Independent Component Analysis for Blind Source Separation

Yu-Jui, Chen B053021023

Abstract

In this experiment, the FastICA algorithm was used to deal with the problem of image separation, and found a demixing matrix which separated the mixing signal close to the original input signal. By converting the pictures into gray scale images, we can transform each picture to standard form. According to the color of the picture, we will compare the color of the picture which is similar and difference in the experiment and observe the separation of signals under different parameters. Finally, based on the results of each experiment, we will discuss the advantages and disadvantages of FastICA, and the difference between independent and non-independent signals.

I. Introduction

Independent Component Analysis, or ICA, is a method of calculating using statistical principles. It is a linear transformation. This transformation separates the data or signal into a statistically independent linear combination of non-Gaussian signal sources. In 1983 Ans, Herault, and Jutten proposed the method of ICA, which mainly solving the neural network model of the source blind deconvolution. In 1995 and 1997, the following two approaches were proposed. First, A.J. Bell and T.J. Sejnowski propose InforMax Principle. Second, FastICA proposed by Aapo Hyvärinen with Erkki Oja. These two methods are the two mainly methods of ICA operation nowadays and widely used in various fields, such as sound signal analysis, image analysis and network communication signal analysis.

In ICA applications, the classic question is the cocktail party problem. The problem describes how to separate the given mixing signal into individual who speaks at the same time in a cocktail party. When there are N sources, it is usually assumed that there are N observation signals (e.g. N microphones or recorders). This assumption implies that the blend matrix is a square, which means J - D, where D is the dimension of the input data and J is the dimension of the system model. The application of voice separation of blind source signals is the most intuitive, and ICA provides an appropriate solution to such problems in the sound source.

II. Methodology

A. Definition

The most important assumption of independent composition analysis is the statistical independence of the signal source. Use a variable model in a random process. As shown in equation (1), suppose n linear lysing independent components, x_i represents the measured mixed signal, and i is the total number of measured mixed signals. a_{ij} indicates the mixing ratio of the original signal, s_j represents the original signal, where j is 1…n, expressed as n original signal, and combined a_{ij} with s_j to get mixed signal x_i.

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n \quad \forall\, i \tag{1}$$

In order to simplify the problem, vector matrix is generally used. We mostly used linear combinations as shown in equation (2). x is measured mixed signal in formula (1), and the extension vector matrix of x_i is used, where i is 1… m, where m is the measured mixed signal. A is the linear combination matrix of signals as randomly generated squares, and s is a vector matrix of the original signal. By combining A and s, we can acquire a mixed signal x.

$$\boldsymbol{x} = \boldsymbol{As} = \textstyle\sum_{i=1}^{n} \boldsymbol{a_i s_i} \tag{2}$$

Therefore, to separate the blind source signal. We need to find A's inverse matrix W (demixing matrix), so that the received signal x can process through W to obtain the original independent component signal s, such as equation (3).

$$\mathbf{Wx} \approx \mathbf{s} \tag{3}$$

W can be determined by using the independence between signals and signals. Signals are each independent. After the conversion of W, if the signal mixed relationship is eliminated, then the conversion signal is also independent, that is, the original signal. This is the basic core concept of ICA.

B. Limitation

Restrictions on signal reception are based on equation (3). In x = As, the dimension of x is m and the dimension of s is n. In the first case of m = n, a set of independent solutions can be obtained. In the second case of m> n , the number of equations is more than the unknown, as long as the dimension of m is reduced to n, and then perform the ICA operation. In the third case of m <n, the equation is less than the unknown, there is not enough information to calculate the original information. The uncertainty of ICA is generated. In the ICA model in equation (2), the following two uncertainties can be known. First, unable to determine the variance of independent components, where A and s are both unknown terms. When s_i is multiplied for a certain multiple, as long as it is divided by the same multiple, the original signal s_i can be obtained, as shown in equation (4).

$$x_i = \textstyle\sum_{i=1}^{n} \left(\frac{a_i}{b_i}\right)(s_i b_i) \tag{4}$$

Therefore, to fix the size of independent components. The most commonly way is to assume that the variance of these units is one, when these independent components are random variables, as in equation (5).

$$E\{s_i^2\} = 1 \quad (5)$$

Second, the order of independent components cannot be determined. This aspect is also related to the fact that A and s are unknown items. Due to A and s, the order of signals cannot be defined, that is, the order of items in the sum can be changed randomly. Therefore, even if W is found, there is no way to get exactly the same original signal. In addition, the signal itself is unknown, and it is even more unlikely that the resolved signal corresponds to the original signal.

There are the following two restrictions on the signal itself. First, the signal itself is statistically independent. The premise of the use of ICA is that the signals are independent and all signals must meet this characteristic. In most of the separation problems, the signals are indeed independent. Statistically, if the random variables s_1, s_2, $s_3 \ldots s_i$ are independent, their joint probability density function can be obtained by multiplying the marginal probability density function, as shown in equation (6).

$$p(s_1, s_2, \ldots s_i) = p(s_1)p(s_2) \ldots p(s_i) \quad (6)$$

Second, the independent elements must be Nongaussian distributions. The key to determining the ICA model is nongaussianity, because the way ICA finds independent elements comes from the central limit theorem. This theorem tells us that the distribution of the sum of nongaussian distribution will be more closer to any gaussian random variables. Therefore, when searching independent components, utilizing this feature to make the signal found different as Gaussian distribution. Therefore, when finding a set of W, the result is not closest to the Gaussian distribution, which is the original signal.

Based on this concept, if the original signal itself is a gaussian distribution and no matter what the result of W is calculated the original signal is still a gaussian distribution. Therefore, there is no way to find the W that makes the result the most nongaussian and we cannot found the original signal. Nevertheless, most of the audio or images received are non-gaussian, which means ICA will not be limited in practical use.

C. Preprocessing

i. Centering

If the average of the received signal and the independent components is zero, many simplifications can be made in the ICA algorithm derivation. In practical applications, the average of most received signals is not zero, so we must subtract the average of these signals to make the overall average zero, shown in equation (7).

$$x_{center} = x - E\{x\} \quad (7)$$

Not only the receiving signal is centralized, but the original signal is also centralized. $E\{s\} = A^{-1}E\{x\} = A^{-1} \times 0 = 0$ 。

ii. Whitening

Before ICA algorithm, pre-processing is used to transform the signal into uncorrelated. The initial signal is closer to independent, and the relationship between the two is whiteness. When a vector xW is white, its component is uncorrelated and the variance is one. In other words, the covariance matrix of xW is an identitical matrix, shown in equation (9).

$$\mathrm{E}\{\mathbf{xWxW^T}\} = \boldsymbol{I} \tag{9}$$

Whitening is looking for a whitening matrix V to perform linear conversion on the received signal xW=Vx. The most common method is eigenvalue decomposition (EVD), shown in equation (10).

$$\mathrm{E}\{\mathbf{xWxW}^{\boldsymbol{T}}\} = \mathbf{EDE^T} \tag{10}$$

Where E is an orthogonal matrix composed of eigenvector, and ET = E-1, D is a diagonal matrix composed of corresponding eigenvalues, then the whitening matrix is shown in equation (11).

$$\mathbf{V} = \mathbf{E}\boldsymbol{D}^{\mathbf{1/2}}\boldsymbol{E}^{\boldsymbol{T}} \tag{11}$$

The converted signal can be further expressed as equation (12).

$$\mathbf{xW} = \mathbf{Vx} = \mathbf{VAs} = \boldsymbol{\acute{A}s} \tag{12}$$

It can be observed from equation (12) that whitening is actually a linear transformation of matrix A, and results in a new mixing matrix $\acute{A}$. Expand the covariance matrix of the whitening signal to obtain in equation (13).

$$\mathrm{E}\{\mathbf{xWxW}^{\boldsymbol{T}}\} = \mathrm{E}\{\boldsymbol{\acute{A}ss^T\acute{A}^T}\} = \boldsymbol{\acute{A}}E\{\boldsymbol{ss^T}\}\boldsymbol{\acute{A}^T} \tag{13}$$

Where the original signal is independent, then $E\{ss^T\} = I$. Therefore, we can conclude in equation (14).

$$\mathrm{E}\{\mathbf{xWxW}^{\boldsymbol{T}}\} = \boldsymbol{\acute{A}\acute{A}^T} = \boldsymbol{I} \tag{14}$$

This also means that the mixing matrix after conversion is an orthogonal matrix, so when looking for the demixing matrix W, you only need to find the vertical vector, which can greatly simplify the operation.

D. ICA algorithm

In the algorithm, we first define an objective function, and calculate its maximum or minimum value. Defining a function that obtains independent elements which is the result of maximum or minimum value. Therefore, ICA can be divided into two parts, objective function and optimization algorithm. The objective function determines the robustness of the ICA algorithm, and the choice of the optimal algorithm is related to the convergence speed of the ICA. In the discussion of two restrictions on the signal itself, because of the central limit theorem. We can find a W that makes the derivation result the most nongaussian, and to obtain independent component terms. Therefore the term finds the independent component term s, which can be expressed as equation

(15).

$$s = w^T x \tag{15}$$

Since we want to get W that makes the derivation result the most nongaussian, we must quantify whether the random variable is gaussian, that is, measure s. There are two typical measurement methods which are kurtosis and negentropy. First, kurtosis is a function that can be input by random variable, shown in equation (16).

$$\text{kurt}(y) = \text{E}\{y^4\} - 3(E\{y^2\})^2 \tag{16}$$

After the signal y is pre-processed, its variance is equal to one, and the above formula can be simplified to $\text{E}\{y^4\} - 3$ 。For gaussian random variables, it can represent as $\text{E}\{y^4\} = 3(E\{y^2\})^2$ .Therefore, kurt(y)=0 。The value of Kurtosis function has positive and negative, when the value is positive, it is called supergaussian; in the other side, when Kurtosis is negative, it is called subgaussian. It can be seen that we can use the square or absolute value of kurtosis to determine the gaussian degree of random variables. The smaller the kurtosis, the higher the gaussian degree. On the contrary, The larger the kurtosis , it is more clise to nongaussian. Second, Neg-entropy is a function that improve some limitation of kurtosis. Although Kurtosis can measure the degree of gaussian of random variables, but Kurtosis is a high-power formula. It is easy to magnify a very small value and have an effect, so Kurtosis's method is not robust. Neg-entropy is developed from information theory. When the uncertainty or randomness is higher, the entropy is larger, which is closely related to the coding length of random variables. When a random vector y has a density of f(y), its entropy H is defined as equation (17).

$$\text{H} = -\int f(\boldsymbol{y})\, log f(\boldsymbol{y}) d\boldsymbol{y} \tag{17}$$

In the random variables of the same variance, it will cause the entropy of gaussian the largest. Therefore, you can use entropy to measure whether the random variable is gaussian or not. In order to measure more conveniently, the function value of the gaussian random variable is expected to be zero and the function values of other random variables are all greater than zero. Therefore, define neg-entropy J, and neg-entropy of the random vector y is shown as equation (18).

$$\text{J}(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{18}$$

Among them $\mathbf{y}_{\text{gauss}}$ and y have the same covariance matrix of gaussian random vector. It can be seen that when y is gaussian, the function value is zero, and the entropy of gaussian is the largest which means that negentropy must be positive. Therefore, while looking for the most nongaussian random variable is to find the largest negentropy. Because the calculation of negentropy is quite complicated, in practical applications, negentropy will be simplified as in equation (19).

$$\text{J}(\mathbf{y}) \approx [E\{G(y)\} - E\{G(v)\}]^2 \tag{19}$$

Where G is a non-quadratic function and v is a gaussian random variable. There are actually three options for G. As shown below, the smoother G is, the more robust the estimate.

$$G_1 = \frac{1}{a_1} log \cosh a_1 \mathrm{y},\ G_2 = -exp(-y^2/2),\ G_3 = y^4 \qquad 1 \le a_1 \le 2$$

E. FastICA algorithm

FastICA was proposed by Aapo Hyvärinen and used negnetropy to measure the degree of nongaussian. To measure $\mathrm{s} = w^T x$, we can be expressed as shown in equation (20).

$$\mathrm{J}(\mathbf{w}) \approx [E\{G(\boldsymbol{w^T x})\} - E\{G(\boldsymbol{v})\}]^2 \tag{20}$$

The maximum value of Negnetropy can be found by the extreme value of $E\{G(\boldsymbol{w^T x})\}$. According to Kuhn-Tucker condition, $E\{G(\boldsymbol{w^T x})\}$ is limited by $\mathrm{E}\{(\boldsymbol{w^T x})^2\} = ||w||^2 = 1$. To get the extreme value, you must satisfy equation (21).

$$\mathrm{E}\{\boldsymbol{x}g(\boldsymbol{w^T x})\} - \beta\mathbf{w} = 0 \tag{21}$$

Where g is the differential form of G. The above formula can use Newton's method to find the solution, and the left formula is represented by F. We can obtained Jacobian matrix JF(w), shown in equation (22).

$$\mathrm{JF}(\mathbf{w}) = E\{\boldsymbol{x}\boldsymbol{x^T}\dot{g}(\boldsymbol{w^T x})\} - \beta\boldsymbol{I} \tag{22}$$

Pre-processed signal can be approximate $E\{\boldsymbol{x}\boldsymbol{x^T}g(\boldsymbol{w^T x})\}$ to $E\{\boldsymbol{x}\boldsymbol{x^T}\dot{g}(\boldsymbol{w^T x})\} \approx E\{\boldsymbol{x}\boldsymbol{x^T}\}\, E\{\dot{g}(\boldsymbol{w^T x})\} = E\{\dot{g}(\boldsymbol{w^T x})\}\boldsymbol{I}$. Therefore, Jacobian matrix will become a diagonal matrix. Using Newton's method again, the following iterations can be obtained, shown as equation(23).

$$\boldsymbol{w}^+ = \boldsymbol{w} - [\mathrm{E}\{\boldsymbol{x}g(\boldsymbol{w^T x})\} - \beta\mathbf{w}]/[E\{\dot{g}(\boldsymbol{w^T x})\} - \beta] \tag{23}$$

Finally, we can multiple equation (23) with $\beta - E\{\dot{g}(\boldsymbol{w^T x})\}$, shown as equation (24).

$$\boldsymbol{w}^+ = \mathrm{E}\{\boldsymbol{x}g(\boldsymbol{w^T x})\} - E\{\dot{g}(\boldsymbol{w^T x})\} \tag{24}$$

The convergence condition is that w is in the same direction as before the update. When more than one independent component is found, according to the previously mentioned, we can look for the demixing matrix and only need to find mutually perpendicular vectors. So when you find n of w, you need to subtract n-1 of w found before. Therefore, the overall process is as follows.

i. Preprocessing the signals. Centering $x_{center} = x - E\{x\}$

ii. Preprocessing the signals. Whitening xW

iii. Set up counter p=1，m is the total amount of independent element.

iv. Randomly choose a vector w

v. Let $w^+ = \mathrm{E}\{xg(w^T x)\} - E\{\dot{g}(w^T x)\}$ and $\mathrm{w} = w^+/\|w^+\|$

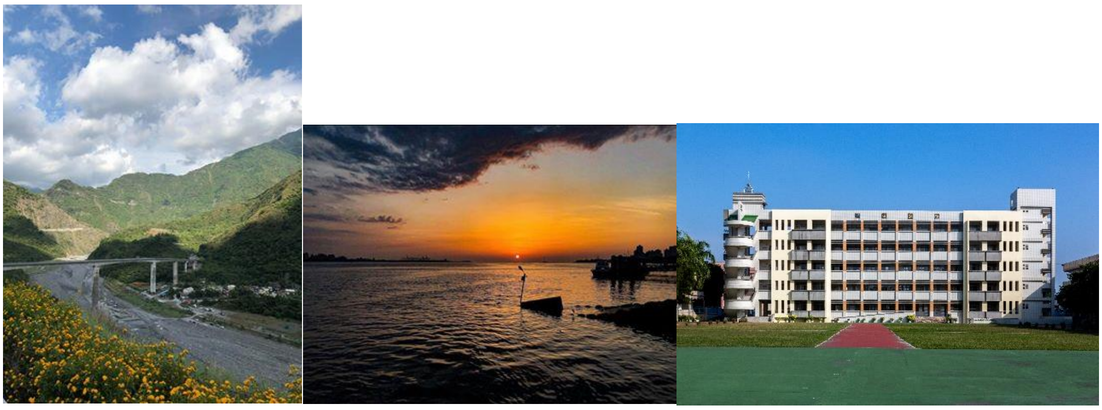
vi. Subtract the direction w which is previously found

vii. If w does not converge, go back to step v viii.Set up $p = p + 1$，if $p < m$，go back to step iv. Until all w is founded.

III. Experiment

Utilizing FastICA to deal with image synthesis problems. First, grayscale the three images, and mix the images. After using the FastICA algorithm to calculate the demixing matrix, three images are separated. The pictures are all taken with a mobile phone. In this experiment, three groups of experiments will be designed. The first group will have three pictures with different colors, including the background color and object color. Picture shown in figure(1). The second group will have two same picture colors which is the background color and one different color. Picture shown in figure(2).The third group will have the same picture colors which is the background color. Picture shown in figure(3).

figure(1)



figure(2)

figure(3)

First, compare three different non-quadratic function shown as follow, and calculate the iteration of each vector. Selecting the minimum value which has ephemeral processing time.

$$G_1 = \frac{1}{a_1} log \cosh a_1 \mathrm{y},\ G_2 = -exp(-y^2/2),\ G_3 = y^4 \qquad 1 \leq a_1 \leq 2$$

Second, compare the demixing matrix of three different non-quadratic function and by compare the W matrix to found out which separate signal is more accurately.

IV. Discussion

In three figure, we obtained different statistic from utilizing different iteration function. The result is shown as chart(1), (2), and (3).

Figure 1		Mean	Standard deviation
G1 function	Iterations	16.7	4.256265
	Correlation coefficient	0.50459	0.028737
G2 function	Iterations	19.85	6.351751
	Correlation coefficient	0.48953	0.040037
G3 function	Iterations	17.8	5.662434
	Correlation coefficient	0.51231	0.022421

Chart(1)

Figure 2		Mean	Standard deviation
G1 function	Iterations	16.5	5.276562
	Correlation coefficient	0.513215	0.032214
G2 function	Iterations	16.25	4.351346
	Correlation	0.50422	0.031754

	coefficient		
G3 function	Iterations	22.9	7.46148
	Correlation coefficient	0.497265	0.032201

Chart(2)

Figure 3		Mean	Standard deviation
G1 function	Iterations	20.15	3.133436
	Correlation coefficient	0.459275	0.021794
G2 function	Iterations	20.1	5.821015
	Correlation coefficient	0.445085	0.040304
G3 function	Iterations	26.6	5.364601
	Correlation coefficient	0.45355	0.037594

Chart(3)

By observing the chart, we can found out in the picture which are more similar need more iterations, and its correlation coefficient are more smaller. From the chart, we can choose the parameters in FastICA are iterated using $G_1 = \frac{1}{a_1} log \cosh a_1 y$, and analyzed in the time domain. In figure (4), (5), and (6) are the results of each experiment.

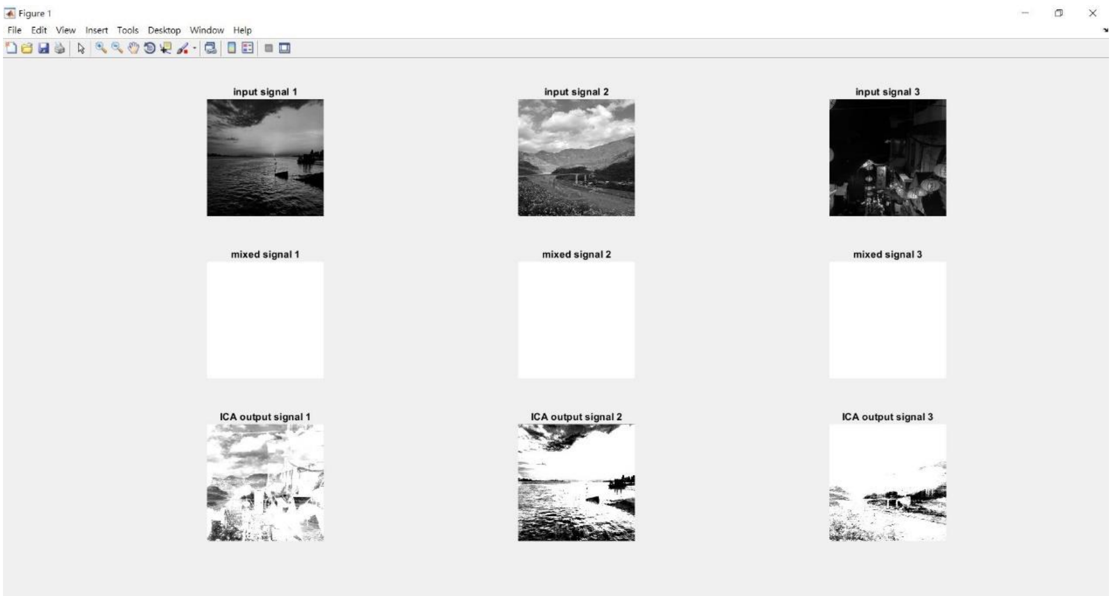


figure (4)

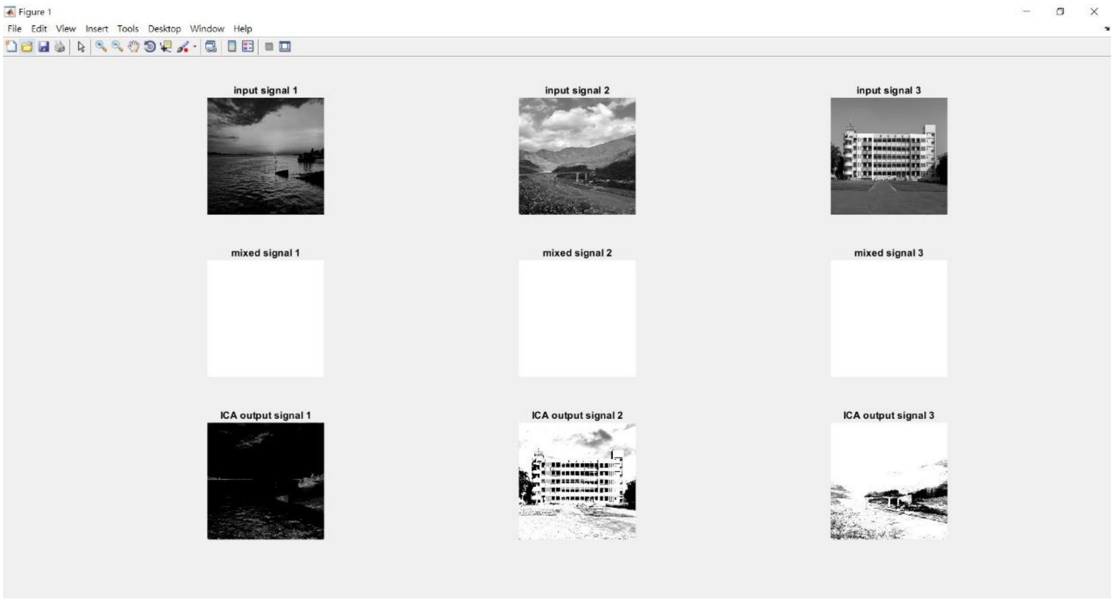


figure (5)

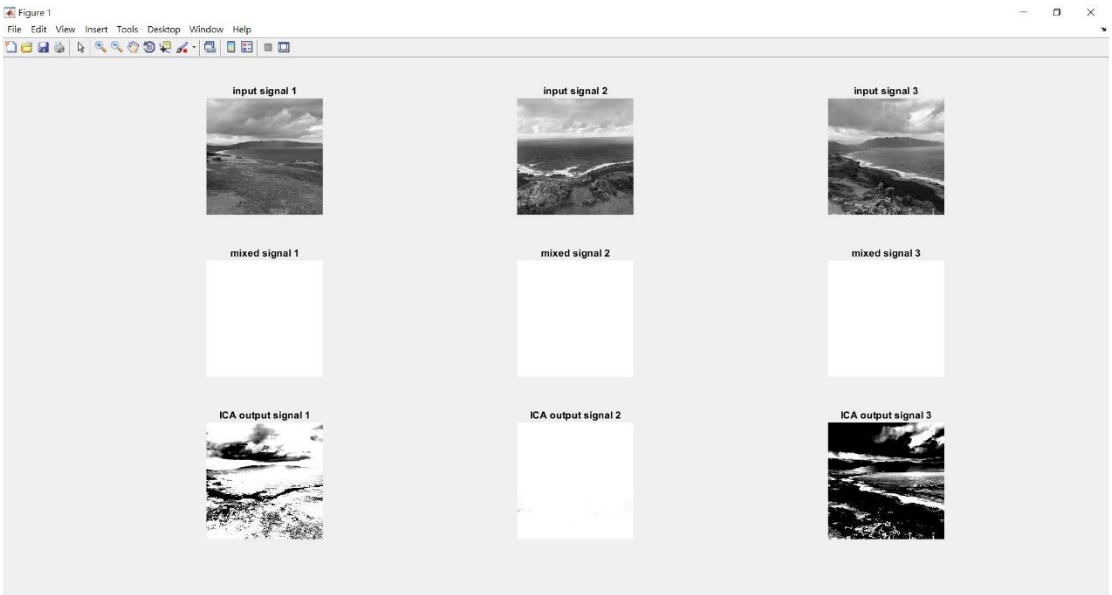


figure (6)

V. Conclusion

This time, FastICA is used to deal with the problem of separating pictures. By finding the demixing matrix, the received signal is separated to achieve the purpose of picture separation. The three signal sources are mixed, and there are three sets of experiments including three different images, two similar images and one different image, three similar images. By analyzing figure (1), (2), and (3) and using three commonly use iteration function, we can explore the advantages and disadvantages of FastICA. We also observe that the different types of signal separation conditions and effects. It is found that as long as there are some correlation in the pictures, the independent signals cannot be distinguished or the degree of separation is not good. The restrictions on the conditions are quite strict. However, in the low correlation independent signals, the results will have a good separation.

VI. Reference

[1]Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Netw*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000.

[2]Ganesh R. Naik "Introduction: Independent Component Analysis" *RMIT University, Melbourne, Australia*

[3] Xingfu Zhang and Xiangmin Ren "Two Dimensional Principal Component Analysis based Independent Component Analysis for face recognition," 2011 International Conference on Multimedia Technology

[4]Yi-Ru, Lian " An investigation of frequency domain ICA for speech signal separation" NCTU, March, 2004.

[5] [Fast ICA code] http://research.ics.aalto.fi/ica/fastica/

[6] https://en.wikipedia.org/wiki/Independent_component_analysis