POLISH ACADEMY OF SCIENCES - MATERIALS SCIENCE COMMITTEE
SILESIAN UNIVERSITY OF TECHNOLOGY OF GLIWICE
INSTITUTE OF ENGINEERING MATERIALS AND BIOMATERIALS
ASSOCIATION OF ALUMNI OF SILESIAN UNIVERSITY OF TECHNOLOGY

Conference Proceedings

AMME'2002

11th INTERNATIONAL SCIENTIFIC CONFERENCE
ACHIEVEMENTS IN MECHANICAL & MATERIALS ENGINEERING

Separation of mixed images by independent component analysis

T. Matsuura

Department of Mechanical System Engineering, Gunma University
1-5-1, Tenjin-cho, Kiryu City, Gunma, 376-8515, Japan

Independent component analysis (ICA) is a new data analysis technique with which modeling of the self-organization in the initial vision of a living body, the application to blind source separation, the analysis of brain activity by electroencephalography and other complex phenomena are studied. The most outstanding feature of ICA is that we only assume the independency between original signals and information except for that is completely unknown to us.

In this report, we investigated whether the original pictures can be restored again by using ICA from the mixed pictures.We analyzed it by using the simple sample pictures and the actual photograph as application.Moreover, the validity of our learning algorithm and its stability were verified and considered.

1. INTRODUCTION

When there are two or more statistically independent sources of signal, generally the signals observed are those with which they were mixed. When those signals are generated by mixing the original signals in addition, the method of separating and identifying the original independent signals from those mixed ones is independent component analysis.

For example, when two or more men are in a certain room and they have talked simultaneously, suppose that those voices are recorded using two or more microphones. Since the voice which each microphone catches is not every person's independent one, it is hard to catch (since two or more men's voices are being mixed). Then, we want to make it easy to separate each man's voice well and to catch. In this case, since we know neither what each man had talked nor the position relation between microphones and speakers, we have to reproduce the original voice signals only from the mixed voice signals.

This is the typical example of the technique called Blind Source Separation. It is formulated as a problem of signal processing which is to separate a specific signal mechanically only paying attention to statistical character. Independent component analysis is the method of restoring the original independent signals under the assumption of the statistical independency between the original signals in such cases.

In this paper, we introduce the index which measures the independency between original sources, and consider whether the relation can be found between the independency of original signals and the level of the separation. Moreover, kurtosis of data (This is important value for guaranteeing the stability of the algorithm of the independent component analysis) is paid attention, and the effectiveness of the algorithm is verified with super-Gaussian data and sub-Gaussian data.

2. INDEPENDENT COMPONENT ANALYSIS

2.1. Formulation of problem

The original signals are $\mathbf{s} = \left(s_1(t), s_2(t), \ldots, s_n(t)\right)$, when there are n sources. We assume that original signals are mutually and statistically independent and those mean values are all zero. That is, each signal is assumed not to depend on time, and follow the probability density function where joint probability density function $P(\mathbf{s})$ can be shown in the form of the product of the probability density function of each signal. Assume that we observe n linear mixtures $x_1,\ldots,x_n$ of n independent components. At this time, the relation between **s** and **x** can be described by the linear model like $\mathbf{x} = \mathbf{As}$ by using mixture matrix **A**. In that case, we assume that the signal is instantaneously mixed and there is no delay through mixing. If **A** is already known and its inverse matrix exists, we ca easily obtain restored signals **y** as follows: $\mathbf{y} = \mathbf{A}^{-1}\mathbf{x}$

The problem is whether original signals **s** can be restored only from mixed signals **x** under the condition that **A**, **s** and their probability density functions are all unknown for us. Because of solving this problem only from the condition of independency of original signals, this is called independent component analysis.

When we make **y=Wx=WAs** by using non-singular matrix **W**, because the element of **s** is statistically independent and the inverse matrix of **A** exists, each element of **y** can be independent only when **WA** becomes a unit matrix **I** or **WA** becomes **DP**. Here, **D** is a diagonal matrix of $n\times n$, and the diagonal element is not 0. Moreover, **P** is a permutation matrix with one in the each row and the each column one by one and other elements of **P** are all zero. In the former case, separation signal **y** becomes equal with original signal **s**. In this case, an original signal is strictly restored including order. In the latter case, the size of the value of restored signal is scaled by **D**, and the order of the components in restored signals is different from the order of original ones.

As described above, if a mixture matrix **A** is known, it is easy to restore original signals. However, because **A** is generally unknown, the problem of the restoration by ICA is how to decide a separation matrix **W** which approximates $\mathbf{A}^{-1}$ or $\mathbf{DPA}^{-1}$ very well.

2.2. Learning algorithm

A general learning algorithm of ICA is given from the maximum likelihood estimation. The probability density function of mixed signal **x** is shown as follows.

$$P(\mathbf{x}) = \left|\det(\mathbf{W})\right| P(\mathbf{y}) \tag{1}$$

where $P(\mathbf{y})$ is a probability density function which presumes $P(\mathbf{s})$. The logarithm likelihood of expression (1) is as follows.

$$L(\mathbf{y},\mathbf{W}) = \log\left|\det(\mathbf{W})\right| + \sum \log P_i(y_i) \tag{2}$$

The learning law of **W** is given by maximizing the logarithm likelihood for **W** as follows

$$\Delta \mathbf{W} \propto [(\mathbf{W}^T)^{-1} - \varphi(\mathbf{y})\mathbf{x}^T] \text{ (steepest descent method)} \tag{3}$$

where

$$\varphi(\mathbf{y}) = -\frac{\partial\varphi(\mathbf{y})/\partial\mathbf{y}}{P(\mathbf{y})} = -\left[\frac{\partial\varphi(y_1)/\partial y_1}{P(y_1)}, \cdots, \frac{\partial\varphi(y_n)/\partial y_n}{P(y_n)}\right]^T \tag{4}$$

3. SIMULATION AND ITS RESULT

The original pictures used for our simulations are 8-bit-gray-scale and those are of the size of 200×200 pixels. After the bias was put so that the average of each image may become 0, as a preprocessing of ICA, PCA was done. The original images were restored by ICA from the mixed images which were made from four original picture images by linear mixing.

As shown in Figure.1, these images have been clearly restored without overlapping. The index of independency of those four original pictures is e=222821.7. In this case, the probability density functions of four original pictures are all sub-Gaussian.

In the case of Figure 2, the index of the independence is e=164990.3. When the independency between signals is kept well, the index of the independence that we introduced here reaches the small value. Therefore, the independency of the combination of original images in Figure.2 is higher than that of the combination of the original images of Figure 1. In ICA, the independency between signals is first assumed, and the restoration of the mixed signals is executed by using the assumption. Therefore, it is forecast that the separation will be done better for the mixed images made from the original pictures with higher independency. Nevertheless, the separation of mixed images in Figure 2 where independency is kept more is not better than that of Figure 1.

In the two examples above, both super-Gaussian and sub-Gaussian were included in the distribution of the original pictures in the latter case, though the distributions of the original pictures were all sub-Gaussian in the former case. In our simulations the non-Gaussian of the original pictures was the same in most examples in which we could succeed separating. Even when the non-Gaussian of the original pictures was not the same, we could get some cases that were separated to a certain degree. But they were not restored to the level of the case of Figure 1. Even when independency was worse (lower) than the case of Figure 1, there were some cases that were separated to significant degree. But we could find the non-Gaussian of them same in all the cases.

However, it cannot be thought that the level of the independency of the original pictures is not related to the success or failure of the separation, because the independency between signals is one of major premises of ICA.

Then, we examined the relationship between the dependency of original images and the level of the separation by using artificial images. Figure 3 shows the result of the separation when four images with complete independence and one random image are used as original pictures. This result shows a complete separation. On the other hand, Figure 4 shows the result of the separation from the original images which were a little different from ones of Figure 4 so as to ruin the independency. Each original picture is composed of a belt which is white or black. However, the belt of the gray appears in the separation result as shown in Figure 4 because it was not able to complete the separation. In both cases, the distribution of the original pictures is all sub-Gaussian. These results from the artificial original images show that the independency of the original images is undoubtedly related to the easiness of the separation.

4. CONCLUSIONS

Paying attention to the statistical character of original signals, we applied independent component analysis to restore original signals from mixed signals and we considered the result of separation. We can summarize our research as follows.

1. The relation is seen between the level of the independency of original signals and the separation result by ICA. That is, when the independency of original signals is higher, the separation of the mixed signals made from them is easier.
2. The quality of separation by ICA depends on the non-Gaussianity in the probability distribution of original signals. That is, when all original signals are same type (super-Gaussian or sub-Gaussian), it is easier for ICA to restore original images from mixed ones. The reason of this fact seems to be from that the non-Gaussian of an original signal is presumed from an imperfect separated signal while learning has not been completed yet, and the learning law has been switched by the result of this defective presumption.
3. Though the goodness of the separation depends on the two conditions mentioned above, there is tendency to depend more on the latter condition.

Figure 1. Result of separation by ICA(fully separated)

Figure 2. Result of separation by ICA(not separated)

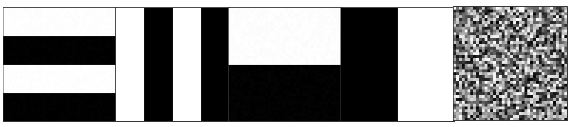

Figure 3. Result of separation by ICA (independent images)

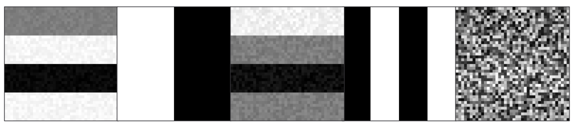

Figure 4. Result of separation by ICA (not independent images)