

複合化学センサシステムにおけるパターン認識の視覚化

吉村 忠与志^{a*}, 青山 義弘^b, 上嶋 晃智^a

^a 福井工業高等専門学校、物質工学科, 〒 916-8507 鯖江市下司町 16-1

^b 福井工業高等専門学校、電子情報工学科, 〒 916-8507 鯖江市下司町 16-1

*e-mail: tadayosi@fukui-nct.ac.jp

(Received: February 25, 2002; Accepted for publication: November 29, 2002; Published on Web: February 12, 2003)

ステレオペア表示は、構造的に重なるの多い分子の立体構造表示などに利用されている。これをグラフの散布図作成に応用したのは本論文の独自な部分である。ステレオペアの散布図は、一般的な表計算ソフト Excel/VBA により作成することができ、近接のクラス同士を 3 次元空間にプロットするのに有効である。主成分のスコアは数次元のデータに及ぶが、人間の視覚に訴えるものとしては 3 次元が限度であり、ステレオペア散布図の作成はパターン認識法の視覚化に貢献するものである。

キーワード: ケモメトリックス, パターン認識, 主成分分析法, K-L プロット, ステレオペア散布図

1 はじめに

複合化した化学センサシステムにおいて計測される多変量データの解析に主成分分析法 (PC) がよく用いられる。主成分のスコア値は 3 次元以上のデータ群である。しかし、3 次元データを表示する効果的なグラフィックツールはほとんど無い。本論文では、主成分スコアの 3 次元表示にステレオペア法 [1] の有用性を論じた。

測定対象の特徴を把握するには、多変量・多変数のデータを求めるのが一般的であり、そのようなデータを的確に評価する手法として、多変量解析法がある [2-16]。本報では主成分分析法について簡潔に述べ、総合特性値をどのように得て視覚化をするかを記述する。

多次元データの図示法にはこのほかに、レーダチャート、二次元および三次元空間写像などがある。写像法には、線形写像法 (Karhunen-Loeve 変換) と非線形写像法がある。前者は、多変量解析法の主成分分析法の結果を利用して、二次元あるいは三次元空間にデータの集合を写像することにより図示する手法で、このプロットを K - L プロット (Karhunen-Loeve plot) という。d 次元パターンを 2 次元に写像する例であるが、その写像への信頼性は累積寄与率を 95 % 以上説明で

きる主成分数で図示する。しかし、2 次元での図形表示は必ずしも適切ではなく、3 次元写像を必要とすることが多い。本報では、3 次元写像で用いられる虫ピン散布図の表示と、新たに提案するステレオペア散布図表示における近接するクラスターの分別視の有用性を論じる。

2 複合イオンセンサシステムによるリンゴワインの酒造工程の場合

イオン選択性電極を 7 種複合化したマルチチャンネル計測システムを開発し、化学調味料の味覚評価 [17]、リンゴワインの酒造工程の味覚評価 [18]、竹炭の水浄化過程の測定 [19] における成果を報告した。その時点では、2 次元図示から 3 次元図示の必要性を論じていたが、今後の課題としてきた [20]。

報文 [18] では、リンゴワインの酒造工程をマルチチャンネル計測システムで計測管理した成果を報告した。サンプリングした試料は人による味覚評価も合わせて実施した。複合した 7 種のイオンセンサの出力電圧値は Figure 1 のとおりである。これらの複合した 7 種類のイオンセンサ (NH₃, K, Cl, Na, Ca, H, CO₂) 出力

電圧値を次のように主成分分析を行ない、各主成分 Z_i を求めたのが Table 1 である。

n 個の d 次元データ ($x_{ik}, i = 1, 2, \dots, n; k = 1, 2, \dots, d$) を二次元空間へ写像するために、固有値の最大のもの、2番目のもの、そして3番目のものに対応する固有ベクトル t_1, t_2, t_3 を使って、3次元パターン Z_i に変換する。

$$Z_{i\beta} = \sum x_{ik} t_{k\beta} \quad (\beta = 1 \sim 3)$$

Z_1, Z_2, Z_3 軸はもとのデータ集合の線形結合における第1主成分軸、第2主成分軸、第3主成分軸である。主成分数を決定する基準として寄与率があり、各主成分が元の情報をどの程度反映しているかを表す。累積寄与率は各主成分の寄与率を累積する。

味覚評価 [18] から、第1日目から10日目まではアルコール発酵初期、11日目から20日目まではワイン主発酵期、21日目からはワイン熟成期と判定した。この3つの酒造工程を区別するために、Figure 2 の散布図ではプロット点を変えることにする (Table 1 の表記マークは丸に色別するもので、上から黒丸、白丸、赤角となっている)。

先に示した主成分分析を行なったところ、累積寄与率は第3主成分で86.8%と低く、第5主成分で96.6%となっていた。Excel等の処理では、2次元散布図が Figure 2 のように作成される。第2主成分まででは累積寄与率は74.1%となっていて、Figure 2 では十分にクラスタとしての分別視が出来ないと思われる。

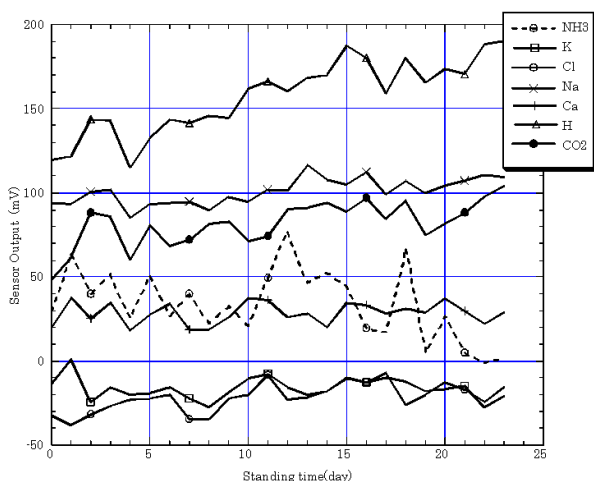


Figure 1. The aging variation of output voltage value (mV) of each ion sensor in the brewing process of apple wine [18].

主成分分析の結果は、教師なし学習による総合特性値を求めるものであり、累積寄与率が95%を越える主成分数で表示しようとする、Figure 2 の場合第5主成分軸まで作成しなければならないが、写像方法では3次元が限界である。そこで、虫ピン3次元散布図を Figure 3 に示す。虫ピン表示方法は一般のグラフ表示として定着している [19]。しかし、Figure 3 のようなデータ集合の場合、プロット点を変えて表示してもクラスタに分別視することはむずかしいものがある。

Table 1. The score of the principal component data.

Day	Mark	Z_1	Z_2	Z_3
1	●	3.256	-0.715	-0.652
2	●	3.323	2.819	0.786
3	●	0.856	-1.423	1.014
4	●	0.630	0.796	0.929
5	●	3.023	-1.462	-1.274
6	●	1.450	-0.108	0.283
7	●	1.160	0.727	-0.977
8	●	2.099	-1.801	0.575
9	●	1.748	-2.726	0.170
10	●	0.669	-0.470	-0.091
11	○	0.532	1.504	-1.114
12	○	-0.288	2.501	-0.550
13	○	0.426	0.410	1.752
14	○	-1.207	-0.171	1.305
15	○	-1.090	0.288	1.008
16	○	-1.382	1.855	0.013
17	○	-2.132	0.878	-0.236
18	○	-0.545	0.899	-1.482
19	○	-0.585	0.973	1.898
20	○	-0.089	-0.364	-1.222
21	■	-0.965	1.127	-0.631
22	■	-1.203	-0.072	-0.870
23	■	-1.907	-2.412	0.009
累積寄与率		0.456	0.741	0.868

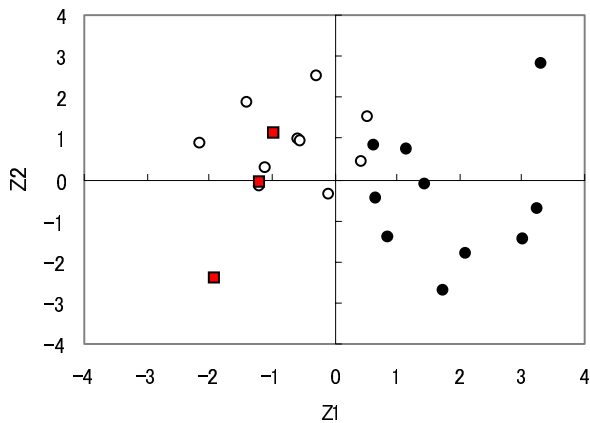


Figure 2. K-L plotting of the first (Z1) and the second (Z2) of the PC data. Mark is in accord with Table 1.

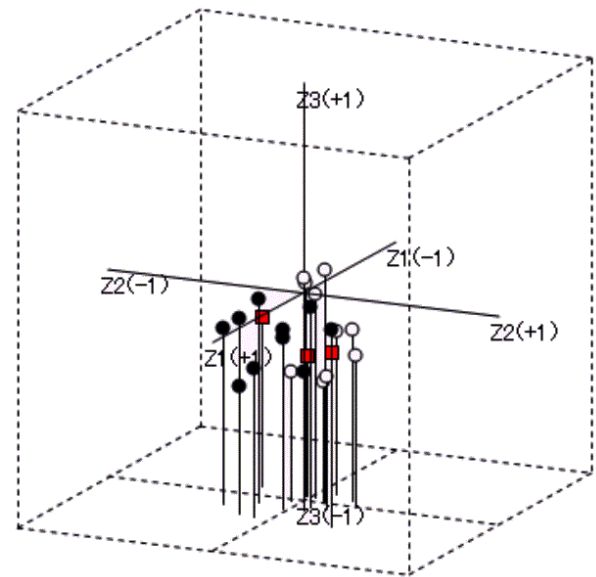


Figure 3. The insect-pin three-dimensional scatter diagram. Mark is in accord with Table 1.

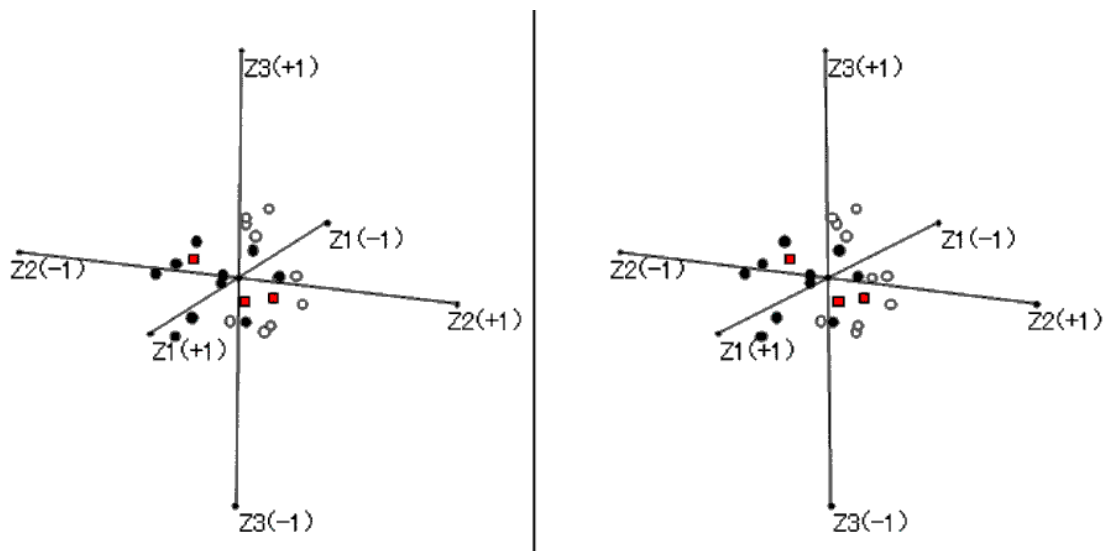


Figure 4. The stereo pair scatter diagram. Mark is in accord with Table 1. ($\phi=25^\circ$, $\theta=15^\circ$, $\psi=3^\circ$)

3 ステレオペア表示による3次元散布図の作成

虫ピン表示による写像法は分離してみたいクラスが近くで集合しているような場合は適切な表現方法とはいえないのである。そこで、3次元散布図をステレオペアで表示することを考えた。パソコン保持者が持つ

ている表計算ソフト Excel/VBA でステレオペアの散布図を作成するプログラムを開発した。ステレオペアの作成には、森口 [21] の方法を引用した。VBA マクロ・プログラムの一部を付録で示す。データ数は 30 個以内と限定した。Call 文でのサブルーチンはグラフィックルーチン [21] で、本文では省略する。

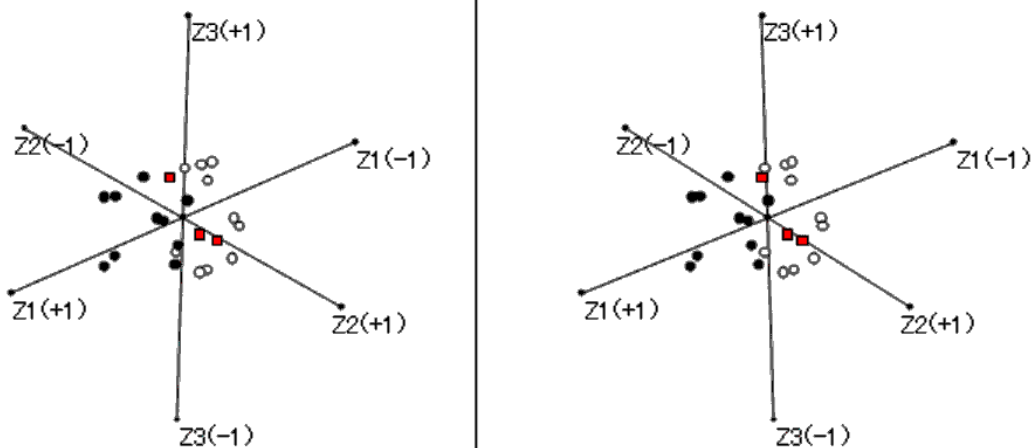


Figure 5. The stereo pair scatter diagram. Mark is in accord with Table 1. ($\phi=50^\circ$, $\theta=30^\circ$, $\psi=3^\circ$)

この散布図は第1主成分 (Z_1) から第3主成分 (Z_3) についてプロットし、軸の幅を ± 1 に換算して表示している。

立体視する角度を変化できるように、 xy 平面を ϕ 、 xz 平面を θ 、そして、変化させた xyz 面を ψ だけ回転した図を作成するようにした。付録リストの Figure 7 の値は $\phi=50^\circ$, $\theta=30^\circ$, $\psi=3^\circ$ で、回転させたものを作図するためである。3次元散布図をステレオペア表示する試みは本報の独自のものである。付録の Figure 6 にシート上での作図の様子を図示する。累積寄与率の悪いデータでもステレオペア立体表示でクラスタの分別視が出来た。累積寄与率が悪い場合、分別視したいクラスタが近接していることが多く、平面投影の虫ピン表示では不可能な分別視がステレオペアで可能となった。

回転の角度を Figure 4 では $\phi=25^\circ$, $\theta=15^\circ$, $\psi=3^\circ$ に、Figure 5 では $\phi=50^\circ$, $\theta=30^\circ$, $\psi=3^\circ$ にセットしたものを作成した。両方の図を立体視で凝視されると、3種類のプロット点が3次元空間で位置している様子が観察される。

4 おわりに

ケモメトリックスにおいて、化学センサを複合して得られた化学情報をパターン認識法で解析する情報処理が多い。主成分分析において主成分軸での散布図の視覚化が検討される中で、近接するクラスタを分別視

する方法は虫ピン散布図までであったが、ステレオペア散布図を提案することにより、分別視できなかったものも視覚化することができ、有用であった。また、よく普及している表計算ソフト Excel/VBA を用いて簡単なプログラムでステレオペア散布図を作成できることを報告した。

5 付録

Excel シート上の作図の様子を Figure 6 に示す。さらに、ステレオペアの散布図を作成するための VBA マクロ・プログラムの一部を Figure 7 に示す。

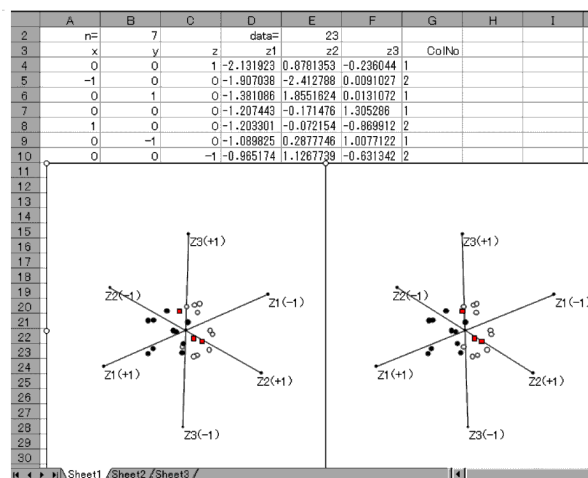


Figure 6. The aspect of the construction on Excel sheet.

```

Sub StereoPairColor()
Dim x(30), y(30), z(30), col(30)
Dim x2(30), y2(30), z2(30), xL(30), xR(30)
Dim xxL(30), xxR(30), zzz(30)
n = Cells(2, 2)
'----- axis rotation -----
DtG = Atn(1) / 45
ph = 50 * DtG: th = 30 * DtG: ps = 3 * DtG
sp = Sin(ph): st = Sin(th): ss = Sin(ps)
cp = Cos(ph): ct = Cos(th): cs = Cos(ps)
a11 = cp * ct: a12 = sp * ct: a13 = st
a21 = -sp: a22 = cp: a23 = 0
a31 = -cp * st: a32 = -sp * st: a33 = ct
xL0 = 175: xR0 = 425: zZ0 = 200: ul = 100
'----- axis -----
For i = 1 To n
    x(i) = Cells(3 + i, 1): y(i) = Cells(3 + i, 2)
    z(i) = Cells(3 + i, 3)
Next i
GoSub Scaler
Call Drectangle(50, 50, 500, 300)
Call Dline(xxL(2), zzz(2), xxL(5), zzz(5), 0)
Call Wlabel(xxL(2), zzz(2), "Z1(-1)")
Call Wlabel(xxL(5), zzz(5), "Z1(+1)")
Call Dline(xxR(2), zzz(2), xxR(5), zzz(5), 0)
Call Wlabel(xxR(2), zzz(2), "Z1(-1)")
Call Wlabel(xxR(5), zzz(5), "Z1(+1)")
Call Dline(xxL(6), zzz(6), xxL(3), zzz(3), 0)
Call Wlabel(xxL(6) - 5, zzz(6), "Z2(-1)")
Call Wlabel(xxL(3) - 5, zzz(3), "Z2(+1)")
Call Dline(xxR(6), zzz(6), xxR(3), zzz(3), 0)
Call Wlabel(xxR(6) - 5, zzz(6), "Z2(-1)")
Call Wlabel(xxR(3) - 5, zzz(3), "Z2(+1)")
Call Dline(xxL(1), zzz(1), xxL(7), zzz(7), 0)
Call Wlabel(xxL(1), zzz(1), "Z3(+1)")
Call Wlabel(xxL(7), zzz(7), "Z3(-1)")
Call Dline(xxR(1), zzz(1), xxR(7), zzz(7), 0)
Call Wlabel(xxR(7), zzz(7), "Z3(-1)")
Call Wlabel(xxR(1), zzz(1), "Z3(+1)")
Call Dline(300, 50, 300, 350, 0)
For i = 1 To n
    Call Dcircle(xxL(i), zzz(i), 1, 0)
    Call Dcircle(xxR(i), zzz(i), 1, 0)
Next i
'--- plot data ---
n = Cells(2, 5): sx = 0: sy = 0: sz = 0: ur = 3
For i = 1 To n
    x(i) = Cells(3 + i, 4): y(i) = Cells(3 + i, 5)
    z(i) = Cells(3 + i, 6)
    col(i) = Cells(3 + i, 7)
    sx = sx + Abs(x(i)): sy = sy + Abs(y(i))
    sz = sz + Abs(z(i))
Next i
For i = 1 To n
    x(i) = x(i) / sx * ur: y(i) = y(i) / sy * ur
    z(i) = z(i) / sz * ur
Next i
GoSub Scaler
For i = 1 To n
    If col(i) = 0 Or col(i) = 1 Then
        Call Dcircle(xxL(i), zzz(i), 2, col(i))
        Call Dcircle(xxR(i), zzz(i), 2, col(i))
    Else
        Call Drectangle(xxL(i) - 3, zzz(i) - 3, 4, 4)
        Selection.ShapeRange.Fill.ForeColor.SchemeColor = 2
        Call Drectangle(xxR(i) - 3, zzz(i) - 3, 4, 4)
        Selection.ShapeRange.Fill.ForeColor.SchemeColor = 2
    End If
Next i
'--- grouping ---
ActiveSheet.Shapes.SelectAll
Selection.ShapeRange.Group
End
Scaler:
For i = 1 To n
    x2(i) = a11 * x(i) + a12 * y(i) + a13 * z(i)
    y2(i) = a21 * x(i) + a22 * y(i) + a23 * z(i)
    z2(i) = a31 * x(i) + a32 * y(i) + a33 * z(i)
    xL(i) = x2(i) * ss + y2(i) * cs
    xR(i) = -x2(i) * ss + y2(i) * cs
    xxL(i) = xL0 + xL(i)*ul:xxR(i) = xR0 + xR(i)*ul
    zzz(i) = zZ0 - z2(i) * ul
Next i
Return
End Sub

```

Figure 7. Micro-program list of the stereo pair scatter diagram.

参考文献

- [1] 大岩正芳, 分子の立体図を描こう, 化学, **35**(1), 38 (1980).
- [2] 吉村忠与志, 青山義弘, 化学のための計測制御入門, 日本ファイルコン (1997).
- [3] 宮下芳勝, 佐々木慎一, ケモメトリックス - 化学パターン認識と多変量解析 -, 共立出版 (1995), p.11.
- [4] 奥野恵一, 久米均, 芳賀敏郎, 吉澤正, 多変量解析法, 日科技連 (1971).
- [5] P. C. Jurs, B. R. Kowalski, T. L. Isenhour, *Anal. Chem.*, **41**, 21 (1969).
- [6] B. R. Kowalski, *Computers in Chemical and Biochemical Research*, vol.2, ed. by C. E. Klopfenstein, C. L. Wilkins, Academic Press, New York (1977).

- [7] 佐々木慎一, 阿部英次, 高橋由雅, 高山千代蔵, 官下芳勝, 化学者のためのパターン認識序説, 東京化学同人 (1984).
- [8] S. Wold, M. Sjostrom, *Chemometrics: Theory and Applications*, ed. by B. R. Kowalski, Am. Chem. Soc., Washington D. C. (1977), p.165.
- [9] 金谷重彦, 吉村忠与志, 官下芳勝, 佐々木慎一, 日本化学会第 53 秋季年会講演予稿集 I, 187 (1986). H. Abe, T. Yoshimura, S. Kanaya, Y. Takahashi, Y. Miyashita, S. Sasaki, *Anal. Chim. Acta*, **194**, 1 (1987).
- [10] 船津公人, *CICSJ Bulletin*, **15(6)**, 3 (1997).
- [11] 官下芳勝, 佐々木慎一, ケモメトリックス - 化学パターン認識と多変量解析 -, 共立出版 (1995), p.143. Y. Takahashi, Y. Miyashita, Y. Tanaka, H. Abe, S. Sasaki, *J. Med. Chem.*, **25**, 1245 (1982).
- [12] 構造活性相関懇話会編, 薬物の構造活性相関: ドラッグデザインと作用機作研究への指針, 南江堂 (1979).
- [13] 吉村忠与志, 竹澤博士, 笠島徹, 加藤浩二, 藤上義弘, *分析化学*, **43**, 873 (1994).
- [14] 吉村忠与志, 斉藤政晴, *分析化学*, **39**, T151 (1990).
- [15] D. Coomans, I. Broeckaert, *Potential Pattern Recognition in Chemical and Medical Decision Marking*, Research Studies Press, England (1986), 59.
- [16] 吉村忠与志, *分析化学*, **37**, 612 (1988).
- [17] Tadayosi Yoshimura, Akira Yoshida, Yoshihiro Aoyama, *Denki Kagaku*, **64**, 1244 (1996).
- [18] 青山義弘, 南友和, 吉村忠与志, 山田明文, *電気化学会技術・教育研究論文誌*, **7**, 139 (1998).
- [19] 吉村忠与志, 土田耕三, 青山義弘, *電気学会論文誌 E(センサ・マイクロマイน์準部門誌)*, **121-E**, 376 (2001).
- [20] 青山義弘, 野原啓介, 吉村忠与志, 山田明文, *電気化学会技術・教育研究論文誌*, **9**, 87 (2000).
- [21] 森口繁一, *Excel/VBA 基礎指南*, 日本規格協会 (2000), p.120.

Visualizing of Pattern Recognition in Plural Chemical Sensors System

Tadayosi YOSHIMURA^{a*}, Yoshihiro AOYAMA^b and Akinori UEJIMA^a

^aDepartment of Chemistry & Biolog Engineering, Fukui National College of Technology

^bDepartment of Electronics & Information Engineering, Fukui National College of Technology

Sabae, Fukui 916-8507, Japan

*e-mail: tadayosi@fukui-nct.ac.jp

The pattern recognition method is used as principal component (PC) analysis in chemistry research on plural sensors systems. The PC score pattern is found from the three-dimensional (3D) data. However, the effective graphical tool for displaying the 3D data is limited. This paper describes the usefulness of the stereo pair in 3D display of the PC data. The scatter diagram of the stereo pair, made by the Excel/VBA macro-program, is useful for the plotting with similar clusters.

Keywords: Chemometrics, Pattern Recognition, Principal Component Analysis Method, K-L Plot, Stereo Pair Scatter Diagram