輔仁學誌—理工類

中華民國九十九年十二月　　第四十四期

目　錄

FU JEN STUDIES

SCIENCE AND ENGINEERING

No.44, Dec.2010

CONTENTS

A Simulation Study on the MLE of G3B Distribution under Progressively Type-II Censoring

Sy-Mien Chen and Yu-Hsiang Hung

Department of Mathematics and Institute of Mathematics
Fu-Jen Catholic University, Taiwan, R.O.C.

Abstract

This paper addresses the maximum likelihood estimator for the three-parameter Generalized Beta (G3B) distribution via a simulation study under progressive type II censoring. Several censoring rates are considered. The simulation results show that when the censoring rate increases, the estimates spread in wider ranges, the mean squared errors become larger, the accuracy and the precision become worse. In addition, estimators are pairwise linearly correlated when there are more than one unknown estimators. The relative biases are found to be small. The MSEs and SEs (precision) diminish as the sample size is increasing and while the censoring rate are decreasing. The sample size needed for asymptotic unbiasedness to hold is smaller than what needed for the asymptotic normality.

Keywords: Trigamma function; Incomplete Beta function; Simulated annealing; MLE.

Corresponding author: Sy-Mien Chen
E-mail:smchen@math.fju.edu.tw

1. Introduction

In 1982, Libby and Novick (1982) proposed a multivariate generalized beta distributions of the first kind with applications to utility assessment. For the univariate case, the three-parameter generalized beta distribution is named G3B. G3B is very versatile and a variety of uncertainties can be usefully modeled by them, and it is more flexible than the regular Beta distribution. It was renamed as the Sahinoglu-Libby (SL) pdf due to the fact that the distribution was pioneered in 1981 by Sahinoglu (1981). Sahinoglu (2000) used the distribution to study the reliability of integrated software network when the soft-ware failure and recovery data are insufficient. Sahinoglu and Libby (2003) applied this distribution to formulate the density function of the unavailability or variability random variables in order to estimate the network reliability and quality indices for engineering and utility considerations. By assuming that the forced outage ratio of an embedded hardware component follows G3B, Sahinoglu, Libby and Das (2005) evaluated availability index with small samples for component and network reliability. Nadarajah (2005) discussed some probability properties of G3B which includes the moment generation function, expectation, variance and the Rényi entropy. Nadarajah (2005) also considered the densities of the sums, products and ratio of two random variates from G3B. So far most of the results in the literature are about probability properties. However, in the real world, the knowledge of the objects which we are interested in are estimated from data collected, it is important to know the statistical properties of the estimator in order to prevent misleading interpretations. Chen and Chen (2008) compared the maximum likelihood estimator with Bayes estimators under squared error loss function and Linex loss function while there are either one or two unknown parameters. Chen (2010) lays out the conditions for the central limit theorem of maximum likelihood estimators to hold.

It is important to know the reliability of a product, hence a life testing and reliability experiments are always carried. However, it is not uncommon that units in a life testing or a reliability experiment are lost or removed from an experiment before failure due to some reasons, censoring scheme is therefore considered. A sample is censored if out of n items placed on a test or an experiment, only part of them are actually observed to fail. If units are removed from test or experiment at points other than the final termination point, then a progressive censoring should be considered. It was Herd (1956) who first discussed

the idea of progressive censoring in his Ph.D. thesis at Iowa State College Clifford. Cohen (1963)(1965) (1975) (1976) then went deeper. Balakrishnan and Aggarwala [10] had a very extensive discussion on this topic. In additoin, Childs and Balakrishnan (1996)(2000), Balakrishnan and Kannan (2001), Balakrishnan et. al (2003) discussed estimation problem for Laplace distribution, Logistic distribution and Gaussian distribution, respectively, based on progressive Type II censored data. Recently, Sarkan and Abuammoh (2008) discussed statistical inference using progressively type-II censored data with random scheme.

In this paper, we are going to study the estimating problem of G3B while we have progressively type-II censored data on hand. In section 2, the three parameter generalized beta distribution and progressively type-II censoring are reviewed and a maximum likelihood approach is considered. A numerical and simulation study is given in section 3. A concluding remark is given in section 4.

2. MLE of parameters in G3B

Let X be a random variable from a three-parameter Generalized Beta distribution, denoted as $G3B(\alpha, \beta, \gamma)$. Then for $0 < x < 1, \alpha, \beta, \gamma > 0$, the probability density function is defined by

$$f_X(x;\alpha,\beta,\gamma) = \begin{cases} \dfrac{\gamma^{\alpha} x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)\,[1-(1-\gamma)x]^{\alpha+\beta}} & \text{if } 0 < x < 1, \\ 0 & \text{else} \end{cases}$$

where α, β are the shape parameters, γ the scale parameter, and $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ is the complete beta function, see Figure 1. Let $I_x(\alpha,\beta) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt}{B(\alpha,\beta)}$ be the incomplete beta function (Müller (1931)), then the cumulative distribution function of X is defined by

$$F_X(x) = \begin{cases} 0 & \text{if } x \le 0 \\ I_{\frac{\gamma x}{1+(\gamma-1)x}}(\alpha,\beta) & \text{if } 0 < x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

Notice that when $\gamma = 1$, $G3B(\alpha, \beta, \gamma)$ becomes the standard beta distribution $Beta(\alpha, \beta)$.

It is not uncommon that industrial units may be lost or removed before failure from a life test or a reliability experiment due to causes other than the normal failure mechanism, a progressively type II censoring may be considered under such cases.

In Progressively Type-II (PTII) censoring, assume there are n units placed on an experiment at time 0. Immediately following the first failure at X_1, R_1 surviving units are removed (or censored) from the experiment at random. The process continues until, at the time X_m of the m^{th} observed failure, the remaining $R_m = n - m - (R_1 + R_2 + \cdots + R_{m-1})$ are all removed from the experiment. By design, X_i are all random, and R_i can be fixed or random.

Let $\mathbf{R} = (R_1, R_2, \cdots, R_m)$, and $\mathbf{r} = (r_1, r_2, \cdots, r_m)$, then while the number of removals are pre-determined, i.e. $R_1 = r_1, R_2 = r_2, \cdots, R_m = r_m$, the joint pdf of the progressively Type II censored order statistics $X = (X_1, X_2, \cdots, X_m)$ can be written as

$$f_\theta(\mathbf{x} \mid \mathbf{R} = \mathbf{r}) = c^* \prod_{i=1}^{m} f_\theta(xi)[S(x_i)]^{r_i}$$

which is also the conditional likelihood function $L(\theta; \mathbf{x} \mid \mathbf{R} = \mathbf{r})$, where $\theta = (\alpha, \beta, \gamma)$, $\mathbf{x} = (x_1, x_2, \cdots, x_m)$, $0 < x_1 < \ldots < x_m < 1$, $c^* = n\,(n - r_1 - 1)(n - r_1 - r_2 - 2) \cdots (n - r_1 - r_2 - \cdots - r_m - m + 1)$, for $i = 1, 2, \cdots, m - 1$, $0 \le r_i \le (n - r_1 - r_2 - \cdots - r_{i-1})$, and for $j = 1, 2, \cdots, m$, $S(x_j) = 1 - \frac{\gamma^{\alpha}}{B(\alpha,\beta)} \int_0^{x_j} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{[1-(1-\gamma)t]^{\alpha+\beta}}\, dt$ (See Cohen(1963)).

Assume that an individual unit being removed from an experiment right after the i^{th} failure is independent of the other removing items but with equal probability p, without loss of generality, let's assume that the number R_i of units removed at the i^{th} failure follows a binomial distribution with parameters $n - m - \sum_{l=1}^{i-1} r_l$ and p, for $i = 1, 2, \cdots, m - 1$. Then,

$$P(R_l = r_l) = \binom{n - m}{\gamma_l} p^{r_l} (1 - p)^{n - m - r_l},$$

and

$$P\left(R_j = r_j \mid R_{j-1} = r_{j-1}, \cdots, R_1 = r_1\right) = \binom{n-m-\sum_{l=1}^{i-1} r_l}{\gamma_1} p^{r_j} (1-p)^{n-m-\sum_{l=1}^{i-1} r_l},$$

for $j = 2, \cdots, m-1$, and $p \in (0,1)$.

The full log-likelihood function $l(\theta, p; \mathbf{x}, \mathbf{r})$ takes the following form:

$$\begin{aligned} l(\theta, p; \mathbf{x}, \mathbf{r}) &= lnA - ln(B(\alpha, \beta)) + (\alpha - 1) \sum_{i=1}^{m} lnx_i + (\beta - 1) \sum_{i=1}^{m} ln(1 - x_i) \\ &\quad - (\alpha + \beta) \sum_{i=1}^{m} ln[1 - (1-\gamma) x_i] + \sum_{i=1}^{m} r_i ln[S(x_i)] \\ &\quad + \sum_{i=1}^{m-1} [r_i lnp - (m-i) r_i ln(1-p)] + (n-m)(m-1) ln(1-p) \stackrel{\triangle}{=} l(\alpha, \beta, \gamma), \end{aligned}$$

where

$$A = \frac{c^* (n-m)!}{(n - m - \sum_{i=1}^{m-1} r_i)! \prod_{i=1}^{m-1} r_i!} .$$

By differentiating the log-likelihood function with respect to α, β, and γ, respectively and set to 0 we have the following estimating equations,:

$$\left.\begin{aligned} \frac{\partial l}{\partial \alpha} &= m[ln\gamma - \Psi_1(\alpha) + \Psi_1(\alpha+\beta)] + \sum_{i=1}^{m} lnx_i - \sum_{i=1}^{m} ln[1-(1-\gamma)x_i] \\ &\quad - \sum_{i=1}^{m} r_i \frac{F(x_i)}{s(x_i)} \{lnt + ln\gamma - \Psi_1(\alpha) + \Psi_1(\alpha+\beta) - ln[1-(1-\gamma)t]\} \stackrel{\triangle}{=} 0 \\ \frac{\partial l}{\partial \beta} &= m[-\Psi_1(\beta) + \Psi_1(\alpha+\beta)] + \sum_{i=1}^{m} ln(1-x_i) - \sum_{i=1}^{m} ln[1-(1-\gamma)x_i] \\ &\quad - \sum_{i=1}^{m} r_i \frac{F(x_i)}{s(x_i)} \{ln(1-t) - \Psi_1(\beta) + \Psi_1(\alpha+\beta) - ln[1-(1-\gamma)t]\} \stackrel{\triangle}{=} 0 \\ \frac{\partial l}{\partial \gamma} &= m[\frac{\alpha}{\gamma}] - (\alpha+\beta) \sum_{i=1}^{m} \frac{x_i}{1-(1-\gamma)x_i} + \sum_{i=1} r_s \frac{s'(x_i)}{s(xi)} = 0 \end{aligned}\right\} \quad (1)$$

where $\Psi_1(m) = \frac{d^2 ln\Gamma(m)}{dm^2}$ is the trigamma function, and

$$S'(x_i) = \frac{\partial l}{\partial \gamma}$$

$$= -\frac{\alpha \gamma^{\alpha+1}}{B(\alpha,\beta)} \int_0^{x_i} \frac{t^{\alpha+1}(1-t)^{\beta-1}}{[1-(1-\gamma)t]^{\alpha+\beta}}\, dt + (\alpha+\beta)\frac{\gamma^{\alpha}}{B(\alpha,\beta)} \int_0^{x_i} \frac{t^{\alpha}(1-t)^{\beta-1}}{[1-(1-\gamma)t]^{\alpha+\beta+1}}\, dt.$$

The second derivatives are much more cumbersome than the first derivatives, and are not given here. Theoretically speaking, by solving the system of estimating equations in (1), we can get the possible maximum likelihood estimators for the unknown parameters. However, from the above expression, it is hopeless to get an explicit form for these es-timators. Therefore, in this research, we will study the MLE of parameters based on a simulation study along with a Monte Carlo method.

3.Simulation study

3.1 Parameter setting

In the simulation, to generate a random sample from G3B(α, β, γ), first we generate a random sample $T_1, T_2, \cdots, T_n$ from Beta(α, β), then a random sample $\underline{Y} = (Y_1, Y_2, \cdots, Y_n)$ from G3B(α, β, γ) can be generated by taking the transformation $Y_k = \frac{T_k}{(1-\gamma)T_k + \gamma}$, for $k = 1, 2, \cdots, n$ (Libby and Novick (1982)).

For a progressively censored type-II sample, first let $X_1 = Y_{(1)}$, and then R_1 items are randomly removed from the n-1 remainders. Let X_i be the smallest one in the remaining $n - (i-1) - \sum_{j=1}^{i=1} R_j$ items after the $(i-1)^{th}$ step. Then R_i items are randomly removed from the remaining n - i - $\sum_{j=1}^{i=1} R_j$ items right after X_i is observed. Continue the procedure for $i = 2, 3, \cdots, m$, a progressively censored type-II random sample $X_1, X_2, \cdots, X_m$ is then obtained.

Nowaday, in industrial practice, lot sizes may be running into the thousands, and the sample size might be in the hundreds rather than a small number such as 10. On the other hand, MLE performs better for samples with larger size. Hence in this research, we consider

the following cases:

Unknown parameters	α	β	γ	α, β	α, γ	β, γ	α, β, γ
Sample size	300 3000	400 1600	400 1600 3000	400 3000	400 3000	1000 3000	1000 2000

As one can see, it is hopeless to solve those non-linear equations given in (1) analytically, to find the possible MLEs a numerical method is used instead. Since the parameters and the digamma function involved in the estimating equations are all positive, the Simulated Annealing method (SA)(Robert and Casella (2004)) is preferred to the Newton-Raphson method. The temperature function $W(t)$ we use to control the cooling is $\frac{1}{100\ln(1+t)}$ and the size of the interval around the current point is 0.5. The accuracy is up to the fourth decimal point. For each different combination of parameters, the simulations are repeated 20000 times.

Assume that at the t^{th} iteration the algorithm is at $((\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)}), l(\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)}))$, a simple algorithm of the SA method is given as the following:

1. Simulate $\underline{u} = (u_1, u_2, u_3)$ from uniform distributions, i.e., let
   $u_1 \sim Uniform\ (g_1(t), k_1(t))$,
   $u_2 \sim Uniform\ (g_2(t), k_2(t))$,
   $u_3 \sim Uniform\ (g_3(t), k_3(t))$,
   where
   $g_1(t) = max\{\alpha^{(t)} - 0.5, 0\}$, $k_1(t) = \alpha^{(t)} + 0.5$,
   $g_2(t) = max\{\beta^{(t)} - 0.5, 0\}$, $k_2(t) = \beta^{(t)} + 0.5$,
   $g_3(t) = max\{\gamma^{(t)} - 0.5, 0\}$, $k_3(t) = \gamma^{(t)} + 0.5$.

2. If $l(\underline{u}) \geq l(\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)})$ then take $(\alpha^{(t+1)}, \beta^{(t+1)}, \gamma^{(t+1)}) = \underline{u}$.
   If $l(\underline{u}) < l(\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)})$, *let* $\rho^{(t)} = min\{\exp(\frac{l(u) - l((\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)}))}{W_t}), 1\}$, and
   generate p from Uniform(0,1),
   if $p \in (0, \rho^{(t)})$ then take $(\alpha^{(t+1)}, \beta^{(t+1)}, \gamma^{(t+1)}) = (\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)})$.
   otherwise go to 3

3. Update W_t to W_{t+1}, then repeat step 1 to step 3.

The true values of unknown parameters used in the simulation are 0.7, 1.0, 3.0, and 10.0 to cover very small to very large cases. For the parameters which are known, we take 0.6, 0.8, 1.0, 1.5, 2.5 to cover the cases while they are smaller than one, equal to one, and larger than one.

In order to see the effect of censoring rate to the estimation, we consider few censored rates, eg. 0% (complete sample), 50%, 75% and 90%.

In the simulation study, IMSL along with FORTRAN is used to generate the random sample and to compute the point estimates of parameters. MATLAB is used for some graphing. In addition, SAS package is used to do the hypothesis testing for normality and unbiasedness.

Different combinations of parameters are chosen from 0.7, 1.0, 3.0 and 10. Since the simulation results are all similar, to save the space, in the following, we only list partial results in figures and tables.

3.2 Simulation Results

3.2.1 When there is only one unknown parameter

Basically, the variability of estimators becomes larger when the censoring rate and the true value of unknown parameter $\hat{\alpha}$, $\hat{\beta}$, or $\hat{\gamma}$ are increased. It can be seen from Figure 2 that more outliers occur in the empirical distributions of $\hat{\beta}$ and $\hat{\gamma}$.

The range of the ratio of the sample mean squared errors MSE under the higher censoring rate over the sample MSE under the lower censoring rate are given in Table 1. Information relates to the relative bias and standard deviation are also given in Table 1. From the table, we can see how much the estimating error can be varied when the censoring rate is increased. In addition, the effects of censoring rate on standard deviation and mean squared error are stronger for $\hat{\beta}$ and $\hat{\gamma}$ than the effect for $\hat{\alpha}$. See Table 2.

Based on the theorem, an MLE is asymptotically unbiased and asymptotically normally distributed under some regularity conditions. Since it is hard to check those conditions in the current distribution, here we use SAS package to do the job instead. When the sample size is 300, $\hat{\alpha}$ passes normality test except when the true value is 0.7 and the censoring rate is 90%. $\hat{\beta}$ performs worse and $\hat{\gamma}$ performs the worst. But when the sample size is increased to 2000 and above, the performance of estimators are improved dramatically. The estimators pass the normality test for all but one case which occurs for $\hat{\gamma}$ at level 0.05 and 90% rate. And all three estimators are asymptotic unbiased. See Table 3.

The simulating time needed for estimating parameter γ is longer than what are needed for the parameters α and β. However, under each case, the simulation time is faster while the two known parameters are both equal to one or both greater than one.

3.2.2 When there are two unknown parameters

In general, smaller MSE, smaller SD, and smaller SMSE are reached under smaller censoring rate and also under the case when the unknown parameters have smaller true value.

In each pair of unknown parameters, if one of them is fixed, then when the other one varies from 0.7 to 1, the increment of the MSE of the corresponding estimator is about double. However, the pattern no longer exist when the true values are getting larger. When both α and β are unknown and β is fixed, the increment of the MSE of $\hat{\alpha}$ is increased no less than ten times when α is varied from 1 to 3 to 10. Same thing happened to $\hat{\beta}$ when α is fixed. When both α and γ are unknown and α is fixed, the increment of the MSE of $\hat{\gamma}$ becomes no less than double when γ is varied from 1 to 3. When γ is varied from 3 to 10 the increment of the MSE of $\hat{\gamma}$ is no less than 10 times. When both β and γ are unknown and γ is fixed, the increment of the MSE of $\hat{\beta}$ becomes no less than ten times when β is varied from 1 to 3, some of them are even as large as forty times. When β is varied from 3 to 10 the increment of the MSE of $\hat{\beta}$ is even worse. One other thing we can see in this case is that the higher the censoring rate, the less the increment of the MSE of $\hat{\beta}$ When β is fixed, the increments of the MSE of $\hat{\gamma}$ are no less than three times when γ is varied from 1 to 3. When γ is varied from 3 to 10 the increment of the MSE of $\hat{\gamma}$ is no less than ten times See Table 4.

Similarly, the relative bias of each estimator increases as the censoring rate or the true value of unknown parameters are increased. However, the increments are quite small, it is no larger than the second decimal point when the rate is increased to more than half. The relative bias for $\hat{\alpha}$ and the relative bias for $\hat{\beta}$ both have no special pattern when the censored rate is increased. However, the relative bias for $\hat{\alpha}$ and the relative bias for $\hat{\gamma}$ are both increased when the true value of the unknown parameters are increased. See Table 5.

The estimators $\hat{\alpha}$ and $\hat{\beta}$ are positive correlated, the same as the correlation between $\hat{\alpha}$ and $\hat{\gamma}$ Their correlation coeffcients are within 0.51 ~ 0.9586 and 0.67 ~ 0.88, respectively. But, $\hat{\beta}$ and $\hat{\gamma}$ are negative correlated and the value of the correlation coeffcients are within -0.96 ~ -0.67. The larger the true value of parameters, the stronger the linear relation between $\hat{\alpha}$ and $\hat{\beta}$, especially when they ar both large, the correlation coeffcient reaches the maximum. For fixed α , corr($\hat{\alpha}$, $\hat{\beta}$) < corr($\hat{\alpha}$, $\hat{\gamma}$) except when α = 1, β = γ =10 and censoring rate = 90%. $\hat{\beta}$ and $\hat{\gamma}$ are negative correlated, and the magnitute is within -0.6705 ~ -0.9586.

The effect of the censoring rate on the MSE and on the SD for $\hat{\gamma}$ are stronger than the effect on what for $\hat{\alpha}$ The effect on $\hat{\beta}$ is stronger than the effect on $\hat{\gamma}$. The effect on $\hat{\beta}$ is stronger than the effect on $\hat{\alpha}$. Overall speaking, the censoring rate affects $\hat{\beta}$ the most, and $\hat{\alpha}$ the least.

3.2.3 When there are three unknown parameters

In the real life, it is quite common that none of the parameters of G3B(α, β, γ) distribution is known in advance. As one can expect, the sample size needed under such situation will be larger than the size needed while there are less unknown parameters. Hence the sample size we use for this case will be n = 1000 and 2000. The simulation results are similar to what we have in previous cases.

The SD, the bias, the MSE and the SMSE increase for higher censoring rates, which is not surprised. The estimator of the parameter that has the largest true value always performs the worst. When the true value are all the same, then $\hat{\gamma}$ always performs the worst, and $\hat{\alpha}$ performs better than $\hat{\beta}$ while the censoring rate is larger. At fixed censoring rate, unless the

true value of α is the largest among three paramters, $\hat{\alpha}$ performs the best regardless the true values of paramters. When n = 1000, the range of the relative bias for all three unknown parameters are $\alpha \in$ (.0009,.079), $\beta \in$ (.0028,.2128) and $\gamma \in$ (.0145,.1569). When n=2000, the relative bias for all three unknown parameters are improved such that $\alpha \in$ (.0004,.0191.), $\beta \in$ (.0470,.1253) and $\gamma \in$ (.0023,.1551), which tells us the sample size does improve the estimation quite a lot.

The correlation coefficients between any pair of these estimators have the following properties: Corr($\hat{\alpha}, \hat{\beta}$) $\in$ (- 0.73, - 0.28), and Corr($\hat{\beta}, \hat{\gamma}$) $\in$ (- 0.93, - 0.47), which indicate that both pairs are negative correlated. Corr($\hat{\alpha}, \hat{\gamma}$) $\in$ (0.63, 0.97) which shows strong positive correlation between $\hat{\alpha}$ and $\hat{\gamma}$. Notice that the relation between $\hat{\alpha}$ and $\hat{\beta}$ has been changed compare to the case while there are only two unknown parameters.

4 Conclusion

G3B distribution is very useful in describing many real world cases. Due to the reality that data may not be collected completely, a progressively type data were discussed thoroughly in the literature. In this research, we conduct a simulation study on the maximum likelihood estimator of parameters in G3B distribution based on progressively type II censored data. Several censoring rates are discussed and the results were compared to the results that a complete data set may lead to.

According to the simulation results, all the estimators have similar pattern regardless the number of unknown parameters in the distribution, i.e. for G3B distribution, when the censoring rate increases, the estimates spread in wider ranges, the mean squared errors become larger, the accuracy and the precision become worse. See Table 6, 7, 8.

We summarize the main contributions of the current research as the following:

1.Estimators are pairwise linearly correlated when there are more than one unknown estimators. The larger the true value of the unknown parameters, the stronger the linear

relation between their estimators. It is worth noting that the sign of the correlation between α and β are changed under two and three unknown cases. From the simulation, we see that $\hat{\alpha}$ and $\hat{\gamma}$ are positive correlated and $\hat{\beta}$ and $\hat{\gamma}$ are negative correlated in all cases, but $\hat{\alpha}$ and $\hat{\beta}$ are negative correlated while all three estimators are unknown, and they are positive correlated when only α and β are unknown.

2.The relative biases of $\hat{\alpha}, \hat{\beta}$ and $\hat{\gamma}$ are found to be small, and the biases do not exceed 6.1% for α, 5.5% for β and 5.4% for γ when only one of the parameter is unknown. When there are two unknown parameters, the biases do not exceed 5.1% for α, 7.5% for β ; do not exceed 5% for α, 2.1% for γ ; do not exceed 3.5% for β, 18% for γ. And do not exceed 2% for α, 13% for β, 18% for γ. when all the three parameters are all unknown. It is easy to see that the accuracy of $\hat{\gamma}$ is the worst. Fortunately, the accuracy (bias) can be improved by decreasing the censoring rate. However, the sample size increases is the principal reason for the decrease in bias.

3.MSE and SD (precision) diminishes as the sample size is increasing and while the censoring rate are decreasing. The sample size needed for asymptotical unbiasedness to hold is smaller than what needed for the asymptotic normality. Over all speaking, for all the MLE's, the sample sizes needed for each estimator to reach the same results satisfy $n_{\hat{\alpha}} < n_{\hat{\beta}} < n_{\hat{\gamma}}$ and $n_{\hat{\alpha},\hat{\beta}} < n_{\hat{\alpha},\hat{\gamma}} < n_{\hat{\beta},\hat{\gamma}}$.

References

1. N. Balakrishnan and R. Aggarwala (2000), Progressive Censoring Theory, Methods, and Applications, second edition, Springer.
2. N. Balakrishnanm and N. Kannan (2001), Point and interval estimation for parameters of the logistic distribution based on progressively Type-II censored samples, in Handbook of statistical N Balakrishnanm and C.R. Rao, Eds. Amsterdam: NorthHoland, 20, 431-456.
3. N. Balakrishnanm, N. Kannan, C. T. Lin and H. Ng (2003), Point and interval estimation for gaussian distribution based on progressively Type-II censored samples, *IEEE Trans. Reliab.*, 52, 90-95.
4. W. T. Chen and S. M. Chen (2008), An Estimating Problem on Generalized Beta Distribution, *Fu-Jen Studies* (Science and Engineering), 42, 31-52.
5. S. M. Chen (2010), Maximum Likelihood Estimation of G3B distribution via Simulated Annealing Method, submitted.
6. A. Childs and N. Balakrishnanm (1996), Conditional inference procedures for the Laplace distribution based on Type II right censored samples, *Statistics Probability Letters*, 31, 31-39.
7. A. Childs and N. Balakrishnanm (2000), Conditional inference procedures for the Laplace distribution when the observed samples are progressively censored, *Metrika,* 52, 253-265.
8. A. C. Cohen (1963), Progressively Censored Samples in Life Testing, *Technometrics*, 5(3), 327-339.
9. A. C. Cohen (1965), Maximum Likelihood Estimator in the Weibull Distribution Based on Complete and on Censored Samples, *Technometrics*, 7(4), 579-588.
10. A. C. Cohen (1975), Multi-Censored Sampling in the Three Prameter Weibull Distribution, *Technometrics*, 17(3), 347-351.
11. A. C. Cohen (1976), Progressively Censored Samples in Three Prameter Log-Normal Distribution, *Technometrics*, 18(1), 99-103.
12. D.L. Libby and M.R. Novick (1982), Multivariate generalized beta-distributions with applications to utility assessment, *Journal of Educational Statistics* 7(4): 271-294.
13. D.L. Libby and S. R. Das (2005), Measuring availability indexes with small samples for component and network reliabiity using the Sahinoglu-Libby probability model, *IEEE Transactions on Instrumentation and Measurement*, 54(3), 1283-1295.
14. S. Nadarajah (2005a), Exponentiated Beta Distributions, *Computers & Mathematics with*

Applications, 49: 1029-1035.

15. S. Nadarajah (2005b), Sums, Products and Ratios of Genearalized Beta Variables, Statistical Papers, 47: 69-90.
16. Robert, C. P. and Casella, G. (2004). Monte Carlo Statistical Methods, 2nd Ed; Springer.
17. M. Sahinoglu (1981), Statistical Inference on the Reliability Performance Index for Electric Power Generation Systems, Ph.D. Dissertation, TAMU.
18. M. Sahinoglu (2000), Reliability Index Evaluations of Integrated Software Systems (Internet) for Insuffcient Software Failure and Recovery Data. SpringerLink, 363-373.
19. M. Sahinoglu and D.L. Libby (2003), Sahinoglu-Libby (SL) Probability Density Function Component Reliability Applications in Integrated Networks, 2003 Proceedings Annual Reliability and Maintainability Symposium, 280-287.
20. M. Sahinoglu, D. L. Libby and Das S.R. (2005), Measuring Availability Indexes with Small Samples for Component and Network Reliability Using the Sahinoglu-Libby Probability Model, *IEEE Transactions on Instrumentation and Measurement*, 54(3) : 1283-1295.
21. M. Sarkan and A. Abuammoh (2008), Statistical Inference Using Progressively Type-II Censored Data with Random Scheme, *International Mathematical Forum*, 3, 35, 1713 -1725.

Table 1: Range about the ratio of higher censoring rate over lower censoring rate

Unknown parameters	α	β	γ
sample size	300	400	400
Range of "ratio of sd"	(.6024, .9465)	(.2738, .7383)	(.3350, .7634)
Range of "ratio of MSE"	(.3449, .8990)	(.0717, .5401)	(.1078, .6377)
Range of "relative bias"	(.0001, .0061)	(.0012, .0548)	(.0016, .0538)

Table 2: The standard error, bias and mean squared error of each individual estimator.

			0%	50%	75%	100%
SD	$\hat{\alpha}$	0.7	0.0388	0.0508	0.0545	0.0612
	n=300	1	0.0529	0.0674	0.0818	0.0879
		3	0.1509	0.1938	0.2274	0.2581
		10	0.4904	0.5977	0.7328	0.7743
	$\hat{\beta}$	0.7	0.0378	0.0548	0.0763	0.1156
	n=400	1	0.0511	0.0718	0.1074	0.1619
		3	0.1886	0.2555	0.4055	0.6887
		10	0.6293	0.8980	1.3027	2.2388
	$\hat{\gamma}$	0.7	0.0552	0.0823	0.1103	0.1637
	n=400	1	0.0882	0.1155	0.1675	0.2335
		3	0.2502	0.3663	0.4846	0.7298
		10	0.8679	1.1844	1.6806	2.5911
MSE	$\hat{\alpha}$	0.7	0.0015	0.0026	0.0030	0.0038
	n=300	1	0.0028	0.0046	0.0067	0.0077
		3	0.0230	0.0377	0.0517	0.0666
		10	0.2423	0.3627	0.5394	0.6000
	$\hat{\beta}$	0.7	0.0014	0.0030	0.0060	0.0136
	n=400	1	0.0026	0.0052	0.0118	0.0267
		3	0.0356	0.0659	0.1677	0.4966
		10	0.4043	0.8161	1.7665	5.3363
	$\hat{\gamma}$	0.7	0.0030	0.0069	0.0124	0.0272
	n=400	1	0.0078	0.0134	0.0281	0.0554
		3	0.0626	0.1345	0.2364	0.5447
		10	0.7549	1.4153	2.8802	7.0034

Table 2 (Continous):

			0%	50%	75%	100%
bias	$\hat{\alpha}$	0.7	0.0010	0.0013	0.0022	0.0018
	n=300		(0.0014)	(0.0019)	(0.0032)	(0.0026)
		1	0.0045	0.0030	0.0029	0.0042
			(0.0045)	(0.0019)	(0.0032)	(0.0026)
		3	0.0144	0.0138	0.0155	0.0041
			(0.0048)	(0.0046)	(0.0052)	(0.0014)
		10	0.0432	0.0741	0.0488	0.0231
			(0.0034)	(0.0071)	(0.0049)	(0.0023)
	$\hat{\beta}$	0.7	0.0019	0.0044	0.0123	0.0154
	n=400		(0.0028)	(0.0063)	(0.0175)	(0.0221)
		1	0.0041	0.0066	0.0148	0.0218
			(0.0041)	(0.0066)	(0.0148)	(0.0218)
		3	0.0052	0.0257	0.0569	0.1491
			(0.0017)	(0.0086)	(0.0190)	(0.0497)
		10	0.0912	0.0987	0.2636	0.5693
			(0.0091)	(0.0098)	(0.0263)	(0.0569)
	$\hat{\gamma}$	0.7	0.0032	0.0064	0.0154	0.0197
	n=400		(0.0045)	(0.0091)	(0.0020)	(0.0281)
		1	0.0013	0.0031	0.0097	0.0304
			(0.0013)	(0.0045)	(0.0139)	(0.0434)
		3	-0.0049	0.0166	0.0389	0.1099
			(-0.0016)	(0.0055)	(0.0130)	(0.0366)
		10	0.0409	0.1113	0.2365	0.5381
			(0.0041)	(0.0111)	(0.0237)	(0.0538)

Remark: Numbers inside the parentheses are the relative bias.

Table 3(a): Some statistical inference results for $\sqrt{m}(\hat{\alpha} - \alpha)$ at different level with sample size=3000.

test for unbiasedness	0%	50%	75%	90%
at level 0.05	y	y	y	y
at level 0.01	y	y	y	y
test for normality	0%	50%	75%	90%
at level 0.05	y	y	y	y
at level 0.01	y	y	y	y

Table 3(b): Some statistical inference results for $\sqrt{m}(\hat{\beta} - \beta)$ at different level with sample size=2000.

test for unbiasedness	0%	50%	75%	90%
at level 0.05	y	y	y	y
at level 0.01	y	y	y	y
test for normality	0%	50%	75%	90%
at level 0.05	y	y	y	y
at level 0.01	y	y	y	y

Table 3(c): Some statistical inference results for $\sqrt{m}(\hat{\gamma} - \gamma)$ at different level with sample size=3000.

test for unbiasedness	0%	50%	75%	90%
at level 0.05	y	y	y	y
at level 0.01	y	y	y	y
test for normality	0%	50%	75%	90%
at level 0.05	y	y	y	n
at level 0.01	y	y	y	y

Table 4(a): MSE for $\hat{\alpha}, \hat{\beta}$ and MSE for $\hat{\alpha}, \hat{\gamma}$ with n=400 under different censoring rate α, β unknown, γ =1.5, and α, γ unknown, β =0.8

rate	$\hat{\beta}/\hat{\alpha}$	0.7	1	3	10	$\hat{\gamma}/\hat{\alpha}$	0.7	1
0%	0.7	0.0020	0.0047	0.0545	0.7849	0.7	0.0049	0.0127
		(0.0018)	(0.0021)	(0.0017)	(0.0018)		(0.0205)	(0.0210)
	1	0.0021	0.0043	0.0521	0.5913	1	0.0043	0.0163
		(0.0052)	(0.0043)	(0.0046)	(0.0036)		(0.0378)	(0.0555)
	3	0.0017	0.0042	0.0399	0.5167	3	0.0035	0.0105
		(0.0529)	(0.0535)	(0.0409)	(0.0425)		(0.2561)	(0.2963)
	10.0	0.0019	0.0041	0.0393	0.5670	10.0	0.0043	0.0106
		(0.6532)	(0.6155)	(0.5147)	(0.5718)		(3.6585)	(2.7665)
50%	0.7	0.0033	0.0066	0.0828	1.1083	0.7	0.0060	0.0182
		(0.0041)	(0.0040)	(0.0035)	(0.0035)		(0.0335)	(0.0403)
	1	0.0033	0.0070	0.0784	0.8729	1	0.0066	0.0218
		(0.0094)	(0.0092)	(0.0082)	(0.0069)		(0.0754)	(0.0862)
	3	0.0028	0.0066	0.0693	0.8337	3	0.0046	0.0131
		(0.1059)	(0.1037)	(0.0802)	(0.0788)		(0.3750)	(0.4109)
	10.0	0.0030	0.0065	0.0675	0.8455	10.0	0.0065	0.0140
		(1.4098)	(1.1296)	(0.9506)	(0.9095)		(7.7193)	(3.9801)
75%	0.7	0.0044	0.0107	0.1151	1.4534	0.7	0.0091	0.0240
		(0.0090)	(0.0099)	(0.0079)	(0.0061)		(0.0719)	(0.0654)
	1	0.0041	0.0107	0.0991	1.2929	1	0.0089	0.0276
		(0.0206)	(0.0173)	(0.0148)	(0.0160)		(0.1416)	(0.1527)
	3	0.0041	0.0091	0.1006	1.2460	3	0.0055	0.0147
		(0.1995)	(0.1876)	(0.1495)	(0.1426)		(0.5787)	(0.5391)
	10.0	0.0043	0.0088	0.0983	1.2237	10.0	0.0079	0.0160
		(2.9529)	(1.9830)	(1.7306)	(1.5109)		(13.5279)	(5.5655)
90%	0.7	0.0061	0.0137	0.1864	2.2489	0.7	0.0142	0.0393
		(0.0310)	(0.0249)	(0.0227)	(0.0209)		(0.2400)	(0.1848)
	1	0.0114	0.0143	0.1729	2.3149	1	0.0130	0.0424
		(0.0732)	(0.0572)	(0.0524)	(0.0446)		(0.3989)	(0.3980)
	3	0.0046	0.0125	0.1367	2.3274	3	0.0059	0.0182
		(0.4170)	(0.3899)	(0.2891)	(0.3888)		(0.7509)	(0.8332)
	10.0	0.0060	0.0100	0.1582	1.9185	10.0	0.0100	0.0180
		(9.0996)	(4.0025)	(4.2040)	(2.9550)		(25.1195)	(7.6293)

Remark* : Number in parenthesis corresponds to the MSE of the estimator * in */**

Table 4(b):MSE for $\hat{\beta}, \hat{\gamma}$ with n=1000 under different censoring rate, β, γ unknown, α =1.0.

censored rate	$\hat{\gamma}/\hat{\beta}$	0.7	1	3	10
0%	0.7	0.0106	0.0042	0.1574	20.5852
		(2.7665)	(0.0063)	(0.0125)	(0.0614)
	1	0.0015	0.0108	0.1865	19.3677
		(0.0110)	(0.0338)	(0.0278)	(0.1234)
	3	0.0014	0.0039	0.1678	19.8230
		(0.09996)	(0.1059)	(0.2432)	(1.0958)
	10.0	0.0016	0.0045	0.1986	19.5172
		(1.2398)	(1.2206)	(2.8649)	(11.5001)
50%	0.7	0.0140	0.0081	0.3321	31.8945
		(3.9801)	(0.0141)	(0.0228)	(0.1034)
	1	0.0030	0.0152	0.4078	29.0965
		(0.0219)	(0.0605)	(0.0506)	(0.2216)
	3	0.0030	0.0094	0.4261	29.1866
		(0.2011)	(0.2343)	(0.4895)	(1.4511)
	10.0	0.0036	0.0090	0.4444	31.3856
		(2.5943)	(2.4098)	(5.5843)	(18.7389)
75%	0.7	0.0160	0.0216	0.7062	42.7801
		(5.5655)	(0.0026)	(0.0505)	(0.1746)
	1	0.0064	0.0188	0.8237	40.1429
		(0.0455)	(0.0498)	(0.1153)	(0.3404)
	3	0.0060	0.0208	0.7816	41.5055
		(0.3204)	(0.4753)	(0.8521)	(2.0112)
	10.0	0.0069	0.0211	0.7734	37.8411
		(4.6363)	(4.8434)	(9.7377)	(29.0608)
90%	0.7	0.0180	0.0807	1.4194	53.2996
		(7.6293)	(0.0619)	(0.0903)	(0.4334)
	1	0.0223	0.0634	1.3608	51.5882
		(0.1161)	(0.1181)	(0.2369)	(0.8973)
	3	0.0252	0.1049	1.3326	46.2267
		(0.5685)	(0.8971)	(1.3098)	(2.2192)
	10.0	0.0207	0.0542	1.3945	48.8978
		(10.6877)	(10.0310)	(19.4805)	(35.5256)

Table 5(a): The absolute relative bias (arb) under censoring rate 0%, 50%, 75%, 100%.
when there are two unknown and n=400.

α, β, γ	arb	0%	50%	75%	90%
(0.7, 3, 1.5)	$\lvert bias(\hat{\alpha})/\alpha \rvert$	0.0031	0.0043	0.0137	0.0051
	$\lvert bias(\hat{\beta})/\beta \rvert$	0.0045	0.0133	0.0317	0.0428
(0.7, 0.8, 3)	$\lvert bias(\hat{\alpha})/\alpha \rvert$	0.0036	0.0047	0.0031	0.0044
	$\lvert bias(\hat{\gamma})/\gamma \rvert$	0.0146	0.0017	0.0055	0.0053
(1, 0.7, 3)	$\lvert bias(\hat{\beta})/\beta \rvert$	0.0063	0.0691	0.0143	0.0819
	$\lvert bias(\hat{\gamma})/\gamma \rvert$	0.0020	0.0009	0.0066	0.0393

Table 5(b): The absolute relative bias (arb) under censoring rate 0%, 50%, 75%, 100%.
when there are two unknown and n=3000.

α, β, γ	arb	0%	50%	75%	90%
(0.7, 3, 1.5)	$\lvert bias(\hat{\alpha})/\alpha \rvert$	0.0020	0.0020	0.0035	0.0085
	$\lvert bias(\hat{\beta})/\beta \rvert$	0.0013	0.0045	0.0077	0.0102
(0.7, 0.8, 3)	$\lvert bias(\hat{\alpha})/\alpha \rvert$	0.0043	0.0014	0.0017	0.002
	$\lvert bias(\hat{\gamma})/\gamma \rvert$	0.0121	0.0038	0.0001	0.0158
(1, 0.7, 3)	$\lvert bias(\hat{\beta})/\beta \rvert$	0.0094	0.0150	0.0390	0.0979
	$\lvert bias(\hat{\gamma})/\gamma \rvert$	0.0031	0.0006	0.0050	0.0063

Table 6(a): $\hat{\alpha}$ when β =1.5, γ =2.5 (the true value of α =3.0)

censoring rate	99%	97.5%	95%	90%	75%	50%	0%
sample size	30	75	150	300	750	1500	3000
MSE	0.1735	0.0781	0.0359	0.0171	0.0072	0.044	0.0022
BIAS	0.0711	0.0341	-0.0117	0.0026	-0.0017	-0.0059	-0.0007
SD	0.4105	0.2773	0.1892	0.1308	0.0846	0.0658	0.0469

Table 6(b): $\hat{\beta}$ when α =1.5, γ =2.5 (the true value of β =3.0)

censoring rate	97.5%	93.75%	90%	87.5%	75%	50%	0%
sample size	40	100	160	200	400	800	1600
MSE	0.1735	0.0781	0.0488	0.0359	0.0177	0.0086	0.0045
BIAS	0.0711	0.0341	0.0204	-0.0117	0.0029	0.0049	0.0017
SD	0.4105	0.2773	0.2200	0.1892	0.1330	0.0925	0.0669

Table 6(c): $\hat{\gamma}$ when α =0.6, β =1.5 (the true value o γ =3.0)

censoring rate	98.67%	96.67%	90.00%	86.67%	75%	50%	0%
sample size	40	100	300	400	750	1500	3000
MSE	0.8701	0.3003	0.1337	0.0755	0.0367	0.0212	0.0105
BIAS	0.1647	0.0552	0.0215	0.0016	0.0259	0.0248	0.0078
SD	0.9182	0.5452	0.3651	0.2749	0.1899	0.1434	0.1019

Table 7(a): $\hat{\alpha}$ and $\hat{\beta}$ when γ =1.5 (the true value of α =0.7, β =3.0)

censoring rate	98.67%	96.67%	94.67%	86.67%	75%	50%	0%
sample size	40	100	200	400	750	1500	3000
MSE($\hat{\alpha}$)	0.0046	0.0041	0.0028	0.0017	0.0008	0.0004	0.0003
MSE($\hat{\beta}$)	0.4170	0.1995	0.1059	0.0529	0.0328	0.0127	0.0078
SMSE	0.4216	0.2036	0.1087	0.0546	0.0336	0.0132	0.0080
BIAS($\hat{\alpha}$)	0.0036	0.0096	0.0030	0.0022	-0.0025	-0.0014	-0.0014
BIAS($\hat{\beta}$)	0.1283	0.0952	0.0398	0.0136	-0.0232	-0.0137	-0.0040
SD($\hat{\alpha}$)	0.0674	0.0631	0.0525	0.0413	0.0277	0.0205	0.0160
SD($\hat{\beta}$)	0.6329	0.4364	0.3230	0.2296	0.1797	0.1120	0.0881

Table 7(b): $\hat{\alpha}$ and $\hat{\gamma}$ when β =0.8 (the true value of α =0.7, γ =3.0)

censoring rate	98.67%	96.67%	94.67%	86.67%	75%	50%	0%
sample size	40	100	200	400	750	1500	3000
MSE($\hat{\alpha}$)	0.0059	0.0055	0.0046	0.0035	0.0015	0.0011	0.0006
MSE($\hat{\gamma}$)	0.7509	0.5787	0.3750	0.2561	0.1409	0.0911	0.0474
SMSE	0.7568	0.5842	0.3796	0.2596	0.1424	0.0922	0.0480
BIAS($\hat{\alpha}$)	-0.0031	-0.0022	-0.0033	0.0025	-0.0012	0.0010	0.0030
BIAS($\hat{\gamma}$)	0.1193	0.0166	-0.0051	0.0437	-0.0004	0.0115	0.0363
SD($\hat{\alpha}$)	0.0771	0.0741	0.0679	0.0588	0.0382	0.0334	0.0253
SD($\hat{\gamma}$)	0.8583	0.7605	0.6123	0.5042	0.3754	0.3015	0.2146

Table 7(c): $\hat{\beta}$ and $\hat{\gamma}$ when α=1.0 (the true value of β=3.0, γ=0.7)

censoring rate	96.67%	91.67%	83.33%	75%	66.67%	50%	0%
sample size	100	250	500	750	1000	1500	3000
MSE($\hat{\beta}$)	1.4194	0.7062	0.3321	0.2884	0.1574	0.1102	0.0511
MSE($\hat{\gamma}$)	0.0903	0.0505	0.0228	0.0185	0.0125	0.0087	0.0043
SMSE	1.5097	0.7567	0.3549	0.3069	0.1699	0.1189	0.0554
BIAS($\hat{\beta}$)	0.5023	0.2170	0.1265	0.1171	0.0508	0.0450	0.0281
BIAS($\hat{\gamma}$)	-0.0173	0.0036	-0.0047	-0.0035	0.0012	0.0004	-0.0022
SD($\hat{\beta}$)	1.0803	0.8119	0.5622	0.5241	0.3935	0.3289	0.2243
SD($\hat{\gamma}$)	0.3000	0.2246	0.1511	0.1359	0.1117	0.0932	0.0652

Table 8: $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ (true value α=3.0, β= 10.0, and γ=1.0) with n=2000

censoring rate	90%	75%	50%	0%
sample size	200	500	1000	2000
MSE($\hat{\alpha}$)	0.1264	0.0725	0.0567	0.0338
MSE($\hat{\beta}$)	22.4657	11.7343	8.3414	4.3942
MSE($\hat{\gamma}$)	0.5123	0.1868	0.1178	0.0670
SMSE	23.1045	11.9937	8.5159	4.4949
BIAS($\hat{\alpha}$)	0.0573	0.0218	0.0018	-0.0012
BIAS($\hat{\beta}$)	1.2526	0.8330	0.7595	0.4696
BIAS($\hat{\gamma}$)	0.1551	0.0530	0.0138	0.0023
SD($\hat{\alpha}$)	0.3509	0.2683	0.2381	0.1837
SD($\hat{\beta}$)	4.5713	3.3227	2.7865	2.0430
SD($\hat{\gamma}$)	0.6988	0.4290	0.3430	0.2587

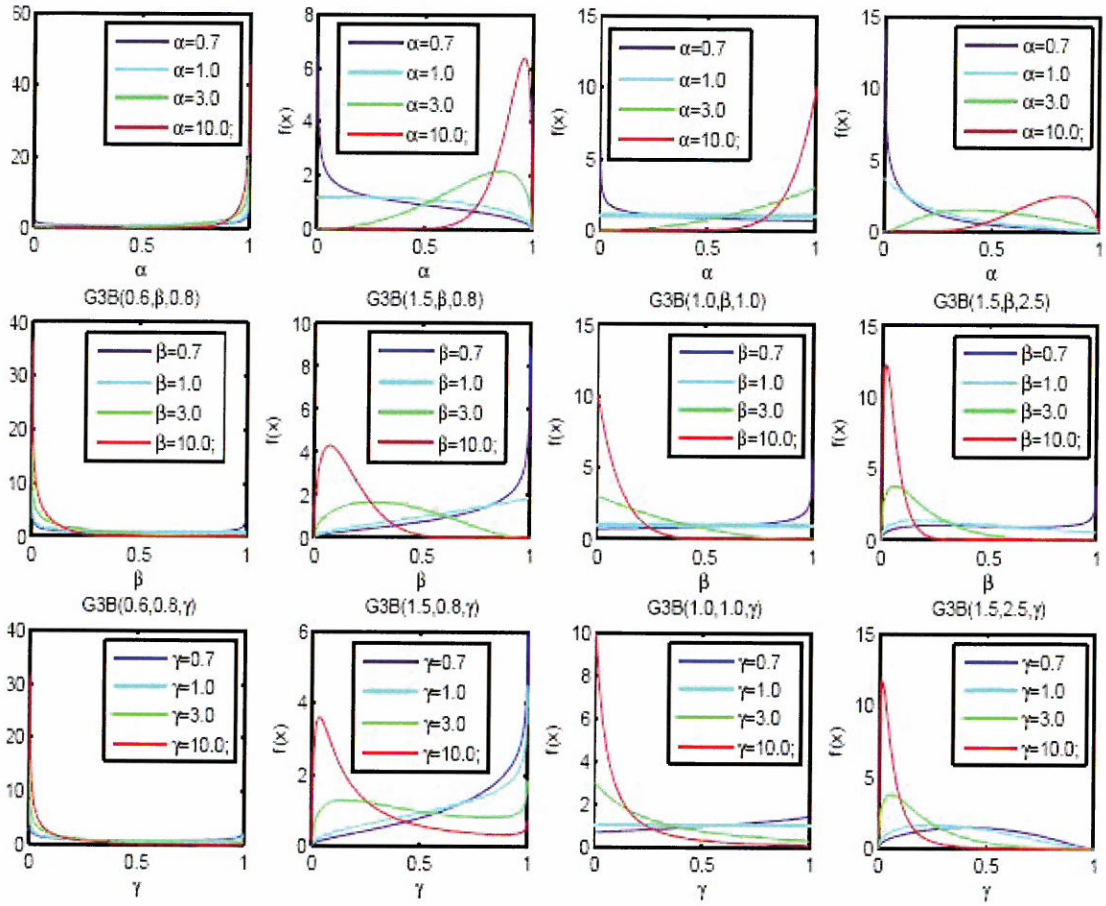


Figure 1: The pdf of G3B(α, β, γ)

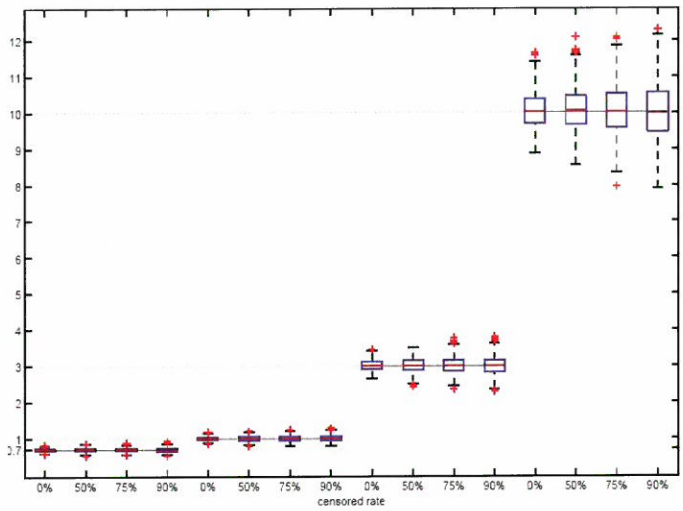


Figure 2(a): Box plot of $\hat{\alpha}$ when α=0.7, 10.0, β= 1.5, γ= 0.8 and α=1.0, 3.0, β= 1.5, γ= 2.5, the horizontal line indicate the place for the true value of α with sample size=300.

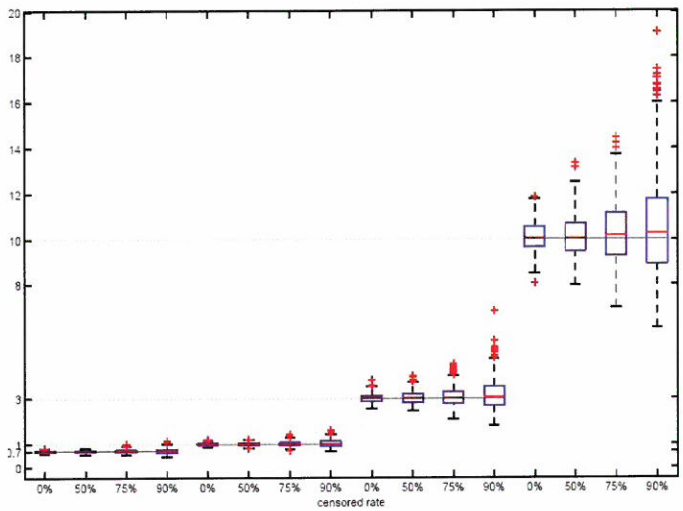


Figure2(b): Box plot of $\hat{\beta}$ when β=0.7, 1.0, α= 1.0, γ= 1.0 and β=3.0, 10.0, α= 0.6, γ= 1.5, the horizontal line indicate the place for the true value of β with sample size=400.

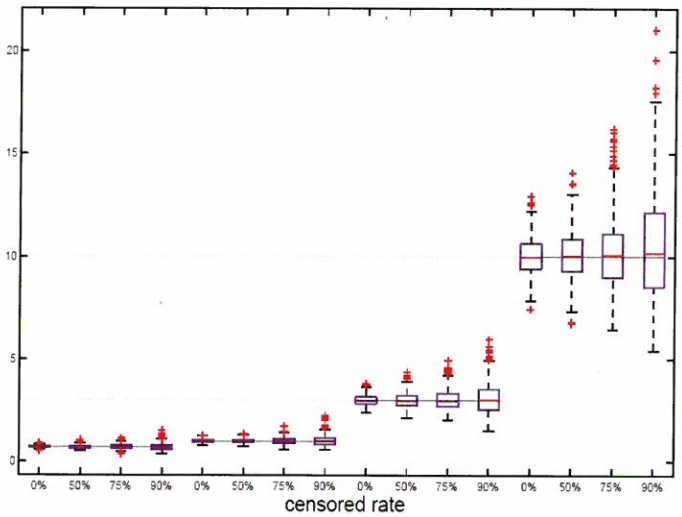


Figure 2(c): Box plot of $\hat{\gamma}$ when α = 1.0, β = 1.0, the horizontal line indicate the place for the true value of γ with sample size=400.

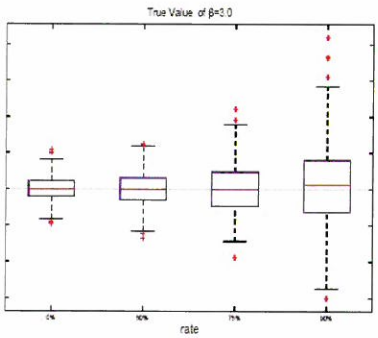


Figure 2(d): Box plot of $\hat{\beta}$ when α =2.5, γ =2.5 with sample size=1600, the horizontal line indicate the place for the true value of β true value of β

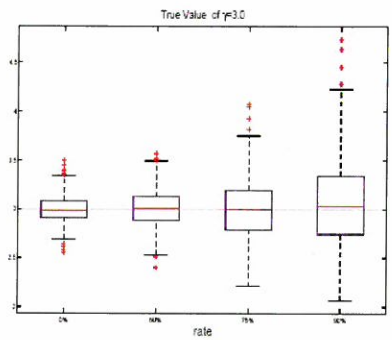
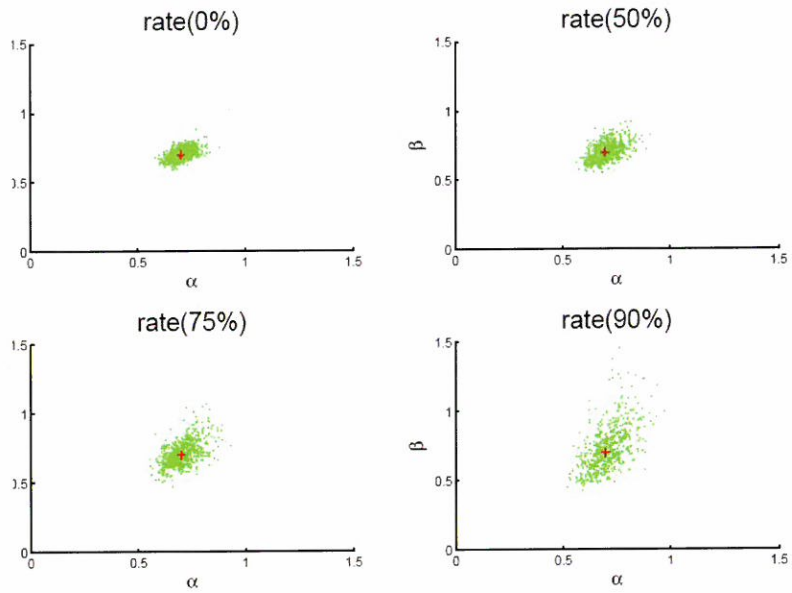
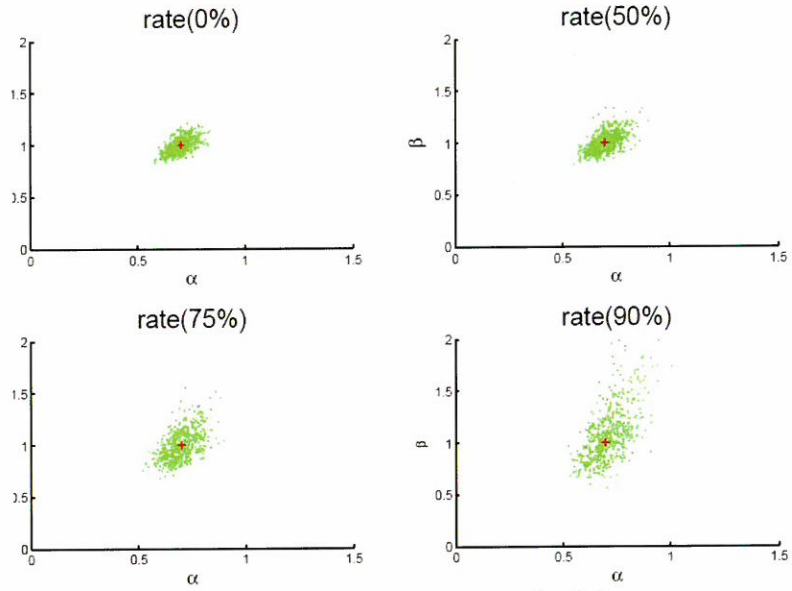


Figure 2(e): Box plot of $\hat{\gamma}$ when α = 0.6, β =1.5 with sample size=1600 , the horizontalline indicate the place for the true value of γ true value of γ

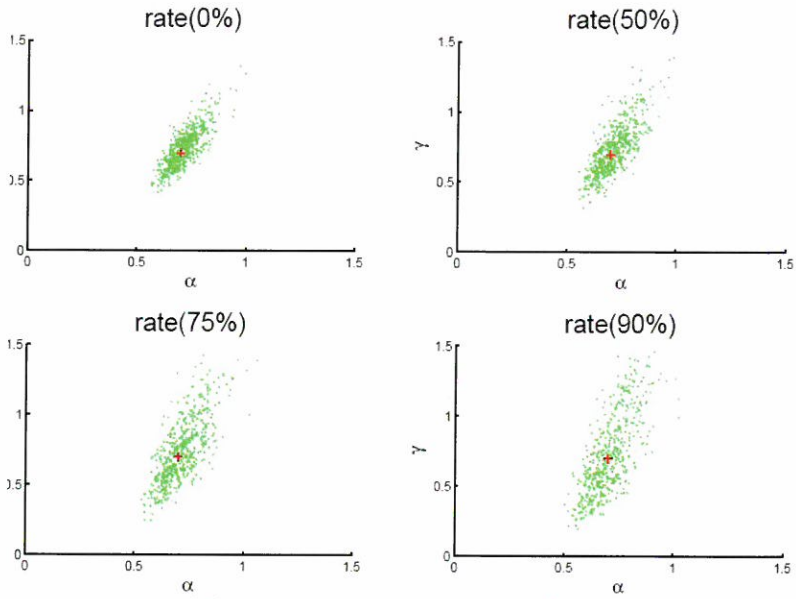


(a) true value α=0.7, β=0.7

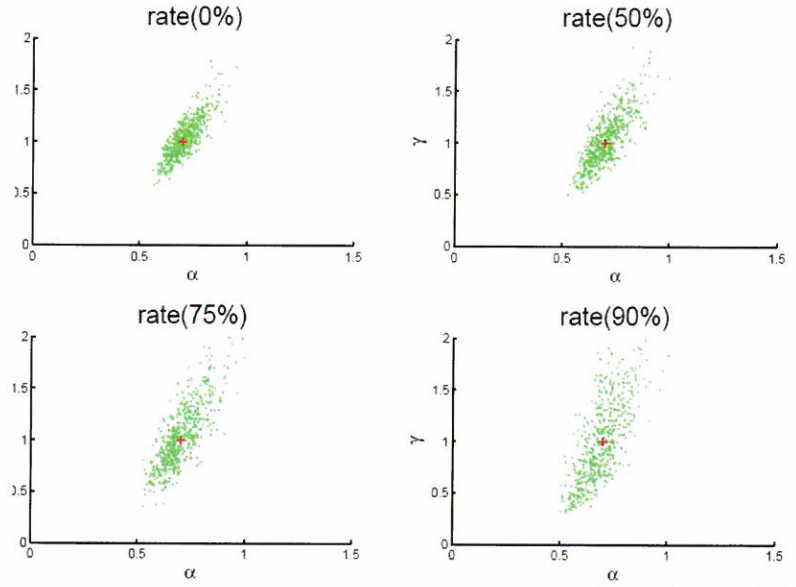


(b) true value α=0.7, β=1.0

Figure 3-1: Scatter plot of $\hat{\alpha}$ and $\hat{\beta}$ with γ=1.5 and sample size=400.

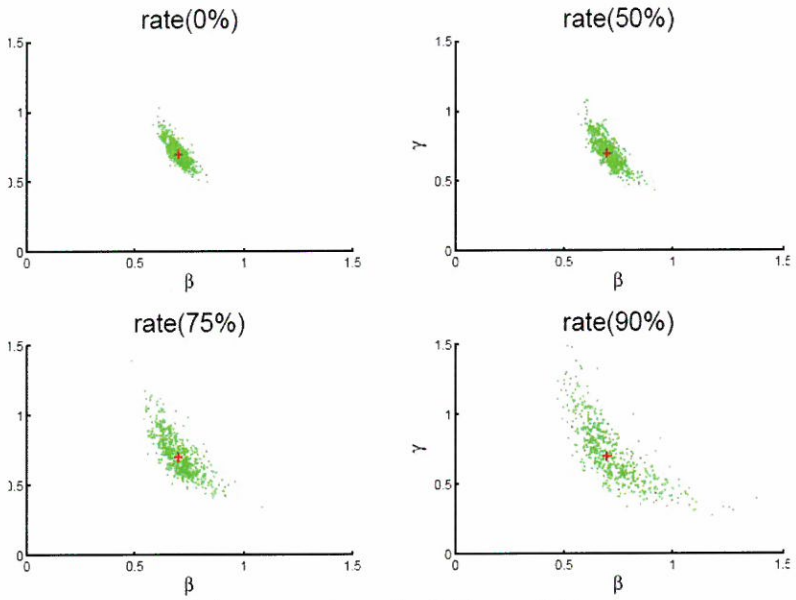


(a) true value α=0.7, γ=0.7

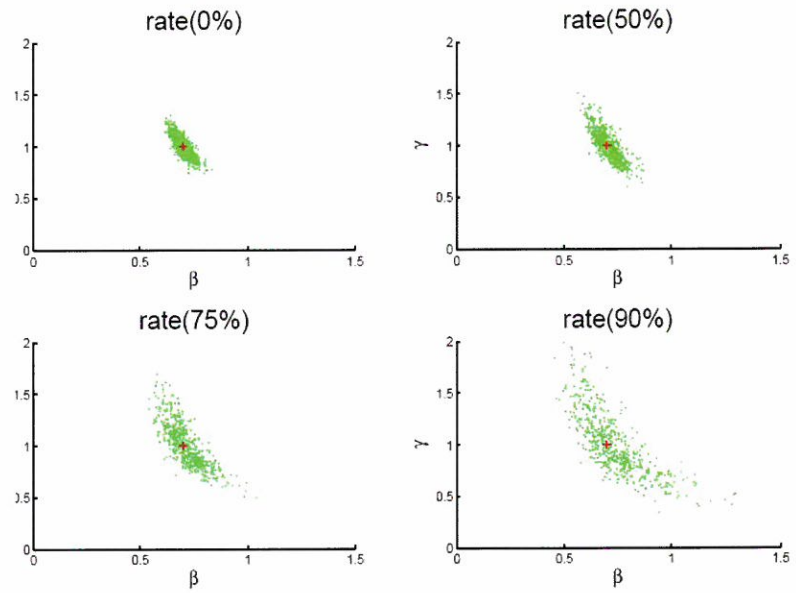


(b) true value α=0.7, γ=1.0

Figure 3-2: Scatter plot of $\hat{\alpha}$ and $\hat{\gamma}$ with β=0.8 and sample size=400.

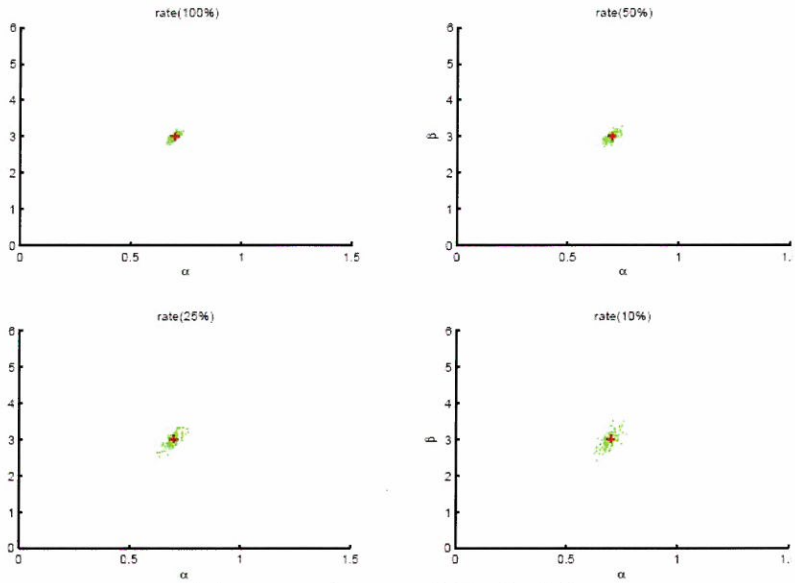


(a) true value β=0.7, γ=0.7



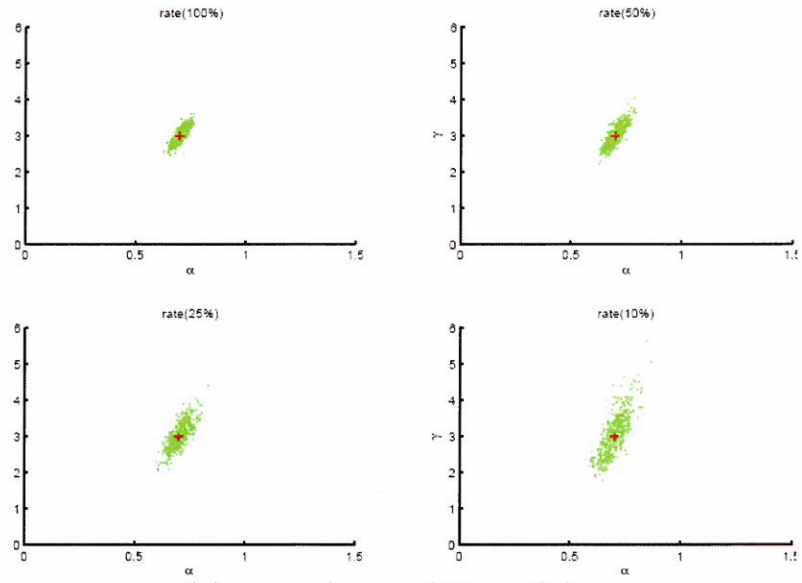
(b) true value β=0.7, γ=1.0

Figure 3-3: Scatter plot of $\hat{\beta}$ and $\hat{\gamma}$ with α=1.0 and sample size=400.



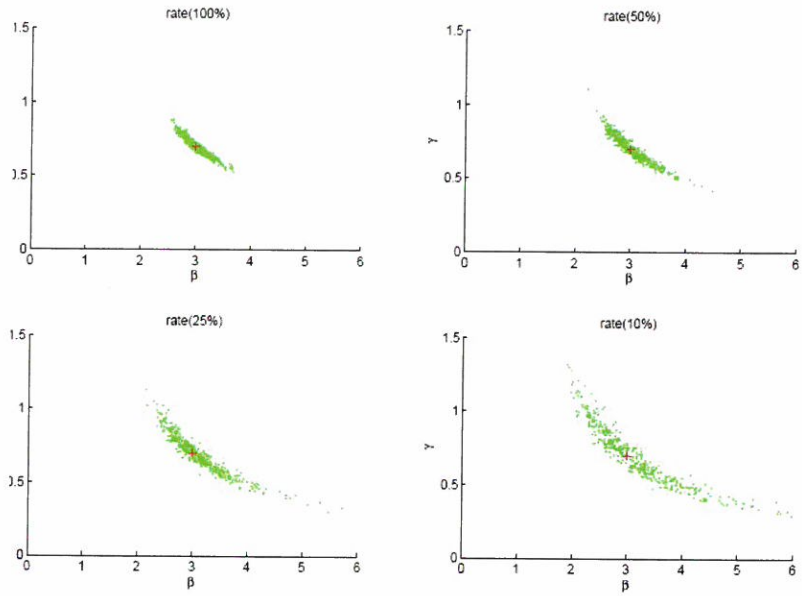
(a) true value α=0.7, β=3.0

Figure 4-1: Scatter plot of $\hat{\alpha}$ and $\hat{\beta}$ with γ=1.5 and sample size=3000.



(a) true value α=0.7, γ=3.0

Figure 4-2: Scatter plot of $\hat{\alpha}$ and $\hat{\gamma}$ with β=0.8 and sample size=3000.



(a) true value β=0.7, γ=3.0

Figure 4-3: Scatter plot of $\hat{\beta}$ and $\hat{\gamma}$ with α=1.0 and sample size=3000.

Received September 1, 2010
Revised November 18, 2010
Accepted December 8, 2010

A Simulation Study on the MLE of G3B Distribution under Progressively Type-II Censoring

陳思勉　洪郁翔

輔仁大學數學系

摘　要

本文探討廣義三參數貝他分布之參數在型二設限資料下之最大概似估計式。然因廣義三參數貝他分布之最大概似函數所得之系統方程式相當複雜，無法求得最大概似估計式之具體解，因此本文採用統計模擬及數值模擬冶煉法探討之。模擬結果顯示最大概似估計式之相對偏差值小，均方差及標準誤在樣本夠大時幾乎可忽略。

關鍵字： *Trigamma 函數；不完全貝他函數；模擬冶煉法；最大概似估計式。*

光纖中 Faraday 旋轉的測量－信號調變法

蕭光志　賴志松
輔仁大學物理學系

摘　要

光纖中 Faraday 旋轉的測量常伴隨線性雙折射的干擾，為解決此問題，本文提出一個新的方法—信號調變法，並以實驗證明其可行性。此方法包含，對信號做已知量值的小幅調變，測量調變前後的信號值，再由測量數據解出 Faraday 旋轉角與雙折射。

關鍵詞：*Faraday 旋轉、雙折射、光纖。*

1. 簡　介

Faraday 效應 [1] 是磁場在介質中感生的圓性雙折射 (circular birefringence)，在磁場中，線偏光在介質中傳輸時，其偏振面將旋轉，角度為

$$\varphi = \int V\mathbf{H} \cdot d\mathbf{l}, \tag{1}$$

其中 H 為磁場強度，V 為介質的 Verdet 常數，積分沿光行進的路徑計算。Verdet 常數主要決定於介質特性、光波長及溫度。若磁場由載流迴路所生，光路徑為圍繞電流的封閉環路，由 Ampere 環路定律可得

$$\varphi = VN_1N_2I, \tag{2}$$

式中 I 為電流，N_1、N_2 分別為電流與光環繞的圈數。

Faraday 效應可以應用於磁場或電流的測量 [2]，介質可選用塊材或光纖。本實驗採用熔矽光纖，優點包括體積小、裝置具彈性、電磁干擾極小等等，此外，反磁性介

Corresponding author: 蕭光志
E-mail:002109@mail.fju.edu.tw

質的 Verdet 常數較不受溫度變化影響。主要缺點則是，光纖的線性雙折射通常較大。光纖製程中產生的非圓柱對稱性結構導致本質雙折射，不易改變；另，彎曲、側壓、電場、溫度等等因素引起感應雙折射。經驗顯示，在未經特殊處理的光纖線圈中，線性雙折射往往遠大於 Faraday 旋轉，因而干擾後者的測量。

2. 信號與信號調變

假設線性雙折射 δ 及 Faraday 旋轉 φ 均勻分布於長度 z 的光纖中，則其光學偏振效應可以用一個等效 Jones 矩陣表示 [3]，輸出及輸入 Jones 向量的關係為

(3)
$$\begin{bmatrix} E_x(z) \\ E_y(z) \end{bmatrix} = \begin{bmatrix} A & -B \\ B & A^* \end{bmatrix} \begin{bmatrix} E_x(0) \\ E_y(0) \end{bmatrix}$$

式中

(4) $$A = \cos(\alpha/2) + i \sin(\alpha/2) \cos\chi ,$$

(5) $$B = \sin(\alpha/2) \sin\chi ,$$

(6) $$\alpha = [(2\varphi)^2 + \delta^2]^{1/2}.$$

(7) $$\tan\chi = 2\varphi/\delta .$$

輸出態受 δ 及 φ 二參數的影響，無法由單一數據直接求得 φ 值，本實驗採用調變 Faraday 旋轉的方法解決此問題。

對於小信號 ($|\varphi| \ll |\delta|$)，以 y- 方向線偏光輸入，

(8) $$\begin{bmatrix} E_x(0) \\ E_y(0) \end{bmatrix} = \begin{bmatrix} 0 \\ E_0 \end{bmatrix}$$

令

(9) $$E_1 = 2^{-1/2}(E_x + E_y),$$
$$E_2 = 2^{-1/2}(E_x - E_y),$$

定義信號為

(10) $$S = (|E_1|^2 - |E_2|^2)/(|E_1|^2 + |E_2|^2),$$
$$= (2\varphi/\alpha)\sin\alpha,$$

此信號在 φ=0 時靈敏度最高。S 對 φ 的導數

(11) $$dS/d\varphi = (2/\alpha)\sin\alpha + (2\varphi^2/\alpha^3)(\alpha\cos\alpha - \sin\alpha).$$

當 $|\varphi| \ll |\delta|$，在線性近似下，

(12) $$S = (2\varphi/\delta)\sin\delta,$$

(13) $$dS/d\varphi = (2/\delta)\sin\delta,$$

在測量信號 S 之後，對 φ 施以已知小量 $\Delta\varphi$ 的調變，再測量信號變化 ΔS，以 $\Delta S/\Delta\varphi$ 近似 $dS/d\varphi$，則 φ 及 δ 可以由 (12) 及 (13) 二式聯立求出。

3. 實　驗

圖一為實驗裝置示意圖，光源 (L) 為 HeNe 雷射，波長 632.8nm，5mW；入射光

經 Glan-Thompson 偏振器 (P) 後，用半波片 (λ /2) 調成 y- 方向的線偏光，其偏振比小於 10^{-5}，再經 10X 顯微鏡物鏡 (M) 導入光纖 (F) 中。光纖為單模熔矽光纖，纖心直徑 5 μ m，折射率 1.4627，長約 10m，繞成 12 圈半徑 35cm 的環路。

Faraday 旋轉由電流的磁場產生，光纖環上纏繞電流線圈 (I，248 圈，0-20A，DC) 及調變電流線圈 (ΔI，250 圈，0-2A，AC)。輸出光用 Wollaston 稜鏡 (W) 檢偏，偏振比小於 10^{-5}；光感測器用矽光二極體，接示波器、鎖相放大器及電腦 (此 3 項未顯示於圖上) 讀取、紀錄並計算信號。實驗過程中保持室溫 25℃，並儘量避免光纖受到擾動。

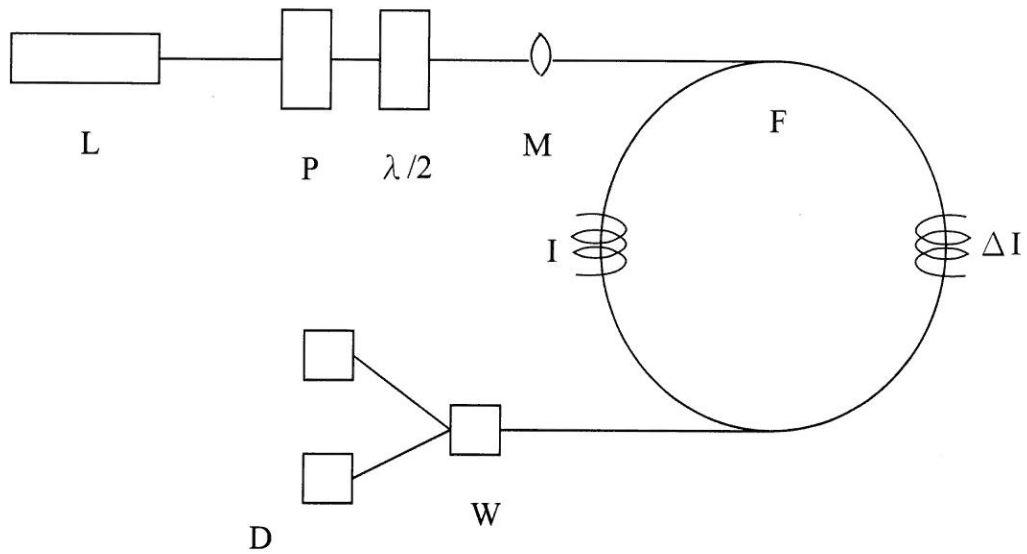


圖 1. 實驗裝置示意圖
L—光源，P—偏振器，λ/2—半波片，
M—顯微物鏡，F—光纖，I—電流線圈，
ΔI—調變電流，W—Wollaston 稜鏡，
D—光感測器；詳見本文。

4. 結果及討論

Faraday 旋轉用 1~16A 的直流電流產生，測量信號 S，平均值如 表 1，信號調變用 1A 交流電流，信號調變之測量平均值為

(14) $\Delta S = 1.12\text{x}10^{-2}$,

以 V=4.68x10-6A-1[4] 計算，

(15) $\Delta \varphi = 1.40\text{x}10^{-2}$,

(16) $\Delta \varphi / \Delta S = 1.25$,

由 (13) 式求出的雙折射值為

(17) $\delta = 2.125 = 122^{\circ}$,

帶入 (12) 式計算得到 Faraday 旋轉角如表 1 及圖 2。

表 1. 信號測量值及 Faraday 旋轉角

I(A)	S(rad)	φ (rad)	I(A)	S(rad)	φ (rad)
1.0	0.0128	0.0160	9.0	0.106	0.132
2.0	0.0225	0.0282	10	0.120	0.150
3.0	0.0400	0.0424	11	0.131	0.163
4.0	0.0458	0.0571	12	0.144	0.179
5.0	0.0580	0.0725	13	0.157	0.196
6.0	0.0682	0.0853	14	0.169	0.212
7.0	0.0809	0.101	15	0.182	0.228
8.0	0.0935	0.117	16	0.195	0.244

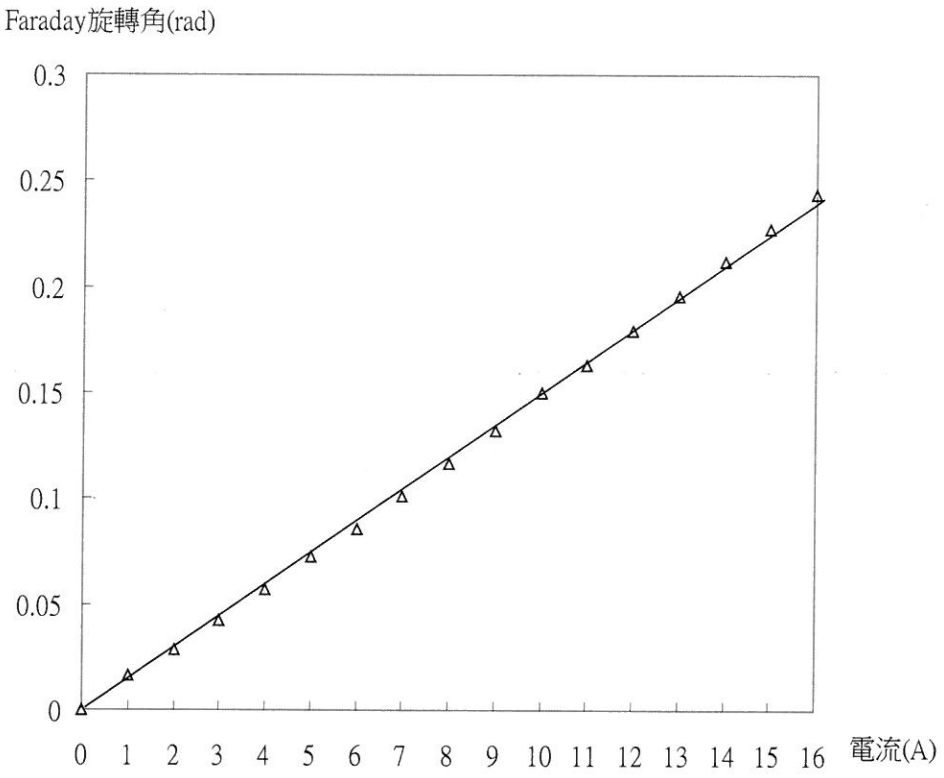


圖 2. Faraday 旋轉角

結果顯示 Faraday 旋轉角與電流接近線性關係，表示在所測量的電流範圍內，(12) 式是合理的近似。

由 (10) 式可以看出，當 α (或線性近似下 δ) 值接近 π 的正整數倍時，測量靈敏度趨於 0，實驗中必須避免此一狀況。雙折射 δ 受光纖長度、光纖環半徑等等多項因素影響，極易改變，故前述要求並不對測量造成實質性的限制。

5. 結　論

由於一般物質的 Verdet 常數值較小，Faraday 旋轉的測量有時並不容易，尤其用光

纖導光時，彎曲等等因素所生的雙折射效應，經常遠大於 Faraday 旋轉，而使直接量測不可能。本文所提的信號調變法，概念簡單，操作容易，適合於雙折射干擾之下，測量 Faraday 旋轉。在某些應用，如大電流的測量，極具潛力。

參考文獻

1. E. Hecht, Optics, 2e., Addison-Wesley, Reading, MA, USA(1987).
2. 例如 A. M. Smith, Appl. Opt., 17, 52(1978); A. J. Rogers, Intern. J. Optoelectronics, 3, 391(1988); R. C. Gauthier, Elect. Lett., 28, 2167(1992).
3. W. J. Tabor and F. S. Chen, Appl. Phys., 140, 276(1969).
4. R. Ulrich and H. Harms, Appl. Opt., 19,3729(1980).

Received September 2, 2010
Revised November 18, 2010
Accepted December 9, 2010

Measurement of Faraday Rotation in an Optical Fiber : Method of Signal Modulation

K. C. Hsiao and C. S. Lai

Department of Physics, Fu Jen Catholic University

Abstract

Measurement of Faraday rotation in an optical fiber is always interfered by the existence of linear birefringence. To solve this problem, the authors proposed and showed experimentally a new method, signal modulation. The signal is modulated with a known small value. Both the original and modulated signals are measured. Faraday rotation and linear birefringence can then be calculated from the measured data.

Keywords: Faraday rotation, linear birefringence, optical fiber.

A Note on All-Ones Problem

Hong-Min Shaw

Department of Mathematics Fu Jen Catholic University

Abstract

K. Sutner introduced the All-Ones Problem and showed that a solution for the All-Ones Problem exists to any undirected graph. The study of this and its related problems intermingles the realms of graph theory and matrix theory. It is known that the number of solutions for the All-Ones Problem to any undirected graph of n vertices would be 2^m for some m, $0 \leqq m \leqq (n\text{-}1)$. Undirected graphs whose number of solutions is 2^{n-1} (or 2^{n-2}) are classified.

Keywords: All-Ones Problem, odd-parity cover.

1. Introduction

In this note a graph always means a finite simple undirected graph whose vertices are labeled by $\{1, 2, \ldots, n\}$. Here n is the number of vertices and simple means no loops and no multiple edges. Consider that each vertex of a graph is equipped with a light bulb, and the state of a vertex is either 0 or 1 depending on that equipped (light) bulb is off or on respectively. Each time when a bulb (vertex) is touched, the state of the corresponding vertex and each of its neighbors will be inverted. For instance in the path $P_5 = (1,2,3,4,5)$ with states (0,1,1,0,1), a touch of vertex (bulb) *3* will invert the states of vertices *2,3,4*; thus the states of P_5 becomes (0,0,0,1,1). Note that (0,0,0,1,1)=(0,1,1,0,1)+(0,1,1,1,0) over the Galois field Z_2. K. Sutner [4] introduced the All-Ones Problem asking that to a given graph with states of all vertices being 0, is it possible to touch a sequence of vertices so that in the end the states of all vertices being 1?

Corresponding author: Tel:02-2905-3546 Fax:02-29044509
E-mail:hmshaw@math.fju.edu.tw

Clearly when a vertex is touched twice, the state of every vertex remains unchanged. Thus a solution of the All-Ones Problem, if exists, can be represented as the characteristic vector of some subset of $\{1, 2, \ldots, n\}$. This subset Q, as observed in [4], satisfies that for any vertex v, $|N[v] \cap Q|$ is odd, where $N[v]$ is the closed neighborhood of v. Such Q is called an odd-parity cover of G. For instance both $\{1, 4\}$ and $\{2, 5\}$ are odd-parity covers of $P_5 = (1,2,3,4,5)$. In any complete graph K_n, notice $N[v] = \{1, 2, \ldots, n\}$ for any vertex v. Hence Q is an odd-parity cover of K_n if and only if Q contains odd number of vertices. So the number of solutions for the All-Ones Problem to the complete graph K_n is 2^{n-1}. Using matrix theory, Sutner [4] proved that the All-Ones Problem is solvable to any graph. He stated it in terms of graphs as follows.

Theorem 1 (Sutner: graph version) *Every graph G has an odd-parity cover.* á

In [2] a similar proof to a more general result was provided by Lossers. Gallai (see [1]) proved that for any graph G, there is a partition $\{X, Y\}$ of $V(G)$ so that in both induced subgraphs $G[X]$ and $G[Y]$ all degrees are even; $\{X, Y\}$ is called a Gallai's partition. For example $\{\{1,3,5\}, \{2,4\}\}$ is a Gallai's partition of $P_5 = (1,2,3,4,5)$. We refer to the book of Lovász [3] for a proof. Caro [1] established that for any graph the existence of an odd parity cover and that of a Gallai's partition are equivalent. Therefore Sutner's theorem is a consequence of Gallai's theorem (and vice versa). Caro [1] also gave a proof using linear algebra.

All these proofs in terms of matrix are based on the following simple observation by Sutner [4]. First we identify each vertex of a labeled graph G with its light bulb. Next we name a binary (column) vector indicating the on-off status of the vertices a state vector of G. Let A be the adjacency matrix of the graph G (labeled by $\{1, 2, \ldots, n\}$) and q be the characteristic (column) vector of a subset Q of $V(G)$. Observe that if initially the state vector of G is a, then $a + (A + I)q$ is the state vector obtained after all vertices of Q are touched. Denote the zero (column) vector by $\mathbf{0}$ and the all-one (column) vector by $\mathbf{1}$. Then Sutner's thoerem can be restated as.

Theorem 2 (Sutner: matrix version) *Let A be the adjacency matrix of a graph G labeled by $\{1, 2, \ldots, n\}$ and I be the identity matrix. The matrix equation $(A+I)\,x = \mathbf{1}$ has a*

solution over the Galois field Z_2. In other words, $\mathbf{1} \in$ im(A+I), the image (or column space) of A+I.

For example the adjacency matrix A for the path $P_5 = (1,2,3,4,5)$ is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$. Solve $$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$ $x = \mathbf{1}$, we get all solutions

$x = (1,0,0,1,0)^T$, $(0,1,0,0,1)^T$ (T stands for transpose). Thus $\{1, 4\}$ and $\{2, 5\}$ are all odd-parity covers of P_5.

The following concept seems to be a quite natural extension. Two state vectors a, b of a graph G are said to be equivalent, denoted by $a \sim b$, if the status of G becomes b after certain vertices of G (with the state vector a) are touched. With this notion, the third version to state Sutner's theorem is

Theorem 3 (Sutner: algebra version) *For any labeled graph, $\mathbf{0} \sim \mathbf{1}$.*

Due to the earlier observation (by Sutner), $a \sim b$ means $a+(A+I)q=b$ for some q, i.e., $b-a(=b+a) \in$ im$(A+I)$. Since im$(A+I)$ is a subspace of Z_2^n, it is a routine to verify the following proposition.

Proposition 4 *Let $\sim$ be the binary relation defined on the state vectors.*
(1) $\sim$ is an equivalence relation.
(2) If $a \sim b$ and $c \sim d$, then $(a+c) \sim (b+d)$.

To any binary vector a, let $\bar{a}$ be the binary vector obtained by inverting every component of a. Namely $\bar{a} = a+\mathbf{1}$ under the arithmatic of Z_2. Sometimes we say a and $\bar{a}$ are complementary (to each other). By Sutner's theorem $\mathbf{0} \sim \mathbf{1}$, adding a to both sides results in

Corollary 5 *For any labeled graph, $a \sim \bar{a}$.*

We have discussed in several ways the existence of a solution to the All-One Problem.

In fact the number of solutions can also be determined by finding the rank of a matrix. Notice that the equivalence class $[\mathbf{0}]$ is equal to the subspace im$(A+I)$. Under the vector addition, $[\mathbf{0}]$ is a subgroup of the abelian group Z_2^n. So each coset of $[\mathbf{0}]$ is an equivalence class under $\sim$. Thus the size of all equivalence classes are the same. In fact for any state vector a, $|[a]|=|[\mathbf{0}]|=2^r$, here r is the rank of $A+I$. If r is determined and $a\sim b$, then the number of solutions to transit a into b is 2^{n-r}.

It seems quite difficult to classify those graphs whose All-Ones Problem has a unique solution. In next section we classify graphs whose number of solutions to the All-One Problem is quite a bunch, 2^{n-1} or 2^{n-2}.

2.Results

Let G be a graph with $V(G) = \{1, 2, \ldots, n\}$, $A(G)$ be its adjacency matrix and I be the identity matrix. Given a positive integer r, the object is to find all graphs G so that the associated matrix $A(G) + I$ has rank r.

For convenience let $B(G) = A(G) + I$ and r denote the rank of $B(G)$. Often G is omitted, i.e., $B = B(G)$. Let $B_{\cdot j}$ denote the j-th column and $B_{i\cdot}$ denote the i-th row of B. Note that B is a symmetric (0,1) -matrix with the main diagonal $\mathbf{1}$. Thus B contains no zero vector and so $r \geqq 1$. A basic fact in linear algebra is that any set of maximal linearly indepndent columns in B has r columns, which from a basis for the column subspace im(B). Hence by Sutner's theorem, $\mathbf{1}$ is a linear combination of these columns. Over Z_2, it means $\mathbf{1}$ is a sum of some of theses columns.

When $r=1$, each equivalence class $[a]$ contains exactly two complementary state vectors a and $\bar{a}$. This class contains only complete graphs.

Theorem 6 *Let G be a graph. The rank of B(G) is 1 if and only if G is a complete graph.*

Proof. Let J be the all-one matrix. G is the complete graph K_n if and only if the adjacency

matrix of G is J-I if and only if $B(G)=J$. The rank of J is clearly 1.

Conversely if the rank of B is 1, all nonzero columns of B are the same, say a. Then for all i =1,⋯,n, a_i = 1 because B_{ii} = 1. So a =$\mathbf{1}$. Since B has no zero vector, B = J. QED

When r =2, $[a]$ = $\{a, a+b, a+c, a+b+c\}$, where im(B)=span$\{b,c\}$. This class contains only disjoint union of two complete graphs. To prove this, a simple and useful lemma is introduced.

Lemma 7 *Let* 0_a (1_a) *denote the set of indices i such that* a_i=*0* (a_i=*1*). *Suppose* $\mathrm{U}_{j=1}^{n}\ 0_{B.J}$ = *{1, ..., n}* . *Then* $\mathbf{1}$ *will not appear as a column in B.*
In particular $\mathbf{1}$ *will not appear as a column in B, when B contains two complementary columns.*

Proof. If $B_{.j}$ = $\mathbf{1}$, then $B_{.j}$ = $\mathbf{1}^T$. Since $j\in$ {1, ⋯, n}= $\mathrm{U}_{i=1}\ 0_{B.i}$, $j\in 0_{B.k}$ for some k. Thus B_{jk} = 0. But B_{jp} = 1, for all p, a contradiction.
QED

Theorem 8 *Let G be a graph. The rank of B(G) is 2 if and only if G is a disjoint union of two cliques.*

Proof. Suppose G is disconnected with components G_1,⋯,G_p. Let B_i = $A(G_i)$+I, i=1,...,p. Note that the order of the identity matrix I here is the same as the order of the component G_i. Then the matrix

$$B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_p \end{pmatrix}$$

has rank r_1 + ⋯ + r_p , where r_i is the rank of B_i. Hence when G is a disjoint union of two cliques, the rank of B is 1+1=2 by Theorem 6.

Conversely let $\{u,v\}$ be a set of maximal linearly independent columns of B.

Then $\mathrm{im}(B) = \{\boldsymbol{0},u,v,w\}$, where $u \neq v$ and $w = u+v$. By Sutner's theorem, $\boldsymbol{1} \in \{u,v,w\}$.
Since $w+u = u+v+u = v$, $v+w = v+u+v = u$, we may assume $\boldsymbol{1}=w$ without loss of generality.
Thus u,v are complementary. Namely $0_u = 1_v$ and $1_u = 0_v$. By the above lemma, $\boldsymbol{1}$ is not a column in B. Thus each column of B is either u or v. Notice that 1_u and 1_v form a partition of $V(G) = \{1,\ldots,n\}$. Since the main diagonal is $\boldsymbol{1}$, the subgraphs induced by 1_u and by 1_v are both complete graphs. Since $1_u = 0_v$ and $0_u = 1_v$, no edge joins these two complete graphs. Therefore G is a disjoint union of two cliques.
QED

References

1. Caro, Y. Simple proofs to three parity theorems, Ars Combinatoria, Vol. 42, pp. 175-180 (1996).
2. Lossers, O.P. Solution to problem 10197, The American Mathematical Monthly, Vol. 100, October, pp. 806-807 (1993).
3. Lovász, L. Combinatorial problems and exercises, North-Holland, Amsterdam. (1979).
4. Sutner, K. Linear cellular automata and the Garden-of-Eden, The Mathematical Intelligencer, Vol. 11, No.2, pp. 49-53 (1989).

Received September 14, 2010
Revised November 16, 2010
Accepted December 12, 2010

全一問題札記

蕭鴻銘

輔仁大學 數學系

摘　要

K. Sutner 介紹了全一問題並證明了對任意圖而言此問題均有解。這方面的研究混合了圖論與矩陣理論。已知對任意一個 n 點圖，全一問題的解數必為 2 的一個次方 2^m $(0 \leqq m \leqq n\text{-}1)$。本文刻劃所有解數為 2^{n-1} 的圖以及所有解數為 2^{n-2} 的圖。

關鍵字：全一問題 , odd-parity cover。

混層複合材料機械手臂牙叉之撓曲分析

蔡昆恊[1] 黃宗立[2] 曾淯瑄[3] 黃鈺珊[3]
邱長塤[1] 吳昌謀[1]

[1] 逢甲大學纖維與複合材料學系 [2] 逢甲大學機械與電腦輔助工程學系
[3] 逢甲大學紡織工程研究所 台中市文華路 100 號

摘 要

本文利用兩種趨近方式分析混層複合材料機械手臂牙叉之撓曲位移，此混層複合材料之堆疊順序為 [W/±45°/ 0°/±45°/ W]，其中 W 代表梭織物。首先，利用混合法則與勁度平均法求得梭織物與各層單方向複材之工程模數，另外利用勁度平均法將上述混層複材之工程模數於厚度方向均質化。其次，將機械手臂牙叉視為一懸臂箱型樑，分別將各層與均質化後之工程模數輸入有限元素 ANSYS 軟體，建構在厚度方向分別為五層與單層之兩個有限元素分析模型，然後分別計算其撓曲位移。經比較驗證，分析求得撓曲位移與實際量測結果有相當好的關聯。值得注意的是使用單層法分析並作適度修正，不僅可以可提供合理的撓曲位移同時也能節省計算時間。此外，開孔會增加混層複材箱型樑的撓曲位移，且開孔效應距離固定端越近越明顯。

關鍵字：混層複合材料、機械手臂、箱型樑、撓曲

Corresponding author:Tel:(04)2451-7250 Ext:3442 Fax:(04)2451-4625
E-mail:khtsai@fcu.edu.tw

1. 前　言

近年來，薄膜電晶體液晶顯示器 (Thin Film Transistor Liquid Crystal Display, TFT-LCD) 的發展相當迅速，其尺寸隨著消費者的要求也逐漸增加。然而，大尺寸玻璃面板不容易用人工搬運，為了防止粉塵汙染作業環境且增加面板的品質和生產率，製造商乃利用機械手臂取代人力來搬運面板。

早期工業用機械手臂的結構是以鋁合金製成的實心方型樑為主，為了使機械手臂能夠承受更大的負荷，通常改變其厚度來增加它的剛性。但近年來隨著液晶顯示器工業的迅速發展，面板尺寸越來越大，為了支撐與搬運大尺寸玻璃面板，機械手臂需要進一步使用質輕且剛性大之材料。此外，機械手臂在搬運玻璃面板時，必需承受週期性負載，基於安全要求，機械手臂結構也需要有優良之耐疲勞性能。在輕量化、高剛性、高強度及耐疲勞性能等多重因素考量下，目前碳纖維補強複合材料機械手臂遂成為各大相關業者的研發重點。

複合材料結構件的性能探討可採用傳統實驗或電腦模擬分析。由於複合材料的設計自由度高使得其結構件之型態種類繁多，若以傳統實驗將每一種 結構件製成試片再進行測試，不僅費時費力且會增加許多成本；而利用電腦程式進行材料性質的預測及複合材料工件的結構分析，則不僅可減少試件製作的成本浪費，更重要的是可以降低設計的時間。因此，目前相關業者均相當重視複合材料結構件之電腦模擬技術的開發，以彌補傳統經驗及技術的不足。

近二十年，許多學者致力於複合材料機械手臂及箱型樑之運動與變形行為的相關研究。1988 年 F. Gordaninejad[1] 等人探討積層板複合材料機械手臂的動態性能與結構設計。1990 年 Chandra[2] 等人設計、製造並量測石墨纖維 / 環氧樹脂複合材料箱型樑受彎曲、扭曲及拉伸作用力時之變形量，結果與簡單樑理論吻合。1991 年 Lee[3] 設計並製造碳纖 / 環氧樹脂複合材料的 SCARA 型的筆直驅動機械手臂，另外也製作相同尺寸的鋁合金機械手臂，研究結果證實，複合材料機械手臂的靜態和動態性質都較鋁合金機械手臂佳。同年，Smith[4] 等人提出複合材料箱型樑之理論分析方法以預測其勁度與變形。1994 年 Seong Min Jeon[5] 等人利用樑的大撓曲變形理論研究複合材料料箱型樑的靜態和動態行為，文中亦討論到纖維方向與疊層順序對箱型樑靜態變形與振動行為之影響。2007 年 Wu[6] 等人探討複合材料箱型樑受彎曲載重時的應力與

撓曲位移分佈。同年，Suresh[7] 等人利用粒子群最佳演算法 (PSO)，設計直升機旋翼中之複材箱型樑，主要目標為尋找結構的最佳參數和疊層順序，使其可符合剛性的要求，並使箱型樑能達到最大的比勁度值和彈性穩定度。除上述研究之外，許多學者進一步將有限元素分析與最佳化程序用於複合材料機械手臂之設計。1998 年 Manet[8] 利用 ANSYS 分析三明治複合材料樑受均佈載重之位移與應力。1999 年 Oh 等人 [9] 使用碳纖維 / 環氧樹脂複合材料的蜂巢三明治結構製作機械手臂牙叉，並且利用有限元素分析與最佳化程序來設計與分析，結果顯示此複合材料牙叉不僅比鋁製的重量減少了 50% 而且靜態剛性也提升了 95%。2002 年 Lee 等人 [10] 利用碳纖 / 環氧樹脂、玻纖 / 環氧樹脂複合材料和 PU 發泡材組成三明治結構製作雙臂型機械手臂之腕夾，並利用有限元素法與最佳化程序決定複合材料之厚度和疊層角度，分析結果顯示複合材料三明治結構比鋁製的腕夾重量減少了 50%，且在靜態與動態測試中其性能亦比鋁製的好。2004 年 Jung 等人 [11] 研發一面板缺陷偵測儀器，其樑結構由高模數碳纖維 / 環氧樹脂複合材料與鋁混製而成，藉以提升其勁度和阻尼，而此結構的動態性質是利用有限元素法分析，計算其彎曲自然頻率和最大下沉量。同年，Lee 等人 [12] 利用一體成型方法製造搬運 LCD 玻璃面板的機械手臂，其結構為三明治發泡材複合材料，並利用有限元素分析及最佳化程序來設計三明治複合材料的手臂結構。由上述文獻回顧，可知使用複合材料製作之機械手臂來搬運大型 LCD 面板目前已成為相關產業的趨勢。

由於目前搬運大型 LCD 面板之機械手臂牙叉之主要結構為梭織物與單方向複合材料混層之箱型樑，且各層使用不同噸數碳纖維所組成，這種梭織物與單方向複材混層結構無法直接使用現有以積層板理論為基礎之有限元素軟體作分析。因此本研究之目的乃在於利用勁度平均法及均質化原理發展出相關分析技術，以探討混層複合材料牙叉承受載重時之撓曲位移。

2. 相關背景及理論

2.1 機械手臂簡介

如圖一所示，搬運大型 LCD 面板之機械手臂的組成主要是由腕夾 (Wrist) 做為固定與支撐，以箱型樑深入卡匣或機台內取 / 放玻璃基板。在移動的過程中，藉由機械手臂上的真空吸盤讓玻璃基板與箱型樑之間不會有相對滑動，待機械手臂到達卡匣肋 (Rib) 或製程機台之確定位置時，再開啟或解除真空吸盤做取 / 放之動作。當箱型樑承

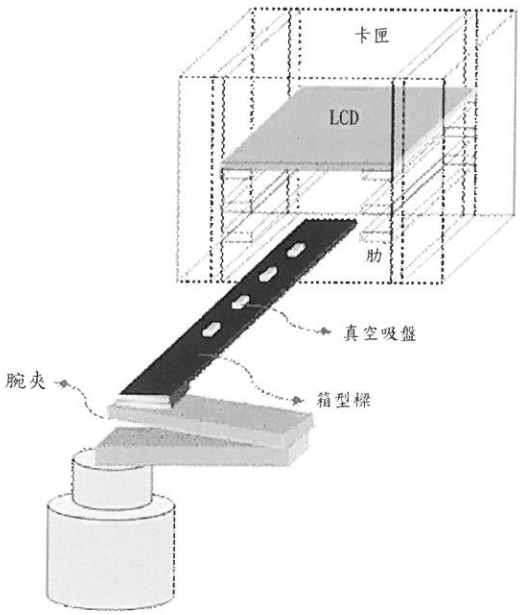
載玻璃基板時產生之撓曲位移為 δ total，包含了箱型樑本身重量的撓曲位移為 δ box 與承受玻璃基板重量的撓曲位移 δ LCD。箱型樑之截面形狀尺寸與疊層結構如圖二所示，箱型樑高度為 14.1mm、寬度為 50mm、上下厚度分別為 2.87mm 與 3.4mm，箱型樑由 [W/±45°/ 0°/±45°/ W] 堆疊順序所組成，其中 W 代表梭織物。

2.2 箱型樑複合材料工程模數之計算

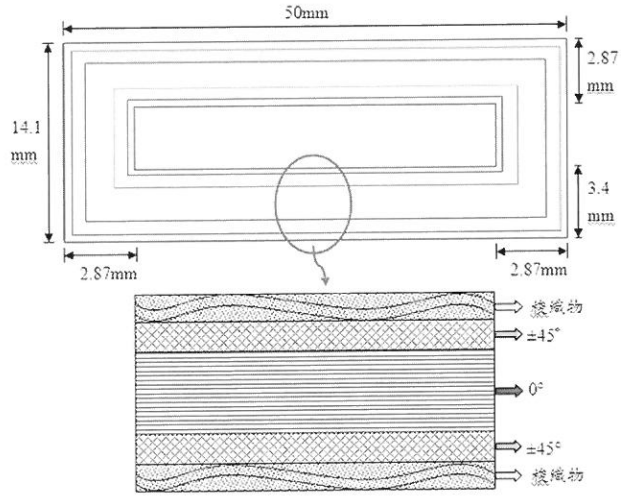
計算箱型樑複合材料之工程模數需先個別計算梭織物及單方向複材之工程模數，然後將之輸入有限元素模型進行分析。其中計算梭織物之工程模數時需先設定單位格子。而後續建構單層模型時，需使用均質化原理求出箱型樑複合材料之等效工程模數。

2.2.1 梭織物單位格子的設定

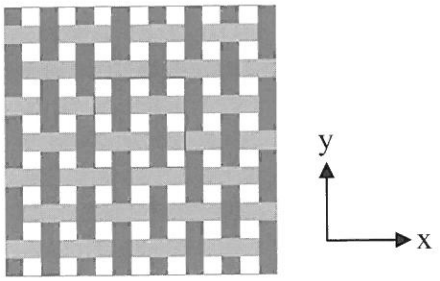
如圖三所示，x 軸及 y 軸分別代表梭織物之經向與緯向，為了求得此梭織物複合材料之工程模數，必須先決定能代表此材料性質的單位格子，如圖四所示。設定單位格子後，為了表示紗束波動情形，將紗束視為正弦波，同時加上單位格子的高 (H) 及長 (L) 以描述單位格子。



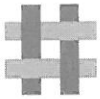
圖一：機械手臂及卡匣組成示意圖



圖二：箱型樑剖面及截面示意圖



圖三：梭織物複合材料示意圖



圖四：單位格子示意圖

2.2.2 單方向複合材料與梭織紗束之工程模數

利用混合法則可計算出單方向複合材料與單位格子內之梭織紗束之工程模數：

$$E_1 = E_{f1}V_{Yf} + E_mV_{Ym} \tag{1}$$

$$\frac{1}{E_2} = \frac{V_{Yf}}{E_{f2}} + \frac{V_{Ym}}{E_m} \tag{2}$$

$$\nu_{12} = V_{Yf}\nu_{f12} + V_{Ym}\nu_m \tag{3}$$

$$\frac{1}{G_{12}} = \frac{V_{Yf}}{G_{f12}} + \frac{V_{Ym}}{G_m} \tag{4}$$

其中 E_{f1}、E_{f2}、E_m 分別代表纖維之軸向、側向和樹脂之楊式模數，G_{f12}、G_m 分別為纖維和樹脂之剪切模數，V_{Yf}、V_{Ym} 為紗束內纖維與樹脂之體積含有率，ν_{f12}、ν_m 為纖維和樹脂之柏松比。

2.2.3 勁度平均法

本文利用勁度平均法 (stiffness averaging method) 來計算箱型樑中梭織物複合材料之工程模數，將單位格子內正弦波紗束分成許多直線段及樹脂，這些直線紗束段及樹脂可視為單位格子的不同組成。再假設單位格子有等應變 (iso-strain) 變形。經推導後，梭織物複合材料勁度矩陣等於將每段組成之勁度矩陣乘上該段之體積含有率後相加，如式 (5) 所示：

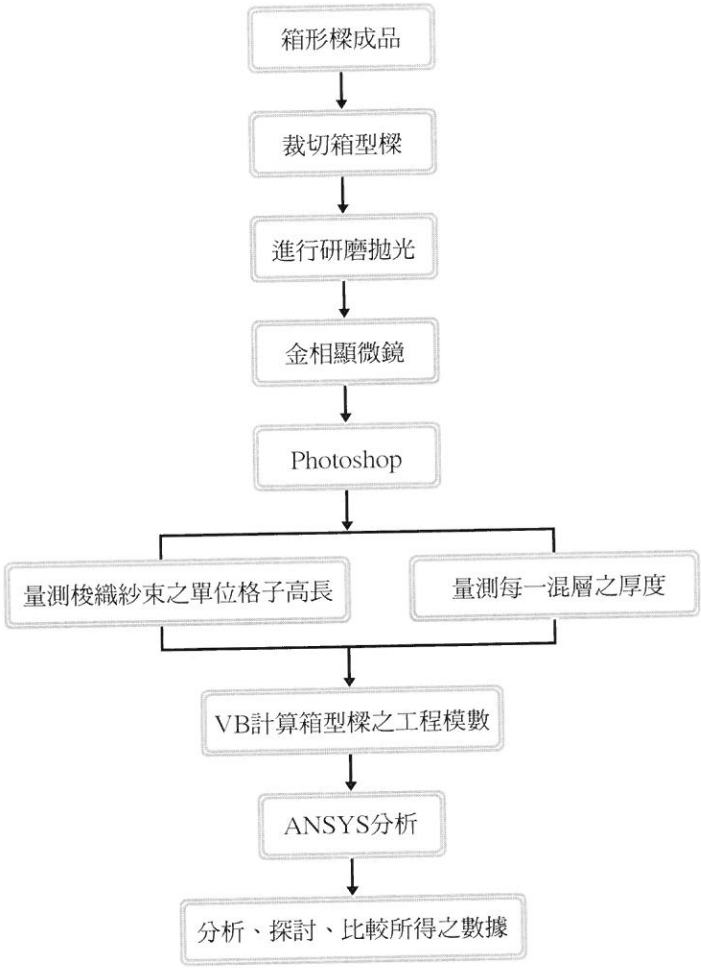
$$C_t = \sum_{i=1}^{n} W_i C_g^i \tag{5}$$

其中 C_t 為梭織物之勁度矩陣，W_i 為第 i 段組成之體積含有率，C_g^i 為第 i 段組成之勁度矩陣。

3. 研究流程與有限元素分析

3.1 研究流程

本研究之實驗與分析流程如圖五所示。

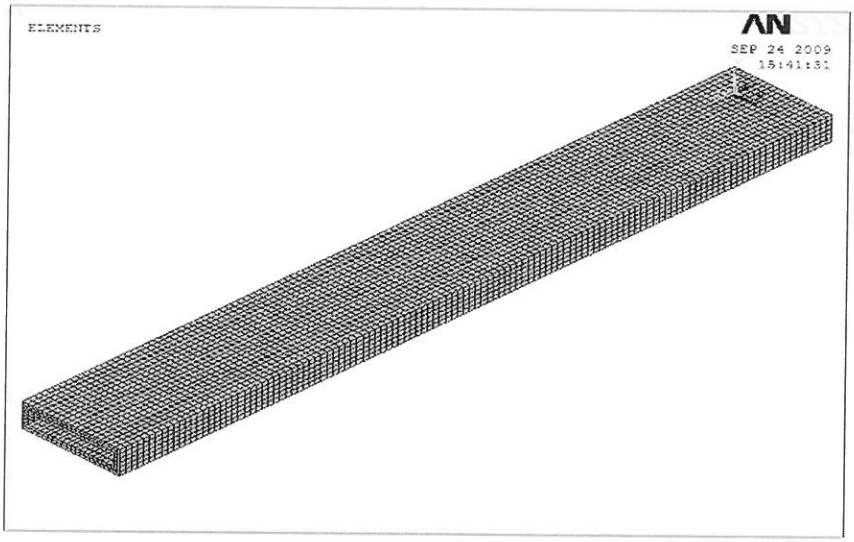


圖五：實驗流程圖

3.2 有限元素分析

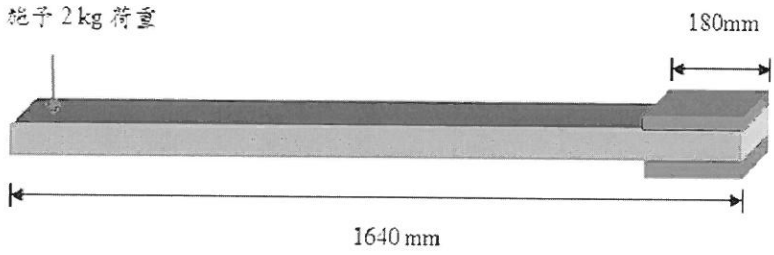
本研究之有限元素分析建模可分為五層與單層兩種方法。其中，五層法是將梭織物與單方向複合材料 (0˚及 ±45 ˚) 之工程模數分別帶入有限元素分析軟體 ANSYS，分層建立分析模型；而單層法是為了簡化分析，將梭織與單方向複合材料的材料性質

在厚度方向均質化，視為一層而建立分析模型。利用 SOLID185 元素完成之有限元素模型，如圖六所示。



圖六：箱型樑有限元素模型（多層）

本研究依照實際測量參數進行模擬分析，此箱型樑總長度為 1640mm，夾持 180mm，故模擬分析時，則總長設定為 1460mm，在前端截面設定拘束，於離末端 20mm 處施予 2kg 荷重後進行求解，如圖七所示。



圖七：箱型樑單點荷重測試示意圖

4. 結果與討論

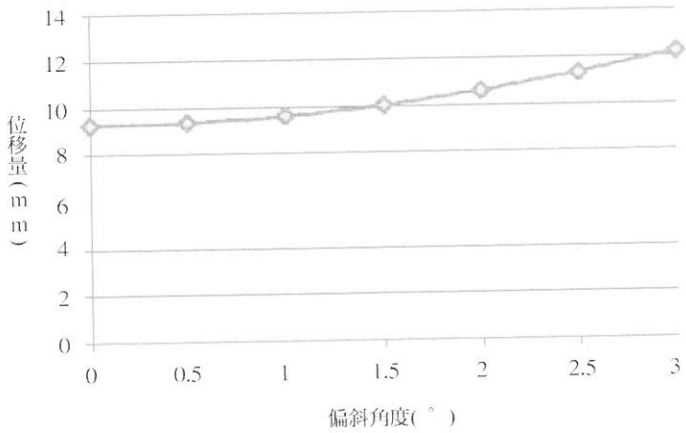
4.1 計算箱型樑撓曲位移與量測值比較驗證

本研究之箱型樑各層使用不同碳纖維之楊氏模數與纖維體積含有率如表一所示。

表一：各層之之楊氏模數與纖維體積含有率

	梭織層	0°層	±45° 層
纖維楊式模數 (GPa)	235	620	235
纖維體積含有率	40%	54%	44%

雖然箱型樑由不同疊層角度所組成，但是 0° 層堆疊層數較多且纖維模數高，故箱型樑受力時，主要影響撓曲位移者為 0° 層碳纖維。由於 0° 層碳纖維在製作過程中一定會有些許偏斜或歪曲，不管角度的正負，都會造成箱型樑撓曲剛性之降低。為了估計 0° 層纖維之實際可能偏斜角度，本文利用五層分析方法計算各個不同偏斜角度時之撓曲位移，如圖八所示。圖中顯示偏斜角度在 2.5° 時之位移量為 11.33mm，與實驗數據 11.61mm(利用紅外線量測儀測得) 相當接近，誤差僅 2.5%，故本研究選定 2.5° 為 0° 層碳纖維之偏斜角度，並用於以下各節之分析。



圖八：纖維偏移角度對撓曲位移之關係

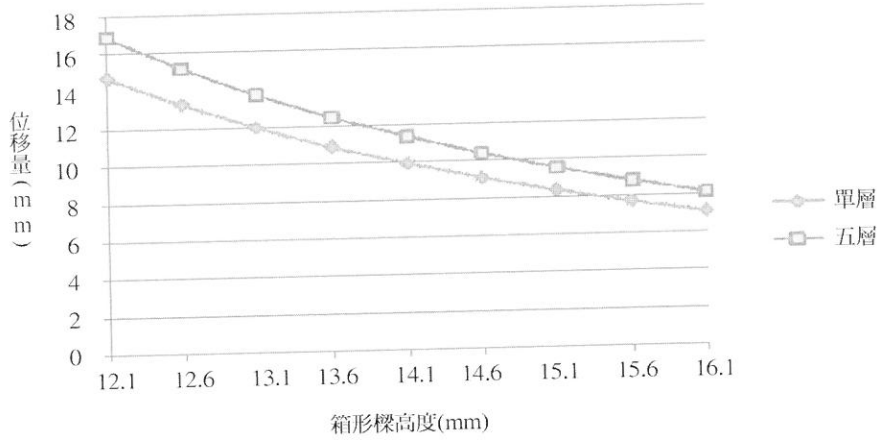
4.2 箱型樑參數改變對撓曲位移之影響

4.2.1 箱型樑高度對撓曲位移之影響

為探討截面高度對箱型樑撓曲位移之影響，在此固定箱型樑厚度並改變樑之高度（原始高度為 14.1 mm），分別以五層與單層兩種分析方法計算其受力後之撓曲位移並比較其之。如表二所示，當高度未改變時，單層與五層之撓曲位移分別為 9.897mm 與 11.333mm。當箱型樑高度增加 0.5mm 時，撓曲位移分別減少至 9.05mm 與 10.36mm，即各減少 8.52%，而當箱型樑高度減少 0.5mm 時，撓曲位移分別增加至 10.86mm 與 12.43mm，即各增加 9.75%。不管箱型樑高度增加或減少，兩種分析方法之平均誤差都為 12.68%，但平均分析計算時間五層法 (8 分 10 秒) 為單層法 (1 分 25 秒) 的 5.76 倍。上述計算時間差異之主要原因為此兩種分析方法建立有限元素模型時，五層的網格數較單層多而造成計算時間的增加。因此，為節省時間以提升分析效率，建議可以使用單層法作分析，然後再將所得之撓曲位移向上修正 12.68% 即可。兩種方法所得之撓曲位移與樑高度的關係如圖九所示，顯示當高度增加時撓曲位移會逐漸下降。

表二：箱型樑撓曲位移與高度之關係

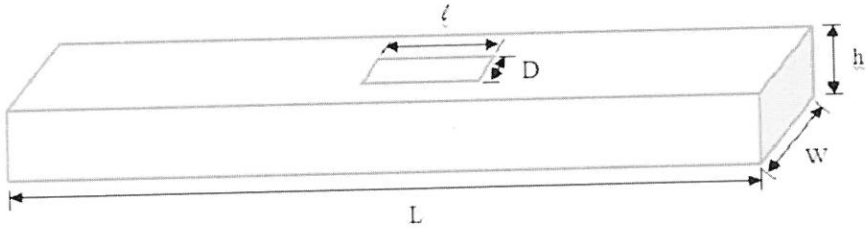
高度 h (mm)	五層位移 (mm)	單層位移 (mm)	位移變化率 (%)	誤差 (%)
12.1	16.89	14.73	49.04	-12.77
12.6	15.17	13.24	33.93	-12.73
13.1	13.70	11.96	20.97	-12.71
13.6	12.43	10.86	9.75	-12.68
14.1	11.33	9.89	0	-12.67
14.6	10.36	9.05	-8.52	-12.67
15.1	9.51	8.31	-16.03	-12.66
15.6	8.76	7.65	-22.67	-12.66
16.1	8.09	7.07	-28.56	-12.67



圖九：箱型樑撓曲位移與高度之關係圖

4.2.2 箱型樑開孔對撓曲位移之影響

箱型樑於實際應用時需開孔裝置真空吸盤，以吸著並固定面板，所以在此探討開孔及其位置對撓曲位移之影響。首先在箱型樑上表面開一長方形孔，如圖十所示，孔長寬分別設定為 45mm 與 30mm，此孔距離拘束面位置分別為 200、400、600、800、1000 與 1200mm。然後使用五層分析方法計算撓曲位移，結果如表三所示，當開孔位置距離拘束面為 200mm 與 1200mm 時，其撓曲位移分別為 13.79mm 與 12.06mm，較未開恐時分別增加了 21.71% 和 6.41%，而且開孔位置離拘束面越近其撓曲位移越大，也就是說當開孔位置越接近拘束面時，箱型樑之剛性變得越低，抵抗撓曲之能力也變得越差。



圖九：箱型樑撓曲位移與高度之關係

表三：箱型樑開孔位置與位移量之關係

距離拘束面位置 (mm)	五層位移 (mm)	位移增加率 (%)
未開孔	11.333	0
200	13.793	21.71
400	13.222	16.67
600	12.754	12.54
800	12.465	9.99
1000	12.146	7.17
1200	12.060	6.41

5. 結　論

本文利用單位格子、勁度平均法及均質化原理開發出五層法與單層法，用以分析混層複合材料箱型樑之撓曲位移。結果顯示，五層法與經修正之單層法均可以正確估算箱型樑之撓曲位移，而平均分析計算時間五層法為單層法的 5.76 倍。因此，為節省時間以提升分析效率，建議可以使用單層法作分析，然後再將所得之撓曲位移向上修正 12.68% 即可。此外，若箱型樑具有一大小為 45×30mm2 之開孔，其位置距離拘束面分別為 200mm 與 1200mm 時，其撓曲位移分別較未開孔時增加了 21.71% 和 6.41%，即開孔位置離拘束面越近其撓曲位移越大，也就是開孔位置越接近拘束面時，箱型樑之剛性變得越低，抵抗撓曲之能力也變得越差。

誌　謝

本研究之執行承國科會計畫 NSC99-2622-E-035-009-CC3 及明安股份有限公司在實驗及試片上之幫助才得以完成，特此誌謝。

參考文獻

1. F. Gordaninejad, A. Ghazavi and N. G. Chalhoub, Dynamic stress analysis of a laminated composite-material flexible robot arm, Recent Advances in the Macro and Micro Mechanics of Composite Materials Structures; ASME Winter Annual Meeting, 27, pp. 213-218, (1988).
2. R. Chandra, A. D. Stemple and I. Chopra, Thin-walled composite beams under-bending, torsional and extensional loads. J. Aircr, 27, pp.619-626, (1990).
3. DG Lee, KS Kim, YK Kwak, Manufacturing of a SCARA type direct-drive robot with graphite fiber epoxy composite material, Robotica , 9, pp.219-229, (1991).
4. EC Smith, I Chopra, Formulation and evaluation of an analytical model for composite box-beams, J. Am. Helicopter Soc., 36, pp.23-25, (1991).
5. S. M. Jeon, M. H. Cho, I. Lee, Static and dynamic analysis of composite box beam using large deflection theory, Computer & Structure, 57, pp. 635-642, (1995).
6. Y. Wu, X. Wang, Q. Su, L. Lin, A solution for laminated box beams under bending loads using the principle of complementary energy, Composite Structures,79, pp.376-380, (2007).
7. S. Suresh , P.B. Sujit , A.K. Rao, Particle swarm optimization approach for multi-objective composite box-beam design, Compos Struct, 81, pp.598–605, (2007).
8. Vincent Manet, The use of ANSYS to calculate the behavior of sandwich structures, Composite Science and Technology,58, pp.1899-1905, (1998).
9. J. H. Oh , D. G. Lee , H. S. Kim, Composite robot end effector for manipulating large LCD glass panels, Composite structures, 47, pp.497–506, (1999).
10. C. S. Lee , D. G. Lee , J. H. Oh , H. S. Kim , Composite wrist blocks for double arm type robots for handling large LCD glass panels, Composite Structures, 57, pp.345–55, (2002).
11. S. C. Jung, J. E. Lee, S. H. Chang, Design of inspecting machine for next generation LCD glass panel with high modulus carbon epoxy composites, Composite Structures, 66, pp.439-447, (2004).
12. C. S. Lee, D. G. Lee , Manufacturing of composite sandwich robot structures using the co-cure bonding method, Composite Structures,65, pp.307-318, (2004).

Received September 23, 2010
Received November 24, 2010
Accepted December 14, 2010

Deflection Analysis of a Hybrid Composite Robot Fork

K. H. Tsai[1], C. L. Hwan[2], Y. S. Zeng[3], Y. S. Huang[3], C. H. Chiu, C. M. Wu[1]

[1] *Department of Fiber and Composite Materials, Feng-Chia University*
[2] *Department of Mechanical and Computer Aided Engineering, Feng-Chia University*
[3] *Graduate School of Textile Engineering, Feng-Chia University*
100 Wenhwa Rd. Taichung, Taiwan

Abstract

In this study, two approaches were adopted to analyze the deflection of a hybrid composite robotic fork with stacking sequence [W/±45°/0°/±45°/W] under loading, in which W denotes woven fabrics. First, using the rule of mixture and the stiffness averaging method, the engineering moduli of the woven fabric and unidirectional composites were calculated. Furthermore, the engineering moduli of composite were homogenized in the thickness direction. Next, these engineering moduli were put into ANSYS to build two finite element models both with five-layers and a single-layer respectively in the thickness direction, and the deflections of the cantilever box-beams under loading were then calculated. The computed deflections were correlated well with the experimental results. It is noted that the simplified one layer method with some correction can provide a reasonable prediction of the deflection of the hybrid box beam in the shorter computing time. Besides, an open hole tends to increase the deflection of the hybrid box-beam and the hole effect becomes more obviously as the hole locates closer to the fixed end.

Keywords: hybrid composites, robot, box beam, deflection.

Experimental Studies on Cooling Performance of Desktop PCs

[1] Feng-Chyi Duh [2] Tsung-Han Li

[1] *Department of Mechatronics Engineering, Ta Hwa Institute of Technology No.1, Tahwa Road, Qionglin, Hsinchu, Taiwan 307, ROC*
[2] *Graduate School of Mechanical Engineering, National Chiao Tung University 1001 University Road, Hsinchu, Taiwan 300, ROC*

Abstract

This study investigated the cooling performance of desktop computer heat sinks under varying conditions, including heat load and axial load between heat sink and central processing unit (CPU), thermal interface material, and fan speed. In the condition of 3000 rpm fan speed and axial load F=18kgf, the surface temperature of CPU simulator T_j was reduced by as much as 67.5%. Quantitative data comparisons indicated that, under the condition of axial load F=18kgf and fan speed N=3000rpm, increasing effective heat flux q'' achieved a maximum ratio of 50%. The data confirm that the most efficient solution for the CPU heat problem is to simplify the attachment mechanism between the CPU and heat sink so as to ensure the effectiveness of thermal interface material between their surfaces. This solution is simpler and more economical than improving heat sink fins.

Keywords: Heat load, axial load, thermal interface material, fan speed.

Corresponding author:Tel:(03)592-7700 Ext:2654 Fax:(03)592-1047
E-mail:aedfc@thit.edu.tw

In modern desktop computers, improved thermal solutions are required due to the extremely small contact area between the CPU and the heat sink. Increasing CPU power and reduced die sizes are also problematic. Achieving perfectly flat CPU and heat sink surfaces is not possible. Therefore, small gaps in their interface negatively affect heat transfer since air conducts heat poorly. Therefore, an interface material with high thermal conductivity is needed to improve heat conductivity between the CPU and heat sink by filling these gaps. Meanwhile, for acceptable thermal transfer, the axial load between the two surfaces must be high. This study analyzes the cooling performance of a desktop PC heat sink. In doing so, the objective is to obtain experimental data that could be used to improve heat sink cooling performance. Since the heat sink is clamped to the CPU by an attachment mechanism, the structural design uses a high clip stiffness that resists local board curvature under the heat sink, which is designed to maintain a uniform contact surface between the simulated CPU and the heat sink. The axial load is analyzed by using a compression/tensile tester between the CPU simulator and heat sink. The heat sink attachment mechanism is designed to provide a strong axial load but still be reasonably easy to install. The extent to which axial load, thermal interface materials, and fan speed affect thermal dissipative performance can then be determined.

2. Experimental facility and method

2.1 Facility description

Figure 1 schematically depicts the experimental setup in this study. The facility included an attachment mechanism, a cooling unit, and a CPU simulator unit. The cooling unit and the CPU simulator unit were secured to an attachment mechanism on the tester with four fasteners. In this study, cooling performance measurements were performed in an enclosed test chamber to ensure maximum reliability.

The attachment mechanism consisted of a fixture, a digital force gauge, and a manual type compression/tensile tester capable of measuring axial loads from 8kgf to 20kgf with $\pm 0.5\%$ accuracy. The major functions of the attachment mechanism were to secure the

cooling unit and the CPU simulator unit and to apply and maintain the desired pressure on the TIM by applying an axial load to the heat sink base. The cooling unit consisted of a heat sink and an axial-flow fan. The CPU simulator unit consisted of three specimens: a copper plate (33mm×33mm×2.9mm, k=401W/mK), an electrical heater (33mm×33mm×1mm), and an adiabatic base. The adiabatic base was fabricated with bakelite (k=1.4W/mK).

The cooling performance simulation was run on a desktop PC with an Intel Pentium IV processor. The top surface of the CPU simulator unit was designed to interface with the heat sink of the cooling unit. The experimental thermal solution was an active-cooling design with a fan mounted on the heat sink. The apparatus contained no other heat-generating components. Most of the heater power was dissipated through the cooling unit, and the power dissipated as heat through the adiabatic base and into the compression/tensile tester was usually minimal.

The heat generated by the CPU simulator unit within the chassis (245mm×265mm ×510mm) of the experimental apparatus was minimized to provide a suitable operating environment. Heat generated by the CPU simulator was dissipated by maintaining airflow from the external ambient environment through the experimental apparatus. The chassis was designed to obtain reliable experimental results and precise data.

2.2 Testing procedures

Tests were conducted under normal environmental conditions, including a 20.8±1.0℃ ambient temperature. Table 1 shows the contact interface of CPU simulator and heat sink with/without TIM that was selected for each parameter (axial load F and fan speed N). Heat generated by the CPU was simulated using an electrical heater. The cooling unit was tested on a simulator with a 100W capacity. The simulator was insulated from external heat. Three temperatures were determined by type K (chromel–alumel) thermocouples. One thermocouple was set at the contact interface of the CPU simulator, and another was set at the other side of the copper core at a distance of 3.5mm ($\Delta z_{2\text{-}3}$=3.5mm). These thermocouples were secured with high conductivity thermal grease (k=4.8W/mK). The simulator temperature was determined at the geometric center of the contact surface. A thermocouple was also placed

8 mm above the fan (Δz_A=8mm) to monitor ambient temperature. All thermocouples were measured simultaneously at 1.25 Hz.

The CPU simulator was characterized at step-wise increases in a heat load ranging from 15.5W to 87.6W and with a time lag in stationary regimes. The maximum heat load was limited by an admissible simulator temperature of about 80℃ and its limiting capacity. The experimental thermal solution was an Intel radial curved bifurcated fin heat sink (RCBFH) (diameter 89.4mm, height 40.6mm) with aluminum fins, aluminum base, and copper core . The cooling unit had the following axial fan specifications: rated voltage 12VDC, dimensions L80mm×W80mm×H25mm, noise level 35.5dB, rating speed 3500rpm. The fan was tested at 2500rpm and 3000rpm.

The objectives were to determine temperature characteristics such as thermal resistance of the cooling unit under varying conditions such as heat load. For the condition of contact interface without TIM, only the effects of axial load F and fan speed N were studied in a heat load range lower than 61W. The full characterization of TIM included tests of heat loads up to 87.6W and all possible combinations of input parameters (eight total). Eight series of experiments were performed to test different combinations of TIM, axial load, and fan speed (Table 1). All tests were performed with the simulator in a thermally stable condition. In this case, heat losses from the simulator into the environment were considered negligible and were excluded from the calculations.

2.3 Theoretical background

A theoretical model of thermal management was developed to quantify how TIM, axial load, and fan speed affect heat dissipation from a desktop PC. The thermal profile parameter defined the maximum case temperature (MCT) as a function of processor heat load as described in the Intel reference design [24]. The thermal design power (TDP) and MCT were defined as the maximum values of the thermal profile. The thermal solutions were intended to satisfy the thermal profile requirements for all system operating conditions and processor power levels. The thermal solution performance was expressed as the slope on the thermal profile, which represented the thermal resistance of the heat sink attached to the CPU.

The thermal resistance of the heat sink/CPU assembly, which is the conventional measure of heat dissipation in a CPU cooling unit, refers to the ratio of the temperature difference between the heat sink and CPU simulator to the total heat load.

$$R = \frac{\Delta T_{1-2}}{q} = \frac{T_1 - T_2}{q} \tag{1}$$

where ΔT_{1-2} denotes the temperature difference between surface temperature of CPU simulator T_1 and base temperature of heat sink T_2, q denotes the heat load of the CPU simulator, which is the product of electric current I and electric voltage V produced by the electrical heater in this study. The R is commonly presented in units of ℃ /W. According to the equation for R, contact resistance approaches the minimum as ΔT_{1-2} approaches zero.

Heat load and temperature were determined to validate the thermal solutions. Here, thermal characterization parameter (TCP) ψ was compared among all thermal solutions

$$\psi = \frac{\Delta T_{1-A}}{q} = \frac{T_1 - T_A}{q} \tag{2}$$

where T_1 denotes the surface temperature of CPU simulator, T_A denotes the ambient temperature above the fan, and q denotes the heat load of the CPU simulator. Unlike thermal resistance, the TCP is estimated according to total CPU power. The three essential measurement parameters for characterizing thermal properties were surface temperature of the CPU simulator, heat load of the CPU simulator, and ambient temperature.

Heat is removed from the CPU simulator via conduction across the interface of the CPU simulator surface, through a TIM, into the heat sink and then to the environment via convection. In the condition of contact interface with TIM, the heat flux through the heat sink is calculated by

$$q'' = -k\frac{\Delta T_{2-3}}{\Delta z_{2-3}} = -k\frac{T_2 - T_3}{\Delta z_{2-3}} \tag{3}$$

where k denotes the thermal conductivity of the copper core of heat sink. In this study, k

(copper) is $401 W/mK$. The T_2 and T_3 were the surface temperatures at the two ends of the copper core and are determined on the same vertical central-axis. The value $\Delta z_{2\text{-}3}$ was the distance between T_2 and T_3 (in this case, $\Delta z_{2\text{-}3}$=3.5mm).

2.4 Uncertainty analysis

The experimental analyses of desktop PC cooling performance were subject to errors in instrumentation, methodology, and procedure, which were quantified to assess the confidence in the results. The relative uncertainty in the TCP [25] is given by

$$\frac{\delta p_i}{p_i} = \left\{ \left[\frac{\delta y_i}{y_i} \right]^2 + \left[\frac{\delta x_i}{x_i} \right]^2 \right\}^{1/2} \tag{4}$$

where δp represents the parameter uncertainty in the result, and δy_i and δx_i denote the uncertainties in the variables y_i and x_i, respectively. In this equation, $\delta p_i / p_i$ represents the relative uncertainty in the result, and the factors $\delta y_i / y_i$ and $\delta x_i / x_i$ are the relative uncertainties of each variable. The parameters in this experimental study included thermal resistance R, thermal characterization parameter ψ , and heat flux q'' as defined in Eqs. (1)~(3), respectively. The variables y_i and x_i are described in Eqs. (1)~(3).

In this experimental study, estimated maximum relative uncertainties in thermal resistance R , thermal characterization parameter ψ , and heat flux q'' were 5.3%, 5.3%, and 5.2%, respectively.

3. Results and discussion

Thermal resistance refers to the ability of a cooling unit to transfer heat between different surfaces. Reducing thermal resistance significantly affects overall thermal performance. A higher thermal resistance implies a larger temperature drop across the interface. Additionally, a more efficient thermal solution is required to achieve the desired cooling.

Three thermocouples were used to monitor temperature history and to obtain regularized sequential trends that represent thermal resistance and other TCP measures.

3.1 Bare junction

To reduce thermal resistance between the heat sink and CPU interface, thermal resistance is typically minimized by filling the interfacial gap with TIM. A TIM was used to eliminate gaps in the contact interface by improving the conformance between the mating surfaces. A direct conduction path from the CPU to the heat sink and use of TIM with increased thermal conductivity can also improve cooling performance. Specifically, the quality of the contact interface between the CPU and the heat sink has a higher impact on overall solution performance as CPU cooling requirements become stricter.

The analysis first compared different TIMs. This experimental study considered only two limiting cases, contact interfaces with and without TIM coatings. The thermal grease applied in this study had a thermal conductivity of k=4.8W/mK . In Fig. 2, the vertical coordinate is the measured surface temperature of CPU simulator T_1 , and the horizontal coordinate is the heat load q. The figure shows that fan speeds were 2000rpm, 2500rpm, and 3000rpm when the CPU and cooling unit were integrated under an axial load. The surface temperature of CPU simulator T_1 generally increased linearly with heat load. Table 2 shows the results of a linear, least-squares data fit, which can be simply expressed by the following correlation

$$T_1 = aq + b \tag{5}$$

where coefficients a and b are experimental constants. For coefficients a and b, Table 2 tabulates the best fit for different cases as determined by least squares method.

Notably, the change in T_1 was larger in the cases of uncoated TIM than in the cases of coated TIM. For example, the change in heat load q from 15.5W to 87.6W resulted in a 46℃ increase in T_1 (experimental number #7) whereas an increase in q from 15.5W to 39.2W obtained a 32.7 ℃ increase in T_1 (experimental number #2), which was significantly larger than that in the previous case. All such properties were attributable to the TIM. Under

heat load condition q=39.2W, the use of TIM had the greatest effect on cooling performance (64.3% reduction in T_1). The data in Table 2 provide further confirmation. In all tests of coated TIM, coefficient a was 0.62℃/W to 0.68℃/W whereas, for all tests of uncoated TIM, coefficients a was 1.38℃/W to 1.57℃/W, which was more than double that observed in the previous cases.

Experimental data for the case of uncoated TIM were compared with the Intel data [24]. In the low heat load condition (q<15.5W), the heat sink dissipated the heat generated by the CPU. Therefore, the experimental data were lower than the data obtained by Intel. However, as heat load increased, thermal resistance gradually increased until it significantly disrupted heat dissipation, and the discrepancies with the Intel data increased accordingly. Immediately after reaching a load of 39.2W, the CPU simulator temperature exceeded 76.9 ℃ under the two different axial loads whereas the corresponding temperature reported by Intel was only 53.3℃ . As larger axial load reduces the temperature of the CPU simulator, heat dissipation increases. These data reveal the important effect of thermal resistance.

In the next experiment, the same two axial loads were applied at fan speeds of 2000rpm, 2500rpm, and 3000rpm. The only difference was that the contact surface between the bottom of the CPU simulator and the heat sink was coated with TIM to fill the gaps between the contact surfaces in order to improve thermal resistance. Figure 2 shows the experimental results, which confirm that the temperature of CPU simulator T_1 was significantly reduced after TIM coating was applied. This phenomenon implies that the TIM can weaken thermal resistance, which improves heat transfer from the CPU through the interface and to the heat sink. Additionally, the TIM not only efficiently reduced the temperature of the CPU simulator, it also improved heat dissipation performance at higher temperature ranges. Further, the data in Fig. 2 show that, although the temperature of the CPU simulator at the three different fan speeds did not substantially differ, fan speed was clearly inversely related to simulator temperature. This confirms the improved heat dissipation achieved by faster fan speeds.

Further analysis of the data distribution in Fig.2 reveals significant differences in the availability of coated TIM, especially in the case of high-efficiency cooling performance. Quantitative data comparisons indicated that, although the thermal conductivity of the TIM

(k=4.8W/mK) was much lower than the copper core (k=401W/mK), the coating on the heat sink and CPU simulator still weaken the contact resistance significantly. Under heat load condition q=39.2W, axial load F=18kgf, and fan speeds of 2000rpm, 2500rpm, and 3000rpm (experimental number #6, #7, and #8), the effective temperature reductions were 57.6%, 64.3%, and 67.5%, respectively.

3.2 Axial load effects

In the next series of tests, the heat sink was attached directly to the CPU. The attachment mechanism was designed to apply a axial load on the CPU. When measuring the thermal performance of the TIM between the heat sink and the CPU, the axial load was maintained within the minimum/maximum range specified in the CPU. Due to asperity deformation under the axial load, the number of contact spots increased, which in turn decreased thermal resistance. An appropriate axial load can optimize thermal contact conductance at the CPU/ heat sink interface. An overly low axial load causes poor contact, and an overly high axial load causes CPU/heat sink assembly damage.

In further studies of the effects of axial load, the four groups in Fig.2 represent the T_j value of the CPU simulator under the conditions of 12kgf and 18kgf axial load. The data distribution in Fig.2 indicates that, although all CPU simulator temperatures were very similar, axial load (experimental numbers #6, #7 and #8) was negatively related to the T_j value of CPU simulator. This occurs because the axial load increases the pressure on the contact surfaces, which then reduces the air gap on its surface, improves thermal resistance, and reduces the T_j value of the CPU simulator.

Thermal resistance was further measured under a steady-state, one-dimensional axial heat flow condition. Figure 3 shows the thermal resistance R [Eq.(1)] distribution for varying heat load q. Thermal resistance correlated with heat load and reached a maximum of 0.182℃/W at 87.6W (experimental number #3). Such a high thermal resistance for traditional active cooling is due to axial load and static load. Nevertheless, the effect of fan speed is significantly smaller than those of axial load and static load. For a given fan speed, thermal resistance decreased as axial load increased due to enhanced thermal contact conductance between the

heat sink and CPU simulator.

As the TCP reflects cooling performance, the experimental data were transformed into the parameters and drew it with the heat load q in the diagram, as Fig. 4 shows. According to Eqs. (1) and (2), the thermal resistance R and the TCP have the same unit. In physical terms, the TCP determines thermal resistance and generally increases as heat dissipation increases; under larger axial loads, the TCP is slightly lower. Although the variation in TCP is small, the effect of axial load is still distinguishable. This implies that, given a consistent fan speed, axial load is inversely related to thermal resistance.

Figure 4 plots the TCP (ψ) for varying heat load q, which was determined by calculating the temperature difference between T_I and T_A [Eq.(2)]. For a heat load range of 15.5-87.6W, TCP values approximated 0.5 ℃/W. Notably, axial load and fan speed provided efficient cooling performance by increasing thermal contact conductance and forced convection, which then reduced the temperature difference between T_I and T_A . Quantitative analysis indicated that, as heat load q increased from 15.5W to 87.6W, TCP increased to 19.6% and decreased to 13.2% in two of the limiting cases (experiments #3 and #8, respectively).

The next experiment examined heat flux q'' by the heat sink cooling unit. The distribution of six sets of experimental data in Fig.5 clearly show that axial load correlated with heat flux through the heat sink.

3.3 Fan performance

The examined cooling unit consisted of a heat sink with radial curved bifurcated fin and an axial-flow fan. The primary function of the fan was to enhance heat transfer from the CPU by efficiently transferring heat out of the CPU to an attached cooling device. Convective heat transfer between the airflow and the surface exposed to the flow is characterized by the local air temperature above the fan T_A and the local air velocity via the surface. The cooler the air, the more efficient the resulting cooling unit.

The effects of fan speed on cooling performance were then examined at fan speeds of 2000rpm, 2500rpm, and 3000rpm. Figure 2 shows that fan speed was inversely related to temperature T_1. At speeds of 2000rpm to 2500rpm, average reduction in T_1 was 3.24%, and, at 3000rpm, T_1 increased to 5.69%. Further analysis of the data distribution shown Figs. 3 and 4 revealed that both thermal resistance and TCP correlated negatively with fan speed. The TCP results were consistent with the Intel data [24]. Intel reported that, in the reference design performance of Intel RCBFH-3, the TCP was 0.29 ℃/W at 3600rpm and 0.325 ℃/W at 2400rpm and that temperature reduction reached 12.07% [24]. In the current experimental study, a fan speed increase from 2000rpm to 2500rpm caused an average decrease in ψ of 0.042℃/W (8.45%). Further, a fan speed increase from 2500rpm to 3000rpm caused an average decrease in ψ of 0.057℃ (13.3%).

Finally, Fig. 5 summarizes the impact of fan speed on conductive heat flux. The experimental results clearly reveal that heat load q correlated with heat flux q''. This finding suggests that heat transfer through the copper increased. A data comparison between the three different fan speeds (2000rpm, 2500rpm and 3000rpm) clearly shows that, given a constant axial load of 18kgf, fan speed correlated with heat flux q'', especially under the condition of high heat load q. Quantitative data comparisons also indicated that, under axial load condition F=18kgf and fan speeds of 2000rpm, 2500rpm, and 3000rpm in experiments #6, #7, and #8, respectively, the effective heat flux q'' increased 37.5%, 37.5%, and 50%, respectively.

Figure 6 shows the relation between the simulated CPU surface temperature T_1 and fan speed N based on the different heat loads. The value T_1 was negatively associated with N. This implies that heat dissipates via the heat sink as a function of fan speed. The plot also indicates that cooling performance through the heat sink requires high volume flow-rate for active cooling unit due to high fan speed. Quantitative data analysis indicated that, when the speed N increased from 2000rpm to 3000rpm, T_1 decreased from 3.42% to 7.31%. Although the three different fan speeds revealed only small differences in CPU temperature, the comparison confirms that faster fan speeds improve heat dissipation.

4. Conclusion

Thermal solutions that optimize cooling performance can be selected based on the results of this experimental study of heat sink cooling performance in desktop PCs. This study is the first to analyze the interrelationships of axial load, thermal interface materials, and fan speed. The conclusions of this experimental study of heat sink cooling performance of desktop PC heat sinks are the following:

The experimental results with or without TIM coating clearly showed decreases in temperature T_1 , which corresponded with heat dissipation. In the condition of 3000rpm fan speed and F=18kgf axial load, the largest reducing range of T_1 reached 67.5%, which indicates that thermal grease reduced contact resistance and facilitated heat transfer from the interface to heat sink. Thus, it significantly reduced the simulated CPU temperature and increased the range of the CPU heat load. Despite the effects of axial load and fan speed, the surface temperature of CPU simulator T_1 revealed a linear increase, which corresponded with the increase in heat load q.

Generally, axial load was negatively associated with simulated CPU temperature. The temperature reduction resulted in increased axial load, which enlarged the contact pressure between the CPU and the heat sink. Thus, it substantially increased pressure on the metal surface and also reduced the air gap between the contact surfaces. Regarding the effect of axial load, simulated CPU temperature did not significantly differ (temperature change was less than 4.05% of maximum) despite the increased axial load from 12kgf to 18kgf (50% increase). Nevertheless, increased axial load decreased simulated CPU temperature T_1 .

This study explored how affects overall cooling performance. The simulation results indicated that over the range of heat load q from 15.5W to 87.6W, the values of TCP were approximately 0.5℃/W. Additionally, larger axial loads increased the pressure at the interface between the CPU and the heat sink. The compression of the metal reduced air gaps between the contact surfaces and therefore reduced thermal resistance. That is, increased heat dissipation reduced TCP. However, at high fan speeds, increased heat dissipation caused by forced convection resulted in a relatively smaller TCP. Compared to axial load, fan speed had

a larger effect on TCP.

Finally, the experimental results indicated that heat load q was positively associated with conductive heat flux q'', which indicates increased heat conduction through the copper. Analysis of the impact of axial load and fan speed revealed that axial load was positively related to heat flux q'' conduction through the copper and that fan speed was positively related to heat flux q''. The impact of axial load on conduction of heat flux q'' was significantly greater than the impact of fan speeds. Quantitative data comparisons indicated that, under the condition of axial load F=18kgf and fan speed N=3000rpm, the increase in effective heat flux q'' achieved a maximum ratio of 50%.

In summary, we conclude that the most efficient solution for the CPU heat problem is to maintain heat transfer efficiency between the CPU and heat sink by simplifying their attachment mechanism. This solution is simpler and more economical than improving heat sink fins.

References

1. Mochizuki, M. 2002. Development of folded fin heat sink with solid base and vapor chamber, Proceedings of the 12th International Heat Pipe Conference, Proc. *12th IHPC*, 369-372.
2. Kim, K.S. 2002. Heat pipe cooling technology for desktop PC CPU, Proceedings of the 12th International Heat Pipe Conference, Proc. *12th IHPC*, 398-401.
3. Lin, S.C., Chuang, F.S., and Chou, C.A. 2005. Experimental study of the heat sink assembly with oblique straight fins, *Experimental Thermal and Fluid Science*, 29: 591-600.
4. Lin, S.C. and Chou, C.A. 2004. Blockage effect of axial-flow fans applied on heat sink assembly, *Applied Thermal Engineering*, 24: 2375-2389.
5. Pastukhov, V.G. and Maydanik, Y.F. 2009. Active coolers based on copper-water LHPs for desktop PC, *Applied Thermal Engineering*, 29: 3140-3143.
6. Joudi, K.A. and Witwit, A.M. 2000. Improvements of gravity assisted wickless heat pipes, *Energy conversion and management*, 41, 18: 2041-2061.

7. Lin, S., Broadbent, J., and McGlen, R. 2005. Numerical study of heat pipe application in heat recovery systems, *Applied Thermal Engineering*, 25, 1: 127-133.
8. Suman, B., De, S., and Das, G.S. 2005. A model of the capillary limit of a micro heat pipe and prediction of the dry-out length, *International Journal of Heat and Fluid Flow*, 26, 3: 495-505.
9. Vasiliev, L.L. 2008. Micro and miniature heat pipes–electronic component coolers, *Applied Thermal Engineering*, 28, 4: 266-273.
10. Wang, J.C., Huang, H.S., and Chen, S.L. 2007. Experimental investigations of thermal resistance of a heat sink with horizontal embedded heat pipes, *International Communications in Heat and Mass Transfer*, 34: 958-970.
11. Wang, Y. and Vafai, K. 2000. An experimental investigation of the thermal performance of an asymmetrical flat plate heat pipe, *International Journal of Heat and Mass Transfer*, 43, 15: 2657-2668.
12. Zhang, H. and Zhuang, J. 2003. Research, development and industrial application of heat pipe technology in China, *Applied Thermal Engineering*, 23, 9: 1067-1083.
13. Mochizuki, M. 2004. Practical application of heat pipe and vapor chamber for cooling performance personal computer, Proceedings of the 13th International Heat Pipe Conference, *Proc. 13th IHPC*, 448-454.
14. Advanced Thermal Solution. 2008. Decreasing thermal contact resistance by using interface materials, *Qpedia Thermal eMagazine*, 2, 6: 16-21.
15. Teertstra, P. 2007. Thermal conductivity and contact resistance measurements for adhesives, Proceedings of IPACK, *Proc. of IPACK2007*: 1-8.
16. Xu, R.P., Xu, L., and Zhao, L.P. 2004. Fractal description of thermal contact resistance between rough surfaces, *Journal of Shanghai Jiaotong University*, 38, 10: 1609-1612.
17. Chung, D.D.L. 2001. Materials for thermal conduction, *Applied Thermal Engineering*, 21: 1593-1605.
18. De Mey, G., Pilarski, J., Wojcik, M., Lasota, M., Banaszczyk, J., Vermersch, B., and Napieralski, A. 2009. Influence of interface materials on the thermal impedance of electronic packages, *International Communications in Heat and Mass Transfer*, 36: 210-212, 2009.
19. Gwinn, J.P. and Webb, R.L. 2003. Performance and testing interface materials, *Microelectronics Journal*, 34: 215-222.
20. Shaikh, S., Lafdi, K., and Silverman, E. 2007. The effect of a CNT interface on the thermal resistance of contacting surfaces, *Carbon*, 45: 695-703.

21. Xu, J. and Fisher, T.S. 2006. Enhancement of the thermal interface materials with carbon nanotube arrays, *International Journal of Heat and Mass Transfer*, 49: 1658-1666.
22. Sartre, V. and Lallemand, M. 2001. Enhancement of thermal contact conductance for electronic systems, *Applied Thermal Engineering*, 21: 221-235.
23. Duh, F. C. 2010. A study on heat dissipative performance for CPU. *Journal of Advanced Engineering*, 5, 1: 69-75. [in Chinese]
24. Intel Corp., 2004, "*Intel Pentium 4 processor on 90nm process in the 775-Land LGA package- thermal design guidelines*" . Rept. 302553-001: 10-58.
25. Kline S.J. 1985. The purpose of uncertainty analysis, *Journal of Fluids Engineering*, 107: 153-160.

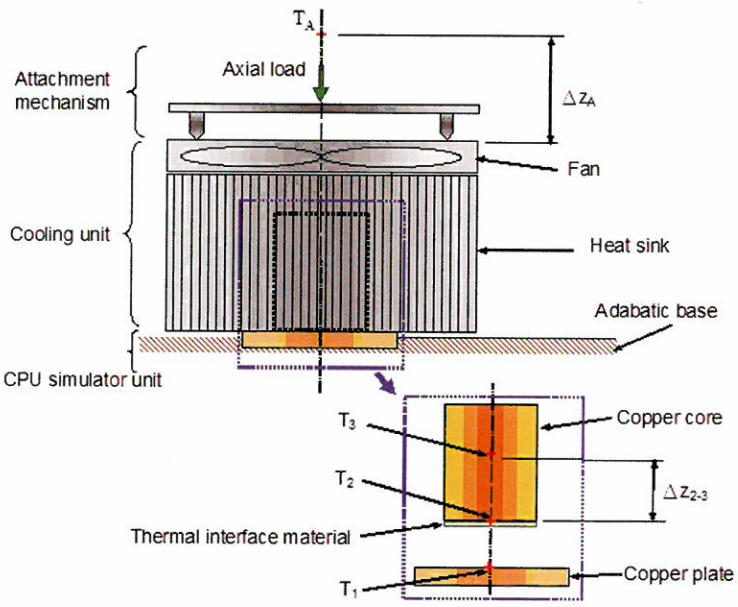


Fig.1 Experimental apparatus

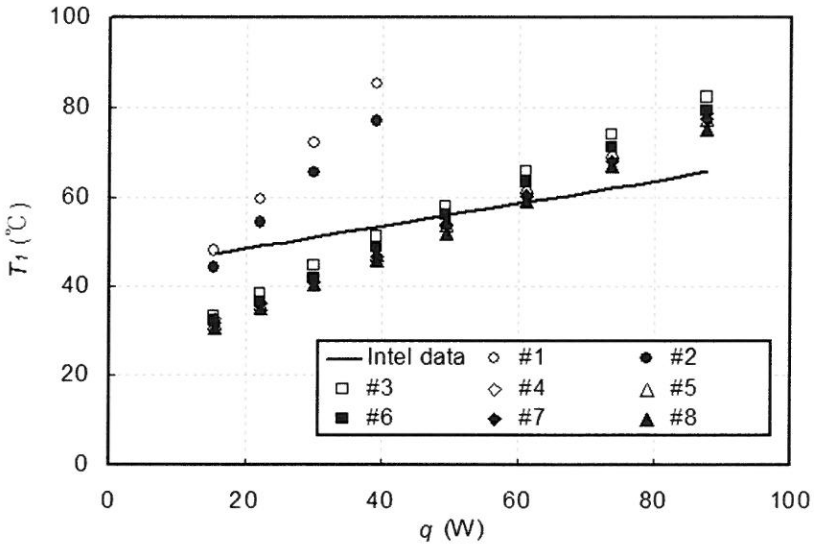


Fig.2 Heat load dependences of surface temperatures of CPU simulator

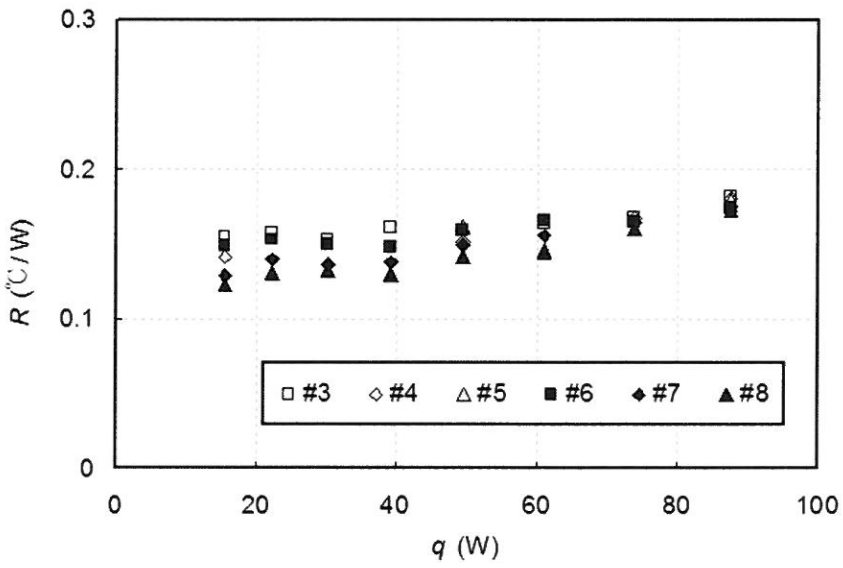


Fig.3 Heat load dependences of thermal resistance

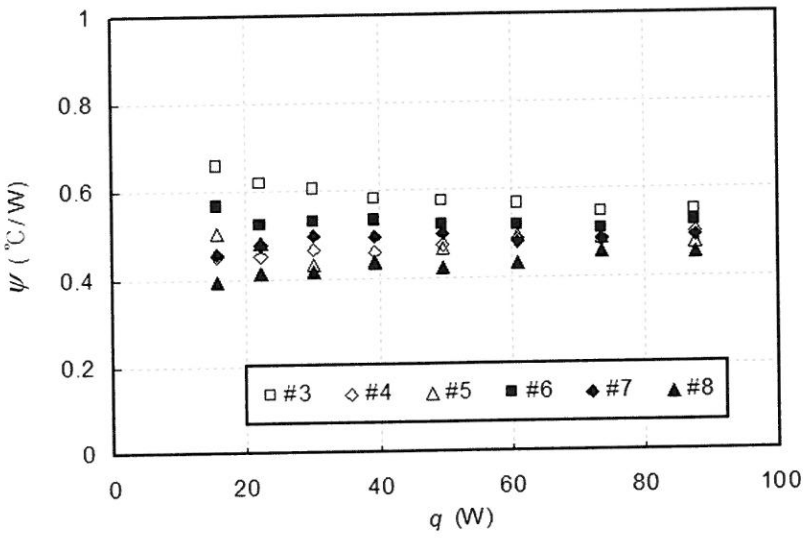


Fig.4 Heat load dependences of thermal characterization parameter

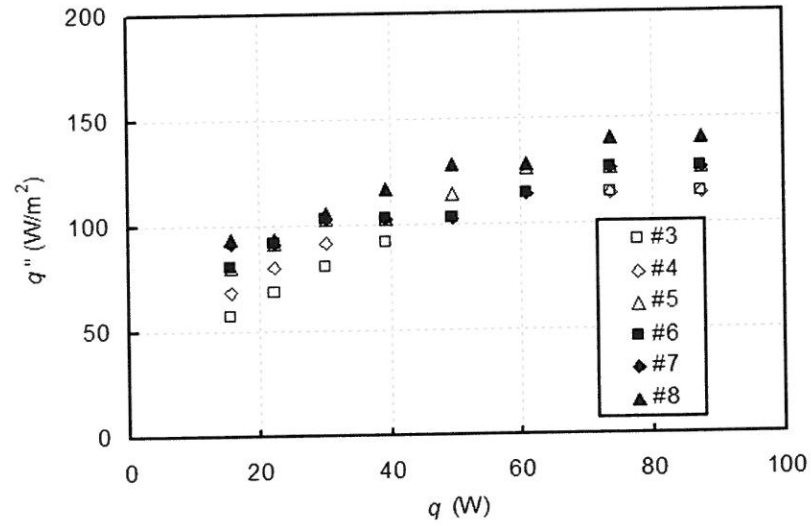


Fig.5 Heat load dependences of CPU simulator heat flux for different fan speed

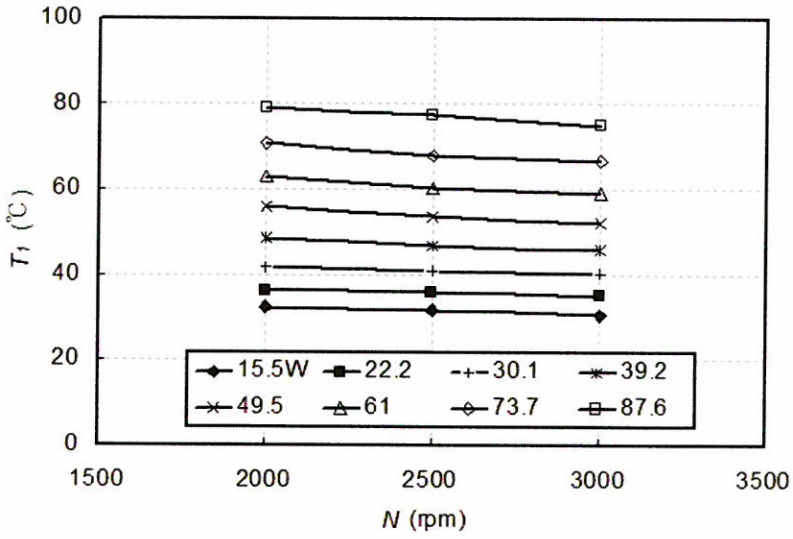


Fig.6 Fan speed dependences of surface temperatures of CPU simulator

Table 1 Experimental parameters

Test #	Contact interface of heat sink/CPU	Fan speed N (rpm)	Axial load F (kgf)
1	Uncoated TIM	2500	12
2	Uncoated TIM	2500	18
3	Coated TIM	2000	12
4	Coated TIM	2500	12
5	Coated TIM	3000	12
6	Coated TIM	2000	18
7	Coated TIM	2500	18
8	Coated TIM	3000	18

Table 2 A linear expression for the surface temperatures of CPU simulator as a function of heat load, $T_i = aq + b$

Test #	Coefficient a (℃/W)	Coefficient b (℃)	Max. error %
1	1.57	24.0	-1.03
2	1.38	23.0	-1.04
3	0.68	23.5	2.53
4	0.64	22.0	-2.08
5	0.64	22.0	1.88
6	0.64	22.0	1.58
7	0.63	22.0	0.93
8	0.62	21.5	1.67

Received September 24, 2010
Revised November 27, 2010
Accepted December 16, 2010

使用實驗方法探討桌上型電腦之冷卻效能

[1] 杜鳳棋　[2] 李宗翰

[1] 大華技術學院機電工程系　307 新竹縣芎林鄉大華路 1 號
[2] 國立交通大學機械工程研究所　300 新竹市大學路 1001 號

摘　要

本論文旨在探討在各種不同的負荷狀況下，桌上型電腦 CPU 藉由散熱器之冷卻效能，主要的考慮因素包括介於散熱器與 CPU 之間的熱負荷、軸向負荷、熱介面物質、風扇轉速等。經由實驗的結果顯示，在風扇轉速 3000rpm 及軸向負荷 18kgf 的操作狀態下，CPU 模擬器的表面溫度 將會降低達到 67.5%。就量化數據比較發現，在軸向負荷 18kgf 及風扇轉速 3000rpm 的運作情況，有效熱通量最大增加將近 50%，從實驗數據肯定只要保持散熱器與 CPU 間的熱介面物質之有效性，並確保散熱器與 CPU 間夾持機構的簡單性，將可大幅的強化 CPU 散熱器之冷卻效能。

關鍵字：熱負荷、軸向負荷、熱介面物質、風扇轉速。

巴金森氏症之基因表現判別分析

康家睿　侯藹玲
輔仁大學生命科學系所

摘　要

本文分析兩組已發表之巴金森氏症 (PD) 的微陣列資料，分別為 47 片的腦部組織樣本和 105 片的血液樣本，利用統計方法之判別分析篩選出顯著性基因，並建立判別函式判別一個新個體是否為 PD 患者，且藉由群集分析佐證顯著性基因的判別力，探討腦部組織中所篩選出來的生物標記是否也能在血液檢體中出現，以便用於未來臨床之診斷，並結合 GO 基因功能分類資料庫，篩選出來的顯著性基因進行基因功能分類，以提供未來生物及醫學研究上之參考依據。

由分析結果找出 10 個在腦部及血液樣本中皆有顯著表現的基因，分別為 SUMO2、ABCE1、PPP3CB、ARFGEF1、STMN2、C5orf22、GCH1、CYP2J2、RGR 和 TH 基因，其中 TH 基因為先前文獻上指出和巴金森氏症有關，而利用此十個基因所建立的判別函式判別腦部組織樣本高達 90% 的判別率，雖然血液樣本的判別率不如腦部樣本，但在可提供更多且更詳細的資料情況下，血液樣本還是具有找出生物標記的潛力存在。

關鍵詞：*巴金森氏症、微陣列、判別分析、群集分析。*

Corresponding author:Tel:(02)2905-2464　Fax:(02)2905-2193
E-mail:022446@mail.fju.edu.tw

1. 前　言

巴金森氏症 (Parkinson’ s disease) 僅次於阿茲海默症是最常見的神經退化性疾病，病徵包括動作緩慢、靜態性顫抖、肢體僵硬和步態不穩等。現今檢測是否罹患巴金森氏症主要仍以臨床診斷為主，而典型巴金森氏症發病期較晚，必須當中腦黑質部 (substantia nigra) 多巴胺 (dopamine) 神經元喪失 50~70% 才會出現病徵，因此，目前可能還有大量罹患此病但尚未表現病徵的病人無法經由臨床診斷出來。而目前藥物治療目的在控制病程發展，無法完全治癒疾病，且藥物劑量隨著病程加重須逐步增加，直至無法控制症狀為止，因此如何達到及早發現並控制病程發展，成為目前金森氏症一項重要議題。

由於巴金森氏症致病原因目前被廣泛接受的理論為多因子理論，而微陣列技術具有同時檢測大量基因表現量的能力，近年也被應用在巴金森氏症的研究上，取自於巴金森氏症病人腦部黑質組織的微陣列資料經由顯著性分析後，發現第一條染色體短臂的 PARK10 基因座上與疾病的關聯密切，而此區域似乎與巴金森氏症的病發年齡有關，進一步分析這些顯著性基因發現，HERPUD1 表現量在巴金森氏症病人中比在正常人中高 (Moran, et al. 2006)。另外，分析取自於巴金森氏症患者血液的樣本微陣列資料，發現巴金森氏症患者的 ST13 基因表現量比正常人低 (Scherzer, et al. 2007)。這些已發表的微陣列資料中，除可用於治病機制的研究外，並可應用於分析疾病與基因表現的關聯性，作為判定是否罹患巴金森症的生物標記 (biomarker)。

本文中利用多變量統計中的判別分析以及 t 檢定分析兩組巴金森氏症的微陣列資料，分別為巴金森氏症早期血液樣本與死後中腦黑質部組織樣本，探討在腦部組織中所篩選出來的生物標記是否與血液樣本中找到的有一致性，以便用於未來臨床上之診斷，提供未來生物及醫學研究上之參考依據。

2. 材料與方法

一、微陣列資料來源

本篇研究之兩組巴金森氏症的微陣列資料皆來自於 NCBI(National Center for Biotechnology Information) 網站的 GEO datasets，編號分別為 GSE6613 和 GSE8397，此兩組微陣列皆為 Affymetrix 公司的 HG U133 A，每片包含 22,283 個 probe sets。

二、微陣列資料簡介

(一) GSE8397

腦部樣本分別為 15 個 PD 患者黑質中間部位、9 個 PD 患者黑質側邊部位和 8 個正常對照組的黑質中間部位、7 個正常對照組 (control) 的黑質側邊部位，而黑質的中間與側邊部位樣本均來自於同一個人，另外，5 個 PD 患者和 3 個正常對照組的前葉大腦皮層樣本也一併加入分析 (Moran et al., 2008)。

(二)GSE6613

本組資料包含 105 個樣本，分別為 50 位巴金森氏症患者、22 位與 PD 患者年齡相符的正常對照組和 33 位與 PD 患者年齡相符的神經退化性疾病對照組的血液樣本，PD 患者均達到 United Kingdom Parkinson' s Disease Society Brain Bank 的臨床診斷標準，為了使偽陽性 (false positive) 的可能性降到最低，患者病徵必須至少達到三個症狀 (bradykinesia and two of rigidity、rest tremor or postural instability) 才可通過診斷標準，正常對照組均無個人或家族遺傳性疾病史，且為了降低環境因素干擾，將近 90% 皆為 PD 患者之配偶 (Scherzer et al., 2007)。

三、研究工具

(一) 程式語言 R

程式語言 R 是一個以物件導向為主資料處理以及統計軟體，其原始碼可自由下載使用 (http://www.r-project.org/)，是免費的公開軟體，因具備撰寫容易且相容於各平台的優點，近年來使用的人越來越多 (林 , 2006)。

而 R 在生物資訊領域上應用甚廣，並針對微陣列資料分析開發出各種相關套件 (package) 組成 Bioconduct 軟體 (http://www.bioconductor.org/)，亦為免費下載的公開資源，提供許多前人研究所研發的成果，對於相關研究者的使用有相當的便利性。

(二) 程式語言 Perl

Perl(Practical Extraction and Report Language) 融合了 awk/sed/cut/grep 等 unix 上常見的工具與 C 語言等語法所創造的一種語言。Perl 和 R 一樣是免費的公開軟體，有現成的套件或程式碼可供使用，也可在不同的使用平台上操作。Perl 對於文字處理功能較強，近年來很受生物資訊學家的喜愛，例如資料的編輯排列、分子序列比對與處理等 (蔣 , 2008)。

四、微陣列資料前處理

從 NCBI 下載後的資料為 .CEL 檔，必須先經過前處理步驟轉換成數值才能做後續統計分析，前處理分為三個步驟，分別為 background adjustment、normalization、summarization。Background adjustment 為校正微陣列資料中的背景干擾雜訊，微陣列資料中由於受到光學的干擾或非專一性的雜合而影響其數值，這些因素統稱為 background noise，而 background adjustment 的目的為校正資料使其呈現原有的數值，normalization 的目的為校正資料使其站在同一水平點上做比較，而 summarization 的目的則是把微陣列上的基因，其含有的數個 probe pairs 數值轉換成一個數值以代表其基因的值 (Wu, et al. 2004)。

本文使用 R 語言中 gcrma 套件的 gcrma 指令做前處理步驟， gcrma 利用 robust multi-array average (RMA) 演算法做前處理，搭配後續的 unequal variance t-test 有較佳的顯著性分析結果 (Vardhanabhuti, et al. 2006)。

五、統計方法

(一) Unequal variance two sample t-test

本文使用 unequal variance two-sample t-test，檢定對照組與病患組間的差異，以 p-value 來篩選表現量具有顯著差異的基因。

(二) FDRs(False Discovery Rates)

在統計檢定中會出現兩種錯誤，分別為第一型錯誤 (Type I error) 和第二型錯誤 (Type II error，β)，適當控制 Type I 及 Type II error，即能間接控制統計的檢定力 (Power，1-β)。控制 Type I error 的方法很多，最常見的有兩種，分別為 family-wise error rates(FWERs) 和 false discovery rates(FDRs)，而本論文所使用的方法為 FDRs(Gentleman, 2005)。

(三) 判別分析 (Discriminant Analysis)

判別分析是對已知群別的族群資料建立判別函數或判別式，然後對未知群別的樣品資料進行判別分類的統計方法，本文所使用的判別法為費氏線性判別法，此判別函數是利用兩群或數群已知的多變數 (或指標) 經線性組合減為單變數後再判別新個體的歸屬，此法並不考慮兩群變數之分布形態，利用判別分析可觀察出篩選出來的顯著性基因是否能夠正確地判別對照組與患者組 (沈 , 2007)。

本文利用 leave-one-out 方法，每次抽出一個樣本當作待測組 (validation set)，其餘樣本則建立判別函式，並對其待測組進行判別，leave-one-out 方法逐一進行，直到每個樣本都被抽出當成待測組後才會終止，由於每個樣本都會被抽出過，可信度高，所以此步驟為目前建立判別模型常用的一種方法 (圖一)。

(四) 集群分析 (Clustering Analysis)

集群分析法是將一大群觀測值依其相似性 (similarity) 或同質性 (homogeneity) 分成數群的方法，使得每群內個體間差異小，而群間之個體差異大。集群法不同於分類法 (classification)，後者是根據已知的資料建立函數關係，作為新個體分配於何群的分類法，而集群分析法並不知道研究的資料有幾群，僅根據資料中每個個體之特徵是否相似而進行分群的方法 (沈 , 2007)，本文使用階層式群集分析 (Hierarchical clustering) 法，觀察判別分析中所篩選出來的顯著性基因是否能夠將對照組與患者組正確地歸類分群。

(五)Gene Ontology 基因功能分類

Gene Ontology 為生物資訊常用的基因功能資料庫，由於生物技術與生物資訊的發展，目前公開的生物相關資料庫為數眾多，往往同一個基因蛋白質產物功能，在不同資料庫中所描述的不盡相同，使得研究者不知該如何判定是否這些資料庫描述的是同一件事，且花了很多時間在網路中搜尋。

而 1998 年開始了 gene ontology 計畫，結合了 Flybase(Drosophila)、Saccharomyces Genome Database(SGD) 和 Mouse Genome Database(MGD) 等基礎模式生物資料庫，共同合作制定基因蛋白質產物的功能描述，並將功能主要分為三大項目，分別為 Gene Ontology Biological Process、Gene Ontology Cellular Component 和 Gene Ontology Molecular Function，並給予基因產物功能一個 GO 編號，如今 Gene Ontology 資料庫

更進一步結合了數個世界知名的資料庫，涵蓋了植物、動物、微生物等基因體資料。(http://www.geneontology.org/GO.doc.shtml)

經由基因表現顯著性分析篩選的多個基因中，各有其功能及註解，透過 GO 資料庫的分類配合超幾何分布的機率，可得知篩選出來的顯著性基因是否偏向某些基因功能。

六、分析流程

此處以 GSE8397 為例介紹，步驟如下：

1. 從 NCBI 網站下載編號為 GSE8397 的巴金森氏症 Affymetrix 微陣列資料，並對其進行資料前處理 (Preprocessing)，將下載下來的 .CEL 檔轉換成數值並對其做校正，此步驟是利用程式語言 R 中的 gcrma 軟件的 gcrma 指令進行。
2. 判別分析一開始將資料隨機分出總片數的 1/4，總共 10 片樣本當作待測組 (test set)，剩下的 37 片樣本資料做為學習組 (learning set) 利用 leave-one-out 方法建立判別函式 (圖一)。
3. 進行 leave-one-out 方法時，將資料中的樣本逐一抽出來當作建立判別函式中的驗證組 (validation set)，而剩餘的 36 片樣本資料 (training set) 則進行顯著水準 α 為 0.05 的 unequal variance two- sample t 檢定分析。
4. 每次篩選出來具有顯著性水準之基因，進行 False discovery rates (FDRs) 檢定，並取其前 500 個具有顯著性水準之基因對先前抽出之樣本 (validation set) 進行判別。
5. 將所有 37 次判別中，至少出現 60% 以上的顯著性基因保留下來，對 10 片測試組 (test) 進行判別並計算其判別率。
6. 將所有 37 次判別中，至少出現 60% 以上的顯著性基因保留下來，利用程式語言 perl 進行資料編排並進行 GO enrichment 分析。
7. 將所有 37 次判別中，至少出現 60% 以上的顯著性基因保留下來，和另一組資料 (GSE6613) 進行比對，取其交集基因，並利用交集基因建立一組判別函式，分別對兩組資料 (GSE6613、GSE8397) 進行判別並觀察其判別率。
8. 將所有 37 次判別中，至少出現 60% 以上的顯著性基因保留下來，進行階層群集分析，並建立 heatmap。
9. 對交集基因進行階層群集分析，並建立 heatmap。

3. 結　果

一、利用建立判別模型找出關聯性基因

(一)GSE8397

47 組微陣列資料中除 10 組測試組，其餘 37 組則利用 leave-one-out 方法建立判別模型，其中過程包括了 unequal variance two-sample t 檢定及 FDRs 檢定，並選取其結果前五百個顯著性基因進行待測組 (validation set) 之判別，總共進行 37 次。

37 次判別中，前五百個顯著性基因的 FDRs 約為 0.02(如表一)，而 37 次判別中，全都判別正確，其判別率為 100%。為了找出重現性高的基因，進一步觀察 37 次中的前五百個顯著性基因，篩選至少出現超過 60% 的基因，總共有 439 個，這表示前五百個顯著性基因中大部分基因的重現率都很高，其中包括了先前文獻上指出跟巴金森氏症有關的 SNCA、NR4A2、TH、HTRA2、DNAJB6、RELN、SLC6A3、SLC18A2 等八個基因，最終利用這 439 個基因建立判別函式，並對當初抽出的測試組樣本 (test) 進行判別分析，而 10 片測試組樣本中，判錯 2 片 (GSM208626、GSM208629)，也就是判錯所屬類別 (PD or Control)，其判別率為 80%。

(二)GSE6613

GSE6613 為巴金森氏症的血液樣本，實驗分析步驟跟 GSE8397 同步，目的是為了後續的比較，資料首先隨機取出四分之一當作未來的測試組，總共 20 片，剩餘四分之三進行判別模式之建立，總共進行 85 次。

85 次判別中，前五百個顯著性基因的 FDRs 約為 0.04~0.1，而 85 次判別中，判錯 17 次，判別率為 80%。進一步觀察 85 次判別中出現超過 60% 的基因，總共有 370 個，包括與巴金森氏症相關的 TH、ST13、GDNF 等三個基因，而利用這 370 個顯著性基因建立判別函式，判別當初抽取的測試組，而 20 片測試組樣本中，判錯了 3 片 (GSM153443、GSM153435、GSM153484)，判別率為 85%。

二、相關性基因之階層群集分析

一般群集分析用於樣本所屬類別不明時，將樣本進行歸類分群，而本試驗利用群集分析輔佐判別分析的結果，觀察經由判別分析篩選出來的顯著性基因，是否確實能

夠拿來分辨 PD 患者與正常人，以間接證實顯著性基因的可信度。

(一)GSE8397

經由判別分析所篩選出來的 439 個顯著性基因，將利用階層群集分析對 47 片樣本資料進行歸類分群，結果如圖二。

經由判別分析所篩選出來的 439 個顯著性基因大致上可以將 PD 患者與正常人正確歸類，除了 5 片 PD 樣本被錯判分類到 Control 組和 1 片 Control 樣本被分類到 PD 組，這結果顯示出經由先前判別分析處理巴金森氏症腦部組織所篩選出來的 439 個顯著性基因確實具有分辨 PD 患者與正常人的能力。

(二)GSE6613

經由判別分析所篩選出來的 370 個顯著性基因，利用階層群集分析對 105 片樣本資料進行歸類分群，結果如圖三所示。

就此群集分析的結果來看，GSE6613 血液樣本所篩選出來的 370 個顯著基因在區分 PD 患者與正常人的判別能力較弱。

三、基因功能性分類

(一)GSE8397

經由判別分析所篩選出來的 439 個相關性基因，為了找出此 439 個基因偏向於哪些基因功能，進行基因功能 GO(gene ontology) 分析，GO 主要將基因功能分為三個部分，分別為 Biology process、Cellular component 和 Molecular function，439 個相關性基因在 Biology process 基因功能方面傾向於 37 種基因功能 (表二 A)，在 Cellular component 基因功能方面傾向於 10 種基因功能 (表二 B)，而在 Molecular function 基因功能方面則傾向於 22 種基因功能 (表二 C)。

(二)GSE6613

經由判別分析所篩選出來的 370 個相關性基因，其在 Biology process 基因功能方面傾向於 18 種基因功能 (表三 A)，在 Cellular component 基因功能方面傾向於 2 種基因功能 (表三 B)，則在 Molecular function 基因功能方面則傾向於 7 種基因功能 (表三 C)。

四、GSE8397 與 GSE6613 相關性基因之比對

進一步比較 GSE8397 和 GSE6613 之相關性基因，總共找到 10 個交集基因 (表四)，分別為 SUMO2、ABCE1、PPP3CB、ARFGEF1、STMN2、C5orf22、GCH1、CYP2J2、RGR 和 TH，其中只有 TH 基因為文獻上指出與巴金森氏症有關，其餘基因皆找不到其與巴金森氏症有關之文獻。

而此十個交集基因所建立的兩組判別函式分別為：

GSE8397 :

$y = -2.6462\,(CYP2J2) + 1.5318\,(ARFGEF1) + 1.2109(SUMO2) + 0.9435\,(ABCE1) - 0.9142\,(C5orf22) + 0.7328\,(PPP3CB) + 0.6605\,(TH) - 0.6015\,(STMN2) - 0.3827\,(GCH1) - 0.2013\,(RGR)$

GSE6613 :

$y = -4.2542\,(STMN2) + 2.4171\,(ABCE1) + 1.9724\,(SUMO2) - 1.6424\,(CYP2J2) - 1.2907\,(RGR) - 1.1453\,(GCH1) - 0.5415\,(C5orf22) + 0.3115\,(ARFGEF1) + 0.2632(PPP3CB) + 0.0313(TH)$

其判別 GSE8397 資料之測試組的判別率為 90%，判別 GSE6613 資料之測試組的判別率僅有 50%，而利用此十個交集基因分別對 GSE8397 和 GSE6613 資料進行分類歸群，結果如圖四、五，此十個交集基因大致上能將 GSE8397 資料中的 PD 與 Control 正確歸類分群，而無法辨別 GSE6613 資料中的 PD 與 Control 組。

4. 討　論

本篇論文利用判別分析處理兩組巴金森氏症微陣列資料 (GSE8397、GSE6613)，藉由兩組資料互相比對，期盼能在血液中尋找生物標記，以利未來臨床上的診斷，在 GSE8397 大腦黑質部組織樣本資料中，共篩選出 439 個顯著性基因，其中包括了先前文獻上指出跟巴金森氏症有關的 SNCA、NR4A2、TH、HTRA2、DNAJB6、RELN、SLC6A3、SLC18A2 等八個基因，SNCA 基因在少數家族性遺傳及偶發的巴

金森氏症患者中發現有 point mutation、duplication 及 triplication 的現象，NR4A2 基因第一個 exon 的兩個位置突變 (-291delT 與 -245T-G) 被認為與家族性巴金森氏症有關 (Le, et al. 2003)，HTRA2 基因產物蛋白質在德國巴金森氏症患者發現有兩個突變位置 (G399S、A141S)，此突變造成粒線體功能缺陷，G399S 位置突變也使細胞較易受壓力而死亡，而 HTRA2 基因產物蛋白質也在偶發性巴金森氏症患者腦部的特有路易氏體 (Lewy bodies) 中發現，被認為與巴金森氏症致病有關 (Strauss, et al. 2005)，TH 基因在遺傳性體染色體隱性的幼兒巴金森氏症患者中發現有同質結合突變現象 (Ludecke, et al. 1996)，DNAJB6 基因則被認為與 protein folding、renaturation of denatured proteins 和 prevention of protein aggregateon 有關，內質網堆積太多不正常蛋白質將導致內質網功能缺陷，而內質網功能缺陷也被認為與巴金森氏症的多巴胺神經元喪失有關 (Moran et al. 2006)，RELN 基因被發現在阿茲海默症患者腦脊液中有表現量增加的現象，而巴金森氏症患者也有同樣的情況 (Botella-Lopez, et al. 2006)，而 SLC6A3 基因的蛋白質產物 DAT 為多巴胺的轉運子 (transporter)，負責調控神經元突觸中多巴胺神經傳導物質的回收，在巴基斯坦幼兒巴金森氏症患者中發現 SLC6A3 基因的第八個 exon 有同質結合突變 (1103T-A)，導致蛋白質序列置換 (L368Q)，而在歐洲幼兒巴金森氏症患者中也有類似突變，不過位置在 SLC6A3 基因的第九個 exon(1184C-T)，造成蛋白質序列置換 (P395L)，這些置換皆使得 DAT 喪失調控多巴胺的能力 (Kurian, et al. 2009)。

利用 439 個相關性基因建立出來的判別函式判別測試組的判別率可達 80%，證明此 439 個相關性基因具有判別 PD 患者與正常人的能力，而此 439 個相關性基因所做出來的階層群集分析結果也輔佐證明了相關性基因的判別力，只有 6 片無法正確分類歸群，然而即使具有高度判別力，卻無法百分之百判別成功，也代表了本資料中還是有少數變異樣本存在。

而此 439 個相關性基因經由 GO 分析後，發現在 Biology process 方面傾向於 ubiquitin-dependent protein catabolic process、apoptosis、vesicle-mediated transport、regulation of apoptosis 等基因功能，而在 Molecular function 方面則與 protein tyrosine kinase activity、transmembrane receptor activity 等基因功能有關，可與在 KEGG 網站中提供的巴金森氏症致病途徑作一映證。

而在 GSE6613 血液樣本資料中，經由判別分析共篩選出 370 個相關性基因，其中包括與巴金森氏症相關的 TH、ST13、GDNF 等三個基因，ST13 基因具有穩定

HSP70(Heat-shock protein 70) 的能力，HSP70 可藉由修正 SNCA 的 misfolding 而抑制 SNCA toxicity(Scherzer et al. 2007)，GDNF 基因被指出可以防止成鼠大腦黑質部中的多巴胺神經元的喪失，把 GDNF 注入成鼠中腦黑質部可防止細胞免於受 MPTP 神經毒素迫害，由於 MPTP 是已知神經毒素會破壞多巴胺神經元，所以 GDNF 也被認為可以用於治療巴金森氏症，而隨後也經證明確實 GDNF 可以用於治療巴金森氏症患者 (Beck et al. 1995; Tomac et al. 1995; Gill, et al. 2003)。

由此 370 個相關性基因所建立的判別函式判別測試組的判別率雖然可達 85%，但其階層群集分析的結果卻顯示此 370 個相關性基因缺乏分辨 PD 患者與 Control 的能力，而建立判別函式的過程中，發現隨機抽取的測試組將大大影響後續的判別率，進一步觀察，得知 GSE6613 血液資料樣本間變異很大，即使是最顯著的基因，在同一類別的樣本資料間，表現量也有很大的差距，此問題導致群集分析結果無法如預期般準確地將 PD 患者與 Control 分類歸群，這也可以間接說明為何 GSE6613 資料在建立判別函式時，85 次的 FDRs 的值會有 0.04~0.1 的範圍出現，由於樣本間的變異過大，造成每次抽出一片樣本作驗證組時，剩餘樣本做的後續分析所篩選出來的顯著基因並不相同，如果同組樣本間變異並不大時，FDRs 的值應該會出現和 GSE8397 資料一樣維持在穩定的數值，而雖然 85 次建立判別函式中，每次篩選的前五百個顯著性基因中，出現超過 60% 的基因也有 370 個，因此相信還是有很多具有判別力的基因並未被篩選，370 個相關性基因中也含有一些不具判別力的基因在裡面，而此 370 個相關性基因所做的 GO 分析，也看不出其傾向的基因功能和巴金森氏症有關聯性存在。

不過值得注意的是，雖然 GSE6613 資料中樣本間變異過大，以致無法準確判別樣本類別，但是在 GSE8397 資料所篩選出來的相關性基因具有很大的判別力，相當具有潛力可提供為後續研究之參考依據，此也證明本試驗中所使用的方法具有找出生物標記的能力，而本試驗所分析的血液樣本資料並無提供更詳細的資料作參考，例如抽血的時間點、樣本的性別資料等，如果能提供抽血時間點，分析資料時就能多加入一個時間的因素，而可更深入去探討不同時間點血液中的基因表現量變化，也許能克服本文所遇到 GSE6613 資料樣本間變異過大的問題，而提供樣本性別也是如此，雖然資料中 90% 的對照組皆為 PD 患者的配偶，但並不知道個別樣本的性別，所以分析時就無法納入性別這個因素了。

由於比對 GSE8397 和 GSE6613 兩組資料之相關性基因，找到了 10 個交集的基因，

其中包含了文獻中指出與巴金森氏症有關的 TH 基因，並發現用此 10 個基因所建立的兩組判別函式中，同一個基因的係數皆為同正或同負，此代表了腦部中這些基因的表現與在血液中的表現趨勢為一致性，而其判別腦部組織的判別率也高達 90%，至於血液部分，若是可以提供更多且更詳細的樣本資料，血液樣本還是具有尋找巴金森氏症生物標記的潛力存在。

參考文獻

1. 沈明來 (2007) 實用多變數分析 , 第二版 . 九州圖書 , 台北 , 台灣 .
2. 林建甫 (2006) 醫學統計與 R. 國立台北大學統計系 , 台北 , 台灣 . (http://web.ntpu.edu.tw/~cflin.).
3. 蔣大偉 (2008) Perl 學習手冊 (編譯), 第五版 , O'reilly, 台北 , 台灣 .
4. Beck, K. D., J. Valverde, T. Alexi, K. Poulsen, B. Moffat, R. A. Vandlen, A. Rosenthal, F. Hefti. (1995) Mesencephalic dopaminergic neurons protected by GDNF from axotomy-induced degeneration in the adult brain. Nature 373(6512): 339-341.
5. Botella-Lopez, A., F. Burgaya, R. Gavín, M. S. García-Ayllón, E. Gómez-Tortosa , J. Peña-Casanova, J. M. Ureña, J.A. Del Río, R. Blesa, E. Soriano, J. Sáez-Valero. (2006). Reelin expression and glycosylation patterns are altered in Alzheimer's disease. *Proc Natl Acad Sci U S A 103*(14): 5573-5578.
6. Gentleman, R. (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer.
7. Gill, S. S., N. K. Patel, G. R. Hotton, K. O'Sullivan, R. McCarter, M. Bunnage, D. J. Brooks, C. N. Svendsen, P. Heywood. (2003). Direct brain infusion of glial cell line-derived neurotrophic factor in Parkinson disease. *Nat Med 9* (5): 589-595.
8. Kurian, M. A., J. Zhen, S.Y. Cheng, Y. Li, S. R. Mordekar, P. Jardine, N. V. Morgan, E. Meyer, L. Tee, S. Pasha, E. Wassmer, S. J. Heales, P. Gissen, M. E. Reith, E. R. Maher. (2009). Homozygous loss-of-function mutations in the gene encoding the dopamine transporter are associated with infantile parkinsonism-dystonia. *J Clin Invest 119*(6): 1595-1603.
9. Le, W. D., P. Xu, J. Jankovic, H. Jiang, S. H. Appel, R. G. Smith, D. K. Vassilatis. (2003). Mutations in NR4A2 associated with familial Parkinson disease. *Nat Genet 33* (1): 85-89.
10. Ludecke, B., P. M. Knappskog, P.T. Clayton, R. A. Surtees, J.D. Clelland, S. J. Heales, M.

P. Brand, K. Bartholomé, T. Flatmark. (1996). Recessively inherited L-DOPA-responsive parkinsonism in infancy caused by a point mutation (L205P) in the tyrosine hydroxylase gene. *Hum Mol Genet 5* (7): 1023-1028.

11. Moran L.B., D. C. Duke, M. Deprez, D. T. Dexter, R. K. Pearce, M. B. Graeber. (2006). Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. *Neurogenetics. 7* (1):1-11.
12. Scherzer, C. R., A. C. Eklund, L.J. Morse, Z. Liao, J. J. Locascio, D. Fefer, M. A. Schwarzschild, M. G. Schlossmacher, M. A. Hauser, J. M. Vance, L.R. Sudarsky , D. G. Standaert, J. H. Growdon, R. V. Jensen, S. R. Gullans. (2007). Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc Natl Acad Sci U S A 104* (3): 955-960.
13. Strauss, K. M., L. M. Martins, H. Plun-Favreau, F. P. Marx, S. Kautzmann, D. Berg, T. Gasser, Z. Wszolek, T. Müller, A. Bornemann, H. Wolburg, J. Downward, O. Riess, J. B. Schulz, R. Krüger. (2005). Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Hum Mol Genet 14* (15): 2099-2111.
14. Tomac, A., E. Lindqvist, L.F. Lin, S. O. Ogren, D. Young, B.J. Hoffer, L. Olson . (1995). Protection and repair of the nigrostriatal dopaminergic system by GDNF in vivo. Nature 373(6512): 335-339.
15. Vardhanabhuti, S., S. J. Blakemore, S. M. Clark, S. Ghosh, R. J. Stephens, D. Rajagopalan. (2006). A comparison of statistical tests for detecting differential expression using Affymetrix oligonucleotide microarrays. *OMICS 10* (4): 555-566.
16. Wu, Z. J., R. A. Irizarry, R. Gentleman, F. Martinez-Murillo, F. Spencer. (2004). A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association 99* (468): 909-917.

表一、37 次判別分析中前五百個顯著基因的 FDRs

次數	FDRs	次數	FDRs
10	.019243	20	0.01825
20	.019111	21	0.017128
30	.017587	22	0.01828
40	.018804	23	0.017036
50	.018804	24	0.016767
60	.018694	25	0.022576
70	.020962	26	0.019825
80	.020387	27	0.0176 85
90	.019495	28	0.015593
10	0.020443	29	0.021061
11	0.019564	30	0.016031
12	0.018836	31	0.019478
13	0.019444	32	0.020884
14	0.018788	33	0.019772
15	0.018058	34	0.019343
16	0.019226	35	0.014348
17	0.017463	36	0.019243
18	0.017436	37	0.020221
19	0.015792		

表二、GSE8397 顯著基因中 GO 分析結果

A. 顯著偏向的 Biology process 基因功能

GO_ID	p.value	Biology process
1525	0.0091	angiogenesis
6281	0.0401	DN A repair
6350	1.55E-08	transcription
6355	6.70E-11	regulation of transcription, DNA - dependent
6366	0.0398	transcription from RNA polymerase II promoter
6412	0.0057	translation
6414	0.0180	translational elongation
6468	0.0030	protein amino acid phosphorylation
6508	0.0019	proteolysis
6511	0.0326	ubiquitin - dependent protein catabolic process
6915	0.0073	apoptosis
6917	0.0414	induction of apoptosis
6935	0.00594	chemotaxis
6954	0.00114	inflammatory response
6955	1.65E-09	immune response
7155	0.0033	cell adhesion
7165	0.0084	signal transduction
7166	0.0272	cell surface receptor linked signal transduction
7169	0.00962	transmembrane receptor protein tyrosine kinase signaling pathway
7267	0.0004	cell -cell signaling
7275	3.09E-06	multicellular organismal development
7283	0.0120	spermatogenesis
7605	0.0251	sensory perception of sound
8152	0.0143	metabolic process
8283	8.60E-05	cell proliferation
8284	0.0016	positive regulation of cell proliferation
9887	0.0267	organ morphogenesis
16192	0.0414	vesicle - mediated transport
16337	0.0429	cell - cell adhesion
16481	0.0050	negative regulation of transcription
30036	0.0467	actin cytoskeleton organization
30154	0.0085	cell differentiation
42981	0.0281	regulation of apoptosis
43066	0.0270	negative regulation of apoptosis
44419	0.0005	interspecies interaction between organisms
45449	8.07E-05	regulation of transcription
50896	0.0459	response to stimulus

B. 顯著偏向的 Cellular component 基因功能

GO_ID	p.value	Cellular component
5576	1.06E-07	extracellular region
5578	0.0183	proteinaceous extracellular matrix
5615	9.14E-06	extracellular space
5622	0.0002	intracellular
5634	2.83E-05	nucleus
5730	0.0256	nucleolus
5840	0.0307	ribosome
5886	0.0024	plasma membrane
9897	0.0438	external side of plasma membrane
16021	0.0058	integral to membrane

C. 顯著偏向的 Molecular function 基因功能

GO_ID	p.value	Molecular function
3676	9.46E-05	nucleic acid binding
3677	1.21E-06	DNA binding
3700	1.18E-06	transcription factor activity
3735	0.0040	structural constituent of ribosome
3779	0.0094	actin binding
3823	0.0206	antigen binding
4222	0.0293	metalloendopeptidase activity
4713	0.0090	protein tyrosine kinase activity
4872	0.0004	receptor activity
4888	0.0112	transmembrane receptor activity
5102	0.0346	receptor binding
5125	0.0438	cytokine activity
5179	0.0210	hormone activity
5201	0.0490	extracellular matrix structural constituent
5515	0.0004	protein binding
8083	0.0334	growth factor activity
8233	0.0264	peptidase activity
8234	0.0267	cysteine - typepeptidase activity
8270	0.0064	zinc ion binding
16853	0.0426	isomerase activity
30528	0.0034	transcription regulator activity
43565	0.0025	sequence - specific DNA binding

表三、GSE6613 顯著基因中 GO 分析結果

A. 顯著偏向的 Biology process 基因功能

GO_ID	p.value	Biology process
6334	0.0279	nucleosome assembly
6350	0.0078	transcription
6355	0.0020	regulation of transcription, DNA - dependent
6366	0.0218	transcriptio n from RNA polymerase II promoter
6412	1.68E-07	translation
6508	0.0389	proteolysis
6810	4.21E-05	transport
6811	0.0264	ion transport
6813	0.0337	potassium ion transport
6955	0.0005	immune response
7165	0.0102	signal transduction
7242	0.0343	intracellular signaling cascade
7605	0.0254	sensory perception of sound
8284	0.0100	positive regulation of cell proliferation
15031	0.0407	protein transport
45449	0.0310	regulation of transcription
45786	0.0244	negative regulation of cell cycle
45944	0.0071	positive regulation of transcription from RNA polymerase II promoter

B. 顯著偏向的 Cellular component 基因功能

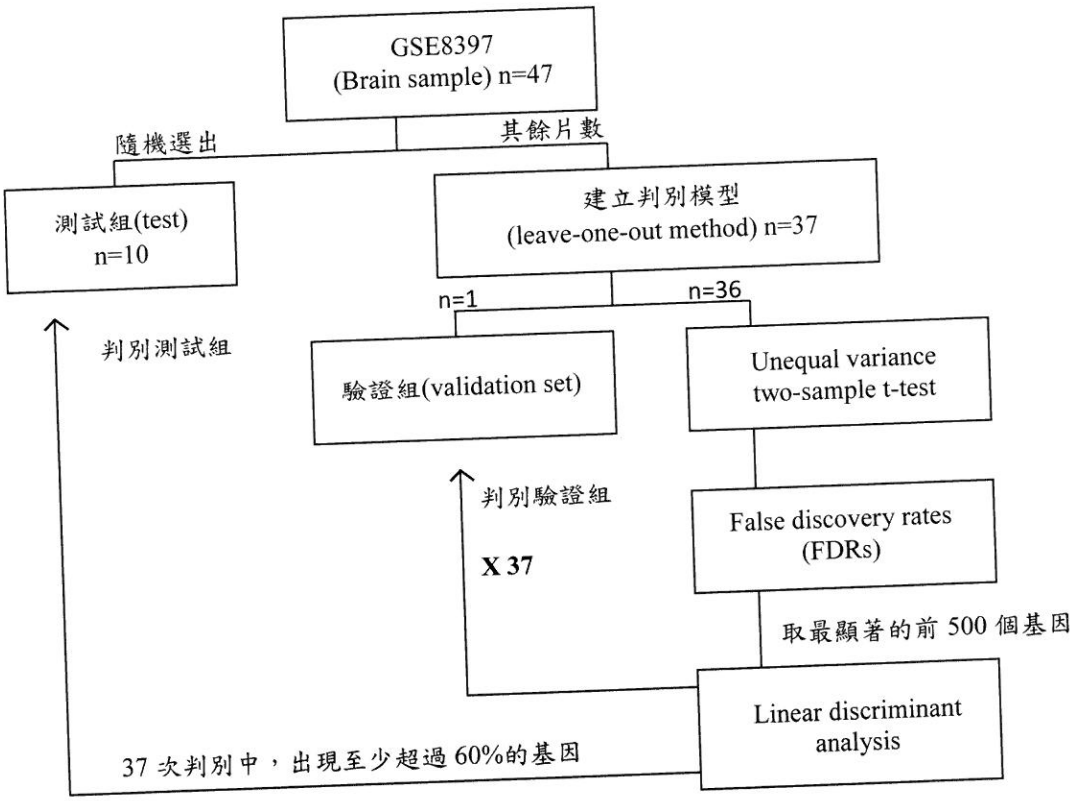
GO_ID	p.value	Cellular component
5622	0.0224	intracellular
16020	0.02714	membrane

C. 顯著偏向的 Molecular function 基因功能

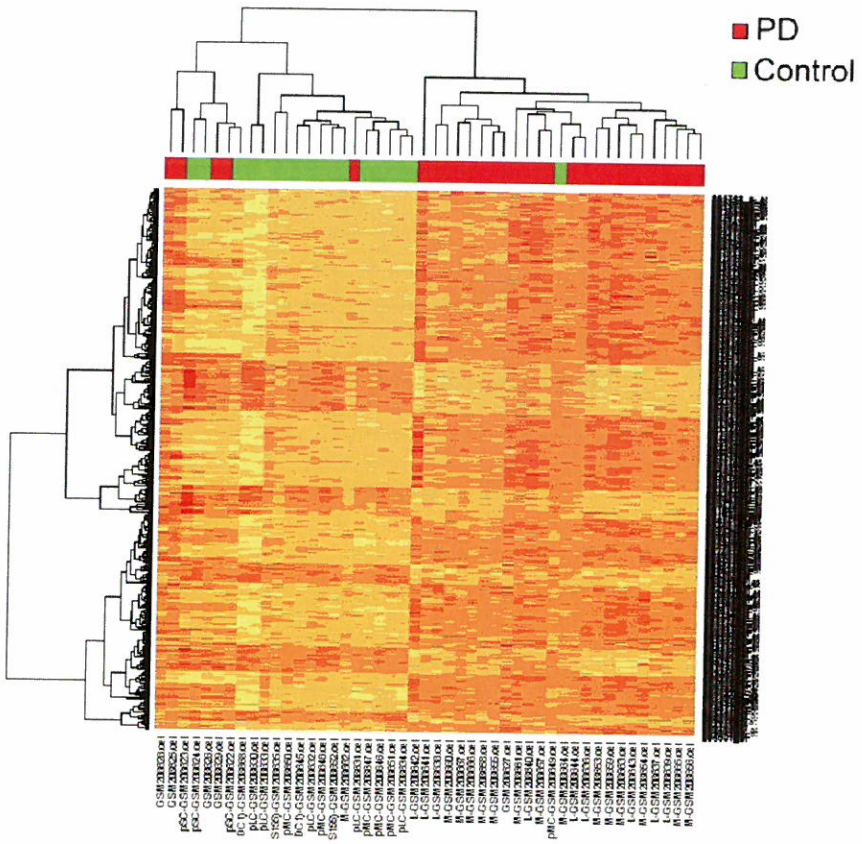
GO_ID	p.value	Molecular function
3700	0.0119	transcription factor activity
3779	0.0110	actin binding
5215	0.0299	transporter activity
5515	0.0018	protein binding
8270	0.0011	zinc ion binding
16787	0.0134	hydrolase activity
46872	0.0132	metal ion binding

表四、GSE8397 與 GSE6613 交集之相關性基因

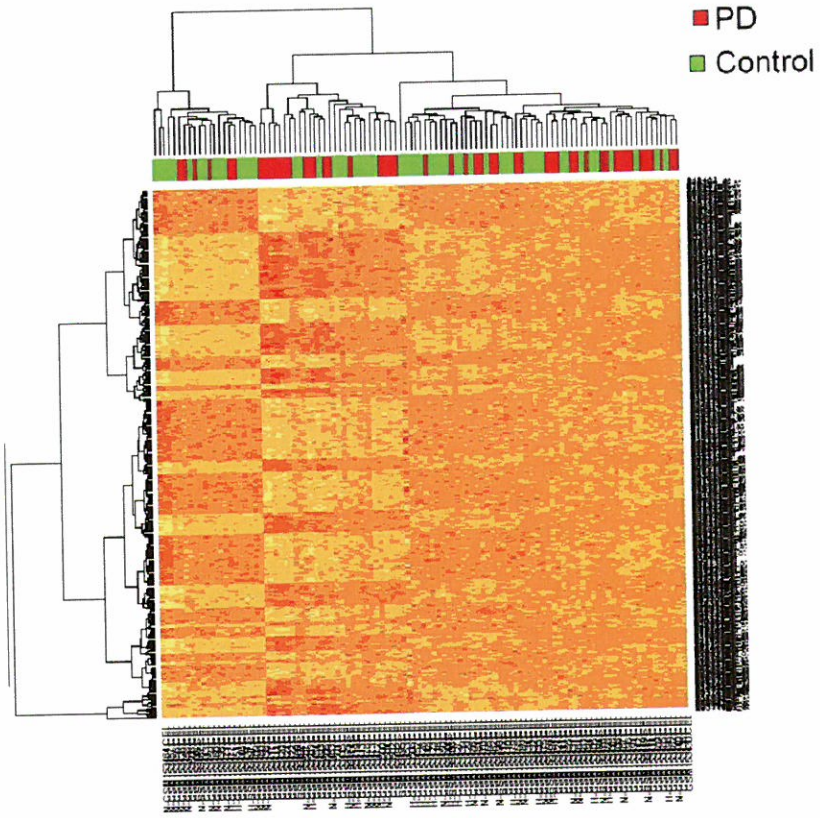
Probe set ID	Gene.Title	Gene.Symbol
200740_s_at	SMT3 suppressor of mif two 3 homolog 2	SUMO2 ; SUMO3
201873_s_at	ATP-binding cassette, sub-family E (OABP), member 1	ABCE1
202432_at	protein phosphatase 3, catalytic subunit, beta isoform	PPP3CB
202956_at	ADP-ribosylation factor guanine nucleotide-exchange factor 1(brefeldin A-inhibited)	ARFGEF1
203001_s_at	stathmin-like 2	STMN2
203738_at	chromosome 5 open reading frame 22	C5orf22
204224_s_at	GTP cyclohydrolase 1	GCH1
205073_at	cytochrome P450, family 2, subfamily J, polypeptide 2	CYP2J2
207070_at	retinal G protein coupled receptor	RGR
208291_s_at	tyrosine hydroxylase	TH



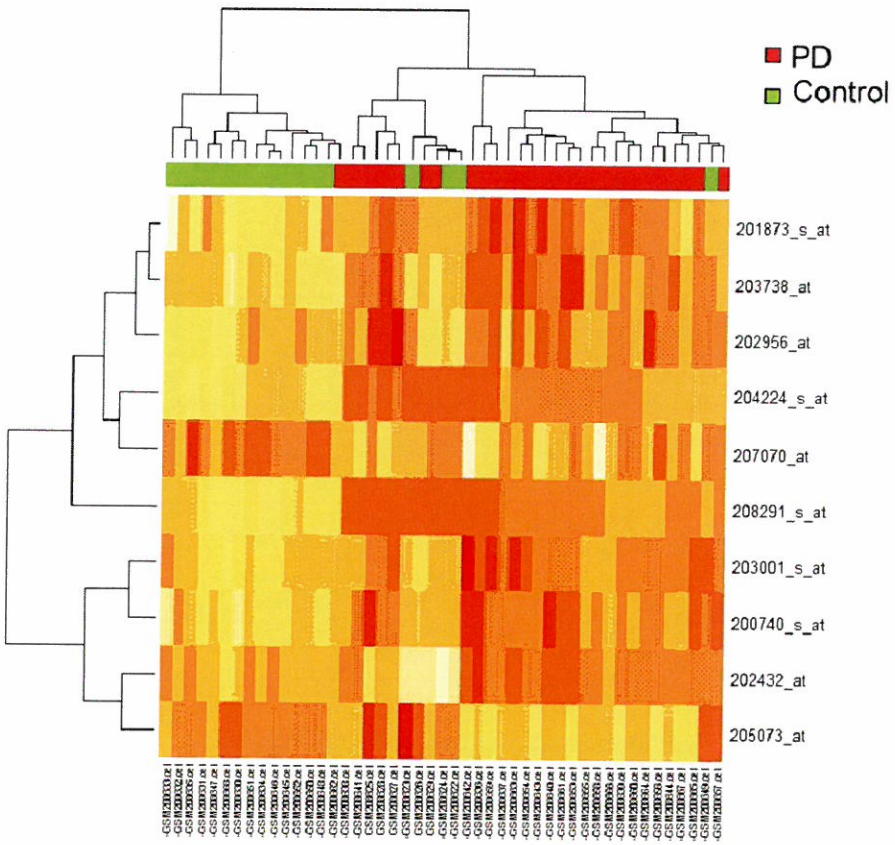
圖一、GSE8397 建立判別模式流程，先將 47 片隨機分成試驗組 (10 片) 與學習組 (37 片)，學習組再以 leave-one-out 方法，逐次抽出一個樣本當作驗證組 (validation set)，其餘樣本則用以挑選最顯著基因並建立判別函式，並對其驗證組進行判別，直到每個樣本都被抽出驗證。在這 37 次的判別分析中，顯著基因出現超過 60% 者再次建立判別式，用以判別 10 片試驗組。



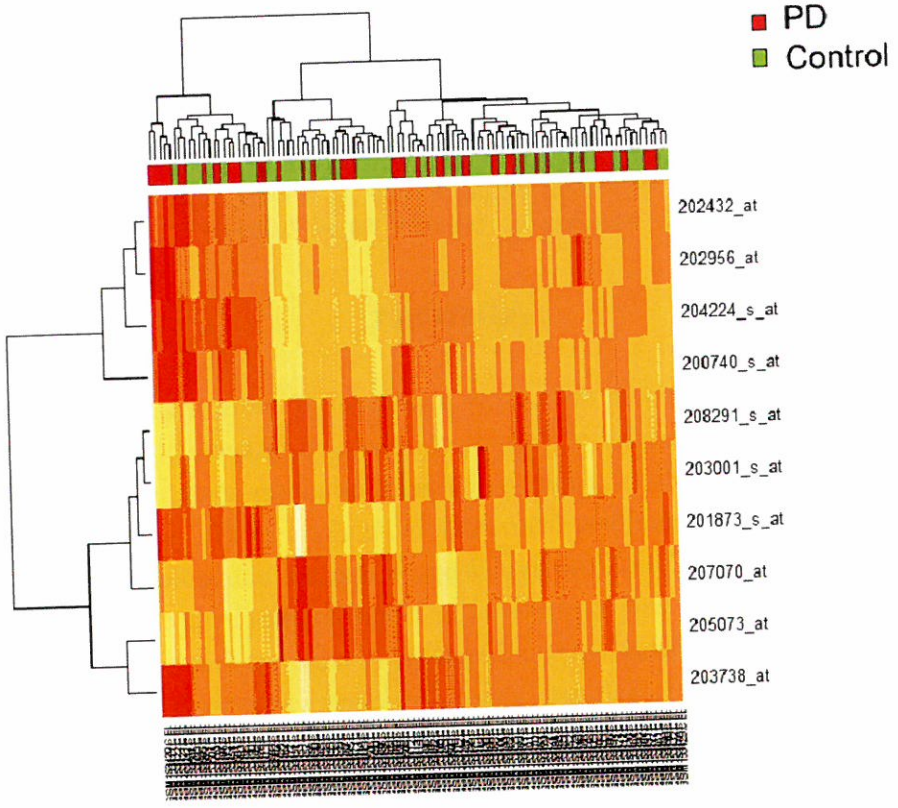
圖二、經由判別分析篩選出 439 個顯著性基因對 GSE8397 資料建立的 heatmap 圖，橫軸為 47 片樣本，縱軸為 439 個顯著基因的 probeset ID，圖中紅色部分代表 PD 患者樣本，綠色部分代表 Control 樣本



圖三、經由判別分析篩選出 370 個顯著性基因對 GSE6613 資料建立的 heatmap 圖，橫軸為 105 片樣本，縱軸為 370 個顯著基因的 probeset ID，圖中紅色部分代表 PD 患者樣本，綠色部分代表 Control 樣本



圖四、利用十個交集基因對 GSE8397 資料建立的 heatmap 圖，橫軸為 47 片樣本，縱軸為十個基因的 probeset ID，圖中紅色部分代表 PD 患者樣本，綠色部分代表 Control 樣本



圖五、利用十個交集基因對 GSE6613 資料建立的 heatmap 圖，橫軸為 105 片樣本，縱軸為十個基因的 probeset ID，圖中紅色部分代表 PD 患者樣本，綠色部分代表 Control 樣本

Received September 27, 2010
Revised November 29, 2010
Accepted December 20, 2010

Discriminant Analysis of Gene Expression in Parkinson's Disease

Ja-Ray Kang and Ai-Ling Hour

Department and Insititute of Lfe Science,
Fu-Jen Catholic University, Taipei, Taiwan.

Abstract

We have analysis microarray data from 47 brain tissue samples and 105 blood samples of Parkinson's disease's respectively. In order to detect patients from normal ones, the discriminant analysis was used to characterize candidate laboratory biomarkers of PD, and prove it by clustering analysis. The hypothesis is that the biomarkers resulted from brain tissue will also present in blood samples. And that will be advantageous in future clinical diagnosis. Finally, we find out the gene ontology terms of candidate genes, provided some useful index for researchers in medicine or biology in the future.

We have found out ten genes (SUMO2、ABCE1、PPP3CB、ARFGEF1、STMN2、C5orf22、GCH1、CYP2J2、RGR、TH) that have significant expression in brain and blood samples. The TH gene is related to the Parkinson's disease in previous papers. The discriminability of model reaches 90% for brain tissue. Although the discriminability of blood model is lower than brain tissue, there will be a potential possibility to find out the biomarkers from blood samples with more specific information included.

Keywords: Parkinson's disease, microarray, discriminant analysis, clustering analysis.

Efficient Relocation and Range Adjustment to Maintain Coverage in Wireless Sensor Networks

Chun-Hsien Lu and Chien-Hung Lin

Dept. of Computer Science and Information Engineering
Fu Jen Catholic University

Abstract

Due to fast development of wireless technology in recent years, wireless sensor network can be widely used in many applications, such as health care, home monitoring, environment monitoring, and battlefield surveillance. The network coverage percentage is an important issue because the information may not be considered useful when the total coverage drops below a certain level. Our goal is to move the sensor nodes around to maintain the network coverage over a given level as long as possible. In this paper, we propose an algorithm to keep good coverage. Whenever a sensor node stops working due to energy outage and a coverage hole appears, we select at most three nodes from its neighbors and instruct them to move toward the dead node. After reaching its target position, each selected node then enlarges its sensor radius to cover both its original area and a portion of the hole. Simulation results show that our method uses energy more efficiently and provides a system lifetime 30% longer than both the VFA [9] and EVFA [10] methods.

Keywords: wireless sensor network, network coverage, energy efficiency

Corresponding author:Tel:(02)2905-3874 Fax:(02)2902-3550
E-mail:jonlu@csie.fju.edu.tw

1. Introduction

Due to the fast development of embedded system technologies in recent years, wireless sensor network (WSN) has become one of the most important research areas. WSN can be used in widespread domains such as prevention of fire or other accidents, health care and environment monitoring, and battlefield surveillance [1-5]. In general, a WSN consists of one base station and many sensor nodes as shown in figure 1. The base station is in charge of collecting data while the sensor nodes are responsible for detecting events and reporting them in data packets sent to the base station.

A sensor node is usually equipped with a small-sized battery which can only provide limited amount of energy. Some sensor nodes will become dead due to zero energy or hardware breakdown after operating for a period of time. Dead sensors can cause holes in coverage or packet routing, which may lead to serious network degradation. [6-8].

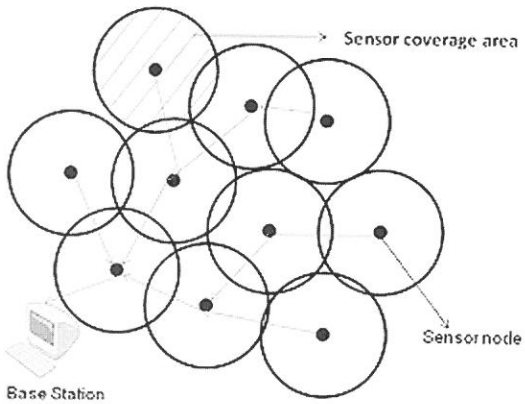


Figure 1. An example of wireless sensor network

The base station will not be able to get enough information when its coverage percentage is too low, and the system cannot perform its desirable function. Therefore, it is very important to maintain a high coverage. We would like to propose a mechanism that primarily focuses on maintaining a high coverage percentage as nodes may start to shut down one after

another due to zero energy. We assume that the sensor nodes are mobile, and can dynamically adjust their sensing ranges. When there is a dead node, a few sensor nodes will be selected to move toward the dead node to cover its sensing area such that a high coverage percentage can be maintained as long as possible.

The remaining of this paper is organized as follows: Chapter two lists the related work. Chapter three describes the mechanism we propose in detail and chapter four shows the performance evaluation. The conclusion is given in chapter five.

2. Related Work

Different mechanisms have been proposed that required the use of mobile sensors to maintain good coverage. One possible approach to find the target location of each sensor is based on the interactions between a sensor and all the other ones [9-11]. In Virtual Force Algorithm (VFA) [9], any sensor is assumed to generate attractive or repulsive force on the other sensors. There will be a repulsive force to separate any two sensors if they are too close, and an attractive force if they are far away. In VFA, the sensing ranges of the sensors are assumed to be all the same, thus nodes with lower energy will exhaust sooner. Energy-Considered Virtual Force Algorithm (EVFA) [10] was a distributed version of VFA. It suggests that a node with higher energy level should sense a larger region, while a node with lower energy should sense a smaller region. EVFA produces a more efficient energy use and prolongs the useful lifetime of the WSN. In both VFA and EVFA, a sensor node may zigzag to the target location in several rounds, which results in longer distance and wastes much energy.

Target Involved VFA (TIVFA) [11] was proposed to increase the detecting accuracy with hot spot areas in the network. A hot spot area represents an important place and thus requires a lot of sensors to sense the area. They present a probability model to ensure the continuity of the detection probability. The detection probability will be higher if the event point is closer to the sensor. They also propose a sensor ranking and protecting algorithm to improve the robustness of the sensor network. A sensor with a higher ranking value means that it is closer

to the hot spot and is considered a key node. If the number of key nodes reduces suddenly, redundant nodes will be added or waken up to cover the region accordingly. This method can provide a higher successful rate of sensing events due to better coverage percentage.

Another set of approaches try to move sensors to predefined regular positions such that they will spread out evenly. ISOmetric GRID-based algorithm (ISOGRID) [12] and Crystal-Lattice Permutation (CLP) algorithm [13] use the hexagon model and suggests the movement of sensors to the vertices of a hexagon. In ISOGRID, a specific node will first become a seed and instructs its neighbors to move to the vertex positions of the surrounding hexagon as shown in figure 2. Upon moving to the vertex position, a node becomes another seed and again starts to ask its neighbors to move to the vertex positions of another hexagon, etc.

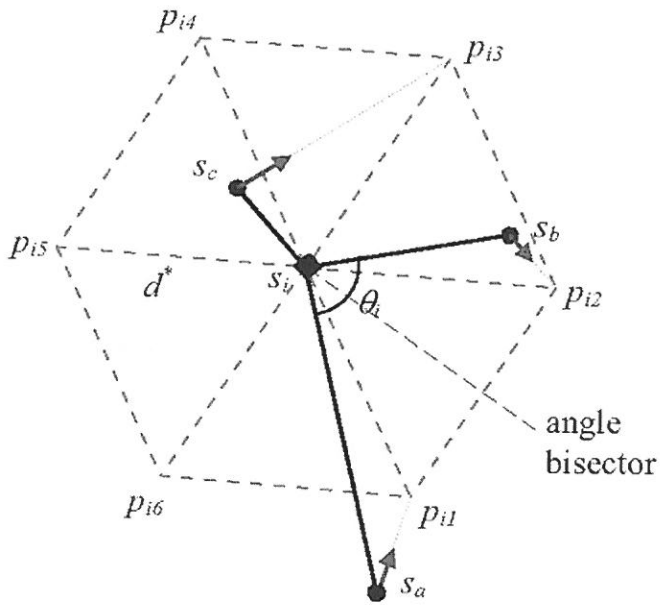


Figure 2. Forming hexagon topology [12]

Some other approaches have adopted both stationary and mobile sensors simultaneously to reduce cost [14][15]. The stationary sensors are randomly deployed first, while the mobile sensors can stay in a grid to decrease the load of stationary nodes, or travel around to move to

areas of low denser density to enhance coverage.

3. Efficient Relocation and Range Adjustment (ERRA) Algorithm

Whenever a sensor becomes dead, methods such as VFA will calculate a new target location for every other sensor in the network to move to. Because such frequent moving takes much energy, we hereby propose a mechanism called Efficient Relocation and Range Adjustment (ERRA) that tries to reduce moving while maintaining the coverage. Our system consists of a base station and many sensors nodes. The base station is stationary while the sensor nodes are all mobile. We assume that every sensor node knows its own location by GPS or any other locationing mechanism. Every sensor node can adjust its own sensing range R_s and communication range R_c. A sensor is assumed to transmit its packets to the base station using multi-hop transmissions along the path of minimum number of hops if such path exists. The sensor is considered disconnected from the base station if there is no path between them. Initially, the sensors can be randomly deployed, or a method based on VFA or hexagon model can be applied to make the sensors spread out more evenly. Every time a sensor node becomes dead and a coverage hole appears, we will instruct at most three neighbor sensors to move to cover the dead node. The new sensing ranges of those moving sensors will also be enlarged to cover the hole as much as possible. The operation of the ERRA mechanism includes the following phases:

A. Information Exchange

Initially each node will broadcast Hello messages to find neighbor nodes within its communication range. A Hello message contains the following information about the node itself: node id, node location, energy level, distance to base station, list of neighbor nodes and their distances, and the designated node in each section. Upon receiving a hello message, each node records the information in a table. After the information exchange, each node will select the neighbor node with the minimum distance to the base station as its next hop to transmit packets toward the base station.

We divide each node's (node g in figure 3, for example) communication region into three sections: The area between $-\pi/6$ and $\pi/2$ is called section_1, the one between $\pi/2$ and $-5\pi/6$ is called section_2, and the one between $-5\pi/6$ and $-\pi/6$ is called section_3. Because the energy for mechanical movement of mobile sensors dominates both the sensing and transmissions costs [10], we divide a sensor's neighborhood into merely three sections instead four or six in order to reduce the number of sensors that have to move to cover a vacancy. Node g calculates each neighbor's *ED*_value, which is defined to be the neighbor's energy value divided by the square of the distance between the neighbor node and node *g*. Node g then selects the neighbor with the maximum *ED*_value in each section as the designated node for that section.

$$ED\_value = \frac{Energy \quad value}{d^2} \tag{1}$$

When Hello messages are exchanged a few times, each node will know whether it has been selected as a designated node by a neighbor node or not, and it records the information in its own table. After the initial network establishment, sensor nodes start sensing events and sending data packets to the base station. Each sensor will also broadcast Hello messages periodically to maintain the network topology.

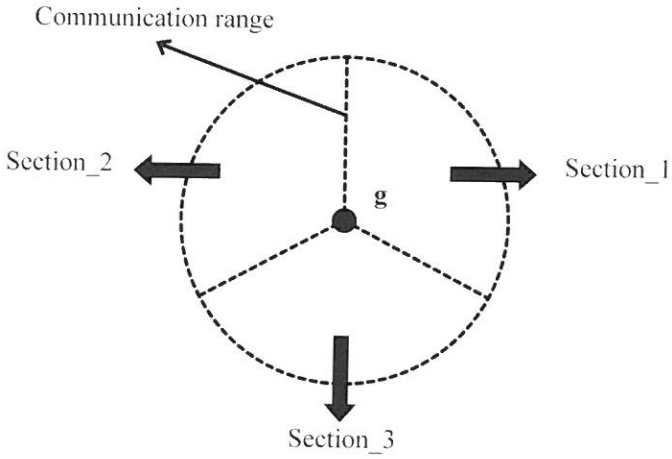


Figure 3. Division of neighborhood into three sections

B. Relocation

When a sensor node (e.g., node g) is about to consume all its energy, it will try to ask the designated neighbors to help. Let k be the total number of designated nodes in all three neighbor sections. If k is zero, there is no node to help; otherwise, k can be from 1 to 3. Node g will send a Cover_Request message to each of the k nodes, and designated node i uses the following formula to calculate the distance that it should move toward node g :

$$\text{Moving distance} = \rho \times dist(i) \times \frac{1}{k+1} \tag{2}$$

where

$$\rho = \frac{neighbor's\_energy}{average\_energy}$$

$dist(i)$ = the distance between designated node i and g

$average\_energy$ = average energy value of the k designated nodes.

If the neighbor nodes do not receive Hello messages from node g periodically, they will assume that node g is down, and will calculate the moving distances and move to the target location automatically. Figure 4 shows an example where node g has selected node a, node b, and node d as the designated node in each section, respectively If node g is about to consume all its energy, it sends a Cover_Request message to each of the three nodes, and each of them will calculate the moving distance individually and move to the target location.

C. Sensing Range Enlargement

After a designated node has received a Cover_Request message and moved to the target location, it has enlarge its sensing range to cover its original region plus a portion of the dead node's. The new sensing range should be set to the old sensing range plus the moving distance such that the original region is still covered. Figure 5 shows the topology where nodes a, b, and d had moved to the new locations and enlarged their sensing ranges. For example, assume that nodes a, b, and d in figure 4 all have the same amount of residual energy and form an equilateral triangle with a side length of x, and node g is located at its center. The distance between nodes a and g equals $(x / \sqrt{3})$, and each node will set its sensing range to $(1 / \sqrt{2}) * (x / \sqrt{3})$. When node g dies, each of nodes a, b, and d will move a distance of (1/4)*

$(x / \sqrt{3})$ toward the center, and changes its sensing range to $(1/4 + 1/\sqrt{2}) * (x / \sqrt{3})$which is about 35% larger than before.

Nodes *a*, *b*, and *d* again start finding new neighbor nodes by broadcasting Hello messages. Those nodes may have to increase their communication ranges large enough to find neighbors as well as routing packets to the next hop node on the path to the base station.

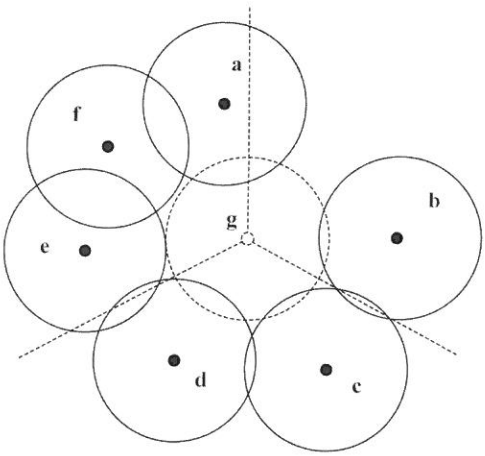


Figure 4. The topology before sensor g dies

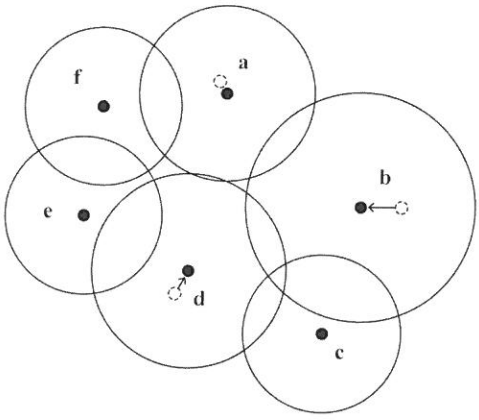


Figure 5. The topology after sensor g dies

4. Performance Evaluation

4.1 Simulation Environment

We implemented a simulator in JAVA platform to evaluate the performance of our method. We assume that the number of sensor nodes is 100 and initially they are uniformly deployed in a network of size 100 m 100 m. The maximum sensing range and communication range are set to 20 m and 40 m, respectively. The initial energy of a sensor node ranges from 150 J to 3000 J, where a node closer to the base station gets more energy because it has heavier traffic going through. Sensor moving consumes 27.96 J per meter [10]. We use the following formulas for the energy consumption [16] :

- Energy consumption for sensing: $E_s = \alpha \times r_s^2$ μJ/round
- Energy consumption for transmitting a packet: $E_t = E_c + \beta \times l \times r_c^2$ μJ/packet

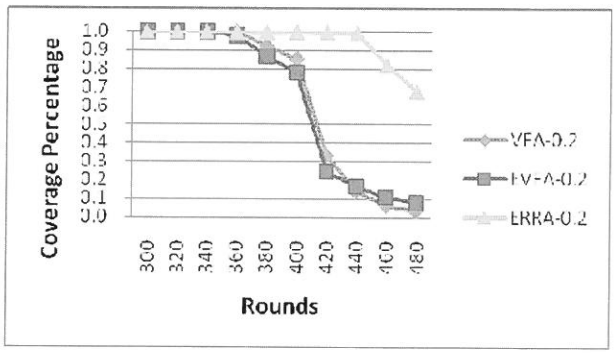
where α (set to 0.01μJ/m^2/round) and β (set to 0.05μJ/bit/m^2) are two adjustment parameters, r_s is the sensing range in meters, r_c is the transmission range in meters, E_c (= 0.05μJ) is the fixed part of energy consumption for transmitting a packet, and l is the packet length fixed at 512 bytes. The energy consumption of receiving a packet is assumed to be a fixed value of 0.205 J per packet.

In our simulation, we compared our method with VFA and EVFA with respect to the following performance metrics: the coverage percentage (which means the fraction of the network area that is covered by at least one sensor), the total residual energy over all the sensors, the number of connected nodes that still can send packets to the base station along a connected path, and the total number of packets successfully received by the base station.

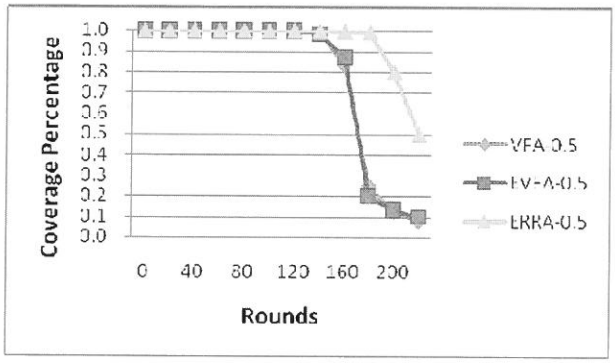
4.2 Simulation Results

Figures 6a to 6c display the coverage percentages when the traffic loads are 0.2, 0.5, and 0.8, respectively. (In figure 6a we only show the results after round 300 because the coverage does not change from round 0 to 300.) As a sufficient number of sensors run out of energy, the coverage percentage starts to drop below 100%. To compute the coverage of a given round, we only add up the areas covered by the sensors that are still connected to the

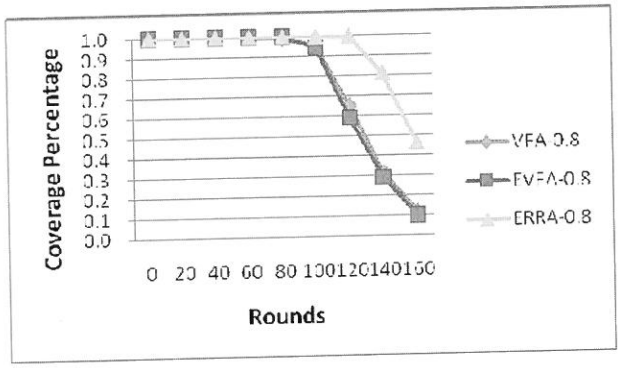
base station in that round. We can see that our ERRA scheme can maintain a good coverage 30% longer than both VFA and EVFA. This is because after moving to the new positions to cover a neighboring dead node, those designated sensors would enlarge their sensing ranges in order to keep the coverage as high as possible. In addition, sensors under ERRA can operate for a longer time since a lot of node moving has been eliminated. This leads to higher residual energy of the entire network for ERRA than the other two schemes as show in figure 7a to 7c.



(a)

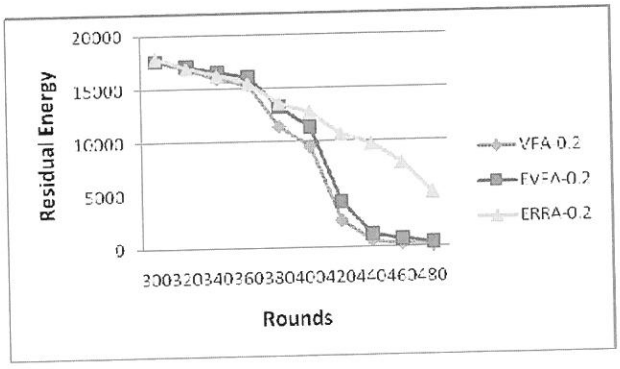


(b)

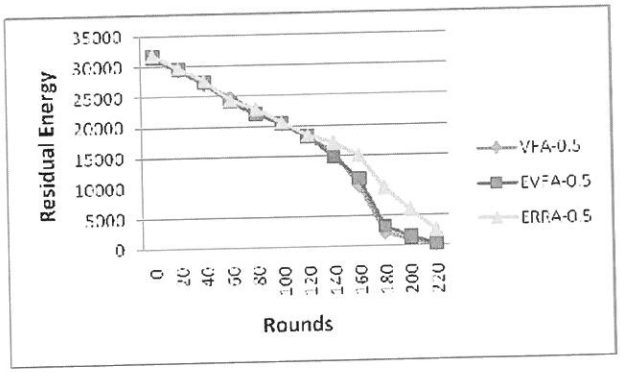


(c)

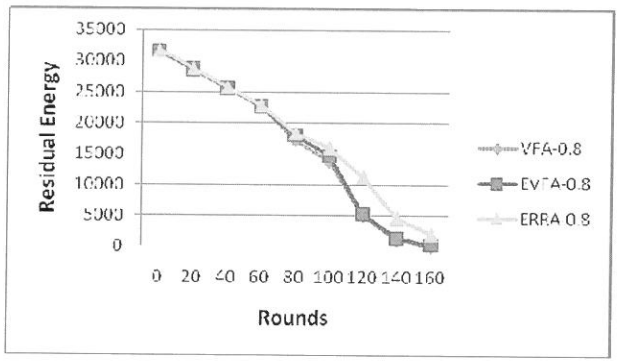
Figure 6. Coverage percentage, (a) load = 0.2, (b) load = 0.5, (c) load = 0.8



(a)



(b)

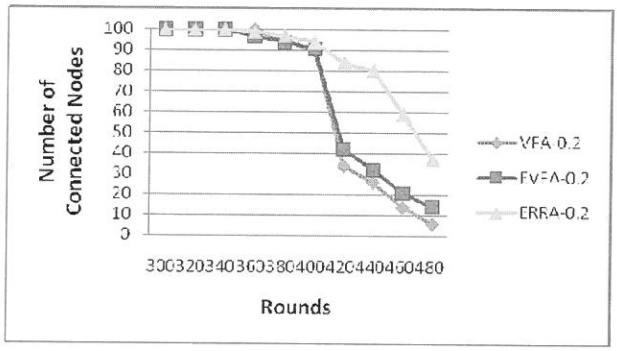


(c)

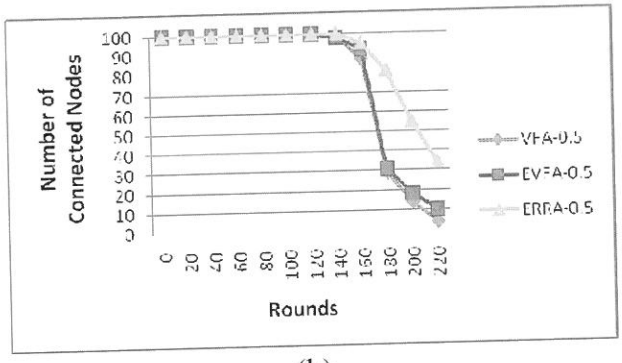
Figure 7. Residual energy, (a) load = 0.2, (b) load = 0.5, (c) load = 0.8

Figures 8a to 8c plot the number of nodes that are still connected to the base station in each round. We can see that ERRA has more connected nodes because its sensors always try to stay connected to the base station by increasing the communication range as large as necessary. Also sensors in EERA have more energy to stay connected longer since they have to do less moving.

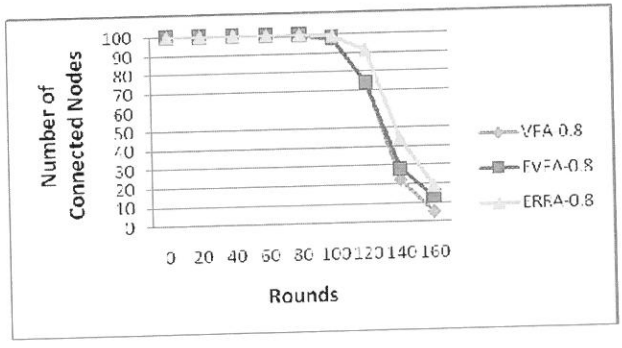
Figure 9 shows the total number of successful packets received by the base station for each method. Again more packets could be received under ERRA since it could keep more nodes connected for a longer time than other methods.



(a)



(b)



(c)

Figure 8. Number of connected nodes (a) load = 0.2, (b) load = 0.5, (c) load = 0.8

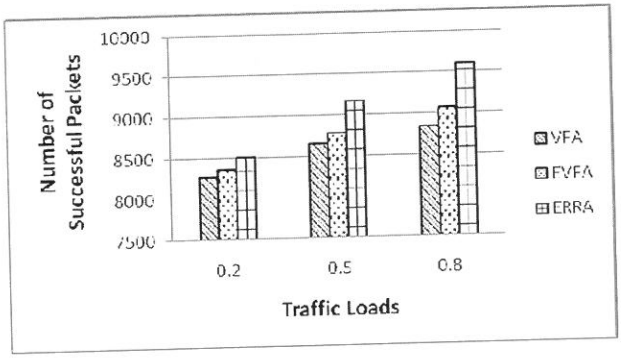


Figure 9. Total number of successful packets

5. Conclusion

Wireless sensor network has been widely used in many important applications, and it is essential to keep a good coverage for the network function properly. We proposed an efficient ERRA method to maintain the total coverage as high as possible when some sensor nodes start to run out of energy and become dead. Whenever a coverage hole appears, we only select at most three nodes with high ED_values to move and enlarge their sensing range to cover the hole. The other nodes save energy by not having to move. The simulation results showed that the ERRA could achieve higher coverage percentage and more successful packets than both the VFA and EVFA schemes. The operating lifetime of the system thus could be effectively prolonged by about 30%.

References

1. Juhyun Cho, Soonmok Kwon, Jaehoon Ko, and Cheeha Kim, "An Efficient Multilevel Coverage Scheme in Mobile Sensor Networks," *4th International Symposium on Wireless Pervasive Computing*, February 2009.
2. Min-Sook Jin, Hosung Park, Fucai Yu, Euisin Lee, Soochang Park, and Sang-Ha Kim, "Energy-Efficient Mechanism for Mobility Guarantee Used Location-Information in Wireless Sensor Networks," *4th International Conference on Wireless Communications, Networking and Mobile Computing*, October 2008.
3. Changsoo Ok, Pransenjit Mitra, Seokcheon Lee, and Soundar Kumara, "Distributed Energy Adaptive Routing for Wireless Sensor Networks," *IEEE International Conference on Automation Science and Engineering*, September 2007.
4. Tomoki Kimura and Iwao Sasase, "Energy Efficient Routing Protocol Adapted to Mobile Sensors for Multiple Sinks in Wireless Sensor Networks," *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2007.
5. Shuo Zhang, Yuheng Liu, Juhua Pu, Xiao Zeng, and Zhang Xiong, "An Enhanced Coverage Control Protocol for Wireless Sensor Networks," *Hawaii International Conference on System Sciences*, January 2009.
6. Lian Li, Limin Sun, Jian Ma, and Canfeng Chen, "A Receiver-based Opportunistic

Forwarding Protocol for Mobile Sensor Networks," *28th International Conference on Distributed Computing Systems Workshops*, June 2008.

7. F. Zabin, S. Misra, I. Woungang , H. F. Rashvand, N.-W. Ma, and M. Ahsan Ali, "REEP: Data-Centric, Energy-Efficient and Reliable Routing Protocol for Wireless Sensor Networks," *IET Communications*, September 2008.
8. Ying-Hong Wang, Chin-Yung Yu, Wei-Ting Chen, and Chun-Xuan Wang, "An Average Energy based Routing Protocol for Mobile Sink in Wireless Sensor Networks," *First IEEE International Conference on Ubi-Media Computing*, August 2008.
9. Yi Zou and Krishnendu Chakrabarty, "Sensor Deployment and Target Localization Based on Virtual Forces," *IEEE INFOCOM*, April 2003.
10. Younghwan Yoo and Dharma P. Agrawal, "Mobile Sensor Relocation to Prolong the Lifetime of Wireless Sensor Networks," *IEEE Vehicular Technology Conference*, May 2008.
11. Shijian Li, Congfu Xu, Weike Pan, and Yunhe Pan, "Sensor Deployment Optimization for Detecting Maneuvering Targets," *8th International Conference on Information Fusion*, July 2005.
12. Miu-ling Lam and Yun-hui Liu, "ISOGRID: an Efficient Algorithm for Coverage Enhancement in Mobile Sensor Networks," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2006.
13. Pang-Chieh Wang, Ting-Wei Hou and Ruei-Hong Yan, "Maintaining Coverage by Progressive Crystal-Lattice Permutation in Mobile Wireless Sensor Networks," *International Conference on Systems and Networks Communications*, October 2006.
14. Dan Wang, Jiangchuan Liu, and Qian Zhang, "Mobility-Assisted Sensor Networking for Field Coverage," I*EEE Global Telecommunications Conference*, November, 2007.
15. Ping Song, Jize Li, Kejie Li, and Li Sui, "Researching on Optimal Distribution of Mobile Nodes in Wireless Sensor Networks being Deployed Randomly," *International Conference on Computer Science and Information Technology*, September 2008.
16. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, October 2002, pp. 660 - 670.

Received September 27, 2010
Revised November 30, 2010
Accepted December 21, 2010

在無線感測網路中有效率的重新配置及半徑調整以維護涵蓋率

呂俊賢　林建宏

輔仁大學 資訊工程系

摘　要

近幾年來，由於無線感測網路迅速發展，可以應用的層面愈形廣泛，像是健康照顧，居家防護，環境監測，及戰場應用等。無線感測網路是由基地台及許多的無線感測器所組成，無線感測器負責感測事件及傳送資料封包至基地台加以收集整理。通常來說，無線感測器主要是使用電池來提供電力，電池通常是無法再充電，因此衍生出許多問題待解決。其中網路涵蓋率為一個重要的研究議題，因為有良好的涵蓋率才可提供充足的資訊。當部分無線感測器耗盡電力時，如何移動及調整其它無線感測器，使得整體網路涵蓋率維持在一定標準之上，是我們主要的研究目標。在本篇論文中，我們提出一個維持良好涵蓋率的演算法。當某一感測器耗盡電力因而造成一個涵蓋空洞時，我們從其鄰近感測器中選取最多三個節點，指揮其向空洞方向移動並放大感測範圍，將空洞覆蓋以達到維持涵蓋率的目的。模擬實驗的數據顯示我們提出的方法與 VFA 及 EVFA 兩種方法比較可以更有效地使用電力，延長系統壽命達 30%。

關鍵字：無線感測網路、網路涵蓋率、能源效率。

影像資料管理平台[1]

鄒濟鴻　徐嘉連[2]
輔仁大學資訊工程學系

摘　要

在多媒體產業日益發達的今日，多種數位商品不斷地問世，其中數位相機更是推陳出新，傳統的相片亦逐漸被數位相片所取代，在這數位化的過程中，傳統的關鍵字搜尋的技術已逐漸不符需求，面對日漸增加的數位影像資料，更需要精確且便捷的系統管理。因此，本篇論文針對此開發一套以 MPEG-7 為資料描述格式之「影像資料管理平台」，並以 MPEG-7 Visual Descriptors 中的 Edge Histogram、Angular Radial Transform（ART）及 Curvature Scale Space（CSS）為資料比對之特徵模型，在實驗中，系統使用這三種特徵所得的 11-Point Average Precision 最高可達九成的準確度，而整體的平均值亦達七成以上。

關鍵詞： MPEG-7、內涵式影像查詢（Contend-Based Image Retrieval）、徑度角度轉換（ART）、曲度空間（CSS）、邊線直方圖（Edge Histogram）。

[1] 本論文，為輔仁大學（計畫編號：409931044034）與
國科會（國科會計畫編號：NSC-98-2622-E-030-001-CC3）之補助研究。

[2] 論文通訊作者，電子郵件：alien@csie.fju.edu.tw

1. 前　言

多媒體產業的產品已大量進行數位化，諸如音樂 CD（WMA、MP3 等），影像（JPEG、BMP、GIF、TIFF、PNG 等）及影片（AVI、WMV、MP4 等），其中數位相機已逐漸取代傳統相機，故傳統的相片更逐漸以數位影像的型式呈現。

但是在這樣的趨勢下，勢必會產生大量的數位影像，因此面對這眾多的資料，如何去搜尋及管理是現在所需面對的課題；目前市面上所有的資料搜尋方式大多是以關鍵字查詢（Query-by-Keyword）來達成，這類搜尋的方式，在使用上使用者必需針對影像鍵入相關的文字標籤（tag）才可能達成，這樣的查詢方式有很大的限制，故內涵式影像查詢（Content-Based Image Retrieval）已成為國內外需求的研究課題。

另外在這一個網路發達的時代，社交活動並不一定要面對面，透過網路的社交平台（如 Facebook、MSN 及 Plurk 等）進行社交活動為目前盛行的網路活動，而在這類的社交平台中分享數位相片亦為主要的活動之一，故如何在不同平台間交換資料也是一項值得考慮的需求，在這樣資料交換的需求下 Moving Picture Experts Group（MPEG）於 2001 年 9 月所提出的 MPEG-7 是一個值得參考的資料描述結構。

MPEG-7 為一套開放式的多媒體資料描述格式（相關資訊描述於之後的章節中），透過這一套標準，使用者可以利用 MPEG-7 所定義的特徵模型來描述多媒體資料，亦可依照 MEPG-7 的格式自行定義新的特徵模型來達成資料交換的目的，因此在考慮到資料管理及交換的需求下，本論文開發一套「影像資料管理平台」，系統使用 MPEG-7 為資料描述（Metadata）格式，提供一個擴充性高及具備跨系統交換資料的資料管理平台，而在影像特徵選擇方面，初期採用 MPEG-7 Visual Descriptors 中的 Angular Radial Transform（ART）、Curvature Scale Space（CSS）及 Edge Histogram 作為內涵式影像查詢之相似度比對模型，未來依需求加入新的影像特徵來增加系統的附加價值。

本篇論文的結構如下，第二章討論目前市面上使用者常用的影像搜尋、管理平台及國內外關於內涵式影像查詢的研究，第三章則著重於 MPEG-7 的介紹，第四章介紹系統架構及特徵擷取演算法的實作，第五章為系統的實驗及結果討論，第六章為本文的結論及未來研究的方向。

2. 相關研究

在這一章中，我們將會介紹目前市面上的影像管理、搜尋系統（以 Facebook 相簿及 Google 圖片搜尋系統為例）及六篇 [12][13][14][15][16][17] 國內外內涵式影像查詢及 MPEG-7 相關研究。

（一）影像管理及搜尋系統

在現今的社會中，透過社群網路系統（如 Facebook、MSN、Plurk 等）聯絡親朋好友已成為許多人在日常生活中重要的活動之一，而其中照片分享亦是這類社群網路系統的基本功能之一，而這樣的相片分享功能主要的管理大多是使用傳統相簿的分類概念（如圖 1），當使用者要尋找其中一張相片時必須從每一本相簿中逐一瀏覽才可以找出所需的相片，系統本身並未提供任何的搜尋工具。

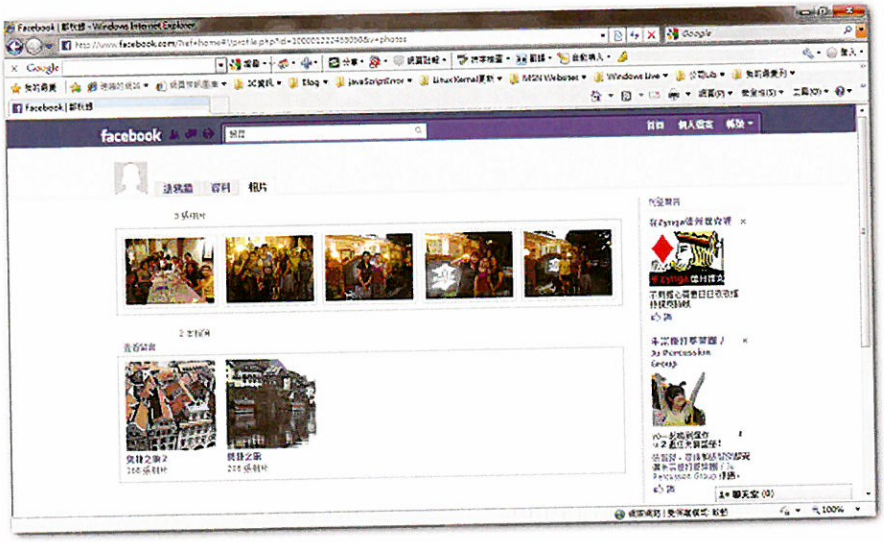


圖 1：Facebook 相簿

另一方面，提到網路搜尋引擎，Google 可說是目前市佔率最高的搜尋引擎，其在圖片搜尋的功能中提供了關鍵字查詢（Query-by-Keyword）及範例式查詢（Query-by-Example）兩種搜尋方式，在搜尋介面的使用上，使用者可於介面上所提供的文字方塊（TextBox）輸入所要查詢的關鍵字，即可取得相關的圖片，而從介面中取得搜尋結果

後，使用者可進一步從左下方的介面中使用其他的影像資料（如大小、類型、色系等）再一次過濾搜尋結果，另外亦可從搜尋結果中選擇其中一張圖片使用“類似圖片”的連結取得範例式查詢的結果。

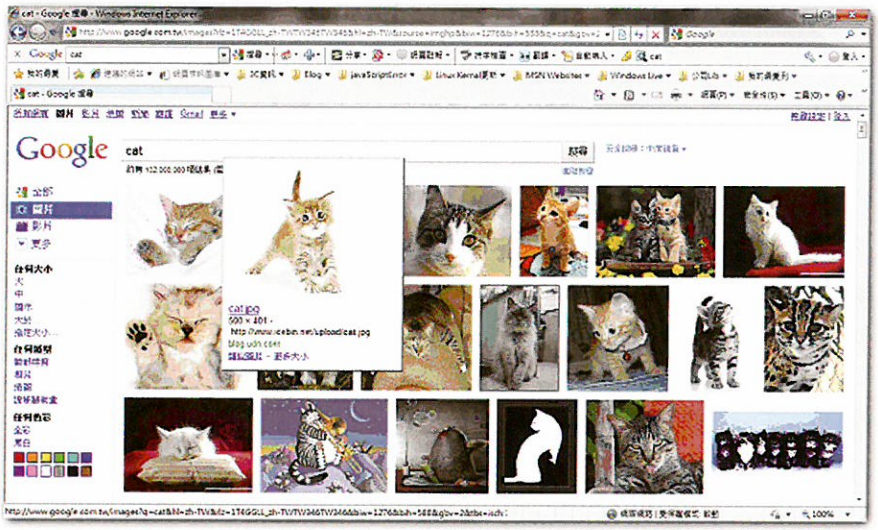


圖 2：Google 圖片搜尋介面

（二）國內外相關研究

在 MPEG-7 及內涵式影像查詢的研究上，[14] 的研究致力於 Descriptor 的設計，目的在於建立新的特徵模型，並套用於相關的應用的研究，在這一項的研究中，作者定義了 Compactness Descriptor（CD），大致的內容是描述物件前景的 Pixels 佔整個物件的比率，再配合 Scalable Color Descriptor（SCD）及 Edge Histogram Descriptor（EHD）將查詢影像與 Training Data Set 中比對，取出相似度最高的前一百張圖片，藉以識別查詢影像是否為色情圖片，另外在 [12][16] 的研究中分別修改 MPEG Visual Descriptors 中的 Angular Radial Transform（ART）及套用 Edge Histogram 來達成商標檢索及手繪影像輸入查詢，且都有很高的辨識度，而在 [13] 的研究中，利用 Angular Radial Transform（ART）及 Curvature Scale Space（CSS）以半自動的方式建立各類蝴蝶影像的特徵樹，在 [16] 的研究目的在於建立一套大型影像資料庫的查詢模型，但是對於相似度比對模型上並未有太多的描述。

（三）系統比較

在 [12][13][14][15][16][17] 的研究中皆針對內涵式影像查詢提出了一個可參考的結構，但是相對的在使用上各有各的限制，此外，Google 率先於搜尋引擎中導入內涵式影像查詢的技術，可謂是在影像搜尋引擎技術中建立一個新的里程碑，但是 Google 並未提供使用者影像輸入介面，使用者必需在 Google 圖片搜尋介面上先以關鍵字查詢取得所需的圖片才可以進一步的使用範例式查詢的功能，在使用上就多了一層限制，現針對上述各項系統的比較整理如表一。

表一：相關系統比較表

系統 / 論文名稱	影像輸入介面	內涵式影像查詢	優點	使用限制
Facebook 相簿	V		一個相片管理的平台	無搜尋介面
Google 圖片搜尋		V	率先於搜尋引擎中導入內涵式影像查詢技術	無影像輸入介面，僅可使用 Google 現有資源
以影像擷取技術為基礎之圖片篩選系統（楊長慎 -中華大學）[14]	V	V	提出了一個以膚色偵測演算法做為前景及背景分離的手段，並提出了一套新的特徵定義法則，為影像檢索系統提出了另一套實作模型	單一色系做為前景或背景的參考值，無法應用在其他類型的影像中
以 MPEG -7 為基礎之蝴蝶影像形狀檢索（黃子癸 -台灣大學）[13]	V	V	針對蝴蝶的細部特徵建立資料分類，可提高影像查詢的辨識度	特徵樹的建立尚需人工的協助
以影像基本圖案及輪廓特徵建立內函商標檢索系統（蔡昕峻-義守大學）[12]	V	V	基本幾何圖形做為初步過濾的手段，提供一個影像資料分類的模型	特徵的定義上需先定義 Training DataSet，再以 Training Data 的 Index 做為特徵，是以特徵向量的大小取決於 Training DataSet的大小
Cortina: a system for loge-scale , content-based web image retrieval（T. Quack,et. al.）[15]	V	V	建立CBIR系統參考的架構	對於相似度的比對及特徵擷取方面並未有太多的說明
A descriptor for large scale image retrieval based on sketched feature line（M Eitz, et. al.）[16]	V	V	建立了一套以手繪圖片為主要查詢的影像檢索系統，在 CBIR 的系統上，提出了一套的架構	系統結構專注於紋理的比對方面，在風景照及素描圖片可以得到很好的結果表現，但是紋理特徵對於旋轉圖片的抵抗力略嫌不足

3. MPEG-7 簡介

MPEG-7 為 MPEG（Moving Picture Experts Group）於 2001 年 9 月所發表的多媒體資料描述格式，為一種開放式的多媒體資料描述語言及多媒體資料交換介面，現將其內容做一個簡介。

（一）Overview

MPEG-7 的正式名稱為 Multimedia Content Description Interface，它提供了一套豐富的資料描述標準，讓不同的系統之間可以透過這一套標準來進行影音資料的交換作業，其中包含了 7 個主要結構，分別為 System、DDL（Description Definition Language）、Visual、Audio、Multimedia Description Schemes、Reference Software 及 Conformance 等，內容可包含圖片、聲音、影片、3D 模組及作品相關資訊，儲存格式可為文字（DDL）或 BiM（binary format for MPEG-7）的形式，其中 DDL 為 MPEG-7 的資料描述定義語言，其內容以 XML 為基底，其結構包含 Descriptor（D）及 Description Scheme（DS），DDL 的結構關係如圖 3 所示，Descriptor 為定義所要描述的多媒體資料的特徵參數（如顏色、外形及紋理等）及內容（圖 3 左下區塊），而 DS 則為這些特徵定義的集合（圖 3 中間下方區塊），其內容可包含多媒體資料特徵參數的集合，與相關的文字資料，如作品名稱、作者等資料；其中要注意的是 Descriptor 的資料定義中相同的特徵參數不可分散於其他的節點（Element）中，必須集中於同一節點中定義，而 DS 則可使用繼承（Extend）及參考（Reference）的概念，故 DS 的資料定義可分散於不同的節點中描述。

而常見的 MPEG-7 DDL 與多媒體系統間的應用（如圖 4），影像應用實例（如圖 5），對應的 XML（如圖 6）。首先系統依預先定義好的 Descriptors 執行特徵擷取，並依 Description Scheme 之定義產出對應的 XML 文件，再進行資料的傳輸及存儲作業，最後依各系統的設定執行後續流程如搜尋、播放等。

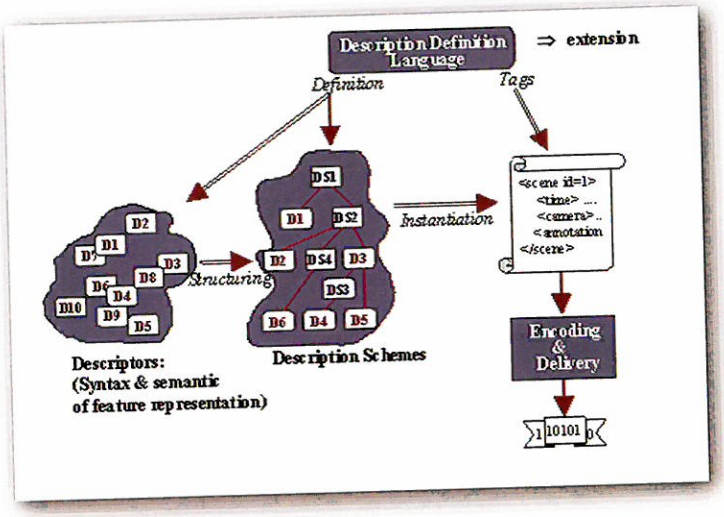


圖 3：DDL 的結構關係圖 [10]

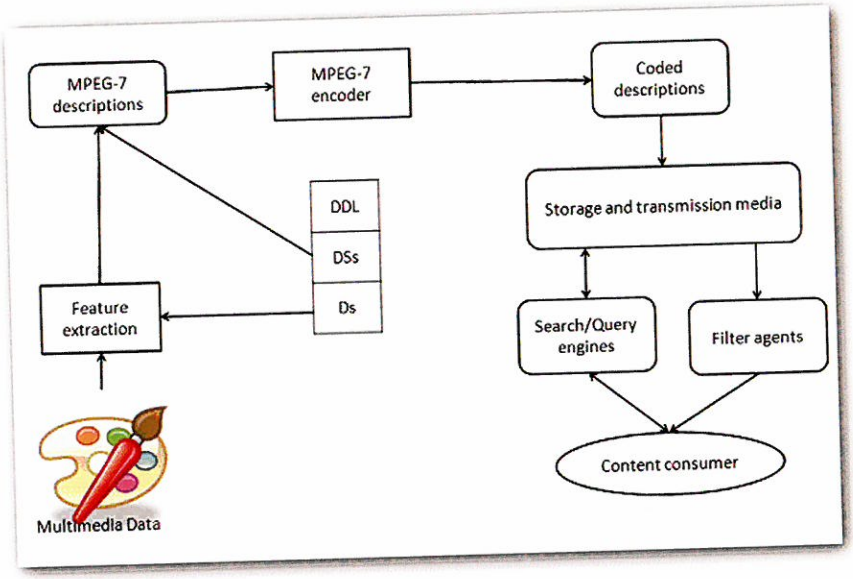


圖 4：MPEG-7 應用結構 [11]

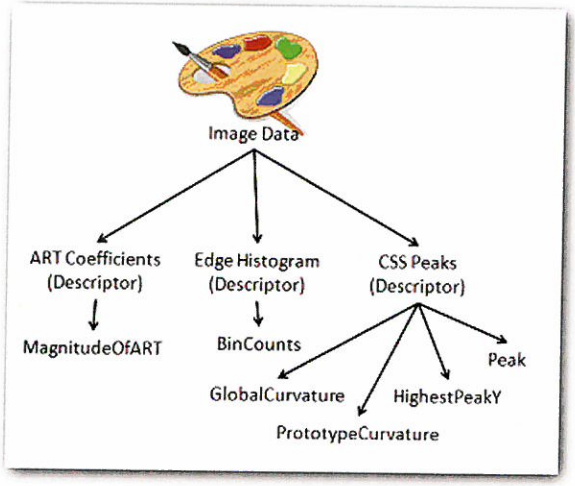


圖 5：MPEG-7 影像資料應用實例

```
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
       xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 .\Mpeg7-2001.xsd">
  <Description xsi:type="ModelDescriptionType">
    <Model xsi:type="DescriptionModelType">
      <Descriptor xsi:type="RegionShapeType">
        <MagnitudeofART>15 5 15 15 5 14 13 2 12 9 1 7 5 4 11 9 6 12 11
                        6 11 10 5 9 8 4 9 8 4 11 10 5 11 10 5</MagnitudeofART>
        <Field>MagnitudeofART</Field>
      </Descriptor>                                          ART Coefficients
      <Descriptor xsi:type="ContourShapeType">
        <GlobalCurvature>63 8</GlobalCurvature>
        <PrototypeCurvature>63 7</PrototypeCurvature>
        <HighestPeakY>95</HighestPeakY>
        <Peak peakX="15" peakY="5" />
        <Peak peakX="8" peakY="6" />
        <Peak peakX="39" peakY="3" />
        <Field>GlobalCurvature</Field>
        <Field>PrototypeCurvature</Field>
        <Field>HighestPeakY</Field>
        <Field>Peak</Field>
      </Descriptor>                                          CSS Peaks
      <Descriptor xsi:type="EdgeHistogramType">
        <BinCounts>3 0 7 6 0 0 6 0 0 0 0 0 6 0 0 1 0 6 2 0 1 4 7 0 0 0 0
                   0 0 0 1 4 7 0 0 0 4 0 0 2 0 0 4 0 0 3 7 0 0 0 0 6 7 0 0
                   2 6 6 0 0 0 0 0 0 4 0 0 0 0 0 7 0 0 0 0 0 0</BinCounts>
        <Field>BinCounts</Field>
      </Descriptor>                                          Edge Histogram
    </Model>
  </Description>
</Mpeg7>
```

圖 6：MPEG-7 Description

(二) Visual Descriptors

MPEG-7 Visual Descriptors 為 MPEG-7 所定義的視覺化的影像特徵，其中整理內涵式影像查詢常用的特徵（Feature），依據特徵屬性分為顏色（Color）、紋理（Texture）、外型（Shape）及動作（Motion），現分別對這幾類特徵集合整理如下（參考表二），並針對幾項特徵模型介紹如下：

表二：Some descriptors of MPEG-7

Descriptors	說明
Color D escr iptors	
	描述色彩與空間的關聯性
Scalable Color Descriptor	描述顏色在影像中分佈特性，資料特徵以Color Histogram呈現，色彩空間為HSV，並以Haar Transform進一步將256階的資料壓縮成16階、32階、64階等
Group -of-Frame or Group -of-Picture Descriptor	用於影片的資料上，其描述一個影片中不同的images或frames之間色彩連結的關係，通常來說images或frames的選擇主要是挑選影片中的關鍵影像（Key image or Key frame）
Color Structure Descriptor	描述影像在整體圖片及區域空間與色彩間的特徵關係
Color Layout Descriptor	描述區域性的顏色特徵，其擷取方式建立於DCT的轉換上
Texture Descriptors	
Homogeneous Texture Descriptor	描述規律性的紋理特徵，利用2-D FFT將2-D平面的頻率空間分為30個頻帶，並計算每一個頻帶的能量及頻帶能量的誤差值
Texture browsing Descriptor	描述影像中規律性紋理特徵的方向，粗細度及規律性
Edge Histogram Descriptor	描述非規律率性的紋理特徵，內容介紹於之後的章節

Descriptors	說明
Shape Descriptors	
Region -Based Shape Descriptor	用以描述區域性的形狀特徵，內容介紹於之後的章節
Contour -Based Shape Descriptor	用以描述物件輪廓的形狀特徵，內容介紹於之後的章節
	描述 3-D 平面的物件特徵
Motion Descriptors	
Motion Activity Descriptor	描述 Motion 的飽和度 (intensity) 、速度 (pace) 、色調 (mood) 等特徵的結構
Camera Motion Descriptor	描述鏡頭的特徵如遠近、視角等結構
Motion Trajectory Descriptor	描述在影片中與視訊圖片沒有相互依存關係的物件，如字幕、聲音等物件

接下來，我們針對本系統中目前所用採用的 Descriptors 做進一步地詳細介紹：

A.Edge Histogram Descriptor

Edge Histogram Descriptor（EHD）為 MPEG-7 針對非規律性的紋理所定義的特徵描述方式，其主要方式系將一張圖片分割成 4X4 的區域（如圖 7），然後針對每一個區域分別統計 0,45,90,135 及不屬於上列之角度之線段數量。

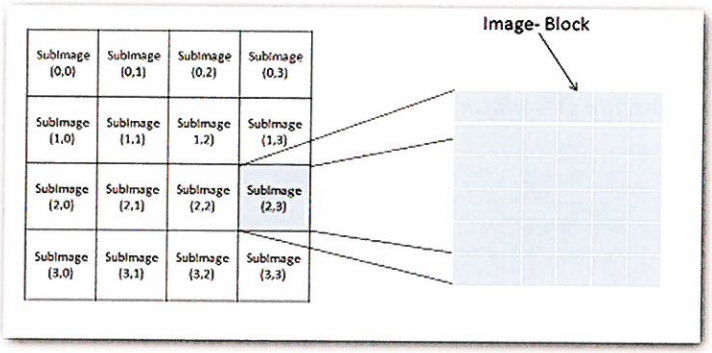


圖 7：EHD 影像分割示意圖 [1]

B.Region-Based Shape Descriptor

Region-Based Shape Descriptor 主要是透過 Angular Radial Transform（ART）Coefficients 來描述影像或物件特徵，其定義如「式 1」，其中 $f(\rho,\theta)$ 為影像在極座標 (ρ,θ) 上的強度，$V_{nm}(\rho,\theta)$ 為 MPEG-7 所定義的 Basic Function，為複數資料，呈現結果如圖 8 所示，（a）為實部資料，（b）為虛部資料，F_{nm} 為即為 ART Coefficients。

$$F_{nm} = (V_{nm}(\rho,\theta), f(\rho,\theta))$$

$$= \int_0^{2\pi}\int_0^1 V_{nm}^*(\rho,\theta), f(\rho,\theta)\rho d\rho d\theta \quad (1)$$

$$V_{nm}(\rho,\theta) = A_m(\theta)R_n(\rho) \quad (2)$$

$$A_m(\theta) = \frac{\exp(jm\theta)}{2\pi} \quad (3)$$

$$R_n(\rho) = \begin{cases} 1 & n = 0 \\ 2\cos(\pi n\rho) & n \neq 0 \end{cases} \quad (4)$$

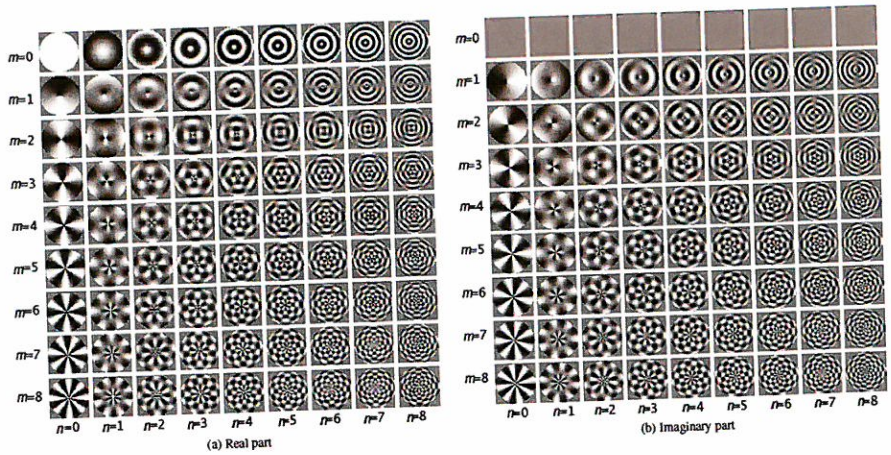


圖 8：ART Basic Function[1]

C.Contour-Based Shape Descriptor

Contour-Based Shape Descriptor 主要透過 Curvature Scale Space（CSS）來呈現，其主要是利用 arc length 原理來取得 2-D 平面上的斜率變化，並將其結果投射在 1-D 的座標上，計算方式概述如下：

1. 取得邊緣點集合 C (u) = {x (u), y (u)}

$$C'(u) = \{x(u,\rho), y(u,\rho)\} \tag{5}$$

$$x(u,\rho) = x(u) * g(u,\rho) \tag{6}$$

$$y(u,\rho) = y(u) * g(u,\rho) \tag{7}$$

其中 g (u,ρ) 為低通濾波器，* 為 convolution 運算

2. 取得斜率變化 k (u,ρ)，計算方式如「式（8）」

$$k(u,\rho) = \frac{x'(u,\rho)\,y''(u,\rho) - x''(u,\rho)\,y'(u,\rho)}{(x'(u,\rho)^2 + y'(u,\rho)^2)^{3/2}} \tag{8}$$

當 k (u,ρ) ≧ 0 and k (u,ρ−1) < 0 則 k (u,ρ) 即為 Peak 資料，如圖 9 表示，左邊的圖形表示原始圖片經過 Low Pass Filter 後的結果，中間的圖形表示 k (u,ρ) 的演進流程，右邊的圖形表示每次經由 Low Pass Filter 後所取得的 Peak 值。

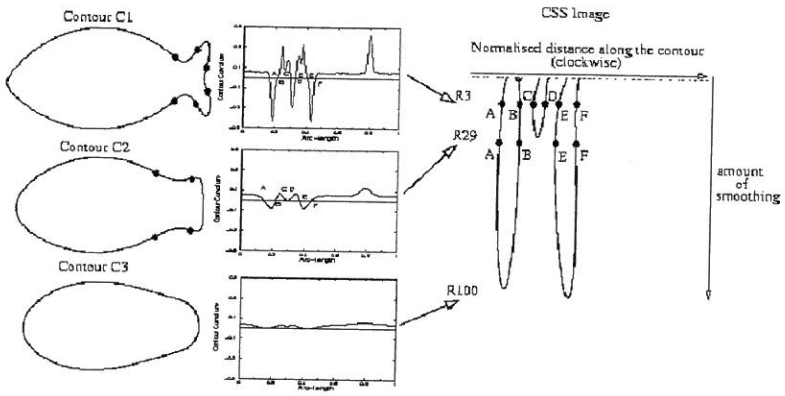


圖 9：CSS 演算示意圖 [1]

4. 系統設計與實作

在這一章中將會介紹系統設計結構及 MPEG -7 Visual Descriptors 演算法實作方式。

（一）系統架構

本系統之系統架構如圖 10 所示，大致可區分四個功能模組及資料庫系統，在影像資料庫方面，系統採用 Microsoft SQL Server，用以儲存影像資料及 MetaData，功能模組分別描述如下：

A. 使用者介面（Web API）

主要目的為提供給使用者的操作介面，其中包含管理模組及查詢模組，在本系統中，使用者介面是以 Web 的方式呈現，使用者透過瀏覽器即可進行系統操作，開發方式採用 Microsoft 之 ASP.NET 為網站開發工具。

B. 前置處理模組（Preprocessing Utility）

主要目的是在進行 MPEG-7 Visual Descriptors Feature Extraction 之前所針對影像的之前置處理，如邊緣偵測（Edge Detection），細化（Thinning）等，用以強化影像中高頻（邊線）的部分以及濾除低頻的部分，藉以提高 Edge Histogram 的辨識度，相關實作細節描述於之後章節中。

C. 特徵擷取（Feature Extraction）模組

為系統之核心模組，主要利用 MPEG-7 DDL 做為影像資料描述語言（Meta Data），特徵採用 MPEG-7 Visual Descriptors 所定義之 Edge Histogram、ART 及 CSS 三種特徵向量，相關實作細節描述於之後章節中。

D. 相似度比對模組（Searching Utility）

相似度比對依存於 MPEG-7 Visual Descriptors Module 之中，因每一個 Descriptor 所呈現的特徵都有所不同，因此所要進行相似度比對的模型也有所不同，在本系統中特徵的選用是採開放式的作法，使用者可依不同的查詢需求選用不同的特徵比對模型來進行相似度比對作業。

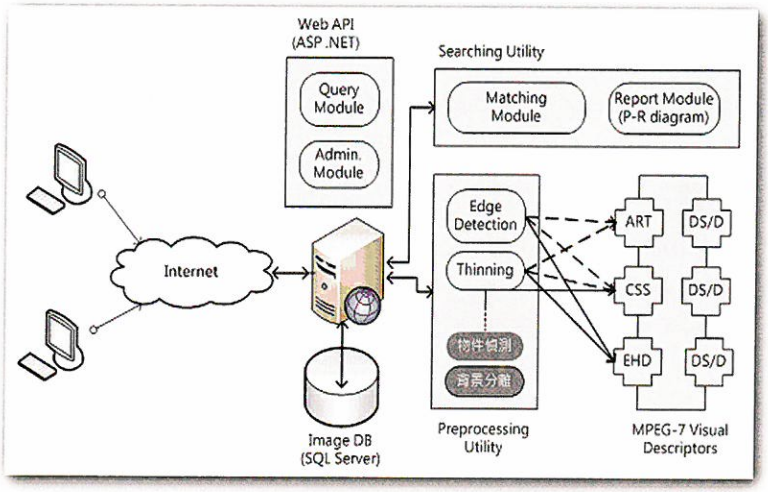


圖 10：系統架構示意圖

（二）統一塑模語言（UML）

為了實現上述的系統架構，我們使用 UML 來執行系統設計，但因篇幅的關係，在此只列出較重要的 Use Case、Class Diagram 及 Sequence Diagram；首先依據使用者可能的行為定義使用案例（Use Case）如圖 11：

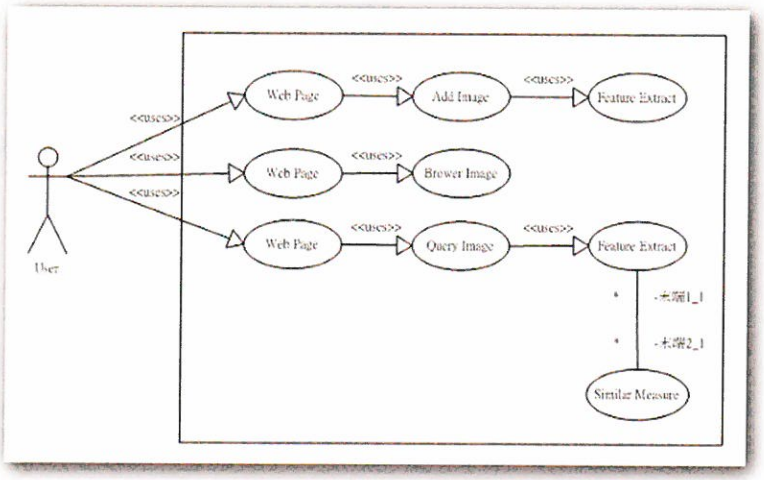


圖 11：影像資料管理平台使用案例

並且依照前置處理模組及特徵擷取模組所需的相關功能規劃之類別圖（Class Diagram）如圖 12、圖 13 所示：

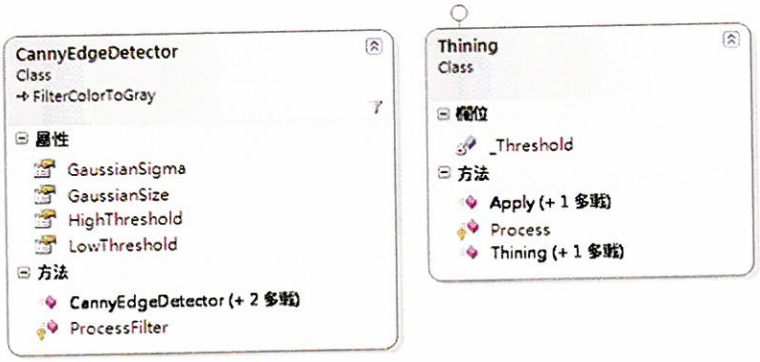


圖 12：前置處理模組類別圖

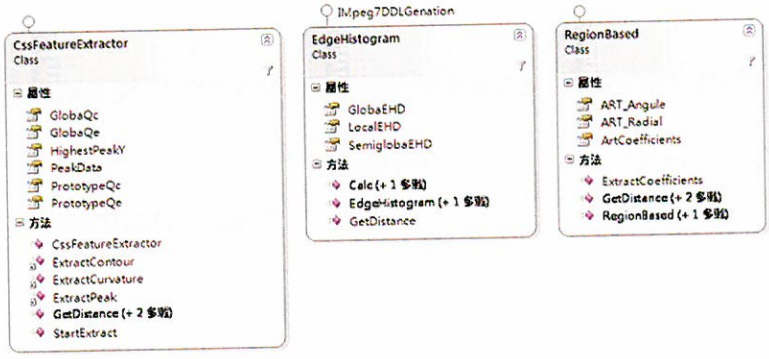


圖 13：特徵截取類別圖

而兩個模組之間的執行順序如圖 14 所示：

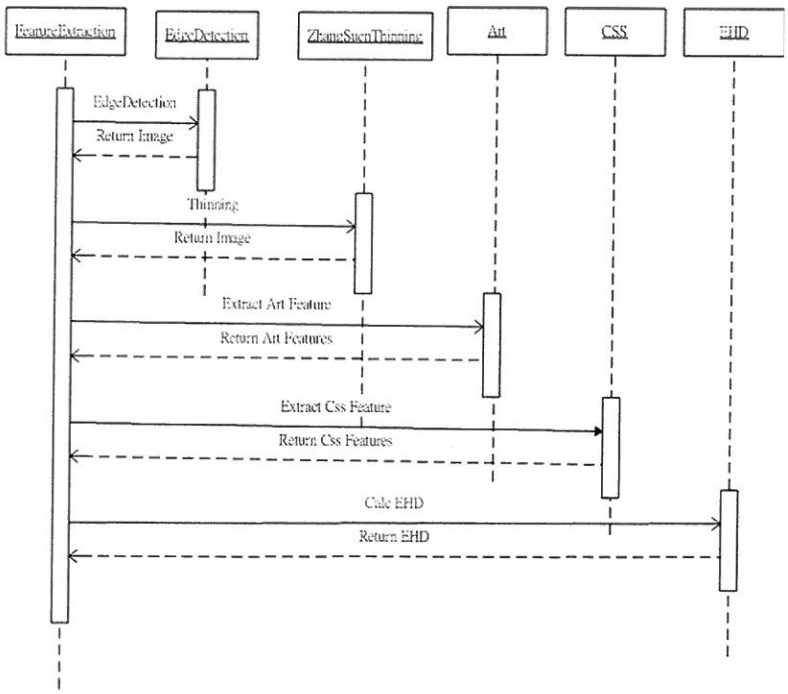


圖 14：特徵擷取循序圖（Sequence Diagram）

（三）特徵擷取（Feature Extraction）與描述子（Descriptors）

在這一節當中所討論的是前置處理模組、特徵擷取模組及各特徵所對應之相似度比對模組的實作方式及演算法介紹。

A. 前置處理模組

A-1. 邊緣偵測（Edge Detection）

在邊緣偵測方面，我們比較了兩個邊緣偵測的方式，分別為 Canny Edge Detection[4] 及 Wavelet Transform。

Canny Edge Detection 為 John F.C Canny 在 1986 年所提出的邊緣偵測的演算法，

其改進了 Sobel Edge Detection 在部分圖片因包含大量雜訊造成邊緣偵測失準的情況，主要方式為在偵測之前先經過一個 Low-Pass Filter 過濾掉大部分的雜訊及採用了 Double Threshold 的方式，其演算法執行步驟如下：

1. Smoothing（均化）：
   目的是消除影像壓縮演算法所造成的雜訊，此使用的是採用高斯濾波器（Gaussian Filter），Kernel 如「式（9）」

$$B = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 2 \\ 2 & 4 & 5 & 4 & 1 \end{bmatrix} \tag{9}$$

2. Finding gradients（取得梯度）：
   目的為偵測邊線資料，在此是使用 Sobel 演算法所定義的兩個矩陣，來判斷是否為邊線資料及線段的角度，計算公式如下：

$$KGx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{5}$$

$$KGy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{6}$$

$$G = \sqrt{KGx^2 + KGy^2} \tag{7}$$

$$\theta = \tan^{-1}\frac{|KGy|}{|KGx|} \tag{8}$$

3. Non-maximum suppression：
   在取得梯度值及角度後，下一個步驟就是將各個梯度向量依角度分類成 0、45、90、135 等四種角度呈現在圖片上。

4. Double Thresholding：
   這一個步驟是移除邊線以外的像素值（設為 0），使用兩個 Threshold T1、T2，且 T1>T2，當像素的灰度值 >T1 時，則該像素為邊緣點，而像素的灰度值 <T2

時則移除，但是灰度值介於 T1 及 T2 時，則必須再比對 8 個方位點，如 8 個方位點包含邊緣點（Pixel > T1）則設為邊緣點。

另一種方式是利用 Wavelet Transform 進行種領域轉換（Domain Transform），其原理為將一來源訊號經由一個低通濾波器（Low-pass Filter）及一個高通濾波器（High-pass Filter）將訊號分為高頻帶及低頻帶的一種演算，如將 wavelet Transform 應用在圖片上則會將原始圖片分為 LL、LH、HL、HH 四個區塊如圖 15，其中 LL 的部分為原始圖片的縮圖。

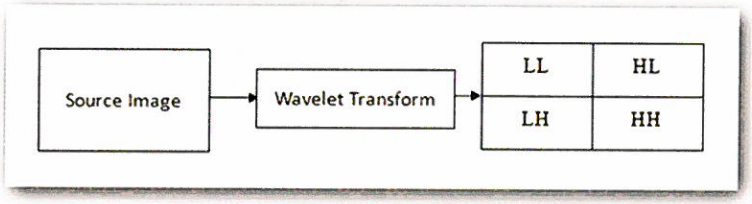


圖 15：Wavelet Transform 示意圖

依照 Wavelet Transform 的特性所定義的邊緣偵測的演算法如下：

```
Image WaveletThinning(Image SrcImg)
{
    Image Img1=RowLowPassFilter(SrcImg);
    Image LL=DownSampling(ColumnLowPassFilter(Img1));
    Image LH=DownSampling(ColumnHighPassFilter(Img1));
    Image Img2=ColumnHighPassFilter(SrcImg);
    Image HL=DownSampling(ColumnLowPassFilter(Img2));
    Image HH=DownSampling(ColumnHighPassFilter(Img2));
    Return Avg(HH,HL,LH);
}
```

在兩個演算法中我們以蘋果的圖片來做實驗，執行結果如圖 16，(a) 為來源圖片、(b) 為 Canny Edge Detection 執行結果，執行時間為 0.141 秒、(c) 為 wavelet 執行結果，執行時間為 0.016 秒。

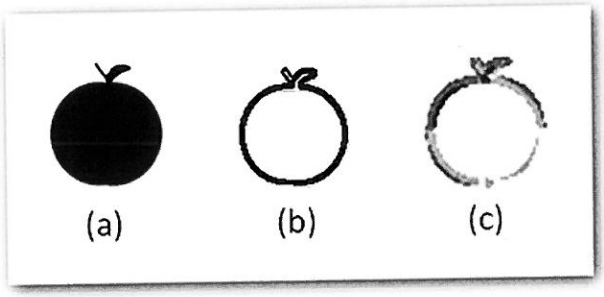


圖 16：邊緣偵測實驗結果

由以上的實驗得知，兩個演算法中以 wavelet 的執行效能比較好，但缺點是 wavelet 會將原本封閉的區間變成開放區間，這種結果會影響 CSS 的特徵擷取的結果，故本系統採用 Canny Edge Detection。

A-2. 細化（Thinning）

在邊緣偵測後所得到的圖片雖已過濾了大部分的低頻訊號（非邊線部分），但為了確保每一條線的度只有一個 pixel，以避免在部分線段度過大造成誤判的情況，因此在邊緣偵測後尚須經過細化的演算，本系統所採用的細化演算法為 Zhang-Suen Thinning Algorithm[5]。

Zhang-Suen Thinning 演算法是參考目標 Pixel，定為 P 的周圍八個方位的 Pixel 值來決定 P 是否應該消除，其關聯圖如圖 17：

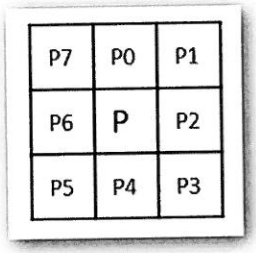


圖 17：Zhang-Suen Thinning Pixel 關聯圖

依據圖 17 定義兩個參數 N（P）為 P0 ～ P7 中 Pixel 為 1 的數量（以 Bin Image 為例）

及 S（P）為 P0 ～ P7 中由 0 變 1 的次數，Thinning 演算法表示如下：

```
Image ZhangSuenThinning(Image SrcImg)
{
    For each Pixel P in Src
    {
        If((2<=N(P)<=6)
        &&(S(P)==1)
        &&(P0&P2&P4==0)
        &&(P0&P4&P6==0)
        P=0;
    }
    For each Pixel P in Src
    {
        If((2<=N(P)<=6)
        &&(S(P)==1)
        &&(P0&P2&P6==0)
        &&(P2&P4&P6==0)
        P=0;
    }
}
```

B. 特徵擷取模組與相似度計算

B-1.Angular Radial Transform（ART）

ART Coefficients 為一組 12x3 之複數陣列（12 Angular x 3 Radial），其中 ART[0][0]，因該參數為 Normalized 的過程中是其他參數的參考值，故在特徵比對時不列入參考，其實作演算法如下：

```
Float[][] ComputeARTCoefficients(image SrcData)
{
    Construct ART Basis Function
    Compute ART Coefficients
    Compute Normalized Magnitude of ART Coefficients
    return Quantization ART Coefficients
}
```

相似度 $D_{ART}(d,q)$ 計算方式計算如「式（14）」，M_A 及 M_B 分別為資料庫影像及查詢影像之 ART Coefficients：

$$D_{ART}(A,B) = \Sigma_i |M_A[i] - M_B[i]| \tag{9}$$

B-2.Curvature Scale Space（CSS）

CSS 之特徵並無固定大小，於 MPEG-7 的規範中只有限制 Peak 的最大總數為 63，但實際 Peak 的數量視物件而定，其實作流程如下：

```
PeakSetComputPeaks(image SrcData)
{
    Get Boundary coordinates x(u),y(u)
    Do
    {
        Compute
        Get Zero-Crossing List
        Convolution x(u),y(u) for kernel (0.25,0.5,0.25)
    }While( Zero_Crossing_List.Count > 0)
    Compute Normalize Peaks
    return Normalized Peaks
}
```

CSS 的 Peak 系數中的第一個永遠為 Peak 集合中 Y 值之最大值，故在進行相似度比對時必須先將 X 值調整到最適合的參考點再取得兩個物件的歐基里德距離 [6]，其演算流程如下：

```
Float GetCSSDistance(PeakSet A,PeakSet B)
{
    Optimized Peak Set A and B for x-coordinates
    return D_CSS(A,B) = Σ √((X_A(i) - X_B(i))² + (Y_A(i) - Y_B(i))²)
}
```

B-3.Edge Histogram Descriptor（EHD）

在 EHD Extraction 是計算影像中各角度（0,45,90,135,non directory）邊線數量，系統採用 DCT 的計算方式 [7] 來取得，演算法實做如下：

```
Float[][] CalcEdgeHistogram(image srcData)
{
    Split Image to 16 SubImages
    ForeachSubImages
    {
        Compute DCT for SubImage
        Mapping AC10 and AC01 to 0,45,90,135,non directory
    }
    Normalize EHD
    return Quantization EHD
}
```

而在相似度計算的部分需 EHD Array 再細分為 Global EHD 及 Semiglobal EHD，其中 Global EHD 為五個角度在 16 個子影像的平均數，而 Semiglobal EHD 會將 16 個子影像再區分為 13 個區域，每個區域為 4 個子影像，並取其平均數，其 13 個區域如圖 18，相似度計算公式如下：

$$D_{ART}(A,B) = \sum_{i=0}^{79} |H_A(i) - H_B(i)| + 5\sum_{i=0}^{5} |H_A^g(i) - H_B^g(i)| + \sum_{i=0}^{64} |H_A^s(i) - H_B^s(i)| \quad (10)$$

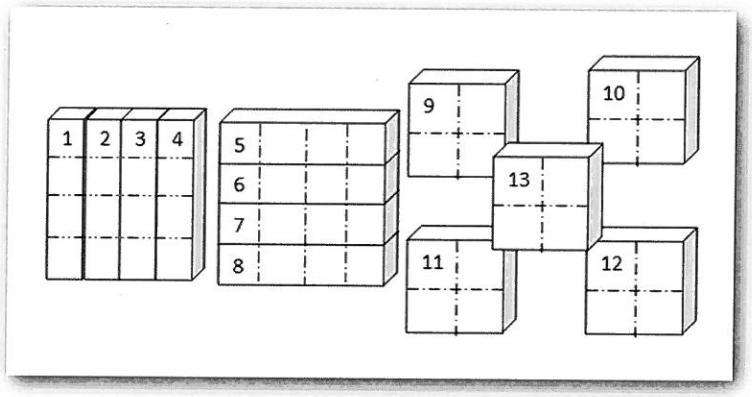


圖 18：EHD 區域分割示意圖 [1]

（四）系統流程與實作

在系統操作方面有新增影像、瀏覽影像及查詢影像三個部分，其流程分別介紹如下：

A. 新增影像

在這一個部分中有三個流程，系統流程如圖 19，操作介面如圖 20，依操作流程，將操作介面分為三個區域，其功能分別介紹如下：

○ 圖 20 之 (1)：上傳影像檔，所接受的檔案格式為 JPEG、GIF、PNG、BMP、TIFF 等影像格式。

○ 圖 20 之 (2)：輸入作者、作品名稱及影像分類名稱等與影像相關的文字資訊，可用以進行關鍵字查詢，在使用上可縮小資料查詢範圍，加速查詢的速度。

○ 圖 20 之 (3)：新增影像，在確定新增後，系統會使用前置處理及特徵擷取模組的順序來取得影像特徵，並將之以 MPEG-7 DDL 格式存入資料庫中。

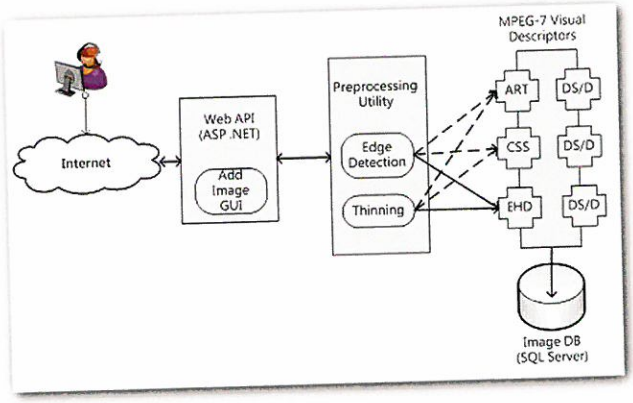


圖 19：新增影像流程

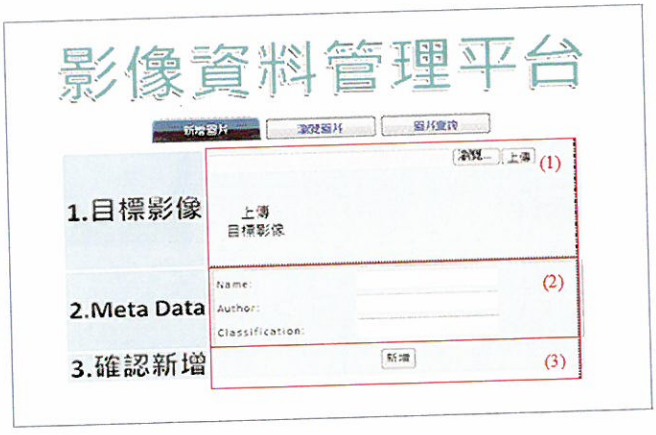


圖 20：新增影像介面

B. 瀏覽影像

這一個部分提供使用者一個以瀏覽的方式檢視現有影像相關資料的介面，操作介面如圖 21，介面說明如下：

○ 圖 21 之 (1)：影像資料一覽，用以顯示新增影像資料時所輸入之文字資料以及選取所要瀏覽的目標影像。

○ 圖 21 之 (2)：用以檢視所選取的影像內容。

○ 圖 21 之 (3)：用以檢視目標影像之 ART Featuresof MPEG-7 之樹狀結構。

○ 圖 21 之 (4)：用以檢視目標影像之 CSS Features of MPEG-7 之樹狀結構。

○ 圖 21 之 (5)：用以檢視目標影像之 Edge Histogram of MPEG-7 之樹狀結構。

圖 21：瀏覽影像介面

C. 查詢影像

提供使用者一個範例式查詢（Query-by-Example）的介面，運作流程如圖 22，操作介面如圖 23、圖 24、圖 25。

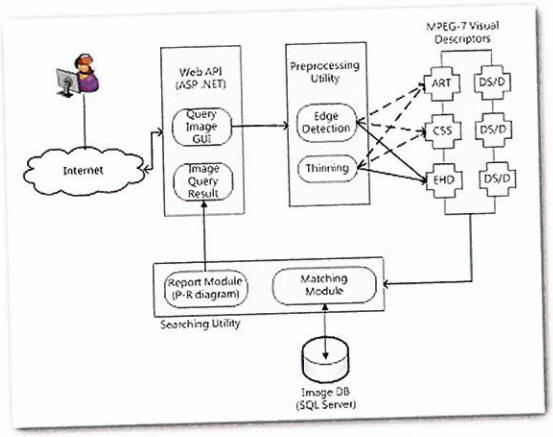


圖 22：影像查詢流程

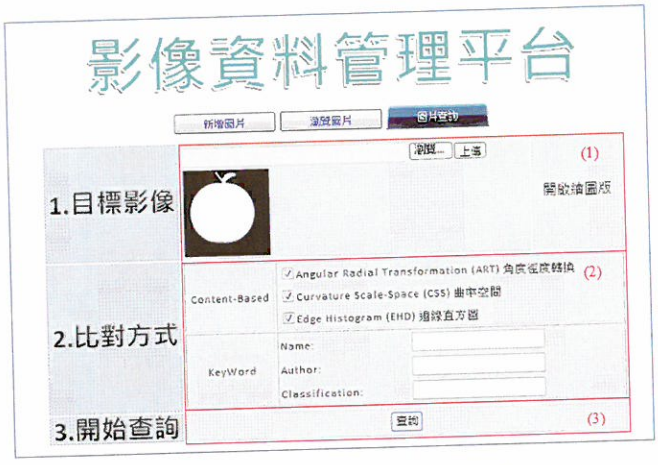


圖 23：影像查詢輸入介面

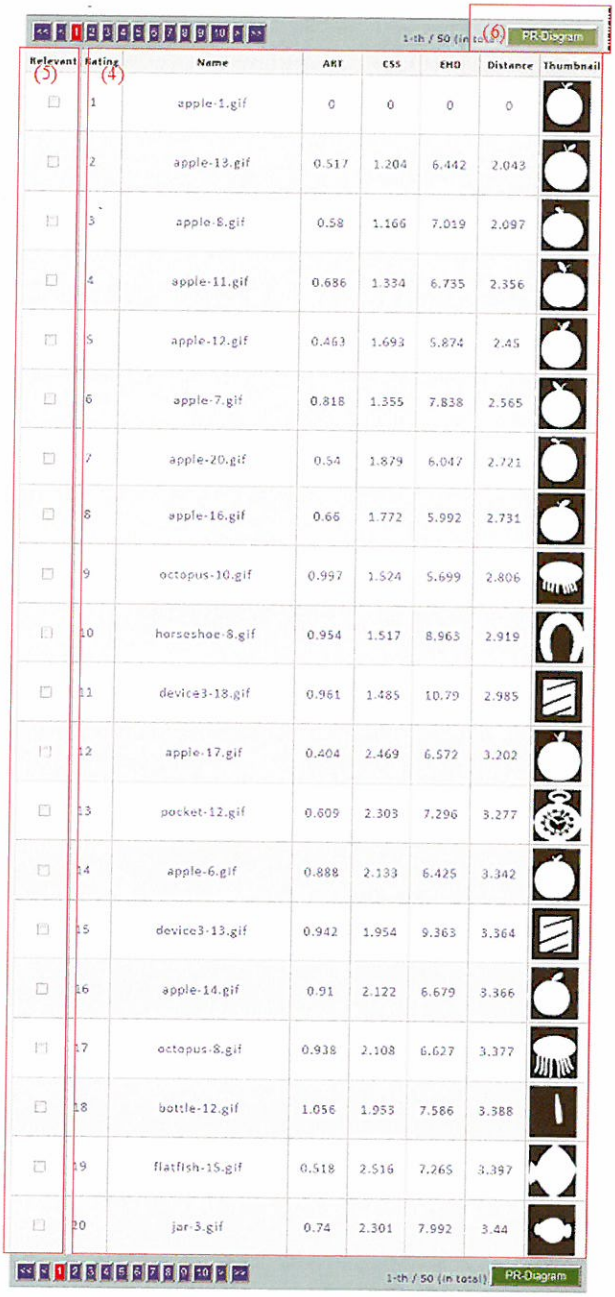


圖 24：查詢結果檢視介面

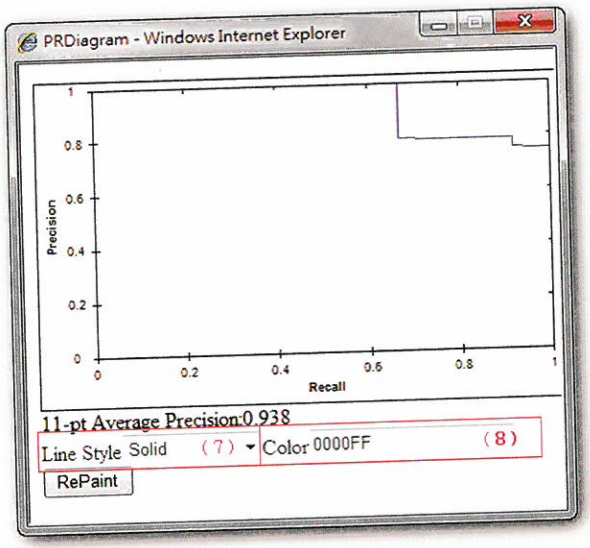


圖 25：PR-Diagram 檢視介面

操作介面依各方塊標示說明如下：

○ 圖 23 之 (1)：選擇並上傳要查詢之影像檔，影像格式同新增影像介面。

○ 圖 23 之 (2)：選擇所要比對的特徵資料，提供使用者選擇所要比對的特徵參數，以及所要過濾的關鍵字資料。

○ 圖 23 之 (3)：查詢影像，系統依使用者所選擇的特徵參數執行相似度運算，並輸出最相似的前 1000 筆資料。

○ 圖 24 之 (4)：查詢結果檢視，用以檢視結果影像的縮圖及相對特徵之相似度（距離）資料。

○ 圖 24 之 (5)：選擇正確之結集合，使用 Check Box 來勾選正確的影像資料，並根據使用者勾選的結果來產生 PR-Diagram。

○ 圖 24 之 (6)：使用 PR-Diagram 鈕，檢視 PR-Diagram。

○ 圖 25 之 (7)：選擇線段樣式來，可改變 PR-Diagram 的線段樣式 。

○ 圖 25 之 (8)：設定線條色彩以 R、G、B 表示。

[5] 圖 25 之 (7)、圖 25 之 (8) 的設計是便於取得實驗數據比較圖，與系統執行結果並無關聯。

5. 實驗與討論

在這一章描述的是系統驗証方式及結果的分析；本系統的實作是採用 Microsoft .Net Framework 3.5 之 C# 為實作語言，使用工具為 Microsoft Visual Studio 2008，測試環境為 Intel Core 2 Due 2.53GHz CPU，4G RAM，作業系統為 Windows 7（64Bit 版）。

（一）實驗設計

實驗資料的選擇方面，我們是使用 MPEG-7 Core Experiment ce-shape-1 的測試資料集（http://www.cis.temple.edu/~latecki/TestData/mpeg7shapeB.tar.gz），其中包含 70 個不同的二值化物件，每個物件有 20 個圖片，總計 1400 張圖片。

在這次的實驗中，我們從實驗資料集中選擇了六個較具代表性的物件為查詢對像，並定義了相似物件的答案，如圖 26 所示，在實驗中我們分別針對每一組查詢使用 ART、CSS、EHD、ART&CSS、ART&EHD、CSS&EHD 及 ART&CSS&EHD 來進行相似度比對，並繪出每一個比對模型之 PR-Diagram 來評估影像查詢的品質，並以 11-Point Average Precision 來判斷整體實驗的精準度。

圖 26：實驗結果示意圖。

（二）實驗結果

依照實驗設計所得的結果（如圖 27、圖 28、圖 29、圖 30、圖 31、圖 32 所示），系統平均執行時間為 1.934 秒，由十字架的圖示（如圖 27）的實驗中，我們得知其中以 ART 為最高的查準確度（以藍底紅字呈現），PR-Diagram 亦呈現在最外圍的曲線（淡紫色線段）；分析原因歸納為 ART 特徵模型上採用極座標的表示法，以物件的中心點為極零座標，並取得物件在 12x3 的矩陣分別代表 12 個角度及 3 個徑度上的強度，在十字架的影像特徵分佈在這一個矩陣中幾乎呈現定值狀態，在相似度比對上呈現了強烈的對比。

而 CSS 在於星狀圖片查詢中得到了最好的結果（如圖 30），其主要的原因在於 CSS 用於相似度比對的模型上是以物件的邊緣輪廓上明顯的斜率變化（凹凸處）為特徵點（Peak），再取得這些特徵點的歐基里德距離，故對於星狀圖形的查詢上，因其特徵點的分佈非常明顯，所以 CSS 的查詢表現就比 ART 及 EHD 要來得好。

從整體實驗的結果看來 EHD 在十字架及機車圖片查詢的驗實（如圖 27、圖 32）中得到較好的數據，且整體的表現中亦不如 ART 及 CSS 兩項特徵，從結果分析來看，因十字架及機車的圖片集合中旋轉圖片及位移的問題較不明顯，且影像資料都分布於圖片的中央，是以 EHD 對於圖片的角度及物件在圖片中的位置變化相當敏感，所以 EHD 在相似度查詢上較適用於類似的風景照片及素描畫等較無旋轉問題及對全域影像關連性較高的圖片。

就各別的實驗來看 ART、CSS 及 EHD 各自對於不同性質的圖片有較好的表現，而本系統主要推薦的查詢方式是採用三種特徵的加權比來判斷影像的之相似度，利用彼此間的長處來補償彼此的不足，就整體的實驗來看，使用複合參數的查詢結果都有不錯的表現，但在某些情況來說使用複合特徵的結果尚不如使用單一特徵來得好，故在於特徵的選用上系統採開放式的角度設計，令使用者可依不同的使用案例決定要採用那些參數來進行相似度比對。

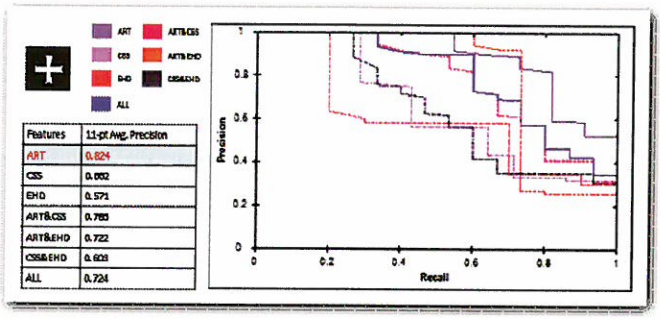


圖 27：Cross 實驗結果

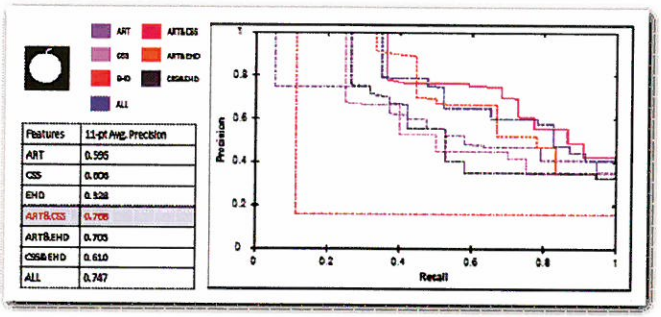


圖 28：Apple 實驗結果

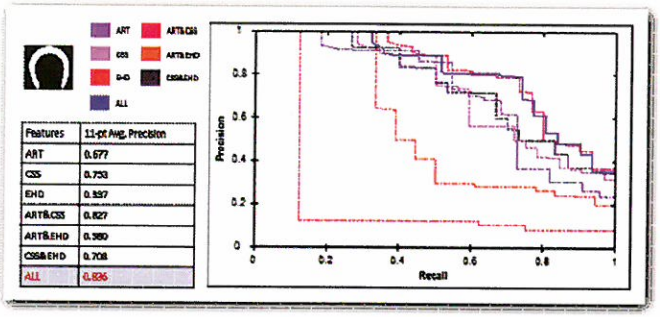


圖 29：Horseshoe 實驗結果

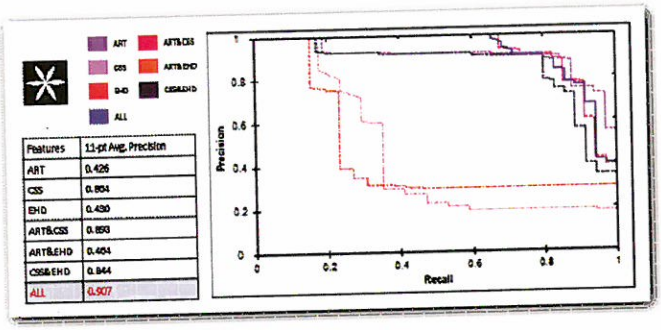


圖 30：Star 實驗結果

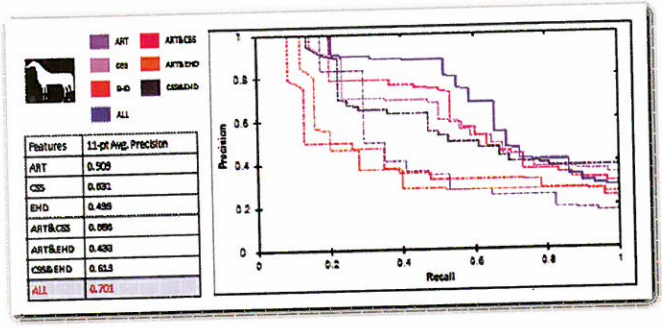


圖 31：Horse 實驗結果

6. 結論

本系統主要目的是建立一個影像資料管理平台，提供一個擴充性高及可跨系統交換資料的平台，故在系統的設計上大多採開放式及模組化設計並使用 MPEG-7 為資料描述格式，未來可依使用需求於前置處理模組及特徵擷取模組中加入新的項目，如於前置處理模組中加入背景分離模組、物件偵測模組等以及在特徵擷取模組中加入顏色特徵或其他外形比對特徵的演算法。

在相似度的比對上，於實驗系統使用 ART、CSS 及 EHD 其 11-Point Average Precision 最高可達九成的準確率，而整體平均也都達七成以上，在外形比對上有不錯的表現，目前系統只針對單一物件的影像作查詢，但是要套用在實際的數位相片的管理，尚有許多不足的地方，如未來可加入背景偵測及物件偵測方面的技術，相信可以為系統提供主題相片的查詢方式，更可滿足使用者在資料管理方面的需求。

於未來，計畫將系統研究成果結合大型儲存媒體（如 NAS，高容量硬碟），開發應用程式管理介面，提供使用者一個影像資料管理及交換的功能，提高資料儲存媒體的附加價值。

參考文獻

1. B.S. Manjunath, T. Sikora, and P. Salembier, *Introduction to MPEG 7: Multimedia Content Description Language,* Wiley, New York, 2002.
2. José M. Martínez, "MPEG-7 Overview" , ISO/IEC JTC1/SC29/WG11N6828 Palma de Mallorca, October 2004.
3. S. Abbasi, F. Mokhtarian, and J. Kittler, "Enhancing CSS-based shape retrieval for objects with shallow concavities," *Image and Vision Computing*, Vol. 18, pp. 199-211, 2000.
4. R. S. Chora, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems" , *International Journal of Biology and Biomedical Engineer*, Vol. 1, Issue 1, 2007.

5. 連國珍，數位影像處理，台北市：儒林，1999.
6. F. Mokhtarian and S. Abbasi, "Shape similarity retrieval under affine transforms," *Pattern Recognition*, vol. 35, pp. 31-41, 2002.
7. M. Eom and Y. Choe, "Fast Extraction of Edge Histogram in DCT Domain based on MPEG-7," in *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 9, Nov., 2005.
8. ISO/IEC/JTC 1/SC 29 , "Information technology-Multimedia content description interface -Part 3:Visual" , ISO/IEC 15938-3,2002.
9. T. Sikora, "The MPEG-7 Visual Standard for Content Description –An Overview," *IEEE Transactions on Circuits and Systems for Video Technology*,Vol. 11,No. 6,June 2001.
10. MPEG-7 overview, http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
11. Ze-Nian Li and M. S. Draw, *Fundamentals of Multimedia*, Prentice-Hall, 2004.
12. 蔡昕峻， "以影像基本圖案及輪廓特徵建立內涵商標檢索系統" ，義守大學，2003.
13. 黃子癸， "以 MPEG-7 為基礎之蝴蝶影像形狀檢索" ，台灣大學，2001.
14. 楊長慎， "以影像擷取技術為基礎之圖片篩選系統" ，中華大學，2007.
15 T. Quack, U.Mönich, L. Thiele, and B. S. Manjunath, "Cortina: a system for loge-scale, content-based web image retrieval," in *Proceedings of the 12th annual ACM International Conference on Multimedia*, 2004.
16. M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, " A descriptor for large scale image retrieval based on sketched feature line," in *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, 2004.
17. M.L. Kherfi and D. Ziou, "Image Retrieval Form the Word Wide Web: Issues, Techniques,and Systems," *ACM Computing Surveys*,Vol. 36, Issue 1, pp.35-67, March, 2004.

Received September 30, 2010
Revised November 30, 2010
Accepted December 22, 2010

A Framework for Image Data Management [3]

Chi-Hung Tsou and Jia-Lien Hsu [4]

Department of Computer Science and Information Engineering
Fu Jen Catholic University

Abstract

When dealing with an increasing and huge amount of digital images, the traditional retrieval architecture associated with the keyword-based searching capabilityis not satisfying anymore, especially for non-experienced and naïve users. In this paper, we propose a framework for image data management, which supports content-based image retrieval. Also, we apply MPEG-7, one of most-popular multimedia content description interface, as the metadata specification of our framework. As preprocessing and extracting features from images, we choose angular radial transform, curvature scale space, edge histogram as features of images to be compared with user queries. In addition to MPEG-7 of sharing metadata,our framework is also followed with open architecture, in which more feature extraction modules and Ds/DS could be easily plugged into. By our experiment and performance evaluation, we show that the efficiency and effectiveness of our framework will be achieved.

Keywords: MPEG-7, content-based image retrieval,
angular radial transform (ART), curvature scale space (CSS),
edge histogram

[3] This paper was supported by Fu Jen Catholic University (Project No.409931044034), and National Science Council (NSC-98-2622-E-030-001-CC3).
[4] Corresponding author. Email: alien@csie.fju.edu.tw