المجلة العراقية للعلوم الاحصائية (10) 2006
ص ص [1–29]

إستخدام بعض أساليب متعدد المتغيرات لتقليل الأبعاد الصورية في تطبيقات علوم الحياة

الدكتورعبد المجيد حمزة الناصر* أسماء غالب الراوي**

الملخص:

يتناول هذا البحث إستخدام خوارزمية تحويل Karhunen loeve وخوارزمية التحليل المتناظر لأستخلاص المعلومات المفيدة في مجموعة الصور المتعددة من خلال تقليل الأبعاد الصورية ، ثم المقارنة بينهما وبيان الطرائق التي تعطي النتائج الفضلى، وتم تطبيق ذلك على مجموعة سلسلة الصور لمراحل إنقسام خلايا حيوانية في علوم الحياة التي بلغت 21 حزمة صورية ملونة .إذ تم الإستنتاج بأن تحليل متعدد المتغيرات يسمح لنا بتمييز مجموعة بيانات متعددة الأبعاد ككل ، ووصفها كعلاقات لمختلف مصادر المعلومات المساهمة في المجموعة.وتم الإستنتاج أن التصوير المتعدد يمكن أن يسجل بيانات متعددة الأبعاد على كمية كبيره من المعلومات التي ربما يكون جزء منها مخفيا ويجب إستخلاصه.

Using some of multivariate analysis methods for the reduction of imaging dimensionality in biological sciences Applications

ABSTRACT

This research uses the Karhunen Loeve (KL) Algorithm and correspondence analysis algorithm for the extraction of the feature and find the region of interest in the sequences multivariate images by dimensionality reduction and then

* استاذ / رئيس جهاز الإشراف والتقويم العلمي / وزارة التعليم العالي والبحث العلمي
** مدرس/ رئاسة جامعة النهرين.

تاريخ التسلم : 8/ 6 /2006 ـــــــــــــ تاريخ القبول : 3/ 10 /2006

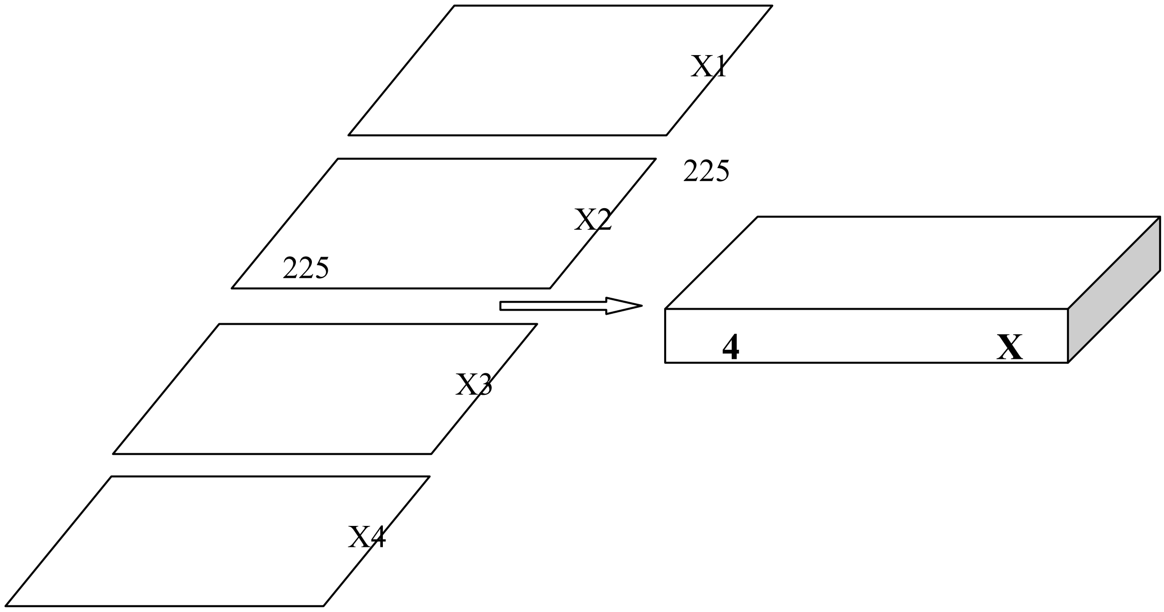
comparing them to show which method gives the best results. The potentials for using this method is illustrated through actual examples dealing with the study of chloride secrtion by airaway epithial cells for 21 bands. We can see that the multivariate analysis allows to distinguish dataset of multivariate dimensionality at once, and describe them as relations for different sources of information in the data. We show that the multivariate imaging can register multivariate datasets for a lot of information that part of it may be hidden and must be extracted.

1- المقدمة

ان وسائل التحليل المرئيMicroanalysis يمكن أن تزودنا بعدة صور لنفس العينة ، أي تسجيل عدة إشارات لمناطق الإستفادة من مجس العينة Aprobe on Specimen فعندما تسجل أكثر من صورة لنفس العينة فإنه يصبح من الصعب ، حتى على نظام الرؤيا الإعتيادية تفسير كل بيانات المجموعة الصورية . وهذا يعود الى سبب أن المعلومات المتعلقة بكل عنصر صورة هي في الحقيقه تتألف من N من الأبعادN- Dimensional إذ أن N تمثل عدد الصور . لكل عنصر صورة ممكن أن يمثل في هذا المجال بالموجه الذي إحداثياته في N من القيم الرمادية Gray Valuesلـــ N من الصور المختلفة [3] .

$$V(x,y) = \{ I_1 (x,y), I_2 (x,y), \dots I_N (x,y) \} \qquad \dots(1.1)$$

مهما يكن فالإرتباط وعدمه بين مختلف الصور، لإبعاد صحيحة تسمى أيضا الأبعاد الحقيقية لمجموعة البيانات هو دائماً صغير بالنسبة إلى عدد الصور المسجلة . وهذا يعني ان مجموعة البيانات تمثل دائما في مجال M أقل من N ومن دون فقدان الكثير من المعلومات المفيدة وتسمى عملية تقليل الأبعاد بالنقلة Mapping والشكل (1.1) يوضح عملية تقليل الأبعاد



الشكل (1.1) : يوضح عملية تقليل الأبعاد الصورية (النقلة)

يهدف البحث إلى عرض كيفية تحديد كمية المعلومات المفيدة التي تحتوي على الكثير من المعلومات الفائضةRedundent Information عن الحاجة التي تؤدي إلى تشوش Disturbs المعلومات المفيدة . إن معظم الطرائق التجريبية الإحصائية المستخدمة لإستخلاص المعلومات عن مجموعات بيانيه واسعة تقع ضمن مجموعتين أساسيتين في مجال التحليل الصوري وتشمل بناء النماذج وإحصاء متعدد المتغيرات [19] . تتألف أساليب متعدد التغيرات من تعريف جديد للمجال الممثل للبيانات أي بإسقاط هذه البيانات في مجال جزئي لتقليل الأبعاد وحساب محاور جديدة للعناصر الصورية يختار بأن يكون ممثلاً لكل البيانات مع عدد صغير من المركبات (M<N) ومن هذه الطرائق هي طريقة المركبات الأساسية أو مايعرف بطريقة تحويل Karhunen Loeve وطريقة التحليل المتناظر [4]. ومن أهم الدراسات والبحوث التي إهتمت بهذا الموضوع هي: في عام 1979 تطرق كل من.Jenson, L.&Waltz, E[13] الى طريقة Karhunen Loeve (KL)لتقليل المعلومات الفائضة التي تؤدي الى وصف بيانات متعددة الأبعاد غير

المرتبطة في أول مركبة أساسية. وفي عام1981قدم ,Heel, M.V.&Frank[11] J. التحليل المتناظر في تحليل صور مجهرية لخلايا حيوان السرطان .إقترح في عام1984 Gambtto, J.P. [8] خوارزمية تعقب عدة أهداف معا في تسلسل صور لتعقب عدة تحركات عسكرية اذ عرض أولا تقنية تقطيع من عدة مستويات لكشف عدة مناطق في كل حقل وبعد ذلك إستخدم التحليل المتناظر بين الحقل الحالي ونموذج المناطق المنجز .في عام 1987 طبق كل من Fuang,T.& [6] Ledrew, E. تحليل المركبات الأساسية لكشف التغيرات في غطاء- الأرضLand- cover مع زمن متعدد لبيانات القمر الصناعي متعدد الأطياف. وفي عام 1996 استخدم Al-Ani, L.A. [1] صور الأقمار الصناعية متعددة الحزم لمنطقة الرمادي غرب العراق وقد طبق اسلوب تحويل Karhunen Loeve لتحسين الصور .في عام1997بيّن كل من ,Baronti, S.Casini [2] Aأنه يمكن استخدام تحليل المركبات الأساسية لغرض تركيزمعنوية المعلومات لمجموعة صور لألواح معرض صور(لالواح زيتية من قماش وصور فوتوغرافية).وفي عام 2004 أيضا نشر Yokoo, T. [21] بحثه الذي هو محاولة لتزويدنا بدارسة نظرية شاملة وبعض الطرائق المقترحة لتحليل متعدد المتغيرات الإحصائي في مجال بيانات التصوير لنشاط خلايا الدماغ Generalized Indicator Function Analysis .

2- الجانب النظري

2-1 تقنية تحليل تحويل Karhunen Loeve [KL]

Karhunen Loeve Transformation Analysis Technique

ان تحليل تحويل [KL] المبني على الخصائص الإحصائية للصورة (ويشار إليه ايضا بتحليل المركبات الأساسية(Principal Components Analysis (PCA وهذا التحويل يسمى تحويل Hotelling اذ أن Hotelling قام بتحويل متغيرات متقطعة إلى معاملات غير مرتبطة وأشار إلى هذا التحويل بتحليل المركبات الأساسية(PCA). تحويل مطابق لهذا التحويل طبق على بيانات مستمرة لمجموعة غير مرتبطة من المعاملات سمي بتحويل[Karhunen Loeve [KL [5] اذ

برهنوا أن التحويل يؤدي إلى تصغير متوسط مربع الخطأ لذلك إستخدم في ضغط حجم البيانات وتقليصها ، وان تحويل بيانات صور ميكروسكوبية بإستخدام (PCA) يمكننا من الحصول على نتائج لمركبات أساسية جديدة أكثر تفسيرا للبيانات الأصلية [15] كما ان إستخدام تحليل الـــ (PCA) يمكن إستخدامة لتقليل المعلومات المكونة في مجاميع الصور إلى معلومات صورية مختزلة في موجهات المركبات الاساسية يمكن تفسيرها بوصفها مدخلا الى المسألة ، إفترض أن ابعاد الصورة f(x,y) هي $N \times N$ قد أرسلت M من المرات على قناة إتصال ما . ولأن أية قناة فيزيائية معّرضة لإضطرابات عشوائية فإن مجموعة الصور المستقبليـــــة $\{f_1(x,y),f_2(x,y),\ldots,f_m(x,y)\}$ تمثل مجموعة إحصائية تتحدد خواصها بخصائص القناة وطبيعة الإضطراب . إن مجموعة الصور المرسلة لنفس المشهد عن طريق مجس فضائي أو ميكروسكوب أو شريط سينمائي تعطي مثالا لهذه المجموعة . في هذه الحالة تفسد الصور بواسطة الإضطرابات الجوية أو الألكترونية والضجيج الكهربائي بين المرسل والمستقبل .

2-1-2- الأساس الرياضي Mathematical Bases

إفرض ان الصوره f(x,y) ، يمكن التعبير عن كل عينة صورة f(x,y) بمتجة ذي بعد واحد X_i لموجه الأبعاد– N^2 Dimensional Vector – N^2 كما يلي[17]:

$$X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iN^2} \end{bmatrix} \quad \ldots(1.2)$$

حيث أنx_{ij} تشير إلى jth من المركبات للموجة i . إن إحدى الطرائق لبناء هذا الموجه هي أن تشكل الـــ N مركبة الأولى من x_i من الصف الأول من f_i (x,y) أي أن:

$[\, x_{i1} = f(0,0)\ , x_{i2} = f(0,1),\ x_{i3} = f(0,2)\ ,\ldots\ x_{iN} = f(0,N-1)]$ والمجموعة الثانية من الـــ N مركبة من الصف الثاني ، وهكذا . هناك طريقه أخرى وهي أن

نستعمل أعمدة $f(x,y)$ بدلا من الصفوف.لذلك بالإمكان تحديد متوسط الموجه بالمعدل حيث أن : $m_x = E(x)$ ... (2.2) اذ أن $E(.)$ هي قيم متوقعة

تعرف مصفوفة التباين والتباين المشترك للمتجهات X كما ياتي

$$C_x = E\{ (x- m_x)(x- m_x)^T \} \quad \ldots(3.2)$$

ممكن تقريب المعادلات (2.2) (3.2) إلى صور معاينة بما ياتي :

$$m_x \cong \frac{1}{\mu}\sum_{i=1}^{\mu} x_i \quad \ldots(4.2)$$

$$C_x \cong \frac{1}{\mu}\sum_{i=1}^{\mu}(x-m_x)(x-m_x)^T \quad \ldots(5\text{-a}.2)$$ أومساوية إلى

$$C_x \cong \frac{1}{\mu}\left[\sum_{i=1}^{\mu}(x_i x_i)^T\right] - m_x m_x^T \quad \ldots(6\text{-b}.2)$$

إذا x لهاحجم N^2 و C_x ستكون مصفوفة $N^2 \times N^2$ ، العناصر C_{ij} لـــ C_x تمثل التباين بين نقاط الصورة ، في حين عناصر القطر C_{ii} تمثل التباين لكل صورة . إفرض e_i و λ_i، $i = 1,2, \ldots N^2$ تمثل موجهات مميزة والقيم المميزة لـــ C_x . مصفوفة التحويل لها صفوف الموجهات المميزه Eigen Vectorلـــ C_x تعطى بالشكل الاتي:

$$A=\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1N^2} \\ e_{21} & e_{22} & \cdots & e_{2N^2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N^21} & e_{N^22} & \cdots & e_{N^2N^2} \end{bmatrix} \quad \ldots (7.2)$$

حيث أن e_{ij} هي j_{th} من المركبات لـــ i_{th} من الموجهات المميزة . A هي$N\times N$ مصفوفة متعامدة Unitary Matrix أي أن $(A^{-1} = A^T)$ الصفوف لـــ A هي N من الموجهات المميزة القياسية Normalized Eigen Vector لـــ C_x .

لغرض إنجاز تحويل [KL] ، مصفوفة التباين يجب أن يكون قطرها بالشكل الاتي

$$C_x = AC_xA^T = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{N^2} \end{bmatrix} \quad \ldots(8.2)$$

حيث أن λ_N^2 ... λ_1 (تباين المركبات الأساسية) هي قيم مميزه لــ C_x ترتب بالشكل $\lambda_1 > \lambda_2 > \ldots \lambda_i$ تمثل تباين الصورة . وعليه فان أصغر قيمة مميزة ستحتوي فقط على المعلومات الخاصة بالضوضاء في الصوره.

تسمى مجموعة عناصر القطر لمصفوفة التباين والتباين- المشترك أثرا يمثل مجموعة كل الجذور المميزة لمصفوفة التباين . لذلك فان نسبة مجموع أول جذور لأثر أول مصفوفة تباين سيكون مقياس النسبة المئويه للطاقة ممثلا بأول موجة مميز [18] . فضلا عن ذلك بيانات التباين ستكتب كما ياتي :

$$\sum_{i=1}^{N^2} \lambda_i = \sum_{i=1}^{N^2} \sigma_{xi}^{\ 2} \ldots(9.2)$$

حيث أن σ_{xi}^2 تمثل تباينات حزمة الطيف اللوني الأصلية للتحويل المتعامد ، X يمكن إعادة بنائها بالعلاقة الاتية

$$X = A.Y + m^T \quad \ldots (10.2)$$

هنا $Y's$ هي كميات من الموجهات الخاصة Characteristic Vectors التي يجب إضافتها إلى متوسطات الموجهات لغرض تقديمها بوصفها تقديراً لموجه الصورة الأصلية، وتسمى مقياس الضرب. وهذا المقياس يحدد لكل موجة معاينة بتوليفة خطية بسيطة

$$Y = \sum_{i=1}^{N^2} W_i (X_i - m_x)^T \quad \ldots(11.2)$$

حيث أن W_i تمثل معامل الأوزان ، نحصل عليها بتقسيم عناصر الموجة V_i على أكبر قيمة له درجة العلاقه التبادلية بين المتغيرين K و 1 تحدد بمعامل الإرتباط r ويعطى كما ياتي:

$$R_{k1} = \frac{Cov_{k1}}{\sigma_k \sigma_1} \quad \ldots(12.2)$$

حيث Cov_{k1} هوالتباين المشترك بين المتغيرات

$\sigma_k\ \sigma_1$ الإنحراف المعياري للمتغيرين K و 1 [9].

وفيما ياتي خوارزمية <u>Karhunen Loeve</u>

3-1-2- الخوارزمية The Algorithm

ان اسلوب تطوير تحليل البيانات الصورية يوصف بالخوارزمية الاتية لتحويل [KL]

الخطوة 1

أدخل بيانات الصور لعينات صورية بعدد n لتشكيل مصفوفة البيانات $X_{n\times r}$

لعدد الصفوف n وعدد الأعمده r ونضع $i=1$.

الخطوة2

إحسب متوسط الصف في المصفوفة $X_{n\times r}$ بأخذ معدل البيانات لكل r من الأعمده.

الخطوة 3

إحسب المصفوفة P للصفوف n والأعمدة r بطرح متوسط الصف من كل صف في المصفوفة $X_{n\times r}$.

المصفوفة P تسمى مصفوفة متوسط البيانات المصحح

. Mean Corrected Data Matrix

الخطوة 4

إحسب مصفوفة التباين والتباين المشترك C_x بضرب مصفوفة التصحيح P Corrected Matrix

بالمبدلة أي أن $C_x = P^T P$.

الخطوة 5

إحسب أثر القيمة $\sum_{i=1}^{R}(P^T P)$

الخطوة 6

القيمة الأولية Initiate Value لموجة الصف هي

$$(U_0)_{1*r} = [1,1,....1]$$

الخطوة 7

إضرب $P^T P$ بموجة الصف U_0

-1.7

نتائج موجه الصف ($U_0 P^T P$) تقاس Normalized بتقسيم كل عنصر من العناصر على أعلى قيمه للعنصر. موجه النتائج يشار إليه بـ U_i

-2.7

إذا

$U_0=U_1$ {مجموع مربعات الخطأ < $|U_i-U_0|$}