \min_{1 \leq i \leq c} d(f, f_i) \\ d(f, f_i) = \sum_j (f - f_i)^2 \quad (3)$$

其中 f 與 f_i 分別為待檢測樣本與找出相似靜脈第 i 個樣本的特徵向量； c 為相似靜脈的群組總數，式(3)中， $d(f, f_i)$ 定義為歐式距離。特徵向量 f 運用相似度測量 $d(f, f_i)$ 找出最短距離值的群組。

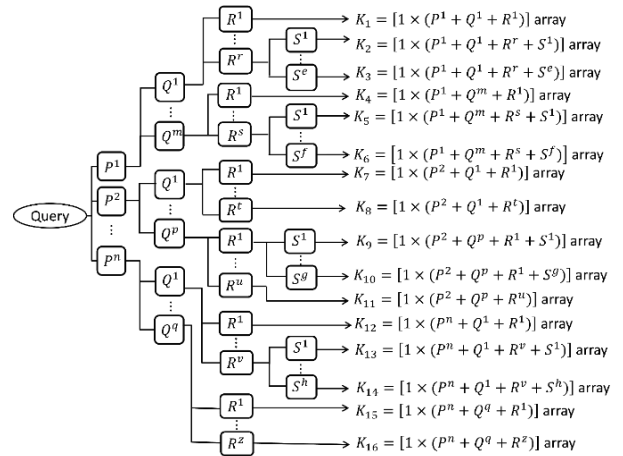


圖 8. 動態模式樹資料結構

五、研究結果

由於目前手背靜脈識別研究尚無具公信力的影像資料庫，因此我們採用本研究團隊於2011年建置之手背靜脈影像資料庫[13]，每檢測者有10張手背靜脈血管紋路影像，共有1290張影像，尺寸為128x128像素的灰階影像。其中某些檢測者之影像品質不均，影像在灰階強度上分佈過高或過低，影響特徵擷取，因此本論文挑選影像資料庫中85位檢測者，每位檢測者使用6張影像，總共510張手背靜脈影像進行實驗。其中每位檢測者有3張手背靜脈影像作為註冊並建立模板，而剩下的3張影像在識別階段時作為模板比對。識別方法開發與實驗環境採用個人電腦 i7-2600 3.4GHz DDR3 4GB 並運用 MATLAB R2014a 執行。

5.1 效能評估

我們針對提出之手背靜脈識別方法實施效能評估，在實驗過程中分為兩種做法：識別(Identification)及驗證(Verification)。在識別階段，採取正確率(Correct recognition rate, CRR)評估整體手背靜脈識別方法效率；於驗證階段，我們採用一般評估生物認證識別演算法之接受器操作特性曲線(Receiver operating characteristic, ROC)描述錯誤接受率(False acceptance rate, FAR)與錯誤拒絕率(False reject rate, FRR)間的相關性。

識別階段實驗中我們採用動態模式樹演算法針對資料庫的手背靜脈影像運用特徵點比對就可比對出唯一的值，若靜脈影像有兩組

或兩組以上之特徵點相同時再運用特徵距離為主的最小距離分類器實施特徵距離比對。本研究進行的實驗中共有 55 位檢測者的影像運用特徵點即被成功比對出。而剩下 30 名特徵點相同之檢測者需要運用 MDC 最小距離分類器實施影像特徵距離比對。

在驗證階段本論文將識別階段實驗結果進行 FAR 與 FRR 之 ROC 曲線評估，並將評估結果採用圖 9 之接受器操作特性曲線圖表示：

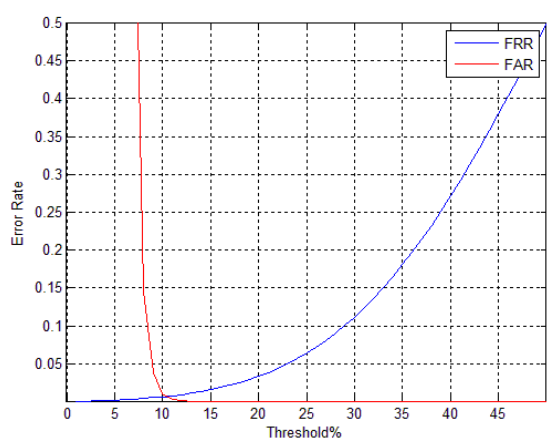


圖 9. 本論文之手背靜脈識別方法 ROC 曲線圖
由圖 9 的 FAR、FRR、EER 關係曲線圖中可看出，當門檻值為 10.45 時，FAR 與 FRR 會在此處相互交會，因此 EER 為 0.78%。

5.2 效能比較

為了驗證與評估識別方法的正確性與有效性，我們將本論文於實驗中採用的手背靜脈資料庫與 EP-Tree[6]、Eigenspace[8]、2D Gabor filter[9]與 CNN 這四種不同手背靜脈識別方法分別進行實驗及效能評估。識別方法之整體辨識率、等誤差率及特徵比對時間及效能比較結果如表 1 所示。

表 1. 手背靜脈識別方法效能比較

Methods	CRR(%)	EER(%)	Matching Time(ms)
EP-Tree	98.31	1.75	0.23
Eigenspace	98.68	1.38	1.45
2D Gabor	98.93	1.12	12.4
CNN	95.30	4.70	2260
Proposed	99.72	0.78	0.11

為了驗證與評估識別方法的正確性與有效性，我們將本論文於實驗中採用的手背靜脈

資料庫與其它四種不同手背靜脈識別方法分別進行實驗及效能評估。

第一種識別方法 EP-Tree 採用模式樹 (Pattern tree) 搜尋方式找尋出相似手背靜脈影像，不過該方法於特徵比對時需要找尋全部三組階層與本研究提出的動態模式樹相比，有些靜脈影像於第一階層即可比對完畢，若採用模式樹比對方式需要繼續搜尋後兩組階層，動態模式樹則不必繼續搜尋，因此兩者互相比較下採用模式樹進行特徵比對需花費較多搜尋時間。

第二種與第三種識別方法分別為改良 $(2D)^2PCA$ 的 Eigenspace 及 2D Gabor filter，採用這兩種識別方法在特徵擷取時花費較多時間，且需要對所有檢測者實施特徵比對，因此這兩種識別方法不論整體辨識率或特徵比對時間都沒有本論文識別方法優秀。

第四種識別方法採用 CNN 針對本論文手背靜脈影像資料庫實施識別與驗證，與本論文提出之識別方法為相同實驗環境，非採用對於 CNN 執行效能較佳的 (Graphics processing unit, GPU) 實驗環境，在特徵比對時間也比本論文提出之識別方法來得慢。此外使用 CNN 進行影像識別需要數量較多的訓練影像提供類神經網路實施訓練，增加 CNN 識別精確性。而本論文手背靜脈影像資料庫每位檢測者影像張數較少，不足以讓 CNN 表現最佳識別效能，因而在 CRR 與 EER 表現低於本論文識別方法。

由表 1 可看出本論文提出之手背靜脈識別方法相較於上述四種識別方法擁有較高辨識率與較低錯誤率，因此本方法具備擁有較佳的識別準確性及實用性。

六、結論

經由本論文實驗結果得知使用動態模式樹實施特徵點比對可成功篩選出資料庫 85 位檢測中的 55 位，後續剩下 30 位檢測者則運用 MDC 最小距離分類器實施特徵距離比對之錯誤率僅有 0.78%，因此本識別方法整體辨識率達到 99.72%。另外，本識別方法透過與其它識別方法比較可以得知，本識別方法效能評估結果相較於過去提出的識別方法來得更好。因此更能夠證明我們提出之手背靜脈識別方法

能夠應用於高安全性環境及驗證多位使用者的身分。

參考文獻

- [1] J. M. Cross, C. L. Smith, "Thermographic Imaging of the Subcutaneous Vascular Network of the Back of the Hand for Biometric Identification", IEEE Proceedings of the Institute of Electrical and Electronics Engineers. 29th Annual 1995 International Carnahan Conference on Security Technology, England, pp.20-35, 1995.
- [2] Sang-Kyun Im, Hyung-Man Park, Young-Woo Kim, "An Biometric Identification System by Extracting Hand Vein Patterns", Journal of the Korean physical Society, Vol.38, No.2, pp.268-272, 2001.
- [3] Chih-Lung Lin, Kuo-Chin Fan, "Biometric Verification Using Thermal Images of Palm-Dorsa Vein Patterns", IEEE Transactions on Circuits and Systems for Video Technology, Vol.14, No.2, pp.199-213, 2004.
- [4] Sang-Kyun Im, Hyung-Man Park, Young-Woo Kim, "An Biometric Identification System by Extracting Hand Vein Patterns", Department of Electronics Engineering, Korea University, 2000.
- [5] Yuhang Ding, Dayan Zhuang, Kejun Wang, "A Study of Hand Vein Recognition Method", Proceedings of the IEEE International Conference on Mechatronics & Automation Niagara Falls, Canada, pp.2106-2110, 2005.
- [6] Jen-Chun Lee, "Dorsal Hand Vein Recognition based on EP-Tree", Machine Vision Applications (MVA), 2015 14th IAPR International Conference on, 2015
- [7] Guangming Lu, David Zhang, Kuanquan Wang, "Palmprint recognition using eigenpalms features," Pattern Recognition Letters, Vol. 24, No. 9, pp. 1463-1467, 2003.
- [8] C. B. Hsu, S. S. Hao and J. C. Lee, Personal Authentication through Dorsal Hand Vein Patterns," Optical Engineering, Vol. 50, No. 8, pp. 087201-1--087201-10, 2011.
- [9] Jen-Chun Lee, "A Novel Biometric System Based on Palm Vein Image," Pattern Recognition Letters, Vol. 33, No. 12, pp. 1520-1528, 2012.
- [10] Kuang-Shyr Wu, Jen-Chun Lee, Tsung-Ming Lo, Ko-Chin Chang, Chien-Ping Chang, "A secure palm vein recognition system," The Journal of Systems and Software, Vol. 86, pp. 2870-2876, 2013.
- [11] LeCun, Yann; Yoshua Bengio; Patrick Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE (Volume: 86, Issue: 11, Nov 1998), pp. 2278-2324, 1998.
- [12] Haipeng Wan, Lei Chen, Hong Song, Jian Yang, "Dorsal hand vein recognition based on convolutional neural networks", Bioinformatics and Biomedicine (BIBM), 2017 IEEE International Conference on, pp.13-16, 2017.
- [13] C. B. Hsu, S. S. Hao and J. C. Lee, Personal Authentication through Dorsal Hand Vein Patterns," Optical Engineering, Vol. 50, No. 8, pp. 087201-1--087201-10, 2011.
- [14] Nobuyuki Otsu., "A Threshold Selection Method from Gray-Level Histograms," Systems, Man and Cybernetics, IEEE Transactions on, vol. 9, pp. 62-66, 1979.
- [15] M. Sonka, et al., "Image processing, analysis, and machine vision", 1999
- [16] 詹閔傑、莊尚仁、黃煌初, "手背靜脈特徵點識別研究", 海軍軍官學校「迎向海洋、逐夢啟航」學術研討會, pp137-146, 2017.

詹閔傑等
手背靜脈影像邊緣特徵識別

硝化三乙二醇最適合成參數及特性之研究

曾文佐¹ 李金樹^{2*} 陸開泰² 葉早發²

¹國防大學理工學院化學工程碩士班

²國防大學理工學院化學及材料工程學系

摘 要

硝化三乙二醇(TEGDN)是三甘醇(TEG)的硝酸酯化合物，由於其熔點及敏感度低，可作為高能塑化劑取代或部分取代硝化甘油(NG)在火炸藥及推進劑的作用，然而，對於硝化三乙二醇最適合成參數的瞭解並不多。本研究利用田口實驗設計法探討硝化三乙二醇的最適合成參數，藉由三控制因子及三水準的 L9(3³)直交表設計 9 組實驗條件，實驗數據經轉換為 S/N (Signal-to-Noise Ratio)值後，經分析評估得到硝化三乙二醇最大產率的實驗參數組合，驗證結果顯示最適的合成參數為反應溫度 15°C、硝酸/硫酸重量比為 60:40 及硝酸/三甘醇的莫耳數比為 4:1，可得 TEGDN 的最大產率為 95.2%。此外，運用傅立葉轉換紅外線光譜儀(FTIR)、液態核磁共振光譜儀(NMR)、元素分析儀(EA)及微差掃描熱卡計(DSC)鑑定合成的 TEGDN，並利用動態流變儀(DSR)及真空安定性試驗(VST)設備釐定其特性。

關鍵詞：硝化三乙二醇，最適合成參數，田口法。

Study on Optimal Synthesizing Parameters and Characteristics of Triethylene Glycol Dinitrate

Wen-Zuo Tseng¹, Jin-Shuh Li^{2*}, Kai-Tai Lu², Tsao-Fa Yeh²

¹ Master Program of Chemical Engineering, Chung Cheng Institute of Technology, National Defense University

² Department of Chemical and Materials Engineering, Chung Cheng Institute of Technology,
National Defense University

ABSTRACT

Triethylene glycol dinitrate (TEGDN) is a nitrated alcohol ester of triethylene glycol (TEG). It can be used as an energetic plasticizer in explosives and propellants to partially or completely replace the nitroglycerin (NG), due to its lower melting point and sensitivity. However, little is known about its optimal synthesizing parameters. In this study, the optimal synthesizing parameters of TEGDN were explored by using Taguchi's experimental design method. A L9 (3³) orthogonal array with three control factors and three levels was used to design nine experimental conditions. The experimental data were transformed into a signal-to-noise (S/N) ratio to analyze and evaluate the experimental

文稿收件日期 106.11.17; 文稿修正後接受日期 107.04.20; *通訊作者

Manuscript received November 17, 2017; revised April 20, 2018; * Corresponding author

condition of the optimal parameter combination for the maximum yield rate of TEGDN. The verification results indicated that the optimal synthesizing parameters were the nitration temperature of 15°C, $\text{HNO}_3/\text{H}_2\text{SO}_4$ weight ratio of 60:40 and HNO_3/TEG molar ratio of 4:1, and then the maximum yield rate of TEGDN was 95.2%. Furthermore, the synthesizing TEGDN was identified by means of FTIR, NMR, EA and DSC, and its characteristics were determined using DSR and VST.

Keywords: Triethylene glycol dinitrate, Optimal synthesizing parameter, Taguchi method.

一、前言

硝酸酯類化合物常作為雙基推進劑及發藥射之主要含能成份，除了提供能量外，另一重要的功能即是塑化作用，使其易於加工符合設計所需的藥型[1,2]。硝化甘油(Nitroglycerin, NG)係甘油的三硝酸酯化合物，是第一個廣泛運用於軍民用炸藥的增塑劑，惟其本質特性敏感，易因外在環境的刺激及生製維保處理不當造成危害，除肇生人員傷亡及裝備損害外，更嚴重耗損部隊戰力。因此，近年來世界各軍事先進國家，均積極開發新式的鈍感含能增塑劑，藉以取代敏感度較高的硝化甘油[3,4]。

硝化三乙二醇的學名為三乙二醇二硝酸酯(Triethyleneglycol Dinitrate, TEGDN)，分子式為 $\text{O}_2\text{NO}-\text{CH}_2\text{CH}_2\text{O}-\text{CH}_2\text{CH}_2\text{O}-\text{CH}_2\text{CH}_2-\text{ONO}_2$ ，結構式如圖 1 所示，係利用混酸對三甘醇(Triethylene Glycol, TEG)進行硝化反應製備而成的硝酸酯類化合物，硝化三乙二醇具揮發性低、機械敏感度低及熱安定性良好等特性，且其對硝化纖維的增塑作用及化學穩定性優於硝化甘油，為硝化甘油替代物的良好選擇之一[5-10]，國外已有諸多針對硝化三乙二醇替代（部份替代）硝化甘油製備發射藥之研究。

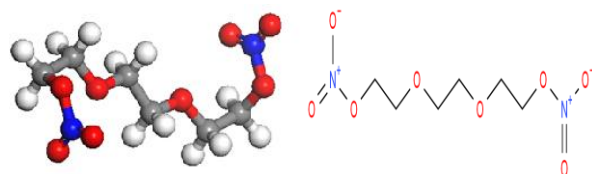


圖 1. TEGDN 結構式（左為 3D 主體模型結構、右為 2D 簡化結構）

硝酸酯類化合物之合成方法，多以硝/硫混酸對醇類進行硝化反應以獲得所需產物，其製備過程有快速放熱的現象具潛在的風險，故反應物多以溶劑稀釋以降低濃度減緩反應速率。TEGDN 製備的方法，鐘顯政等人[11]曾於 1987 年提出，以二氯甲烷(CH_2Cl_2)作為 TEG 的溶劑，以硝/硫混酸（重量百分比 60/40）為硝化劑， HNO_3/TEG 莫耳數比 4.84/1，在反應溫度控制在 15°C 以下的條件，可安全有效地合成 TEGDN，最大的產率可達 94%；韓駿奇等人[12]於 2010 年運用微型反應器進行 TEGDN 的合成，同樣以二氯甲烷(CH_2Cl_2)作為溶劑，以硝/硫混酸（重量百分比 70/30）為硝化劑， HNO_3/TEG 的莫耳數比為 4/1，反應溫度則控制在 22-24°C 之間，可得反應最高產率 82%；2014 年 Nicholas 等人提出專利[13]，以硝酸銨/硫混酸（重量百分比 20/80）為硝化劑，硝酸銨/TEG 的莫耳數比為 4/1，反應溫度控制在 20°C 以下的條件，其反應最高產率為 63%。

本研究綜整文獻所述 TEGDN 的合成方法及製備條件，以 TEG 為原料，硝/硫混酸為硝化劑，二氯甲烷為稀釋溶劑，應用田口實驗設計法，選定反應溫度、硝酸與硫酸重量比及硝酸與 TEG 莫耳比等三個控制因子，每個控制因子各區分三個不同的水準，並利用 $L_9(3^3)$ 直交表的設計規劃 9

組實驗，再藉由 S/N(Signal-to-Noise Ratio)值分析探討硝化三乙二醇的最佳合成參數。此外，運用 FTIR、NMR、EA 及 DSC 實施 TEGDN 合成產物的鑑定，並利用 DSR 及 VST，量測分析 TEGDN 的特性，研究成果將有助於國軍兵工廠開發鈍感含能增塑劑 TEGDN 的生產製程。

二、實驗

2.1 實驗藥品

- (1) 濃硝酸(Nitric Acid, HNO_3)：友和化工股份有限公司商購，試藥級，煙霧褐色液體，純度 95%，與濃硫酸混合作為硝化劑。
- (2) 濃硫酸(Sulfuric Acid, H_2SO_4)：友和貿易股份有限公司商購，試藥級，煙霧無色液體，純度 98%，作為脫水劑使用。
- (3) 三甘醇(Triethylene Glycol, $\text{C}_6\text{H}_{14}\text{O}_4$)：友和貿易股份有限公司商購，試藥級，無色黏稠液體，純度 99%，合成硝化三乙二醇之原料。
- (4) 二氯甲烷(Dichloromethane, CH_2Cl_2)：友和貿易股份有限公司商購，試藥級，無色液體，純度 99.9%，作為溶劑使用。
- (5) 碳酸鈉(Sodium Carbonate, Na_2CO_3)：友和貿易股份有限公司商購，試藥級，白色粉末，密度 2.53g/cm^3 ，純度 99.5%，溶於水後作為鹼洗溶液使用。
- (6) 火炸藥：FOX-7 自行合成，NC、NQ、TNT、RDX 及 HMX 由生製中心第 205 廠提供。

2.2 實驗儀器

- (1) 硝化反應設備（自行組裝）：本實驗設備包含三口雙層反應槽(500 ml)、電動攪拌器(Corning PC-420)、恆溫循環水槽(型式 Uniss RC-10：溫度範圍 $-20\sim 100\text{ }^\circ\text{C}$ ，壓縮機為 1/6 HP)及電子天平(型式 Precisa XS 125A-SCS：最大秤量/靈敏度為 $125\text{ g}/0.0001\text{ g}$)等。
- (2) 傅立葉轉換紅外線光譜儀(Fourier Transform Infrared Spectrometer, FTIR)：本研究使用

Varian 640-IR 型傅立葉轉換紅外線光譜儀，可量測合成產物 TEGDN 官能基的吸收光譜，藉以確認合成產物，掃描範圍 $4000\sim 450\text{ cm}^{-1}$ 。

- (3) 液態核磁共振光譜儀(Nuclear Magnetic Resonance, NMR)：本研究使用臺大貴中心 BRUKER AVIII-500MHz FT-NMR 型液態核磁共振光譜儀，先將合成產物 TEGDN 以氘化氯仿(Deuterated Chloroform, CDCl_3)溶劑溶解後，測定其 ^1H 譜(500MHz)及 ^{13}C 譜(125MHz)，藉以確認合成產物之有機分子結構。
- (4) 元素分析儀(Elemental Analyzer, EA)：本研究使用臺大貴儀中心 Heraeus CHN-O-Rapid 型元素分析儀，可量測合成產物 TEGDN 所含元素，確認其 C、H、N 的重量百分比率。
- (5) 微差掃描熱卡計(Differential Scanning Calorimeter, DSC)：本研究使用 TA WATERS DSC Q-20 型微差掃描熱卡計，可量測合成產物 TEGDN 的熱分解反應起始溫度及放熱量的變化情形，測試樣品盤使用陶瓷坩鍋，樣品測試量 5-10mg，反應溫度量測範圍室溫至 $725\text{ }^\circ\text{C}$ 。
- (6) 動態流變儀(Dynamic Shear Rheometer, DSR)：本研究使用 Haake RheoStress 600 型動態流變儀，可測量合成產物 TEGDN 的黏度特性。
- (7) 真空安定性試驗(Vacuum Stability Test, VST)設備：本實驗設備係自行組裝，由藥品反應器、恆溫加熱槽、壓力感測器及數據擷取系統與軟體等組成，可量測合成產物 TEGDN 與 FOX-7、NC、NQ、TNT、RDX 及 HMX 等火炸藥之相容性。

2.3 實驗規劃

2.3.1 田口法實驗設計

田口法是一種藉由實驗設計來進行系統參數最佳化的方法，因為利用直交表來蒐集資料，故可以較少的實驗次數來獲得所需可靠的品質特性，而品質特性通常分為望大型（品質特性值

越大越好)、望小型(品質特性值越小越好)及望目型(品質特性值逼近目標值)等三種形式。

本研究的目的主要在探討 TEGDN 的最佳合成參數,實驗條件選定三個控制因子包括反應溫度(A)、硝酸/硫酸重量比(B)及硝酸/TEG 莫耳比(C),每個控制因子各區分三個不同的水準,TEGDN 合成實驗參數規劃如表 1 所示,接著利用 $L_9(3^3)$ 直交表的設計規劃 9 組實驗如表 2 所示,並設定產率(%)為品質特性需求,採望大型方式,利用 S/N(Signal-to-Noise Ratio)值分析最佳合成參數結合,最後再實施驗證性實驗確認田口實驗設計法的正確性。

表 1. TEGDN 合成實驗參數規劃表

控制因子	水準		
	1	2	3
A.反應溫度(°C)	15	10	5
B.硝酸/硫酸重量比(wt%)	70:30	60:40	50:50
C.硝酸/TEG 莫耳比	6:1	5:1	4:1

表 2. TEGDN 合成實驗規劃 $L_9(3^3)$ 直交表

實驗編號	A 反應溫度 (°C)	B 硝酸/硫酸重 量比(wt%)	C 硝酸/TEG 莫 耳比
A1	15	70:30	6:1
A2	15	60:40	5:1
A3	15	50:50	4:1
A4	10	70:30	4:1
A5	10	60:40	6:1
A6	10	50:50	5:1
A7	5	70:30	5:1
A8	5	60:40	4:1
A9	5	50:50	6:1

2.3.2 TEGDN 合成實驗步驟

本研究使用硝/硫混酸硝化法進行 TEGDN 的

合成實驗,先將三甘醇(TEG)與二氯甲烷(CH_2Cl_2)溶劑充分混合後,緩慢加入硝/硫混酸溶液進行硝化反應,再經分離、洗滌及過濾烘乾程序,最後獲的 TEGDN 合成產物,實驗步驟如圖 2 所示,簡述如下:

- (1)依實驗規劃秤取所需的 95%濃硝酸,加入 500 ml 三口反應槽內,啟動恆溫循環系統,分別將溫度控制在 5°C 、 10°C 或 15°C ,並伴隨著機械式攪拌裝置,緩慢滴入所需的 95%濃硫酸,以達到充分混合。
- (2)俟反應槽溫度穩定後,將 25 克的 TEG 與 100 毫升的 CH_2Cl_2 溶劑充份混合,緩慢加入反應槽中,藉由機械式攪拌裝置進行硝化反應,以避免局部過熱現象發生,導致危安狀況。
- (3)進料完成後,持續攪拌 5 分鐘使硝化反應完全,再加入冰水中止反應。
- (4)將反應槽完成反應的混合溶液倒入分液漏斗,靜置使反應產物與廢酸溶液分離,此時區分上層為廢酸溶液,呈現淡黃色,下層則為有機層,即為 TEGDN 合成產物。
- (5)汲取下層產物,先以 100 毫升的 10%無水碳酸鈉溶液鹼洗至中性,再以 100 毫升的去離子水清洗兩次。
- (6)最後將溶液過濾置於 65°C 烘箱內 6 小時,以驅離溶劑,所得產物即為無色透明略帶油性之液狀 TEGDN。

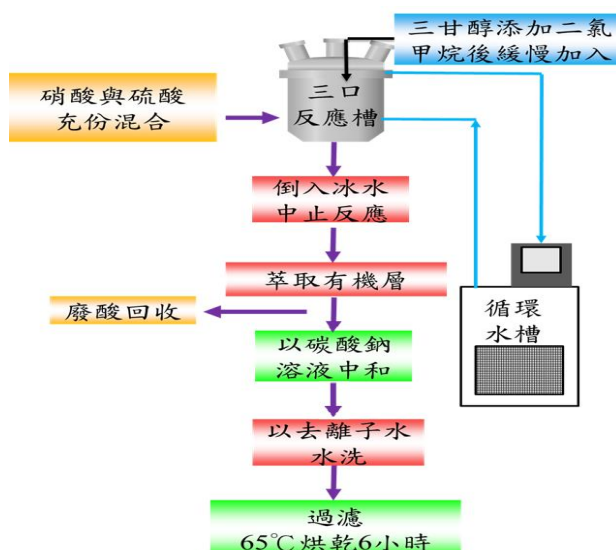


圖 2. 硝/硫酸硝化法合成 TEGDN 實驗程序

2.3.3 合成產物 TEGDN 鑑定及特性分析

後續運用傅立葉轉換紅外線光譜儀 (FTIR)、液態核磁共振光譜儀(NMR)、元素分析儀(EA)及微差掃描熱卡計(DSC)，鑑定硝化合成產物為所需的 TEGDN，再利用動態流變儀(DSR)及真空安定性試驗(VST)設備，量測分析 TEGDN 的特性。

三、結果與討論

3.1 田口實驗規劃分析

TEGDN 合成依田口法規劃的 9 組實驗條件分別完成 3 次實驗，實驗結果取平均值列於表 3，其中以第 A2 組實驗條件(反應溫度 15℃、硝酸/硫酸重量比為 60:40 及硝酸/TEG 的莫耳數比為 5:1)所得 TEGDN 的產率 94.3%最高，產率的計算係以 TEG 經硝化反應完全合成為 TEGDN 的「理論產量」為基準，反應的「實際產量」與「理論產量」的比值即為產率。

表 3. TEGDN 合成實驗結果

實驗編號	A 反應溫度 (℃)	B 硝酸/硫酸重 量比(wt%)	C 硝酸/TEG 莫耳比	產率 (%)
A1	15	70:30	6:1	88.9±0.4
A2	15	60:40	5:1	94.3±0.3
A3	15	50:50	4:1	92.1±0.4
A4	10	70:30	4:1	91.3±0.5
A5	10	60:40	6:1	85.5±0.3
A6	10	50:50	5:1	89.8±0.8
A7	5	70:30	5:1	88.6±0.9
A8	5	60:40	4:1	92.6±0.7

A9	5	50:50	6:1	83.8±0.7
----	---	-------	-----	----------

實驗數據經由田口法分析轉換為 S/N 值，可獲得每個控制因子在不同水準下的產率如表 4 所示，三個控制因子反應溫度(A)、硝/硫重量比(B)及硝酸/TEG 莫耳比(C)的差值分別為 3.42%、5.05%及 2.55%，因此可知各控制因子對反應產率的影響程度為硝/硫重量比>反應溫度>硝酸/TEG 莫耳比，控制因子個別對產率的影響趨勢則如圖 3 所示，其中每個控制因子的產率最大值即為最佳條件，再將三個控制因子的最佳條件加以組合為 A1B2C3 即為最佳實驗條件(反應溫度 15℃、硝酸/硫酸重量比 60:40、硝酸/TEG 莫耳比 4:1)。

表 4. TEGDN 產率 S/N 值分析 (單位：%)

水準	控制因子		
	A	B	C
	反應溫度 (℃)	硝酸/硫酸重 量比(wt%)	硝酸/TEG 莫 耳比
1	91.75	89.96	90.25
2	88.85	92.20	89.45
3	88.33	87.15	92.00
Range	3.42	5.05	2.55
Rank	2	1	3

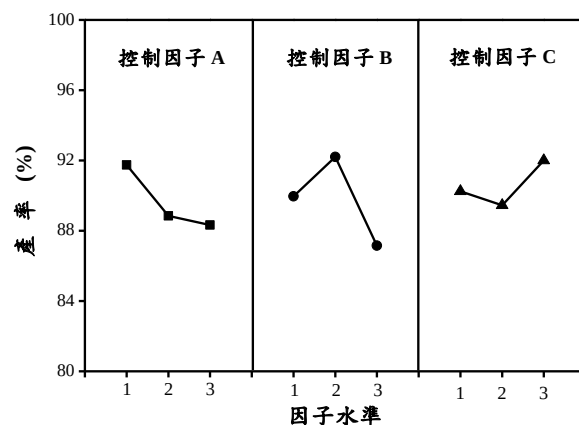


圖 3. 控制因子對產率影響趨勢圖

為了確認田口實驗規劃的準確性必須實施驗證性實驗，以最佳實驗參數組合條件 A1B2C3 進行三組平行實驗，實驗條件及結果如表 5 所示，實驗結果顯示在此最佳實驗參數組合條件下，TEGDN 合成的平均產率為 95.3% 大於原口實驗規劃的 9 組實驗條件。因此，田口實驗規劃法分析所得的最佳實驗參數組合條件得到驗證。

表 5. TEGDN 合成驗證性實驗條件及結果

實驗 編號	A	B	C	產率 (%)
	反應溫度	硝酸/硫酸重	硝酸/TEG	
	(°C)	量比(wt%)	莫耳比	
B1				94.5
B2	15	60:40	4:1	95.8
	(A1)	(B2)	(C3)	
B3				95.5
平均				95.3±0.8

3.2 合成產物 TEGDN 的鑑定及特性分析

實驗合成產物後續利用傅立葉轉換紅外線光譜儀(FTIR)、液態核磁共振光譜儀(NMR)、元素分析儀(EA)及微差掃描熱卡計(DSC)實施鑑定分析。

(1) FTIR 圖譜分析

合成產物之 FTIR 圖譜如圖 4 所示，比對 TEGDN 各官能基：O-CH₂ 官能基在 2892cm⁻¹ 特徵吸收峰位置，NO₂ 官能基在 1626、1279、758、654 cm⁻¹ 特徵吸收峰位置，而 O-N 官能基在 856cm⁻¹ 伸縮振動吸收峰位置，O-H 官能基吸收峰消失，而偏移形成為 O-NO₂ 的特徵吸收峰，此圖譜分析顯示 TEG 已硝化形成 TEGDN。

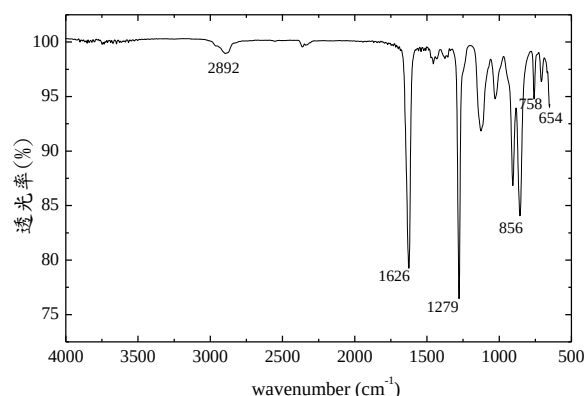


圖 4. 合成產物之 FTIR 圖譜

(2) (¹H、¹³C)-NMR 圖譜分析

將合成產物溶於 CDCl₃，利用液態核磁共振光譜儀(NMR)測定(¹H-NMR、¹³C-NMR)譜圖分析產物結構。¹H-NMR 圖譜如圖 5 所示，其吸收峰的化學位移分別為 δ=4.58~4.6 ppm (m, 4H, -CH₂-ONO₂)、δ=3.74~3.76 ppm (m, 4H, -CH₂-O) 及 δ=3.62 ppm (s, 4H, O-CH₂-CH₂-O)，與文獻 TEGDN 圖譜相符[14]；¹³C-NMR 圖譜如圖 6 所示，其吸收峰的化學位移分別為 δ=72.05ppm (-CH₂-ONO₂)、δ=67.27 ppm (-CH₂-O)及 δ=70.79 ppm (O-CH₂-CH₂-O)，另 δ=76.74-77.24 ppm 則為溶劑 CDCl₃ 的化學位移，亦與文獻相符[14]，另 δ=76.74-77.24 ppm 為溶劑 CDCl₃ 碳的化學位移，藉由(¹H、¹³C)-NMR 圖譜分析亦可研判 TEG 與硝化劑已合成為 TEGDN。

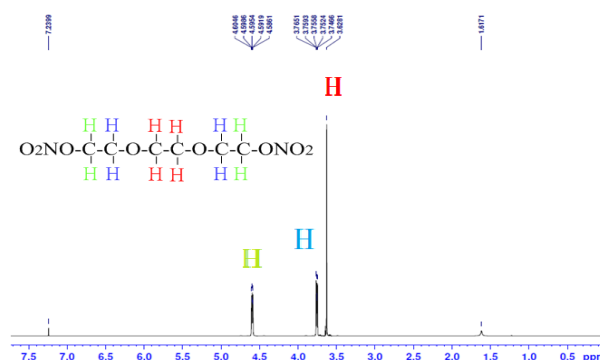


圖 5. 合成產物之 ¹H-NMR 圖譜

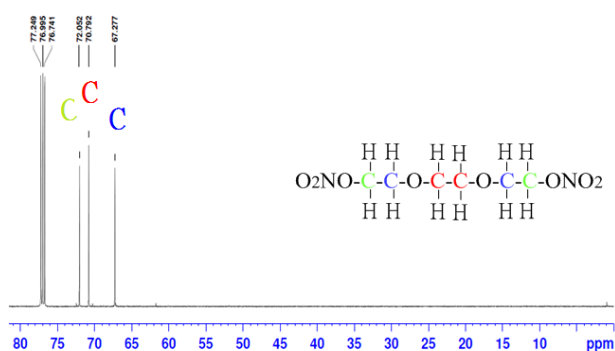


圖 6. 合成產物之 ^{13}C -NMR 圖譜

(3)元素分析

藉由元素分析儀(EA)分析合成產物之 C、H、N、O 元素的重量百分比，分析結果如表 6 所示，TEGDN 各元素重量比率的理論值為 C:30%、H:5%、N:11.67%、O:53.33%，合成產物檢測平均值為 C:30.57%、H:5.14%、N:11.57%、O:52.72%，實驗量測值與理論值相符，亦可佐證合成產物為 TEGDN。

表 6. 合成產物元素分析表

分子式 $\text{C}_6\text{H}_{12}\text{N}_2\text{O}_8$				
區分	C%	H%	N%	O%
樣品 A	30.33	4.89	11.44	53.34
樣品 B	30.62	5.12	11.61	52.65
平均值	30.57	5.14	11.57	52.72
理論值	30.00	5.00	11.67	53.33

(4)熱化學特性分析

量取合成產物 TEGDN 1mg 放入開放式鉗鍋，安置於微差掃描熱卡計，在通入氮氣(流速 50mL/min)氣氛條件，於常壓下分別以升溫速率 2 及 10°C/min 進行熱化學特性分析，DSC 圖譜如圖 7 所示，在升溫速率 2 °C/min 時僅有一個微小的吸熱峰，而升溫速率 10 °C/min 時則在 175-275 °C 之間有一個明顯的放熱峰，由於 TEGDN 的沸點溫度為 89.3°C，而分解起始溫度為 180°C，故在 2°C/min 緩慢的升溫條件下，TEGDN 在未達

分解溫度前即氣化並伴隨緩慢的升溫過程而逸散，故無明顯的分解放熱峰，此實驗結果與文獻相符[15]。

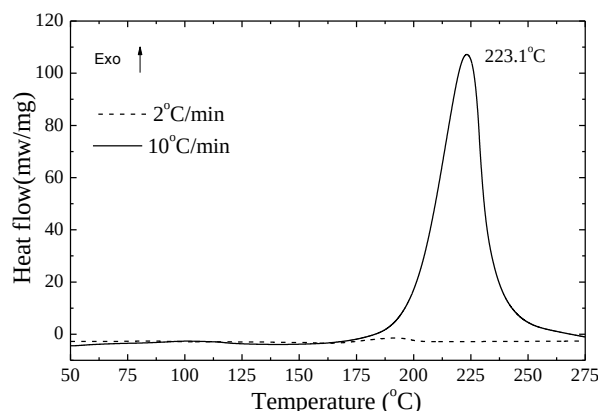


圖 7. 合成產物 TEGDN 的 DSC 圖譜

最後再利用動態流變儀(DSR)及真空安定性試驗(VST)設備量測分析合成產物 TEGDN 的黏度特性及與不同種類火炸藥的相容性。

(5)黏度特性分析

以動態流變儀(DSR)在 20°C 的條件下，量測合成產物 TEGDN 的黏度為 0.013 Pa.s，如圖 8 所示，低於 NG 的黏度 0.036 Pa.s [16]。

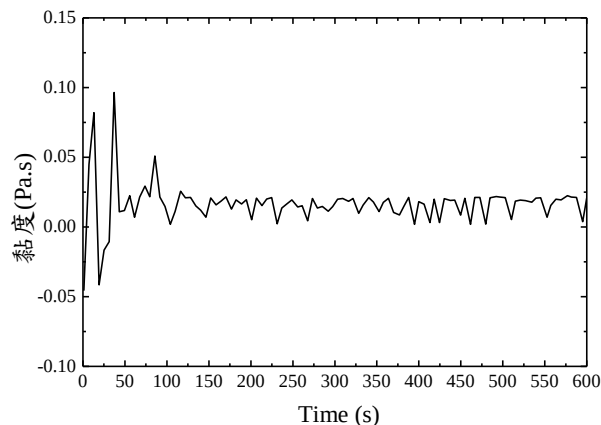


圖 8. 合成產物 TEGDN 黏度測試圖譜

(6)相容性分析

量取合成產物 TEGDN 與各別不同種類的火炸藥(FOX-7、NC、NQ、TNT、RDX 及 HMX)以 50/50wt%混融後，取樣品 1 公克置入真空安定性試驗設備的反應器內，先加熱至 100 °C，並抽真空至 5mmHg 以下，置於恆溫加熱槽在 100°C

的條件下維持 40 小時，觀察並計錄壓力變化，實驗結果如圖 9 所示，再藉由理想氣體方程式，計算在常壓下氣體體積的釋放量，如表 7 所示。實驗結果發現 TEGDN 與 NC、NQ、TNT、RDX、HMX 及 FOX-7 等火炸藥相容性試驗測得釋放之氣體體積均符合 0.6 ml/g 以下之相容性標準，顯示 TEGDN 與上述火炸藥相容性良好，長期儲存不易變質。

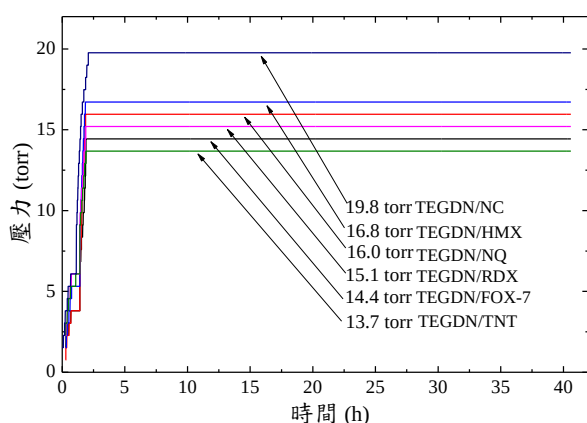


圖 9. 合成產物 TEGDN 與不同單質炸藥相容性試驗壓力變化圖

表 7. 合成產物 TEGDN 與不同單質炸藥相容性試驗釋放體積

炸藥	NC	HMX	NQ	RDX	FOX-7	TNT
產生體積 (ml/g)	0.57	0.44	0.44	0.38	0.38	0.33

五、結論

本研究以硝/硫混酸為硝化劑，將 TEG 經硝化反應生成 TEGDN，並利用田口實驗設計法探討最適合成參數；此外，並對合成產物進行鑑定及特性分析，得到下列結論：

(1) 田口實驗設計法分析所得最適合成參數為反應溫度 15°C、硝酸/硫酸重量比為 60:40 及硝酸/三甘醇的莫耳數比為 4:1，可得 TEGDN 的最大產率為 95.2%。

(2) 經由 FTIR、(¹H、¹³C)-NMR、EA 及 DSC 等儀器分析鑑定，確認合成產物為 TEGDN。
(3) 合成產物 TEGDN 的黏度為 0.013 Pa.s，低於 NG 的 0.036 Pa.s，更有利於製程加工作業。
(4) 將 TEGDN 與相同重量之 NC、NQ、TNT、RDX、HMX 及 FOX-7 均勻混合後，進行相容性實驗顯示相容性均屬良好，故可取代 NG 作為高能低易損性發射藥及推進劑之增塑劑。

致 謝

本研究承蒙軍備局生產製造中心第 205 廠提供火炸藥樣品使研究得以順利完成，特此致謝。

參考文獻

- [1] Agrawal, J. P., High Energy Materials: Propellants, Explosives and Pyrotechnics, John Wiley & Sons Publication, Federal Republic of Germany, pp. 267-271, 2010.
- [2] 王澤山, 火炸藥科學技術, 北京理工大學出版社, 中國北京, 第 284 頁, 2002。
- [3] Chan, M. L., Roy, E. M., Turner, A., Energetic Binder Explosive, US Patent: 5,316,600, 1994.
- [4] 徐復銘, “21 世紀先進發射藥(2): 低敏感高能發射藥—新配方、裝藥、點火和理論模擬技術,” 南京理工大學學報, 中國南京, 第 26 卷, 第 4 期, 第 1-4 頁, 2003。
- [5] Doriath, G., “Energetic insensitive propellants for solid and ducted rockets,” Journal of Propulsion and Power, Vol. 11, pp. 870-882, 1995.
- [6] Meyer, R., Köhler, J., Homburg, A., Explosives, Sixth, Completely Revised Edition, VCH, Wiley-VCH Verlag GmbH, Weinheim, pp. 382, 2007
- [7] Military Explosives (TM 9-1300-214),

Department of the Army, U.S.A., Washington, D.C, pp. 8.1-8.26, 1990.

- [8] 鄭件、王江寧、韓芳、田軍、宋秀鋒、周彥水，“DNTF-CMDB 推進劑的化學安定性”，火炸藥學報，第 33 卷，第 4 期，第 10-13 頁，2010。
- [9] 劉子如，含能材料的熱分所，國防工業出版社，北京，第 65-71 頁，2008。
- [10] Yi, J. H, Zhao, F. Q., Xu, S. Y., Hu, R. Z., “Effects of pressure and TEGDN content on decomposition reaction mechanism and kinetics of DB gun propellant containing the mixed ester of TEGDN and NG,” Journal of Hazardous Materials, Vol.165, pp. 853-859, 2008.
- [11] 鐘顯政、吳銘德、蕭光蜀，“高能可塑劑 DEGDN, TEGDN, TMETN 之合成”，火藥技術，第 3 卷，第 1 期，第 102-107 頁，1987。
- [12] 韓駿奇、孟子暉、孟文君、陳光文、王伯周、葛忠學，“微反應器中合成硝酸酯炸藥”，Chinese Journal Energetic Materials，第 18 卷，第 1 期，第 34-36 頁，2010。
- [13] Nicholas, A. S., Alexander, J.P., Methods of Producing Nitrate Esters, US Patent: 8658818 B2, pp. 8-9, 2014.
- [14] 吳銀秀，“雙基發射藥及其組分的結構與性能研究”，碩士論文，南京理工大學，南京，第 32-33 頁，2013。
- [15] 陳沛、趙鳳起、李上文、陰翠梅，“二縮三乙二醇二硝酸酯的熱分解性能”，火工品，第 3 期，第 6-7 頁，1999。
- [16] Military Explosives (TM 9-1300-214), Department of the Army, U.S.A., Washington, D.C, pp. 8.9, 1990.

曾文佐等

硝化三乙二醇最適合成參數及特性之研究

還原氧化石墨烯/氧化鎳雙層對高分子太陽能電池效率之影響

歐珍方*

國立勤益科技大學化工與材料工程系

摘 要

本研究採用還原氧化石墨烯(r-GO, reduced-GO)與氧化鎳(NiO_x)雙層結構來取代傳統的 PEDOT:PSS(3,4-ethylenedioxythiophene):poly(styrenesulfonate)作為電洞傳輸層來製備高分子太陽能電池，太陽能電池結構為 ITO/r-GO/ NiO_x /P3HT:PC₆₁BM/Ca/Al。我們研究以 r-GO/ NiO_x 取代 PEDOT:PSS 對電池光電特性之影響。由結果發現以 r-GO/ NiO_x 製備之高分子太陽能電池，其短路電流密度(J_{sc})、填充因子(FF)與光電轉換效率(PCE)都明顯比傳統以 PEDOT:PSS 作為電洞傳輸層之電池高。其中以具有 r-GO/ NiO_x -2 雙層之電池表現出最佳的光電特性，它具有最高的短路電流密度(J_{sc}) 8.34 mA/cm²，填充因子(FF) 0.59 與光電轉換效率(PCE) 2.93 %，與傳統參考電池比較，短路電流密度與光電轉換效率分別提升了 51%及 73 %。

關鍵詞：還原氧化石墨烯，高分子太陽能電池，氧化鎳，光電轉換效率

Effect of Reduced Graphene Oxide/Nickel Oxide Bi-layer on Cell Performance of Polymer Solar Cells

Cheng-Fang Ou

Department of Chemical and Materials Engineering, National Chin-Yi University of Technology

ABSTRACT

This work demonstrates the high performance reduced graphene oxide (r-GO)/nickel oxide (NiO_x) double decked hole transport layer (HTL) in the P3HT: PC₆₁BM based bulk heterojunction polymer solar cells (PSCs). The cell structure was ITO/r-GO// NiO_x /P3HT:PC₆₁BM (1:1 weight ratio)/Ca/Al. We study the effect of r-GO/ NiO_x bi-layer on the photovoltaic performance. From the results, the short circuit density (J_{sc})、fill factor (FF) and power conversion efficiency (PCE) of the solar cells employing r-GO/ NiO_x HTL were always higher than those of the reference cell with PEDOT:PSS HTL. The cell employing r-GO/ NiO_x -2 HTL has the highest J_{sc} of 8.34 mA/cm², an increase of 51%, and the highest PCE of 2.93%, an increase of 73%, as compared to the reference polymer solar cells.

Keywords: reduced-graphene oxide, polymer solar cell, nickel oxide, power conversion efficiency

一、前言

再生能源主要包括太陽能、水力能、風力能、地熱能、生質能、海洋能等等，不但取之不盡、用之不竭，更不會有環保方面的顧慮，在這些再生能源中，目前以太陽能最受到矚目。太陽能與其他再生能源相比更有隨手可得的優點，其他再生能源往往受限於區域或地形環境[1]。太陽能電池自 1954 年貝爾實驗室以半導體製程完成並突破 6% 的能量轉換效率以來，隨著製程技術與材料研發的進步，目前太陽能電池大多朝著效率的提升、尺度的縮減及製程的簡化幾個方向進行研究。太陽能電池(solar cell)，依材料大致可分為兩種類型：(1)無機太陽能電池[2]。(2)有機太陽能電池。無機太陽能電池主要包含結晶矽、非晶矽、多晶矽化合物等。有機太陽能電池主要可分成小分子、高分子太陽能電池與染料敏化太陽能電池為主體等類型。若由材料的類型進一步分類，從單晶矽電池到多晶材料再發展到現今的薄膜型無機或有機太陽能電池，元件的設計與材料的選擇上越來越多元化。有機太陽能電池因具有製程簡單、材料的設計較彈性且低成本的優勢，雖然目前效率不高，但在低能量需求或可攜式產品的電力供應上，仍具有應用的潛力。

氧化鎳的結晶結構為 NaCl 結構，結晶的氧化鎳薄膜因為具有極佳的化學穩定性、光學性質、電性及磁性故可以應用在各式各樣不同的工業。如：氧化鎳可作為反鐵磁材料、電致發光顯示器元件材料、化學感測器的感測層及 p-type 的透明導電薄膜。氧化鎳薄膜的導電機構，一般認為是由於其鎳氧比偏離化學計量比所造成。符合化學計量比(stoichiometry, Ni:O=1:1)的氧化鎳薄膜是絕緣體，電阻率在常溫下約為 10^{13} ohm-cm，但當其鎳氧比小於 1 時，氧化鎳薄膜便因結構中具 Ni^{3+} 離子而變成 p-type 半導體所以大幅提高其導電率。氧化鎳的應用大約可分成三方面：(1)催化劑(2)電池與半導體器件：在鎳鎘電池剛發現的時候，也是使用一氧化鎳作電極，但由於低溫性能不好和製造成本較高，現在已經逐漸被鋰電池取代了[3]。近來，一氧化鎳的 p 型半導體（表面含有多餘正離子）特性開始引起研究者的注意，當一氧化鎳吸附氣體或外加電場後，

其表面的鎳離子數目發生變化，從而改變一氧化鎳的導電性和顏色，因此一氧化鎳薄膜可以被用於製造電阻型半導體氣敏元件和電致變色器件。(3)光吸收材料：奈米氧化鎳在光吸收譜上表現為選擇性光吸收。2010 年 Steirer, K. X., Chesin, J. P 等學者將金屬氧化物氧化鎳(NiOx)作為電洞傳輸層，鎳金屬具有寬的帶隙和價電帶能量較接近 P3HT 的 HOMO，可抑制電子重組作為阻擋層，並且降低串聯電阻提高填充因子[4]。2010 年 Park, S. Y 等學者研究採用磁控濺射 p 型鎳金屬氧化物薄膜作為陽極緩衝層，可增加電池轉化效率和穩定性，光電轉化效率和填充因子分別高達 $2.8 \pm 0.1\%$ 和 0.62 ± 0.02 ，且鎳金屬應用壽命比傳統 PEDOT:PSS 高三倍[5]。由前人的研究可以得知，將無機金屬奈米材料應用於有機太陽能電池元件，應用在電洞傳輸層與主動層都可以有效提升太陽能電池元件的穩定性與光電轉換效率。

石墨烯具有非常好的導電、導熱、散熱、強度特性，未來取代現有電子材料的可行性非常大[6-8]。石墨烯既是最薄的材料，也是最強韌的材料，斷裂強度比最好的鋼材還要高出百倍。同時它又有很好的彈性，拉伸幅度能達到自身尺寸的 20%。石墨烯的導電性比銅更好，導熱性遠超一切其他材料。它的電阻只約 $10^{-6} \Omega \cdot \text{cm}$ ，比銅或銀更低，為目前世界上電阻率最小的材料；在室溫下，其電子遷移率(electron mobility)超過 $15000 \text{ cm}^2/\text{V} \cdot \text{s}$ ，是矽晶($1400 \text{ cm}^2/\text{V} \cdot \text{s}$)的 10 倍以上。石墨烯於高分子太陽能電池的應用可分成三個分向：1.以石墨烯為電子受體(acceptor)，2.以石墨烯為界面物質(interface layer)，3.以石墨烯為電極(anode or cathode)[9]。以石墨烯取代無機氧化銦錫(ITO 透明電極)製作成電池之光電轉換效率不高皆低於 2.0%[10-11]。Ryu 等人[12]將 GO(graphene oxide)水溶液，以 1000 rpm 轉速塗佈於 ITO 電極上，然後於手套箱內 200°C 加熱板熱處理 20 分鐘取代傳統電洞傳輸層 PEDOT:PSS 之電池光電轉換效率由 2.33% 提升至 2.75% 提升了 18%。Kim 等人[13]將 GO 與利用 hydrazine 化學還原之 r-GO 當做 HTL，膜厚 5nm 之 r-GO 之電池光電轉換效率為 2.83%。2013 年 Guo 等人[14]研究以不同 GO 濃度(2 與 5 mg/ml)、不同塗佈轉速、不同熱處

理溫度來製備 HTL，最佳電池光電轉換效率為 2.75%。Jeon 等人[15] 研究以 250°C 熱處理之適當還原 GO 取代傳統電洞傳輸層 PEDOT:PSS 之電池光電轉換效率由 3.85% 提升至 3.98%。2013 Jeon 等人[16] 研究以 100, 200, 300, and 400°C 熱處理適當還原 GO，取代傳統電洞傳輸層 PEDOT:PSS，400°C 熱處理 GO 之電池具有最高之光電轉換效率，效率為 3.71%。2017 Rafique 等人[17] 研究以 GO/PEDOT:PSS 雙層為 HTL，以 PCDTBT:PC₇₁BM 為主動層之電池，具有 4.28% 的光電轉換效率。由上述研究知 GO 當 HTL 對電池光電轉換效率影響之因素很多。

傳統有機高分子太陽能電池之電洞傳輸層為 PEDOT:PSS (polyethylene dioxythiophene doped with polystyrene-sulfonic acid)，然而 PEDOT:PSS 具有高酸性，在高溫環境下會腐蝕電池陽極 ITO (indium tin oxide)，因而使電池劣化及降低性能。本研究將藉由改良式 Hummer 法製備適度還原的氧化石墨烯 (r-GO)，然後將還原的氧化石墨烯添加於電池元件中當作電洞傳輸層取代傳統的 PEDOT:PSS 電洞傳輸層，於我們先前之文章[18]，探討 GO 經不同溫度 150、200 及 250°C 熱處理對電池光電轉換效率之影響。本篇論文，我們另外於還原氧化石墨烯上另外再添加一層氧化鎳 (NiO_x)，研究 r-GO/NiO_x 此雙層電洞傳輸層對高分子太陽電池光電性質之影響與探討其可能產生之原因。

二、研究方法

2.1 實驗材料

- (1).ITO (indium tin oxide) glass substrate，片電阻：15 ohm/sq，厚度：120 nm，廠商：Lumtec。
- (2).PEDOT:PSS (polyethylene dioxythiophene doped with polystyrene-sulfonic acid)，型號：BATRON P VP.AI 4083，廠商：H. C. Starck (Bayer AG)，當作電洞傳輸層。
- (3).P3HT:Regioregular poly(3-hexylthiophene) 廠商：Rieke Metals, Inc，當作電子施體 (donor)。
- (4).PCBM: [6,6]-phenyl-C₆₁-butyric acid methyl ester，純度：99.5%，廠商：Nano-C，當作電子受體 (acceptor)。
- (5).石墨：325 mesh，純度 98%，廠商：美國，Alfa Aesar Co, Ward Hill, Massachusetts。

2.2 r-GO 溶液之製備

本研究以改良式 Hummer 法將石墨製備成氧化石墨 (graphene oxide, GO)，首先將石墨片置入濃硫酸與濃硝酸混合溶液中攪拌之，緩緩加入過錳酸鉀氧化劑，氧化反應持續 7 天後，加入過氧化氫以終止反應，接著以 5wt% 鹽酸水溶液清洗並移除多餘之硫酸鹽離子，以逆滲透水清洗產物，經由離心法將懸浮溶液純化，所得固態氧化石墨再重複此程序直到懸浮溶液呈中性，接著於室溫下進行真空乾燥以得到棕色粉體。將 GO 棕色粉體以 DMF (N,N-dimethylformamide) 溶液稀釋成濃度 1.0 mg/ml 之溶液。

2.3 NiO_x 溶液之製備

將 2M 之氫氧化鋰溶液加入水中並攪拌以形成 0.5M 醋酸鎳的溶液，直到 pH 為 9.0 為止，將綠色沉澱物從母液中離心分離。然後，利用冰醋酸對該漿料進行膠溶直至 pH 為 4.5 為止。然後於溶液中加入少量的水，以獲得適當的粘度，再將溶液進行超音波震盪處理並過濾。

2.4 高分子太陽能電池之製備

(1).ITO 玻璃蝕刻與清洗

詳細步驟請參考我們以前之論文[18]。

(2).PEDOT:PSS 電洞傳輸層之製備

詳細步驟請參考論文[18]。

(3).r-GO/NiO_x 電洞傳輸層之製備

將製備完成的 r-GO 分散液以 1000rpm 旋轉塗佈於 ITO 玻璃上，塗佈完成後放置於 250°C 加熱板上退火 10 min，加熱板之溫度經表面溫度計量測確認。NiO_x 分散液以 4000rpm 塗佈於 r-GO 層上，放置於 280°C 加熱板上熱處理 30min。NiO_x 塗佈之層數分別有 1、2 與 3 層，代號為 NiO_x-1、NiO_x-2 與 NiO_x-3。

(4).主動層的製備

P3HT:PC₆₁BM 以重量比 1::1 溶於二氯苯中，在 70°C 下加熱攪拌 12 小時以上，以旋轉塗佈 (step1: 5000 rpm，50 秒；step2: 1600 rpm，1 秒)，旋塗於 r-GO 上。塗佈完成後，靜置於培養皿 30 分鐘，使溶劑以慢速率狀態揮發，此步驟稱之為溶劑退火，完成後再以 120°C 加熱 30 分鐘，將殘餘的溶劑完全揮發。

(5).陰極之製備

詳細步驟請參考論文[18]。製備之高分子太陽能電池結構如圖1。

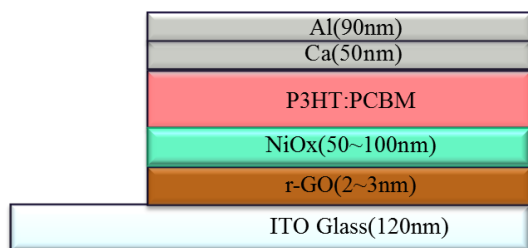


圖 1. 高分子太陽能電池結構圖

2.5 高分子太陽能電池之製備儀器測試

紫外光-可見光光譜儀 (UV-vis)，掃描波長範圍：800-300nm，觀察主動層對光的吸收度，儀器型號 Hitachi，U-2900。場發射掃描式電子顯微鏡(FE-SEM)，利用真空蒸鍍機在待測物表面鍍上金，鍍金時間為 60 秒，儀器型號 JEOL，JSM-7401F。螢光光譜儀(PL)，激發光源波長為 405 nm，掃描波長範圍 550-750nm，儀器型號：Shimadzu，RF-5301PC。太陽光模擬光源系統，系統為 Newport 66901；照光面積：10 mm²，測試光源：功率 100 mW/cm²、A.M 1.5 G。詳細步驟請參考我們以前之論文[15]。透過太陽光模擬器測量出短路電流密度 J_{sc} 、開路電壓 (open-circuit voltage) V_{oc} 、填充因子 FF 及光電轉換效率。

三、結果與討論

3.1 紫外光-可見光光譜分析

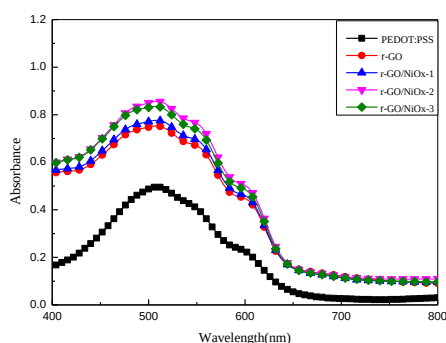


圖 2. 還原氧化石墨烯/與氧化鎳之UV-Vis圖

由圖 2 可看出為還原氧化石墨烯與氧化鎳分散液作為電洞傳輸層後的 UV-Vis 圖；圖中 P3HT 的吸收波長約在 450-550 nm，在 525 nm 附近有較強的吸收譜帶，範圍從可見光到紅外光，能使吸光範圍擴大，吸收光越多能產生更多的電子電洞對，增加光電流，且發現在還原氧化石墨烯與氧化鎳的加入可提高吸收度及吸收光範圍。以 PEDOT:PSS 為電洞傳輸層之最大吸收度為 0.51，以 r-GO 為電洞傳輸層之最大吸收度提高至 0.74，提升了 45%。於 r-GO 上再添加 NiOx 又可再提高最大吸收度。其中以 NiOx 塗佈兩層之 r-GO/NiOx-2，具有最大吸收度，與 PEDOT:PSS 為電洞傳輸層比較，提升了 75%。若與 r-GO 比較，則提升了 20.3%。表示於 r-GO 上再添加 NiOx 可提高電池之光吸收度。我們推測於還原氧化石墨烯上添加氧化鎳時，造成吸光度增加之原因，主要是氧化鎳金屬的粒子表面會與 r-GO 相互填補使表面平坦化，因此增加吸收度。同時也因此降低串聯電阻；減少電子電洞在結合。由此結果知，以 r-GO/NiOx-2 為電洞傳輸層，具有最高的吸收度，吸光效果最佳。

表 1. 不同電洞傳輸層之最大光吸收度

PET/ITO/HTL/P3HT:PCPDTBT:PCBM/Ca/Al		
HTL 成份	最大吸收度 (a.u.)	提升率 (%)
PEDOT:PSS	0.51	--
r-GO	0.74	45
r-GO/NiOx-1	0.83	63
r-GO/NiOx-2	0.89	75
r-GO/NiOx-3	0.76	49

3.2 螢光放射光強度分析

影響表面增強螢光的原因有螢光物質和金屬之間的共振能量轉移，之後螢光物質和金屬之間的交互作用也會造成放光速率和非放光速率的改變。圖3為不同電洞傳輸層組成之螢光光譜圖，淬熄面積比例如表2所示。在電荷轉移過程中，光由P3HT中誘導激子，接著在P3HT和PC₆₁BM之介面間分離，進而產生淬

熄(quenching)現象，當淬熄面積比例增加時，

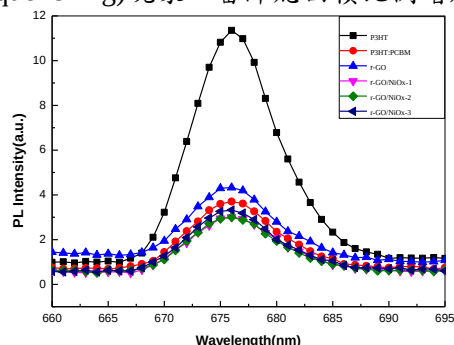


圖3. 不同電洞傳輸層組成之PL螢光光譜圖

表示電荷有效地由P3HT轉移到PC₆₁BM，淬熄面積比例為72%，由圖可看出以r-GO為電洞傳輸層，主動層放射光強度會增加，淬熄面積比例稍微遞減至69%。於r-GO塗佈不同層數之NiOx所產生之電洞傳輸層r-GO/NiOx-1、r-GO/NiOx-2、r-GO/NiOx-3，淬熄面積比例分別為76%、77%及75%。皆比PEDOT:PSS為電洞傳輸層之72%高，而r-GO/NiOx-2具有最高淬熄面積比例77%，表現最佳。由此結果知，表示具NiOx之電洞傳輸層，能提高主動層之能量轉換。

表 2. 不同電洞傳輸層之螢光淬熄面積比率

HTL 成份	quenching area (%)
PEDOT:PSS	72
r-GO	69
r-GO/NiOx-1	76
r-GO/NiOx-2	77
r-GO/NiOx-3	75

螢光淬熄包含三種機制，第一為能量轉移 (energy transfer)、激態複合物的形成(exciplex formation)和電荷轉移(charge transfer, CT)，在電荷轉移過程中，光於P3HT中誘導產生激子，接著在P3HT和還原氧化石墨烯與金屬氧化鎳異質介面之間分離，導致激子發生輻射衰變因而發生quenching 加入石墨烯能改善 $\pi-\pi^*$ 與 P3HT 分子之間的遷移，並提高電子與電洞數的平衡。還原氧化石墨烯增加螢光放射強度，原因有二，其一 PL 強度增加是由於電荷有效地由還原氧化石墨烯轉移到 P3HT 上，

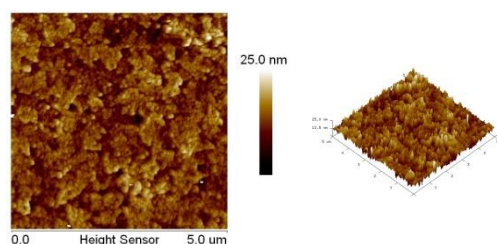
其二因為當奈米粒子置於一螢光分子附近時，利用光激發螢光物質使其內電子躍遷到激態放射出螢光，其中入射光會與奈米粒子產生交互作用，降低電子-電洞對再結合，奈米粒子其具有很大的吸光係數，能夠大幅吸收入射光達到光捕捉的效應，所以 r-GO 可增強放射螢光強度。

於 r-GO 塗佈 1~3 層之 NiOx，螢光淬熄面積比例比 PEDOT:PSS 高，而 r-GO/NiOx-2 具最高螢光淬熄面積比例，與 PEDOT:PSS 比較由 72%提升至 77%，提升了 5%，而 r-GO/NiOx-3 之螢光淬熄面積比例稍微變小至 75%，這是因為塗佈 3 層 NiOx，使氧化鎳含量過多產生聚集，導致能量轉換效果稍減，但螢光淬熄面積比例還是高於 PEDOT:PSS。

3.3 掃描探針顯微鏡(SPM)分析

薄膜表面的形態對電池性能的影響是相當重要的，過於粗糙的表面可能會形成電子的捕抓點，不利於電荷的傳輸，不平整的薄膜也會造成光的損失，使光產生率減少、電子電洞再結合機率增加，圖 4 為不同電洞傳輸層組成之 SPM 表面形態圖。高低差與均方根粗糙度(Rq)彙整於表 3，由表 3 可知 r-GO 之高低差與均方根粗糙度分別為 94 nm 與 3.95 nm。於 r-GO 添加 NiOx 層數 1~3 層，高低差與均方根粗糙度皆降低了，其高低差分別為 37 nm、29 nm、43nm，均方根粗糙度(Rq)分別為 3.53 nm、3.28 nm、3.33 nm。由此可看出，添加 NiOx，可使 r-GO 層平坦化，而 r-GO/NiOx-2，具最小的粗糙度與高低差。但 r-GO/NiOx-3 則高低差與均方根粗糙度皆比 r-GO/NiOx-2 大，因為塗佈 3 層 NiOx，使氧化鎳含量過多產生聚集。

(a) r-GO



(b) r-GO/NiOx-1

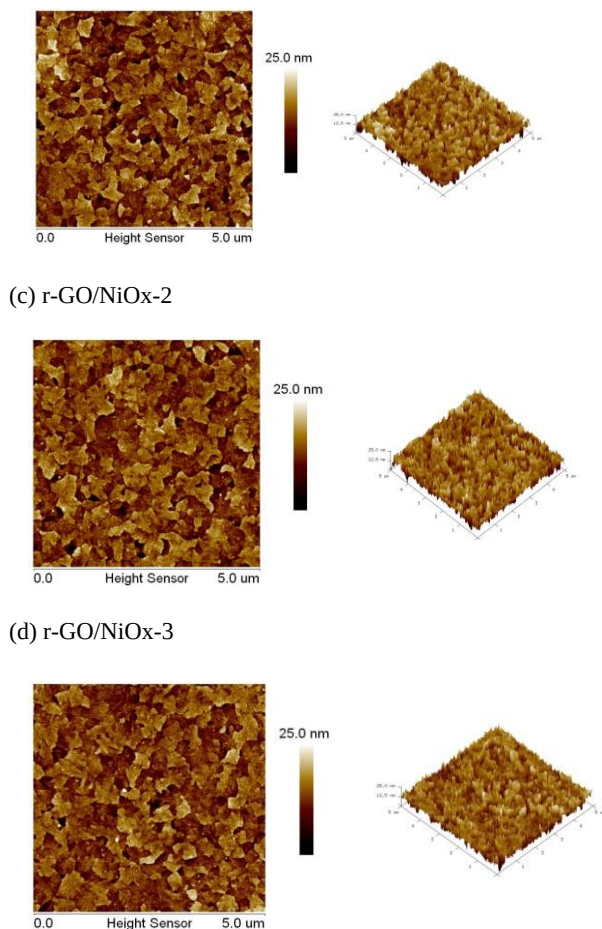


圖 4. 不同電洞傳輸層之 SPM 圖

表 3. 不同電洞傳輸層之表面粗糙度及高低差

表面粗糙度及高低差		
成份	高低差(nm)	Rq (nm)
r-GO	94	3.95
r-GO/NiOx-1	37	3.53
r-GO/NiOx-2	29	3.28
r-GO/NiOx-3	43	3.33

3.4 J-V 特性曲線分析

圖 5 為各種電池之 J-V 曲線圖，其光電特性參數如表 4 所示。由表可觀察到，傳統以 PEDOT:PSS 為電洞傳輸層之參考電池的光電

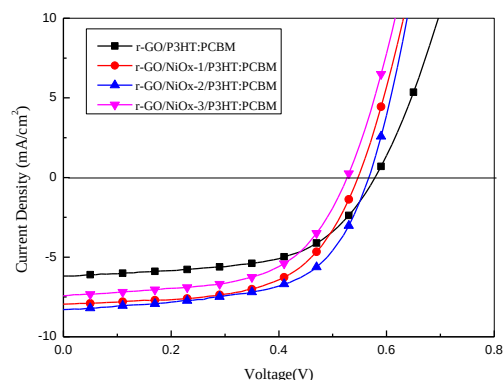


圖 5. 各種電池之 J-V 曲線圖

轉換效率 PCE(%) 為 1.69%。當以 r-GO 為電洞傳輸層之電池的光電轉換效率提升至 2.07%。於 r-GO 添加 NiOx 後，在表 4 可觀察到含 r-GO/NiOx 電池的 J_{sc} (mA/cm^2)、填充因子 (FF) 與 PCE(%) 皆有明顯的再提升。由結果得知，以 r-GO/NiOx-2 具有最佳提升效果，因為此電池具有最高的 J_{sc} 與光電轉換效率，分別為 $8.34 \text{ mA}/\text{cm}^2$ 與 2.93%，與 r-GO 當電洞傳輸層之電池比較分別提升了 34% 與 42%。若與傳統參考電池比較，短路電流密度與光電轉換效率分別提升了 51% 及 73%。

表 4. 各種電池之光電特性

不同電洞傳輸層比較				
Device	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	PCE(%)
PEDOT:PSS	0.56	5.53	0.55	1.69
r-GO	0.58	6.23	0.57	2.07
r-GO/NiOx-1	0.55	7.94	0.58	2.64
r-GO/NiOx-2	0.57	8.34	0.59	2.93
r-GO/NiOx-3	0.53	7.43	0.56	2.33

由上述結果知，以 r-GO/NiOx-2 雙層當電洞傳輸層時可得到最高的光電轉換效率。因為含 r-GO/NiOx-2 雙層之電池，由 UV-Vis 分析 (圖 2) 可知具有最大光吸收度，由螢光光譜圖 (圖 3) 可知具有最大螢光淬熄面積比例，由 SPM 圖 (圖 4) 知具有最小的粗糙度與高低差。由 ref[12] 之內文討論知，於 r-GO 上添加 NiOx

後提升光電轉換效率之主要原因，是 NiOx 平坦化了 r-GO 層，由粗糙度可知此結果。因平坦化可降低串聯電阻而提升了 FF。另一方面，由於 NiOx 之高透光率及能階(energy level)很相配於 P3HT，因此提升了 J_{sc} 與 V_{oc} ，光電轉換效率也就提升了。

由上述結果發現以 r-GO/NiOx 取代 PEDOT:PSS 製備之高分子太陽能電池，其短路電流密度、填充因子與光電轉換效率都有明顯提升。主要可能原因是 NiOx 金屬粒子會使得還原氧化石墨烯表面平坦化，因此增加光吸收度，也因此降低了串聯電阻而提升了 FF；減少電子電洞再結合，因此提升了 J_{sc} 與光電轉換效率。

四、結論

綜合歸納上一節的實驗測試數據分析，我們在本論文中所獲致的結論，簡扼敘述如下幾點：

- (1) 以 r-GO/NiOx 取代傳統 PEDOT:PSS 作為電洞傳輸層，可提高電池之短路電流密度、填充因子與光電轉換效率。
- (2) 以含有 r-GO/NiOx-2 之電池，具有最高短路電流密度、填充因子與光電轉換效率。與傳統參考電池比較，短路電流密度與光電轉換效率分別提升了 51% 及 73%。因為此電池具有最大光吸收度，最大螢光淬熄面積比例與最小的粗糙度與高低差。
- (3) 於 r-GO 層添加 NiOx 可提升電池的性能，主要之原因是 NiOx 可以改善 r-GO 的表面，使得表面平坦化，粗糙度變小，提高了光吸收度，減少電子-電洞對再結合，降低串聯電阻，因此而提高電池之短路電流密度與光電轉換效率。

參考文獻

- [1] Barnham, K. W. J., Mazzer, M., and Clive, B., "Resolving the Energy Crisis: Nuclear or Photovoltaics," *Nature. Materials*, Vol. 5, pp. 161-164, 2006.
- [2] Chang, C. L., Liang, C. W., Syu, J. J., Wang, L., and Leung, M. k., "Triphenylamine-Substituted Methanofullerene Derivatives for Enhanced Open-Circuit Voltages and Efficiencies in Polymer Solar Cells," *Solar Energy Materials and Solar Cells*, Vol. 95, pp. 2371-2379, 2011.
- [3] Nakaoka, K. J., and Ogura, K., "Semiconductor and Electrochromic Properties of Electrochemically Deposited Nickel Oxide Films," *Journal of Electroanalytical Chemistry*, Vol. 571, No. 1, pp. 93-99, 2004.
- [4] Xerxes Steirer, K., Chesin, Jordan. P., Widjonarko, Edwin., Berry, Joseph. J., Miedaner, Alex., Ginley, David. S., and Olson, Dana. C., "Solution Deposited NiO Thin-films as Hole Transport Layers in Organic Photovoltaics," *Organic Electronics*, Vol. 11, pp. 1414-1418, 2010.
- [5] Park, Sun-Young., Kim, Hye. Ri., Kang, Yong. Jin., Kim, Dong. Ho., and Kang, Jae. Wook., "Organic Solar Cells Employing Magnetron Sputtered P-type Nickel Oxide Thin Film as the Anode Buffer Layer," *Solar Energy Materials and Solar Cells*, Vol. 94, pp. 2332-2336, 2010.
- [6] Lee, C., Wei, X., Kysar, J. W., and Hone, J., "Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene," *Science.*, Vol. 321, pp. 385-388, 2008.
- [7] Neto, A. H. C., Guinea, F. Peres, N. M. R., Novoselov, K. S., and Geim, A. K., "The Electronic Properties of Graphene," *Review of Modern Physics*, Vol. 81, pp. 109-162, 2009.
- [8] Schwierz, F., "Graphene Transistors," *Nature. Nanotechnology*, Vol. 5, pp. 487-496, 2010.
- [9] Wang, X., Zhi, L., Tsao, N., Tomovic, Z., Li, J., and Müllen, K., "Transparent Carbon Films as Electrodes in Organic Solar Cells," *Angewandte. Chemie. International. Edition*, Vol. 47, pp. 2990-3002, 2008.
- [10] Wang, Y., Chen, X., Zhong, Y., Zhu, F., and Loh, K. P., "Large Area, Continuous, Few-layered Graphene as Anodes Inorganic Photovoltaic Devices," *Applied. Physics. Letters*, Vol. 95, pp. 063302/1-3, 2009.
- [11] Wu, Z. S., Ren, W., Gao, L., Zhao, J., Chen, Z., and Liu, B., "Synthesis of Graphene Sheets with High Electrical Conductivity and Good Thermal Stability by Hydrogen Arc Discharge Exfoliation," *ACS. Nano*, Vol. 3, pp. 411-417, 2009.
- [12] Ryu, M. S., Jang, J., "Effect of Solution Processed Graphene Oxide/nickel Oxide Bi-layer on Cell Performance of Bulk-heterojunction Organic Photovoltaic," *Solar Energy Materials and Solar Cells*, Vol.

- 95, pp. 2893-2896, 2011.
- [13] Yun, J. M., Yeo, J. S., Kim, J., Jeong, H. J., Kim, D. Y., Noh, Y. J., Kim, S. S., Ku, B. C., and Na, S. I., "Solution-Processable Reduced Graphene Oxide as a Novel Alternative to PEDOT:PSS Hole Transport Layer for Highly Efficient and Stable Polymer Solar Cells," *Advanced. Materials*, Vol. 23, pp. 4923-4928, 2011.
- [14] Liu, X., Kim, H., and Guo, L. J., "Optimization of Thermally Reduced Graphene Oxide for an Efficient Hole Transport Layer in Polymer Solar Cells," *Organic. Electronics*, Vol. 14, pp. 591-598, 2013.
- [15] Jeon, Y. J., Yun, J. M., Kim, D. Y., Na, S. I., and Kim, S. S., "High-performance Polymer Solar Cells with Moderately Reduced Graphene Oxide as an Efficient Hole Transporting Layer", *Solar Energy Materials & Solar cells*, Vol. 105, pp. 96-102, 2012.
- [16] Jeon, Y. J., Yun, J. M., Kim, D. Y., Na, S. I., and Kim, S. S., "Moderately Reduced Graphene Oxide as Hole Transport Layer in Polymer Solar Cells via Thermal Assisted Spray Process" *Applied. Surface Science*, Vol. 296, pp. 140-146, 2014.
- [17] Rafique, S., Abdullah, S. M., Shahid, M. M., Ansari, M. O., and Sulaaman, K., "Significantly Improved Photovoltaic Performance in Polymer Bulk Heterojunction Solar Cells with Graphene Oxide/PEDOT:PSS Double Decked Hole Transport Layer", *Scientific. Reports*, Vol. 7, pp. 39555, 2017.
- [18] Ou, C. F., and Chen, S. Y., "High-Performance Polymer Solar Cells with Graphene Oxide as a Hole Transport Layer," *Journal of Chung Cheng Institute of Technology*, Vol. 44, No. 1, pp. 75-86, 2015.

Implementation of Maximum Power Point Tracking Controller for Photovoltaic Systems

Jeng-Cheng Liu¹, Yuang-Tung Cheng¹, Sheng-Yun Hou², Chih-Lung Tseng³,
Hsien-Sen Hung¹, and Shun-Hsyung Chang^{4*}

¹Department of Electrical Engineering, National Taiwan Ocean University

²Institute of Electronic Engineering, Hwa-Hsia University of Technology

³Department of Electronic Engineering, National Kaohsiung University of Applied Science

⁴Institute of Microelectronic Engineering, National Kaohsiung Marine University

ABSTRACT

In order to increase the output overall conversion efficiency of the photovoltaic(PV) arrays it is used maximum power point tracker (MPPT) control to increase the PV arrays output power. The approach uses a novel controlling scheme to combine perturbation and observation (P&O) method together with three-point weighting comparison (TPWC) method for tracking maximum power points of PV arrays. The advantage of this approach is that it is a simple algorithm conducting on a microprocessor without redundant circuit components and thus the maximum power point of the PV systems will be obtained quickly and stably. According to the practical experiment, the system will produce fast and effective tracking results.

Keywords: photovoltaic arrays, maximum power point tracker, perturbation and observation, three-point weighting comparison

研製新型太陽電池發電系統最大功率追蹤控制器

劉俊成¹ 鄭遠東¹ 何昇運² 曾志隆³ 洪賢昇¹ 張順雄^{4*}

¹國立臺灣海洋大學電機工程學系

²華夏科技大學電子工程系

³國立高雄應用科技大學電子工程系

⁴國立高雄海洋科技大學微電子工程系

摘 要

使太陽電池模組發揮最大輸出功率使其提高整體光電轉換效率，研製新型最大功率追蹤控制器。本文結合擾動觀測法與三點加權比較法的優點，提出一種新的控制方法使太陽電池陣列追蹤最大功率。由於本研究所提出的新型最大功率追蹤控制器沒有冗餘的電路組件在微處理器上，因此可以使演算法簡單化，而讓太陽電池系統快速穩定的得到最大功率輸出。根據實驗結果發現，本文所提出的控制器能產生快速有效的追蹤結果。

關鍵詞：太陽電池，最大功率追蹤，擾動和觀察法，三點加權比較法

文稿收件日期 106.3.22;文稿修正後接受日期 106.11.16; *通訊作者

Manuscript received March 22, 2017; revised November 16, 2017; * Corresponding author

I. INTRODUCTION

Solar cell is a kind of optoelectronics converting energy. It converts optics into electricity after being illuminated by sunlight [1-4]. The solar cell arrays are normally designed to extract maximum electrical power from varying irradiation levels. However, the photovoltaic (PV) power generation have a nonlinear current voltage (I -V) characteristic as their power output depends on the weather conditions [5, 6]. In order to achieve the maximum power from the PV arrays, the control unit needs to have an appropriate strategy for maximum power point tracking (MPPT). Several techniques for MPPT have been used in recent years in which the Perturb & Observe (P&O) algorithm is very popular used in PV industry [7, 8]. A typical disadvantage of this technique is that the slow response speed, oscillations in power output [9, 10], tracking failures in rapid environmental changes, and required parameters less. Several enhanced the P&O algorithm have been proposed [11-13], however they have a higher computational cost, slow down the response speed and increased oscillations which results in weather change conditions. In the tradition P&O algorithm measured power of two points is used to determine movement to the next operating point. Based on the experiments, there are difference in the output power the controller increases or decrease the PV module array output voltage.

This paper proposes implementation of a novel controlling scheme to combine P&O method together with Three-Point Weighting Comparison (TPWC) method [14-17] for tracking maximum power points of PV systems. The basic concept of this approach is that the variation of solar irradiation conditions changes suddenly, the P&O method will be applied for fast tracking of the maximum power point. In the other cases with the solar irradiation is comparatively stable, the TPWC is applied to avoid the perturbation problem around the maximum power point. Therefore, considerable optimal solution is being carried out to implement a microprocessor by combine P&O and TPWC methods for PV systems. The advantage of this approach is that it is a simple algorithm conducting on a microprocessor

without redundant circuit components and also the maximum power point of the PV systems will be obtained quickly and stably. As a result, the PV arrays control is a real time detecting output power and adapting the control algorithm to make the system operate under the optimum work.

II. MATHEMATICAL MODEL

The basic structure of semiconductor solar cell is a p-n junction diode. When the p-type and n-type semiconductor contact to form the p-n junction, the minority carrier of n-type semiconductor will diffuse to the p-type semiconductor to fill the holes. On the p-n junction, electrons and holes combine and form a depletion region, which results in a built-in electric field. An ideal solar cell, without light exposure, can be regarded as the diode, the current-voltage (I-V) characteristics is [1,18-19]:

$$I = I_s (e^{\frac{V}{V_T}} - 1) \quad (1)$$

Where I is current, V is voltage, Is is the saturation current,

$$V_T = \frac{kT}{q} \quad (2)$$

Where kB is the Boltzmann constant (1.38×10^{-23} J/K). q0 is denotes the charges of an electron (1.6×10^{-19} coulomb), temperature of the PV arrays (k). At room temperature, VT is about 0.026V. When the sunlight shines on the solar cells, the built-in electric field direction of the p-n junction is from the n-type to p-type, electrons run to n-type side, and holes to the p-type side, while the photo current flows from n-type to p-type [1,18-19].

$$I = I_s \left[e^{\frac{(V - IR_s)}{V_T}} - 1 \right] + \frac{(V - IR_s)}{R_{sh}} - I_L \quad (3)$$

Where Rs and Rsh represent the series and shunt resistance, IL is negative photo current of the solar cell. Figure 1 shows the solar cell voltage and current (or power) characteristic curve. When load is in short-circuit, the voltage difference is 0 and current is maximum that is called short-circuit current (Isc). When the load

is open circuit, the current is 0 and voltage difference is in maximum that is called the open circuit voltage (Voc). In load changing process from 0 to infinity, there is a particular load which is the highest state of power generation. The power generation at this time is the module's maximum power. The voltage at the maximum power point is called the maximum power voltage (Vmax), and the current is called the maximum power current (Imax). The rectangular area formed by Vmax and Imax at the Maximum power point is the maximum power value. Dotted lines in Figure 1 are the ideal solar cell I-V characteristic, the ideal solar cell hasn't internal resistance and leakage current.

Figure 1 shows a MATLAB/Simulink model of the PV circuit built to obtain the I-V characteristics according to equation (3). This I-V characteristics of solar cell shows that the output curves are non-linear. In addition to that output power influenced by weather changing condition, temperature and load status. Each power voltage (P - V) characteristic curve has a maximum power point, at which the solar cell operates most efficiently.

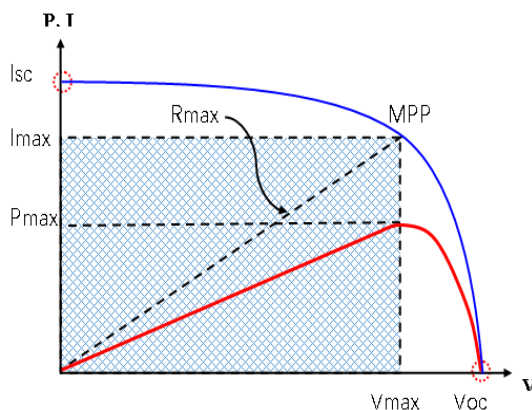


Fig.1. The I-V characteristic relationship of solar cell

III. MPPT System Configuration

The PV system produced output power which is the maximum power that depends on the path of incoming light. In order to achieve the energy generated from PV system during significantly on weather changing condition, the system needs to have a suitable algorithm for

MPPT.

3.1 Perturb and Observe (P&O) Method

The diagram of P&O operation is shown in Figure 2. The operating voltage is sampled and the algorithm changes the operating voltage in the required direction and samples $\Delta P/\Delta V$. The P-V curve can be divided into two characteristic areas. If the operating point lies in left side ($\Delta P/\Delta V > 0$), then the increase of voltage will raise up the perturbed power. However, if operating point lies in the right side ($\Delta P/\Delta V < 0$), then the increase of voltage will decrease the PV system output power. The principle of P&O method will compare one perturbation point to the next, and determine which one is the maximum power point.

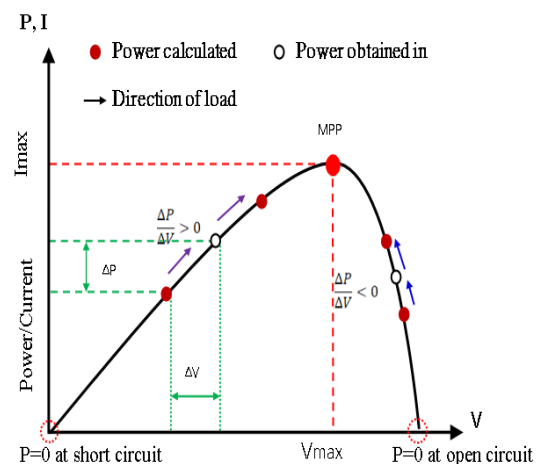


Fig.2. I-V characteristic of PV arrays by operating P&O method

3.2 Three-Point Weighting Comparison (TPWC) Method

TPWC method works similarly to compare the variation of the output power, but with one additional power point to make final determination of maximum power point. The possible states of three perturbation points may be depicted in nine cases which are shown in Figure 3 (with proper weightage '+' or '-' sign). In this method, the voltage and power of the PV systems are adjusted according to the cycle of increasing or decreasing load. At the same time, it will determine the next increasing or

decreasing step of working load by comparing the changed load value of output voltage and power. The three points are the current operation point A, a point, B, perturbed from point A, and a point, C, with doubly perturbed in the opposite direction from point B. The comparison method is weighted by two parts. In these cases one, if point $C > B$, or $C = B$, a positive weight is given, and if $C < B$, then a negative weight is given. Second, if $A > B$ or $A = B$, a negative weight is given, and if $A < B$, then a positive weight is given.

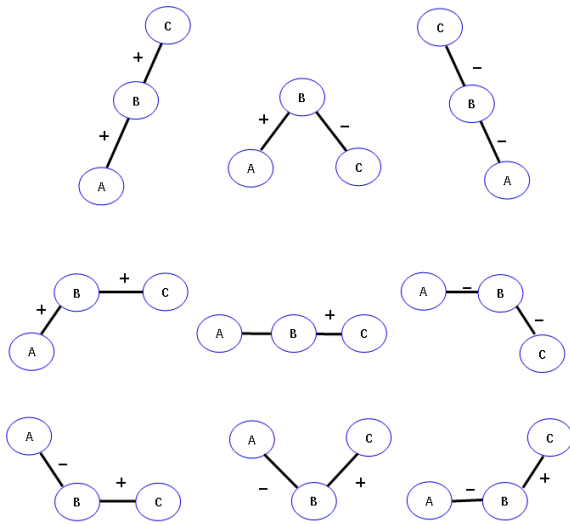


Fig.3. Possible patterns in the direction of step-size in TPWC method

3.3 The Proposed MPPT Method

The P&O method will cause the loss of power and the reduction of the efficiency of the PV arrays. The outcomes of MPPT using classical P&O algorithms are caused by oscillations in power output. Unfortunately, a number of studies to prevent the oscillation problem require complex computational power, resulting in unnecessary power loss of the PV system. In the P&O algorithm, only two points, which are the current operation and the subsequent perturbation point, are observed to see their changes in power. Therefore, this work incorporates the TPWC method to improve the drawback of the P&O method. The P-V characteristics for the proposed MPPT method under different conditions are shown in Figure 4. When the slope value of power is larger than

absolute value $1/3$, which means the variation of solar irradiation is rapid. The current and voltage of PV arrays change during a voltage perturbation. If $\Delta P/\Delta V > 1/3$, the operating voltage of PV arrays is located on the left side in the P-V curve. If $\Delta P/\Delta V < 1/3$, the operating voltage of PV arrays will be located on the right side in the P-V curve. On the other hand, when the slope value of power is equal $1/3$, which means the TPWC method is used to avoid having to move rapidly the operation point or when a disturbance or data reading error occurs.

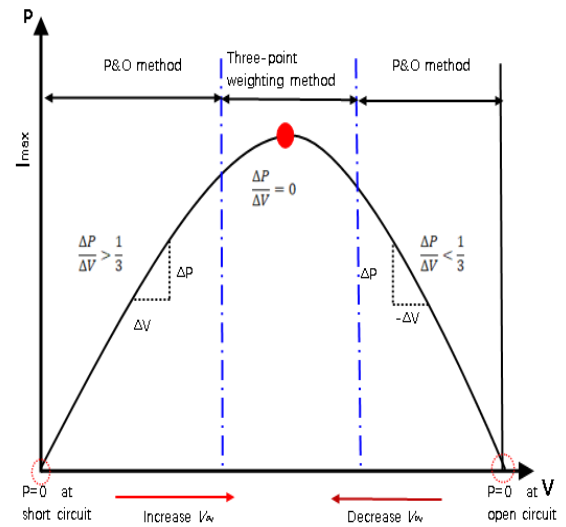


Fig.4. Proposed method sign of $\Delta P/\Delta V$ at different position on the power-voltage characteristic of the PV system

There is unnecessary power loss if only the P&O method is used for tracking the maximum power point. If only the TPWC method is used, the seeking time of the maximum power point will be longer than that of the P&O method when solar irradiation changes rapidly. There are some disadvantages in each method. Therefore, a novel controlling scheme with an adaptive decision mechanism is developed for a PV system to avoid these disadvantages. Figure 5 presents the control flow chart of the proposed control program. The proposed method uses the condition that $\Delta P/\Delta V$ is greater than $1/3$ or not to make the judgment. When the condition $|\Delta P/\Delta V| > 1/3$ is used, the TPWC method is used for tracking the maximum power point. As described in the last paragraph, the P&O method uses the condition $\Delta P/\Delta V > 1/3$ to determine whether the maximum power point has been found or not. The proposed method uses the duty cycle of these switching mode

power interface devices as the judging parameter depend on power output.

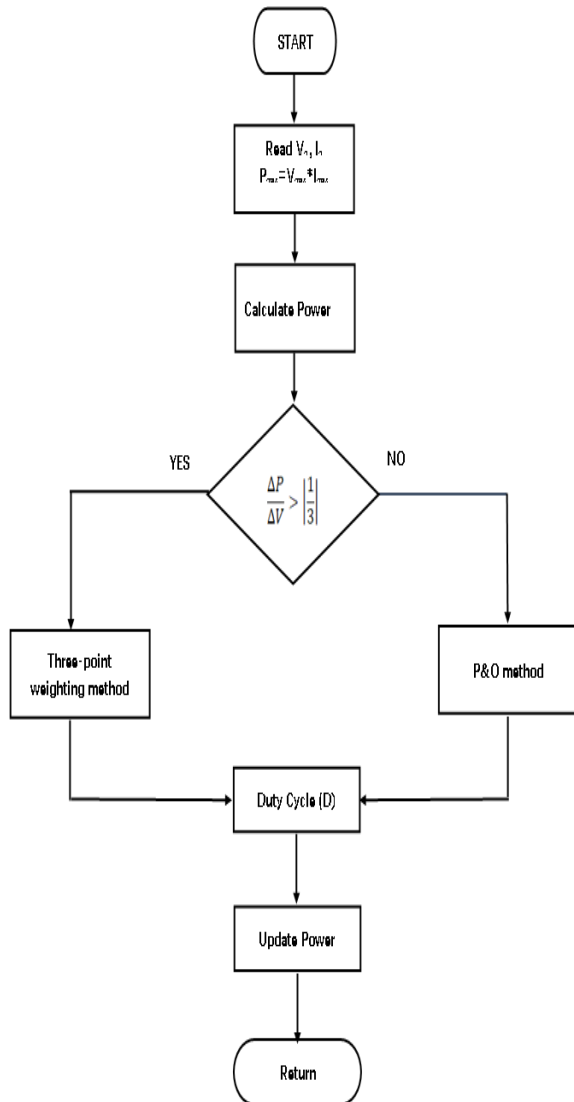


Fig.5. Flowchart of the proposed method

IV. CIRCUIT TOPOLOGY AND OPERATION OF THE BUCK-BOOST DC-DC CONVERTER IN THE PV SYSTEM

From the application of alternating current (AC), voltage control devices are needed to step-up and step-down output voltage. The DC-DC converter can be considered as the DC equivalent to an AC transformer with continuously variable turns-ratio. DC-DC converter is very important in all voltage levels

applications lower power, medium power or high power applications. Many DC-DC converter circuits are considered before design the PV system. Based on this approach, ultimately a buck-boost DC-DC converter was chosen because of this converter an inverting regulator. The buck-boost converter provides an output voltage which may be greater than or less than the input voltage provides an output voltage which may be greater than or less than the input voltage. This converter is a circuit that combines a buck converter topology with a boost converter topology in concatenation. The polarity of input voltage is opposite of the polarity of the output voltage.

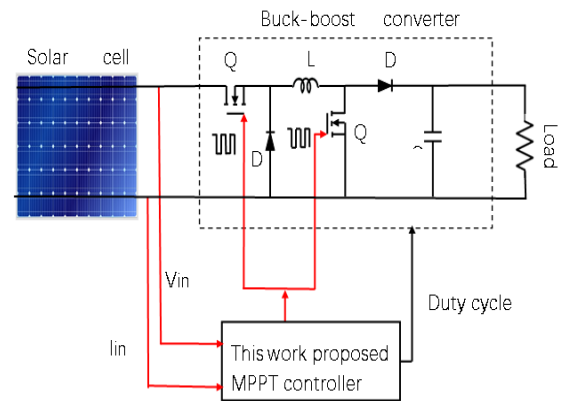


Fig.6. Block diagram of the proposed PV system

Figure 6 shows the topology of the proposed PV system. As shown, the buck-boost DC-DC converter connects the PV modules to the load. The buck-boost converter consists of an inductor (L), two diodes (D1 and D2), capacitor (C) two digital switches (Q1 and Q2) and resistance respectively. The capacitor maintains the load voltage, D2 prevents reverse current, and D1 separates the two modes of operation for this buck-boost topology.

The first mode appears when the transistor is turned on, and diode is reversed biased. During open-switch mode, the input current flows through the transistor and inductor. The second mode appears when the switch is turned off, and the current flows from L through C. The transistor is switched on again in the next cycle when the energy that stored in inductor, the capacitor is added to the converter output to reduce the ripple of its output voltage. Buck converter is a step-down converter that output voltages lower or high than input voltage. The

solar cells load (RL) is connected across PV arrays terminal via buck-boost converter. The buck-boost converter is intended to match the load impedance with source impedance of the PV system to adopted maximum power transfer. The relations between input and output variables of buck-boost converter is given as [20-22]:

$$\frac{V_o}{V_t} = \frac{D}{1-D} = \frac{I_i}{I_o} \quad (4)$$

Where V_o is output voltage, V_i is input voltage of PV system. I_i and I_o are input and output currents, respectively. The power flow is controlled by switch moving during the on/off duty cycle (D). Where duty cycle is the ratio of the on time of the switch to the total switching time. The input/output relation of the voltages can be expressed by equations (5):

$$V_o = \frac{1}{1-D} V_i \quad (5)$$

The buck-boost converter output voltage is controlled by the switch duty-cycle. The duty cycle in terms of the output and the input voltage is given by equation (6). By solving for switch duty cycle:

$$D = \frac{V_o}{V_o + V_i} \quad (6)$$

The algorithm will take this variable as controlled variable for voltage change and then computes the duty cycle from Eq. (5) and (6) as follow:

$$D = \frac{T_o}{T} \quad (7)$$

Where T is switching cycle of power transistor, T_o is turn on period of power transistor. Therefore, we can utilize the buck-boost converter in PV system. During weather conditions change, the PV system will change the duty cycle of power transistor to obtain the maximum power output. In order to maximize the power output of the PV system, a controlling scheme is implemented on a microprocessor to combine with P&O and TPWC methods. There PV system power load output will be regulated to the different weather conditions by controlling the duty cycle of the buck-boost converter.

When controller determines to decrease load, the power transistor's switching duty cycle will gradually decrease the photovoltaic array reaches the maximum power point.

As shown in Figure 6, the proposed MPPT circuit determines an input voltage V_{in} , MPPT corresponding to an atmospheric condition. The PV output voltage is sensed and compared with the V_{in} . Through the proposed controller an appropriate control signal is generated to regulate the PV output voltage so that the buck-boost DC-DC converter draws maximum power from PV arrays. Then, the buck-boost DC-DC converter implant the power into DC bus for DC-distribution applications or into utility via a grid-connection DC-AC inverter. The MPPT controller varies the point of intersection between the load line and I-V curve by varying the duty cycle to achieve an intersection point where maximum power transfer to the load is achieved.

V. SIMULATED AND EXPERIMENTAL RESULTS

5.1 Simulated Results

For implementation purpose a 21V and 4.5A (close circuit current) solar panel is used. It produced 70W at 25°C and 1kW/m² irradiance. To evaluate the performance of the proposed system, Matlab/Simulink is used to implement the tasks of modeling and simulation. The two most important P-V and I-V curves are carried out simulating to validate the performance of the proposed method. These curves show the relation between temperature, irradiation and power generation.

The developed model of PV system is simulated for different temperatures as shown in Figure 7. In the left-hand side of I-V characteristic curves, it shown the output current is nearly constant as the voltage changed from zero to the voltage at the corner point. In the right hand side of I-V characteristic curves, it shown the output voltage is nearly constant as the current falls to zero. Therefore, it can be seen that the I-V characteristics are dependent on the levels of irradiance and the temperature of PV

arrays.

The P-V curves at various temperatures are plotted, as shown in Figure 7. It is clear that the P-V curve has single peak that could be easily found by P&O method. From the above simulations, it is obvious that the factors, array temperature, will affect the generated PV power significantly. To improve PV system efficiency, the proposed MPPT algorithm has to be adopted to draw maximum power from PV arrays.

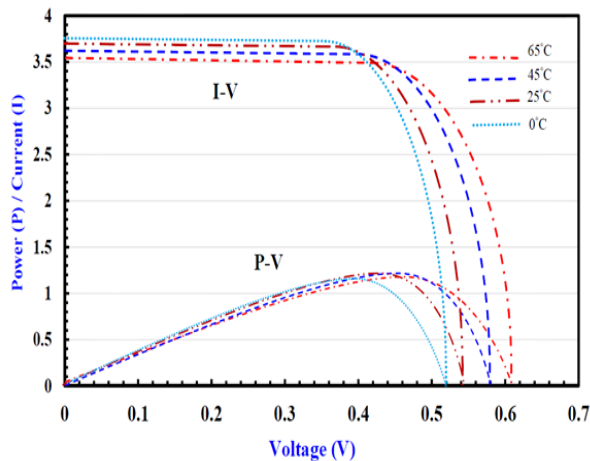


Fig.7. I-V and P-V characteristic curves of a practical photovoltaic device under different temperature (at 1000W/m²)

5.2 Experimental Results

The reliability of power tracking analysis depends primarily on accurate power output data with sufficient time resolution. The data used in this paper are being collected by the DSP laboratory of department of electrical (National Taiwan Ocean University) since 2015 July with 30 days. The available output power of the PV arrays are averaged over several hours, typically AM 09 to PM 05. The sampling time of 30 days assures that all irradiance fluctuations, relevant for different power tracking method, are recorded. There are involve the valid information and useful for tracking evaluation. The cumulative power of PV panel using different methods is shown in Table 1. According to the experiment, it is shown that the proposed method can achieved the best performance by enhancing about 21% efficiency.

The experiment reveals the real variability of power profiles by different methods, as shown in Figure 8. The collection of PV arrays output

power irradiance profiles, shown in this figure, demonstrates the condition of variable-cloudy sunshine, especially in summer. Too small power output is causing by without MPPT, while we proposed method is almost perfect. Power drift is obvious, the irradiance change is the main factor. The deepest fall of power output always occurs for the current and voltage drift at the same time. The quality of PV arrays output power is determined by the terrestrial irradiance. The results represent the energy received by PV arrays, different method tracking efficiency. It is worth noticing, that under highly variable irradiance proposed method can immediate reach steady state.

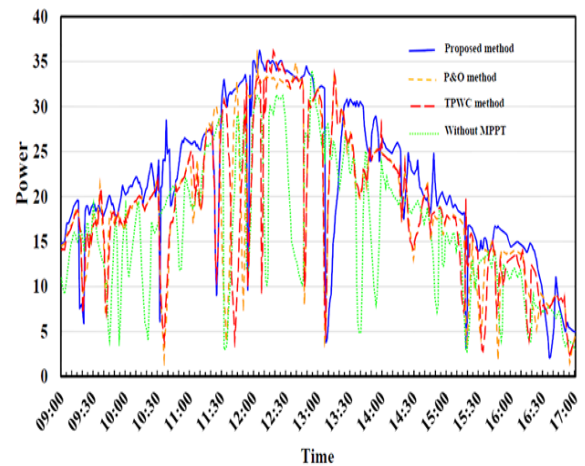


Fig.8. Practical measurement result by different tracking methods. (Remarks: under the condition of irradiance = 850 W/m², temperature = 55°C)

Table.1. Cumulated 30 day's power obtained from different

Method	Cumulated power	Efficiency (%)
Without MPPT	7.21kW	NA
TPWC method	7.9kWf	9.6%
P&O method	7.94kW	10%
Proposed method	8.74kW	21%

VI. CONCLUSION

Maximum power point tracking techniques is always the crucial factor to extract the PV system output power. This paper presents an efficient controlling scheme and developed by

compromising both P&O method and TPWC method for photovoltaic arrays MPPT. This system is implemented in a single chip with a simple algorithm to control these two methods and it will include the advantages of both methods. The experimental results have demonstrated clearly that the proposed method is better than the traditional method, it will track the maximum power point for a PV generation system quickly and stably. The best results are expected for the proposed method can enhance about 21% efficiency. Therefore, we proposed method has the potential in the PV industry application

REFERENCES

- [1] Sze S. M., 2002. Semiconductor Devices-Physics and Technology. 2nd Ed., Chap. 3, John Wiley & Sons, Inc.
- [2] Cheng Y. T., Ho J. J., Lee W., Tsai S. Y., Chen L. Y., Liou J. J., Chang S. H., Shen H., Wang K. L., 2010. Efficiency Improved by H₂ Forming Gas Treatment for Si-Based Solar Cell Applications. International Journal of Photoenergy. 634162: 1-6.
- [3] Outón L. M. P., Lee J. M., Futscher M. H., Kirch A., Tabachnyk M., Friend R. H., Ehrler B., 2017. A Silicon-Singlet Fission Tandem Solar Cell Exceeding 100% External Quantum Efficiency with High Spectral Stability. ACS Energy Lett. 2(2): 476-480.
- [4] Salman K. A., 2017. Effect of surface texturing processes on the performance of crystalline silicon solar cell. Solar Energy. 147: 228-231.
- [5] Biktash L., 2017. Long-term global temperature variations under total solar irradiance, cosmic rays, and volcanic activity. Journal of Advanced Research. 8(4): 329-332.
- [6] Kim B., 2017. Diffuse and direct light solar spectra modeling in PV module performance rating. Solar Energy, 150: 310-316.
- [7] García M., Maruri J. M., Marroyo L., Lorenzo E., M. Pérez, 2008. Partial Shadowing, MPPT Performance and Inverter Configurations: Observations at Tracking PV Plants. Prog. Photovolt: Res. Appl. 16: 529-536.
- [8] Alik R., Jusoh A., 2017. Modified Perturb and Observe (P&O) with checking algorithm under various solar irradiation. Solar Energy. 148: 128-139.
- [9] Shanmugasundaram L. R., Sarbham K., 2015. Load Controlled Adaptive P&O MPPT Controller PV Energy Systems. International Journal of Innovative Research in Science, Engineering and Technology. 4(5): 529-536.
- [10] Mohapatra A., Nayaka B., P. Das, Mohanty K. B., 2017. A review on MPPT techniques of PV system under partial shading condition. Renewable and Sustainable Energy Reviews. 80: 854-867.
- [11] Hohm D. P., Ropp M. E., 2003. Comparative Study of Maximum Power Point Tracking Algorithms. Prog. Photovolt: Res. 11: 47-62.
- [12] Ahmed J., Salam Z., 2015. An improved perturb and observe (P&O) maximum power point tracking (MPPT) algorithm for higher efficiency. Applied Energy. 150: 97-108.
- [13] Karamia N., Moubayed N., Outbib R. 2017. General review and classification of different MPPT Techniques. Renewable and Sustainable Energy Reviews. 68: 1-18.
- [14] Esram T., Chapman P. L., 2007. Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques. IEEE Trans. Energy. Conversion. 22(2): 439-449.
- [15] Karlis A. D., Kottas T. L., Boutalis Y. S., 2007. A novel maximum power point tracking method for PV systems using fuzzy cognitive networks (FCN). Electric Power Systems Research. 77: 315-327.
- [16] Onat N., 2010. Recent Developments in Maximum Power Point Tracking Technologies for Photovoltaic Systems. International Journal of Photoenergy. 2010: 1-11.
- [17] Heydari-doostabad H., Keypour R., Khalghani M. R., Khooban M. H., 2013. A new approach in MPPT for photovoltaic array based on Extremum Seeking Control under uniform and non-uniform irradiances. Solar Energy. 94: 28-36.
- [18] Ali M. F., Hossain M. F. 2017. Influence of Front and Back Contacts on Photovoltaic

Performances of p-n Homojunction Si Solar Cell: Considering an Electron-Blocking Layer. *International Journal of Photoenergy*. 7415851: 1-6.

- [19] Pellet N., Giordano F., Dar M. I., Gregori G., Zakeeruddin S. M., Maier J., Grätzel M. 2017. Hill climbing hysteresis of perovskite-based solar cells: a maximum power point tracking investigation. *Progress in Photovoltaics*, 2: 1-10.
- [20] Femia N., Petrone G., Spagnuolo G., Vitelli M., 2005. Optimization of perturb and observe maximum power point tracking method. *IEEE Trans. Power Electron.* 20(4): 963-973.
- [21] Abouda S., Nollet F., Essounbouli N., Chaari A., Koubaa Y., 2013. Design, Simulation and Voltage Control of Standalone Photovoltaic System Based MPPT: Application to a Pumping System. *Int. J. Renew. Energy Res.* 3: 538–549.
- [22] Killi M., Samanta S., 2015. Modified Perturb and Observe MPPT Algorithm for Drift Avoidance in Photovoltaic Systems. *IEEE Trans. Industrial Electron.* 62(99): 1-10.

Security of a UUP Web Search Protocol with Privacy Preserving

Fuh-Gwo Jeng¹, Bing-Jian Wang², Tzung-Her Chen^{*2}

¹Department of Applied Mathematics, National Chiayi University

²Department of Computer Science and Information Engineering, National Chiayi University

Abstract

Generally, a search engine will keep a record of a user about the websites he ever went and the past searches he had submitted to improve its performance. Similar to a spy tracking and tracing the footpath, a search engine will inevitably violate user's privacy as the record will reveal the user's personal information or the institution he works for. To protect user's privacy, Castellà-Roca et al. proposed a protocol called Useless User Profile (UUP), in which it provided a distorted user profile for a web search engine such that the web search engine cannot generate a real profile of a certain individual. One of the significant advantages lies on that their protocol requires no change in the server side and the server is not required to collaborate with the user. However, to claim security guarantee of new image cryptosystems is meaningful only when the cryptanalysis is taken into consideration. The UUP protocol was claimed to be secure; however, a potential collusion attack is pointed out. In order to benefit the advantages and contribution of Castellà-Roca et al.'s scheme, this paper redesigns a security-improved version by simple modification to remove the possible security concern. Precisely, to correct the shortcoming, the authors suggest the user's query be encrypted firstly by means of the server's public key and then each answer also be encrypted by a session key.

Keywords: privacy preserving, web search engine, private information retrieval

具隱私保護的 UUP 網頁搜尋引擎協定之安全探討

鄭富國¹

王炳兼²

陳宗和^{*2}

¹嘉義大學應用數學系

²嘉義大學資訊工程學系

摘要

一般而言，網路搜尋引擎會紀錄使用的使用動態名為改進效能，然而這樣也違反使用者的個人隱私。因此，Castellà-Roca 等提出 UUP 協定。該協定下，使用者不再需要提供完整的使用資訊，伺服器端得到的並不是某位使用者完整的使用資訊。值得注意的，伺服器端不用修改且不用與使用者配合修改。然而，基於宣稱安全的一個新的密碼技術必須經過安全分析後才有意義，本研究發現該協定有共謀攻擊的可能，當群組中的欺騙者人數高達 $n-1$ 時，最後一位使用將可能受到共謀欺騙。為了維持 Castellà-Roca 等提出 UUP 協定的優點，本文對其進行小而簡單的修改以補強其安全性。主要改進的設計在於使用者先以伺服器的公開金鑰對查詢內容加密，而伺服器回應的內容則以一把會議金鑰加密後傳回。

關鍵詞：隱私保護，網頁搜尋引擎，個人資料取得

文稿收件日期 106.9.20; 文稿修正後接受日期 107.1.30; *通訊作者

Manuscript received September 20, 2017; revised January 30, 2018; * Corresponding author

1. Introduction

In order to protect the confidentiality of sensitive data in outsourcing cloud-computing environments, a well-defined encryption technique is used to encrypt the private and sensitive data stored in the cloud. With issuing a keyword searching on the encrypted data, it unavoidably faces the security problem of how to process the key search without revealing any sensitive information. Especially, the server maintaining the database of encrypted data is not always trusty.

Web search engines can help users to receive a great amount of data they want. However, their private search profiles are possible to be disclosed after submitting queries to a web search engine. This problem of protecting user's privacy can be viewed as a Private Information Retrieval (PIR) problem [8,9,10,11,12]. Generally, a user in a PIR protocol can retrieve a certain amount of data from the database of a server while the server has no idea about which data requested by the user.

In 2004, a public-key encryption with keyword search (PEKS) is first proposed by Boneh et al. [19]. Inspired by Boneh et al.'s scheme, Hwang and Lee [20] proposed another PEKS scheme for multi-receiver in which the concept of proxy re-encryption was later applied in keyword search by Shao et al. [21] and by Yau and Phan [22] as well. Furthermore, Baek et al. [24], on the one hand, demonstrated that outside attackers could perform the test process by collecting the transmitted ciphertexts and trapdoors in the PEKS scheme. Consequentially, attackers are potentially able to further construct the relationship between encrypted data and the given trapdoors of known keywords.

Therefore, Baek et al., on the other hand, proposed their public-key encryption scheme with designated tester (dPEKS) to remove the security problem. Byun et al. [25] presented that Boneh et al.'s design of trapdoors in PEKS suffers off-line keyword-guessing attacks. In such a way, attackers can choose the keywords to test whether the captured trapdoor includes the guessed keyword with the receiver's public key and bilinear map operation, the interested keyword of the receiver is revealed. Unhappily, although Baek et al.'s dPEKS scheme achieves

tester designating, the trapdoor's structure is identical to that in PEKS's. In such way, Baek et al.'s dPEKS scheme cannot prevent off-line keyword-guessing attacks. In 2010, Rhee et al. [26] enhanced the trapdoor security so as to prevent from off-line keyword-guessing attacks existing Baek et al.'s dPEKS scheme [24]. Yet, Wang et al. pointed out the trapdoor design was still on the risk of keyword-guessing attacks especially by malicious servers [27]. After that, there are further searchable encryption schemes taking realistic applications into account, for example, a conjunctive subset keywords search proposed by Zhang et al. [23].

Chor et al. [1, 2] firstly introduced the private information retrieval problem and proposed a protocol. In their protocol, several servers share the same database and these servers are not allowed to communicate to each other. But as mentioned above, Castellà-Roca et al. required one server, the web search engine, and one database in their case. Thus, they looked forward to the single-database PIR protocol proposed firstly by Kushilevitz and Ostrovsky [3]. The single-database PIR schemes are more suitable to apply on web search engines. "However, they suffer from some fundamental problems that make their use unfeasible in communications between a user and a web search engine," Castellà-Roca et al. summarized [5] as follows.

- (1) The single-database PIR schemes are not suited to deal with large databases. With PIR in mind, the single database is usually modeled as a vector. Upon retrieving the value of the i th component of the vector, users wish to keep the index i hidden from the server holding the database. Supposing that the database contains n items, a PIR scheme aims to guarantee maximum server-uncertainty on the i th record retrieved by a user. It seems to be done by accessing to all records in the database. If some user only accesses to a part of them, the server easily lean to know the real interest of this user. And the cost of accessing all records implies a computational complexity of $O(n)$.
- (2) Upon accessing a record in the database, it is reasonably assumed that the user knows its physical location. This assumption is not always realistic because the database is managed by the server. Instead, the user can submit a query consisting on keywords.
- (3) Thirdly, it is assumed that the server, holding

the single database, collaborates with users in the PIR protocol. However, the assumption is not realistic since a server has no motivation to protect the privacy of users. In fact, users should take care of their own privacy by themselves instead of expecting any collaboration from the web search engine.

Consequently, Castellà-Roca et al. proposed the UUP protocol to protect the users' privacy by providing a distorted user profile for a web search engine so that the web search engine cannot generate a real profile of a certain individual. Briefly speaking, there is a central node in their scheme grouping n users who submit a query each and shuffling all queries and finally distributing the queries fairly. When a user receives the assigned query, he submits it to the web search engine and waits for the real answer for his own real query. The answers of the n queries from the search engine are broadcast to all the group users; therefore, a user figures out his answer and ignores the others.

Their scheme improves the performance of existing proposals in terms of the computational cost and communication overhead. To avoid a web search engine profiling a real search record of a certain individual, Castellà-Roca et al. applies the technologies of encryption, re-masking and permutation to achieve their goal and make sure their scheme secure. The UUP protocol is proven able to prevent any attack from a dishonest user, a dishonest central node and a dishonest search engine, i.e. three entities in their protocol, under the assumptions that all the group users follow their protocol and no collusions happen between two of the three entities in their scheme.

In addition to the above-mentioned advantages, the main contribution of Castellà-Roca et al.'s UUP protocol is that the UUP protocol does not require any change in the server side and the server is not required to collaborate with the user.

As security is always the concern for new cryptosystems such as the abovementioned PEKS, dPEKS, etc., all proposed cryptosystems must undergo the scrutiny of the scientific community [13-17]. Unhappily, taking the higher security into account, the UUP scheme cannot avoid the insider collusion attacks, in which the group users plan together to cheat the n th user's privacy profile. It is obvious that the $n-1$ group users can collaborate to analyze the

queries they had submitted and the answers they had gotten, and further they can infer the n th user's authentic search profile.

In order to benefit the advantages and contribution of Castellà-Roca et al.'s scheme, it is worthwhile to re-design the improved version. As C. A. R. Hoare said, "*There are two ways of constructing a software design: One way is to make it so simple that there are obviously no deficiencies, and the other way is to make it so complicated that there are no obvious deficiencies. The first method is far more difficult.*" [18]

In this paper, the authors firstly show that the potential security weakness, i.e. collusion attacks, exists in Castellà-Roca et al.'s scheme. Precisely, their scheme is secure only if the number of dishonest users is less than $n-1$. Unfortunately, if there are $n-1$ dishonest users in the same group, the n -th user encounters the risk of cheating by the others. This security weakness comes from that the $n-1$ group users can collaborate to analyze the queries they had submitted. Upon they had gotten the responses, and further they can infer the n -th user's authentic search profile.

Secondly, a "small and simple" modification to Castellà-Roca et al.'s scheme is proposed. To correct the shortcoming, the authors suggest the user's query be encrypted firstly by means of the server's public key and then each answer also be encrypted by a session key. Inheriting the contribution from Castellà-Roca et al.'s UUP scheme, the main contribution of this paper is to further enhance the security to avoid the collusion attacks.

The rest of the paper is organized as follows. A review of the UUP protocol is given in Section 2. A security improvement and the security analysis are given in Section 3. The conclusions are given in Section 4.

2. Review of useless user profile (UUP) protocol

The main idea of Castellà-Roca et al.'s scheme relies on that each user who intends to submit a query will not send her/his own but a query of another user instead. Simultaneously, her/his query is submitted by just another group user. Considering privacy concerns, the key design relies on that users do not know which query issued by each user based on the assumption that each user submits very different

kinds of queries. There is no clue about that those queries are liable to a certain person.

With the help of n -out-of- n threshold ElGamal encryption [6] and ElGamal re-masking operation, Castellà-Roca et al. proposed the UUP protocol to protect user's privacy by providing a distorted user profile for a web search engine so that the web search engine cannot generate a real profile of a certain individual. On the basis of privacy requirements, their scheme achieves the objective because the link between the wanted query a user submitted originally and the true answer the user had is distorted.

The scenario in the UUP protocol contains the three entities:

- Users (U): Users in the group are the individuals who intend to submit queries to the web search engine but still keep protecting their own privacy in mind.
- The central node (C): The central node takes the responsibility to keep in touch all group users intending to submit their query. That is, it groups users in order to execute the UUP protocol.
- The web search engine (W): This web search engine holds the database but is not always trustworthy. It does not guarantee to preserve users' privacy.

Upon considering the privacy requirements of users, the UUP protocol should satisfies the following properties:

- U_i must not link a certain query with U_j who has generated it.
- C must not link a certain query with U_i who has generated it.
- W must be unable to construct a reliable profile of a certain user U_i .

There are four sub-protocols in their scheme. They are group setup, group key generation, anonymous query retrieval, and query submission and retrieval. The purpose of each sub-protocol and how it works are described in the following.

2.1 Group setup

Assume user wants to submit a query to the web search engine. Firstly, he has to send a message to the central node C for asking to be a group member. The central node receives all the requests from users. As soon as it collects n requests, it sets up a new user group $\{U_1, \dots, U_n\}$

and notifies the n users which group they belong to. A communication channel among them is built up at the same time such that they can talk to each other without the interference of the central node.

2.2 Group key generation

First, all the users $\{U_1, \dots, U_n\}$ in the same group agree on a large prime p where $p = 2q + 1$ and q is a prime too. Then they choose a generator element $g \in Z_q^*$ of the multiplicative group.

Next, user U_i randomly generates his private key $\alpha_i \in Z_q^*$ and publishes $y_i = g^{\alpha_i} \bmod q$. Note that each user should keep his private key secret. Finally, all the users $\{U_1, \dots, U_n\}$ execute altogether the n -out-of- n threshold ElGamal encryption to generate their group public key y , where $y = \prod_{i=1}^n y_i = g^\alpha \bmod q$, and $\alpha = \alpha_1 + \dots + \alpha_n$.

2.3. Anonymous query retrieval

Firstly, user U_i (for $i = 1, \dots, n$) generates a random value $r_i \in Z_q^*$ and encrypts his query m_i with the group key by means of the standard ElGamal encryption function [7], i.e. $E_y(m_i, r_i) = (g^{r_i}, m_i \cdot y^{r_i}) \bmod q = (c1_i, c2_i) = c_i^0$. Next, user U_i (for $i = 1, \dots, n$) sends his cryptogram c_i^0 to the others in his group. In the end of the sending process, each of the group holds the ordered cryptograms $\{c_1^0, \dots, c_n^0\}$.

Then, user U_1 re-masks the cryptograms $\{c_1^0, \dots, c_n^0\}$, which he already holds, to get a re-encrypted version. Then, user U_1 randomly permutes the re-encrypted version to obtain a re-ordered version of cryptograms. Finally, he sends the re-ordered version of cryptograms to user U_2 . Note that it is assumed that the group members are set in order from the first to the n th. Following this way, each of the other users U_i (for $i = 1, \dots, n$) will wait for the re-ordered version of the cryptograms from his immediate predecessor and then goes on the processes of re-masking the cryptograms and randomly permuting the re-encrypted version so as to get a re-ordered version of the cryptograms and finally sending them to the next group member. In the end, User U_n has to broadcast the last

result of the cryptograms $\{e_{\sigma(1)}^n, \dots, e_{\sigma(n)}^n\}$ to all of the group members.

Let $\{e_{\sigma(1)}^n, \dots, e_{\sigma(n)}^n\}$ denote as $\{c_1, \dots, c_n\}$. To decrypt the value c_i , user U_i has to require all the other group members to take part by sending their corresponding shares called $(c1_i)^{\alpha_j}$ from user U_j , where $j = 1, \dots, n$ and $j \neq i$. Finally, user U_i can retrieve the query m^i by computing:

$$m^i = \frac{c2_i}{c1_i^{\alpha_i} (\prod_{j \neq i} c1_i^{\alpha_j})} \bmod q.$$

Note that the value c_i is correspondent to the query m^i , but the query m^i could be generated by one of the other group members.

2.4 Query submission and retrieval

Once user U_i retrieves the query m^i , he submits it to the web search engine W . As soon as he gets the response a^i from the web search engine, he broadcasts it to the other group members. Finally, each user figures out the exact answer from those responses to match his original query.

2.5 Security analysis

Castellà-Roca et al. proposed the UUP protocol by applying the technologies of encryption, re-masking and permutation to preserve the users' privacy when they submit queries to a web search engine. As they defined, a successful attacker is able to know the certain query submitted by a certain user. Their scheme is proven able to prevent any attack from a dishonest user, a dishonest central node and a dishonest search engine, i.e. three entities in their protocol, under the assumptions that all the group users follow their protocol and no collusions happen between two of the three entities in their scheme. And the attackers from external entities cannot get more information than those from the internal entities. Hence, Castellà-Roca et al. perform the security analysis for the internal entities as follows.

2.5.1 Dishonest user

User U_a is supposed to be dishonest. In the end of the cryptogram-sending process, he gets the original ordered cryptograms, which contains all the queries from the group members. To decrypt the cryptograms, user U_a has to

require all the other group members to take part by sending their corresponding shares called $(c1_i)^{\alpha_j}$. Provided that all the other group members contribute their secret keys $(\alpha_1, \dots, \alpha_n)$, he is not able to decrypt the cryptograms $\{c_1^0, \dots, c_n^0\}$. Therefore, their scheme is secure if there is one dishonest user in a group.

2.5.2 Dishonest central node

The job of a central node C is to receive the user's request of being a part of a group and to set up a new group if the number for a group is met. Once a communication channel among the group members is established, it will leave them alone and has no business with the group members any more. Therefore, the central node cannot link any query to any user.

2.5.3 Dishonest web search engine

User U_i submits the assigned query m^i . When the web search engine receives the query, it makes a link between the query m^i and user U_i . Obviously, the web search engine builds a distorted user profile. Because of the re-masking operation and permutation steps, there is less possibility of the query m^i in correspondence to the original query m_i submitted by user U_i . Therefore, the search engine has a useless profile of user U_i .

3. Security improvement

From the security analysis stated above, their scheme is proven able to prevent any attack from a dishonest user, a dishonest central node and a dishonest search engine, and even more, their scheme can be secure if the number of dishonest users is less than $n-1$.

But if there are $n-1$ dishonest users in the same group, the n -th user will encounter the risk of being cheated by them. It is obvious that the $n-1$ group users can collaborate to analyze the queries they had submitted and the answers they had gotten, and further they can infer the n -th user's authentic search profile. Please note that each of the group members can get all the answers to their queries in the final step. For avoiding this kind of collusion attack, we make some security improvements on the UUP protocol.

3.1 The proposed improvement

The proposed improved version encompasses four sub-protocols. The first two, i.e., group setup, and group key generation, are the same as those in Castellà-Roca et al.'s scheme. Thus, they are omitted here. The other two, including anonymous query retrieval, and query submission and retrieval, are described as follows. Figure 1 demonstrates the operations in the anonymous query retrieval phase.

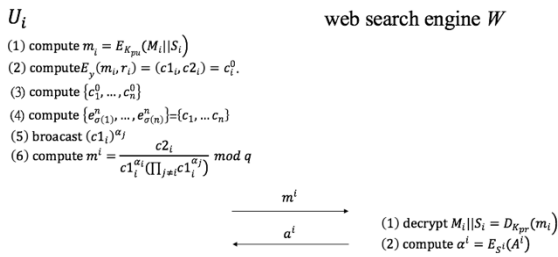


Figure 1: The operations in anonymous query retrieval

3.1.1. Anonymous query retrieval

First, we define M_i to be the real query submitted by user U_i and S_i to be a secret key selected by user U_i . We redefine m_i to be a ciphertext encrypted by the public key of a web search engine, where

$$m_i = E_{K_{pu}}(M_i || S_i). \quad (1)$$

User U_i (for $i = 1, \dots, n$) generates a random value $r_i \in Z_q^*$ and encrypts his query m_i mentioned above with the group key by means of $E_y(m_i, r_i) = (g^{r_i}, m_i \cdot y^{r_i}) \bmod q = (c1_i, c2_i) = c_i^0$. Next, user U_i (for $i = 1, \dots, n$) sends his cryptogram c_i^0 to the others in his group. In the end of the sending process, each of the group holds the ordered cryptograms $\{c_1^0, \dots, c_n^0\}$. Then, user U_1 re-masks the cryptograms $\{c_1^0, \dots, c_n^0\}$ to get a re-encrypted version. Then, user U_1 randomly permutes the re-encrypted version to obtain a re-ordered version of cryptograms. Finally, he sends the re-ordered version of cryptograms to user U_2 . Following the way in Section 2.3, in the end, User U_n has to broadcast the last result of the cryptograms $\{c_1, \dots, c_n\}$ to all of the group members.

To decrypt the value c_i , user U_i has to require all the other group members to take part

by sending their corresponding shares $(c1_i)^{\alpha_j}$. Finally, user U_i retrieves the query m^i by computing:

$$m^i = \frac{c2_i}{c1_i^{\alpha_i} (\prod_{j \neq i} c1_j^{\alpha_j})} \bmod q \quad \text{where} \quad m^i = E_{K_{pu}}(M^i || S^i).$$

3.1.2 Query submission and retrieval

Once user U_i retrieves the query m^i , (s)he submits it to the web search engine W . When the search engine receives the query, it uses its secret key to decrypt it so as to get M^i and S^i . The answer A^i to query M^i is encrypted by the selected secret key S^i .

We denote the encrypted answer as $\alpha^i = E_{S^i}(A^i)$. (2)

Similar to the concept of the UUP protocol, the web search engine has no idea about which user is the original generator of the query M^i and selected secret key S^i . Thus, user U_i receives α^i from the web search engine and broadcasts it to the rest of the group members.

At last, user U_i uses his selected secret key S_i to decrypt all of the encrypted answers to figure out the real answer to his real query.

3.2 Security analysis

The security analysis of $(n-1)$ -collusion-attack-free is given first.

Definition 1 (Collusion-attack-free).

$(n-1)$ -collusion-attack-free is defined as if $(n-1)$ dishonest users in a group with n participants has no feasible way to infer the n th user's authentic search profile by analyzing the queries they had submitted and the answers they had gotten.

Proposition 1: The improved scheme is $(n-1)$ -collusion-attack-free.

Proof.

Each participant in a group has all $m_i = E_{K_{pu}}(M_i || S_i)$ encrypted using the public key of a web search engine by Eq. (1) and all $\alpha^i = E_{S^i}(A^i)$ encrypted using participant's secret key from the search engine by Eq. (2). In such a way, even if $n-1$ collusion attackers in a group have no feasible way to deduce the n th user's authentic search profile without the keys to decrypt all m_i and α^i . Precisely, the $n-1$ group

participants can not collaborate to infer the n -th user's authentic search profile as each of the group members can only obtain their individual answer to their queries in the final step.

Thus, the improved scheme $(n-1)$ -collusion-attack-free. $\square$

Proposition 2: The improved scheme is secure even if there are dishonest users in the group.

Poof.

Suppose U_a is dishonest. In the end of the cryptogram-sending process, (s)he has containing all the queries. In order to decrypt the cryptograms, U_a must ask all the other group members to take part in by sending the shares $(c1_i)^{\alpha_j}$. With all the other group members contribute their secret keys $(\alpha_1, \dots, \alpha_n)$, (s)he is not able to decrypt the cryptograms $\{c_1^0, \dots, c_n^0\}$. Therefore, their scheme is secure if there is one dishonest user in a group. Once there are more than one dishonest users, the improved scheme is still secure by **Proposition 1**. $\square$

Proposition 3: The improved scheme is secure even if the central node is not honest in the group.

Poof.

The proof is the same as that in Section 2.5.2 and thus omitted here. $\square$

Proposition 4: The improved scheme is secure even if the web search engine is not honest in the group.

Poof.

The proof is the same as that in Section 2.5.3 and thus omitted here. $\square$

4. Conclusion

In this paper, only the web search engine can read the query M_i as it can use its secret key to decrypt the ciphertext m_i ; however, there is no link between the real query and the real generator of the query. Moreover, only the original query generator can decrypt and figure out the real answer and read it. The collaboration of $n-1$ group users only can derive a profile of encrypted answers for the n th group member. The improvement relies on the redesigned that the user's query is encrypted by means of the server's public key and then each answer is encrypted by a session key. Therefore, the security improvement proposed here can achieve

the privacy requirements as Castellà-Roca et al. stated and further it can avoid the collusion attack from group members as well.

Acknowledgment

This work was partially supported by Ministry of Science and Technology, Taiwan, R.O.C., under contract by MOST 105-2221-E-415 -012 -.

References

- [1] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," Proceedings of the 36th Annual Symposium on Foundations of Computer Science, pp. 41-50, 1995.
- [2] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," Journal of the ACM, Vol.45, Issue 6, pp. 965-981, 1998.
- [3] E. Kushilevitz, and R. Ostrovsky, "Replication is not needed: single database, computationally-private information retrieval," Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science, pp. 364-373, 1997.
- [4] R. Ostrovsky, and W.E. Skeith III, "A survey of single-database PIR: techniques and applications," Lecture Notes in Computer Science, Vol. 4450, pp. 393-411, 2007.
- [5] J. Castella-Roca, A. Viejo, and J. Herrera-Joancomarti, "Preserving user's privacy in web search engines," Computer Communications, Vol. 32, Issues 13-14, pp. 1541-1551, 2009.
- [6] Y. Desmedt, and Y. Frankel, "Threshold cryptosystems," Advances in Cryptology, Lecture Notes in Computer Science, Vol. 335, pp. 307-315, 1990.
- [7] T. ElGamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms," IEEE Transactions on Information Theory, Vol. 31, pp. 469-472, 1958.
- [8] S. Yekhanin, "Private information retrieval," Communications of the ACM, Vol. 53, pp. 68-73, 2010.
- [9] S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, "Outsourced Symmetric Private Information Retrieval," Proceedings

- of the 2013 ACM SIGSAC conference on Computer & communications security, pp. 875-888, 2013.
- [10] E. Unal and E. Savas, "On Acceleration and Scalability of Number Theoretic Private Information Retrieval," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 27, pp. 1727-1741, 2016.
 - [11] H. Sun, and S. A. Jafar, "The capacity of private information retrieval," to appear *IEEE Transactions on Information Theory*, 2017.
 - [12] Z. Li, C. Ma, D. Wang, and G. Du, "Toward single-server private information retrieval protocol via learning with errors," *Journal of Information Security and Applications*, available online 13 December, 2016.
 - [13] J. Liu, and J. Bi, "Cryptanalysis of a Fast Private Information Retrieval Protocol," *Proceedings of the 3rd ACM International Workshop on ASIA Public-Key Cryptography*, pp. 56-60, 2016.
 - [14] T. Yang, B. Yu, H. Wang, J. Li, and Z. Lv, "Cryptanalysis and improvement of Panda-public auditing for shared data in cloud and internet of things," *Multimedia Tools and Applications*, available online 8 December, 2015.
 - [15] Y. Lu, L. Li, H. Peng, and Y. Yang, "Cryptanalysis and improvement of a chaotic maps-based anonymous authenticated key agreement protocol for multi-server architecture," *Security and Communication Networks*, Vol. 9, pp. 1321-1330, 2016.
 - [16] J. Jia, J. Liu, and H. Zhang, "Cryptanalysis of cryptosystems based on general linear group," *China Communications*, Vol. 13, pp. 217-224, 2016.
 - [17] F.G. Jeng, W.L. Huang, and T.H. Chen, "Cryptanalysis and improvement of two hyper-chaos-based image encryption schemes," *Signal Processing: Image Communication*, Vol. 34, pp. 45-51, 2015.
 - [18] https://en.wikiquote.org/wiki/C._A._R._Ho_are.
 - [19] D. Boneh, G.D. Crescenzo, R. Ostrovsky, G. Persiano, "Public key encryption with keyword search," *Proceedings of EUROCRYPT'04*. LNCS, Vol. 3027, 2004, pp. 506-522.
 - [20] Y.H. Hwang, P.J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," *Proceedings of Pairing 2007*, LNCS, Vol. 4575, 2007, pp. 2-22.
 - [21] J. Shao, Z.F. Cao, X.H. Liang, H. Lin, "Proxy re-encryption with keyword search," *Information Sciences*, Vol. 180, Issue 13, 1 July 2010, pp. 2576-2587.
 - [22] W.C. Yau, R.C.-W. Phan, S.H. Heng, B.M. Goi, "Proxy re-encryption with keyword search: new definitions and algorithms," *Communications in Computer and Information Science*, Vol. 122, 2010, pp. 149-160.
 - [23] B. Zhang, F. Zhang, "An efficient public key encryption with conjunctive-subset keywords search," *Journal of Network and Computer Applications*, Vol. 34, Issue 1, January 2011, pp. 262-267.
 - [24] J. Baek, R. Safavi-Naini, W. Susilo, "Public key encryption with keyword search revisited," *Proceedings of ACIS'06*, 2006.
 - [25] J.W. Byun, H.S. Rhee, H.A. Park, D.H. Lee, "Off -line keyword guessing attacks on recent keyword search schemes over encrypted data," *Proceedings of SDM'06*. LNCS, Vol. 4165, 2006, pp. 75-83.
 - [26] H.S. Rhee, J.H. Park, W. Susilo, D.H. Lee, "Trapdoor security in a searchable public-key encryption scheme with a designated tester," *Journal of Systems and Software*, Vol. 83, Issue 5, May 2010, pp. 763-771.
 - [27] B.J. Wang., T.T. Chen., and F.G. Jeng, "Security improvement against malicious server's attack for a dPEKS Scheme," *Proceedings of International Journal of Information and Education Technology*, Vol. 1, Issue 4, 350, 2011.