Use Principal Component Analysis Technique to Dimensionality Reduction to Multi Source

استخدام تقنيه تحليل المركبات الرئيسيه لتقليل الابعاد الصورية متعددة المصادر

أ.م.د. اسماء غالب الراوي/ جامعة بغداد / كلية الادارة والاقتصاد
م.م. اسيل مسلم عيسى / جامعة سومر/ كلية الادارة والاقتصاد

OPEN ACCESS
CC BY NC SA

P - ISSN 2518 - 5764
E - ISSN 2227 - 703X

Received:16/4/2019
Accepted: 4/7/2019

مستخلص البحث:

يتناول البحث طريقة تحليل المركبات الرئيسية لتقليل الابعاد في حال كون التراكيب خطيه لمعالجة وتحليل الصور الرقمية .تعمل خوارزمية تحليل المركبات الرئيسية على تقليص عدد كبير من المتغيرات الى عدد اقل من المتغيرات غير مترابطة مع ضمان اقل خسارة ممكنه من المعلومات المفيدة .وتم تطبيق هذه الطريقة على مجموعة من صور الاقمار الاصطناعية لمنطقة معينه في محافظة البصرة والتي تمثل مصب نهري دجلة والفرات في شط العرب في محافظة البصره .ولقد تبين من تطبيق تحليل المركبات الرئيسية والذي يعتمد على القيم الكامنة ان تطبيق خوارزمية تحليل المركبات الرئيسية اعطت نتائج عاليه ودقيقه في تحديد افضل حزمة صوريه من بين الحزم الصورية الستة .فقد تم التوصل الى ان الحزمة الرابعة والتي تمتلك اعلى قيمة كامنة هي الافضل وهذا يدل على انها تحتوي على اهم الخواص المستقلة في الصور والتي تستخدم للتحليل كعزل المناطق المائية ونوع التربه والمناطق الزراعيه وغيرها.

المصطلحات الرئيسة للبحث/ الصور ,الصور الرقمية, المعالجة الصورية ,تقليل الابعاد ,تحليل المركبات الرئيسية .

*البحث مستل من رسالة ماجستير

Journal of Economics and Administrative Sciences
2019; Vol. 25, No.115
Pages: 464- 473

1-المقدمـــة

يعد علم الاحصاء من اهم العلوم المتداخلة مع العلوم الاخرى والذي يلعب دورا مهما في كثير من الدراسات والبحوث المختلفة .هنالك بعض الوسائل كتقنيات الاستشعار عن بعد وتقنيات التحليل المجهري وغيرها من التقنيات تلتقط عدد كبير من الصور متعددة الاطياف لغرض معين، وبما ان الصورة الرقمية يمكن ان تتمثل بمصفوفة من العناصر الصورية والتي تعرف بالصيغة $f(x,y)$ لذلك سنحصل على (N)من الابعاد التي تشكلها (M) من الصور المأخوذة لهدف معين وعند دراسة هذا العدد الكبير من الصور الملتقطة يمكن الاعتماد على احد اهم اهداف طرائق التحليل الاحصائي متعدد المتغيرات وهو تقليل الابعاد، ان عملية تقليل الابعاد في مجال معالجة الصور الرقمية تعتبر من اولى المهام المطلوب حلها . ان مفهوم تقليل الابعاد يعتمد على تحويل البيانات المتعددة الموجودة في المساحة ذات الابعاد العالية (High Dimensional Space) الى عرض ذو معنى اوضح ومفيد في المساحة ذات الابعاد الاقل (Low Dimensional Space)[9] ، ولحل هذه المشكلة يتم اللجوء الى طرائق متعدد المتغيرات الاحصائية والتي تكون المرحلة الاولى للمعالجة وبعدها يتم استعمال طرائق المعالجة الصورية من تحسين واستعادة وضغط وتقطيع وغيرها . وان لمفهوم تقليل الابعاد العديد من التقنيات القادرة على تنفيذ هذه المهمة، ويمكن ان تصنف هذه التقنيات الى صنفين هما تقنيات خطية لتقليل الابعاد (Linear Techniques for Dimensions Reduction) وهي التقنيات التي تعتمد على العلاقات الخطية بين المتغيرات ومنها التحليل العاملي (Factor Analysis) وطريقة تحليل المركبات الرئيسية (Principal Component Analysis (PCA)) والارتباط القويم وغيرها الكثير[5]،اما الصنف الثاني فهو تقنيات غير خطية لتقليل الابعاد (Nonlinear Techniques for Dimensions Reduction)[6] وفي الواقع ان اغلب المتغيرات تأخذ شكل المتغيرات غير الخطية ومنها طريقة تحليل المركبات الخطيه اللبيه (Kernel Principal Component Analysis) [7] وطريقة القياس متعدد الابعاد (Multidimensional Scaling) وغيرها .و في بحثنا ستناول تحليل المركبات الرئيسيه((Principal Component Analysis (PCA). وتم تطبيق هذه الطريقة على صور الاقمار الصناعية(LANDSAT-5) لمجموعة متغيرات الترابط الطيفية (correlation spectral) لستة حزم طيفية لمشهد في محافظة البصرة التي يمثل مصب نهري دجلة والفرات في شط العرب وأيضا القنوات المائية المتخللة لمحافظة البصرة والمسطحات المائية المنتشرة حول هذه القنوات . ومن اهم الدراسات و البحوث في هذا المجال :

في عام 2006 قدم الباحثان (الناصر و الراوي)[2]احدى طرائق متعدد المتغيرات في تقليل الابعاد لمعالجة الصور الرقميه ,اذ استخدما خوارزميتان الاولى الخوارزمية تحويل (Karhunen loeve) وخوارزمية الثانية خوارزمية التحليل المتناظر .

في عام 2010 قام الباحثان (حسو، مهى عبد الرحمن) و(صديق، منى جعفر)[1]باستعمال خوارزمية تحليل المركبات الرئيسية على صور الاستشعار عن بعد في تحديد افضل حزمة من بين الحزم المكونة لصور متعددة الاطياف (Multispectral Images) من خلال تحديد اعلى قيمه كامنة و بالاعتماد معيار نسبة اشارة الضوضاء (Signal to Noise Ratio) لتحديد افضل الحزم الناتجة.

استعمل في عام 2012 الباحثان (صالح و احمد)[3] تحليل المركبات رئيسيه (Principal Component Analysis) لتقليل الابعاد الصورية وهي احد طرائق متعدد المتغيرات في معالجة الصور الرقمية في تحسين النسجه.

في العام 2012 ايضا قام الباحث (Santo, R.)[8] بتوضيح تقنية تحليل المركبات الرئيسية في استعمالها في المعالجة الصورية مثل تحديد الانماط والضغط الصوري.

(1-1)هـدف البـحث

يهدف البحث الى استعمال تقنية المركبات الرئيسية لتقليل الابعاد الصورية لبيانات متعدد المصادر في حالة كون التراكيب خطيه واختيار افضل حزمة صور من خلال الاعتماد على اعلى قيمة كامنة .

2- الجاب النظري

(2-1) تحليل المركبات الرئيسية (Principal Component Analysis (PCA))

تحليل المكون الرئيسي او ما يعرف (Karhunen-Loeve أو تحويل (Hotelling [2][4]ينتمي إلى التحويلات الخطية بناءً على التقنيات الإحصائية ويعمل على اختزال المعلومات المتمثله بالتباين وحفظها في مركبات دون خسارة المعلومات. الفكره الاساسية للمركبات الرئيسية هو تحويل المتغيرات المرتبطة خطيا والتي تحتوي على كميه عاليه و مبعثره من البيانات الى مركبات متعامدة غير مرتبطة وتكون مرتبه حسب مقدار التباين , تحوي المركبه الاولى على اكبر نسبة تباين (Variance) من المعلومات اما الثانيه فتكون ثاني اكبر نسبة تباين من المعلومات وهكذا اما الاخيرة فتكون ضوضاء ومشوه لأنها تحتوي اقل نسبة تباين[3]. وهي أداة قوية لتحليل البيانات والتعرف على الأنماط والتي تستعمل غالبًا في معالجة الإشارة والصورة اذ تعمل كتقنية لضغط البيانات ، تقليل أبعاد البيانات أو ارتباطها كذلك. تعتمد على التحليل متعدد المتغيرات أو الشبكات العصبية التي يمكنها إجراء PCA على مجموعة بيانات معينة. الخطوات الرياضية لتقنية تحليل المركبات الرئيسيه تكون كما ياتي :

1- ايجاد مصفوفة variance-covariance للصور المتوفرة والتي تحسب وفق الصيغة الاتية :

$$\Sigma_{ij} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{i=1}^{N}(X_i(\kappa,\iota)-\mu_i)(X_j(\kappa,\iota)-\mu_j) \quad \dots\dots\dots(1)$$

اذ ان

M, N :تمثل عدد الوحدات الصورية في الصوره .

$X_i(\kappa,\iota)$ تمثل المركبه i في موقع النقطه (κ,ι) في الصوره.

$X_j(\kappa,\iota)$ تمثل المركبه j في موقع النقطه (κ,ι) في الصوره.

μ_j, μ_i يمثلان الوسط الحسابي للمركبة $\mu_i = \frac{1}{MN}\sum_{i=1}^{M}\sum_{i=1}^{N}X_i(\kappa,\iota)$

و$\mu_j = \frac{1}{MN}\sum_{i=1}^{M}\sum_{i=1}^{N}X_j(\kappa,\iota)$

2- يتم ايجاد مصفوفة التباين والتباين المشترك من الدرجه n بالصيغة الاتية :

$$\Sigma = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

المصفوفة اعلاه هي مصفوفة متماثلة ومحددها لا يساوي صفر.

3- ايجاد الجذور الكامنة من خلال الصيغة الاتية :

$$|\Sigma - \lambda_j I| = 0 \quad \dots\dots\dots(2)$$

where $\sum_{j=1}^{k}\lambda_j = k$

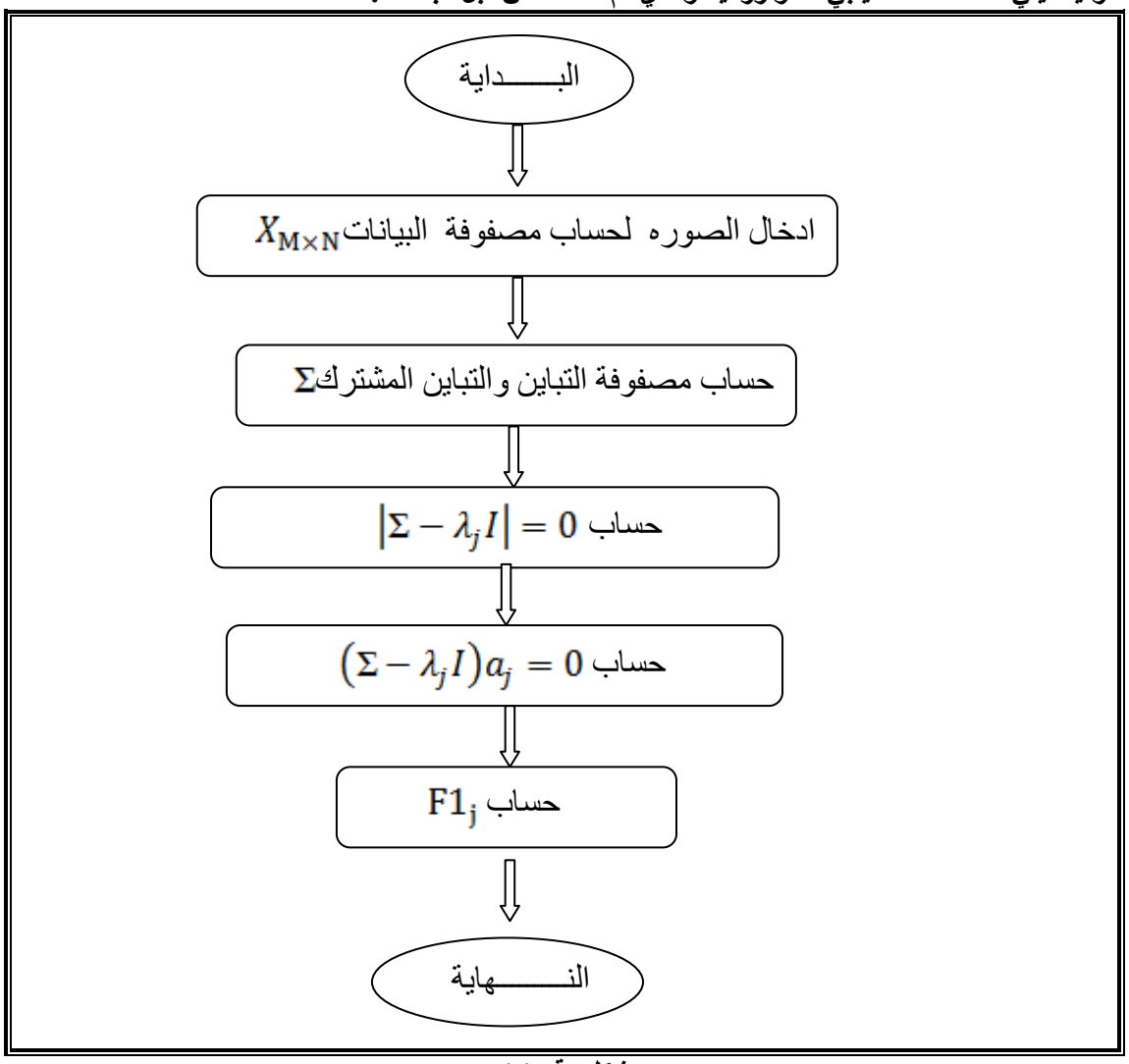
4- وبعدها يتم ايجاد الموجهات الكامنة من خلال المعادله كما ياتي :

$$(\Sigma - \lambda_j I)a_j = 0 \quad \dots\dots\dots(3)$$

5- و اخيرا يتم ايجاد المركبات الرئيسية كما يأتي :

$$F1_j = a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{kj}X_k \quad \ldots\ldots\ldots(4)$$

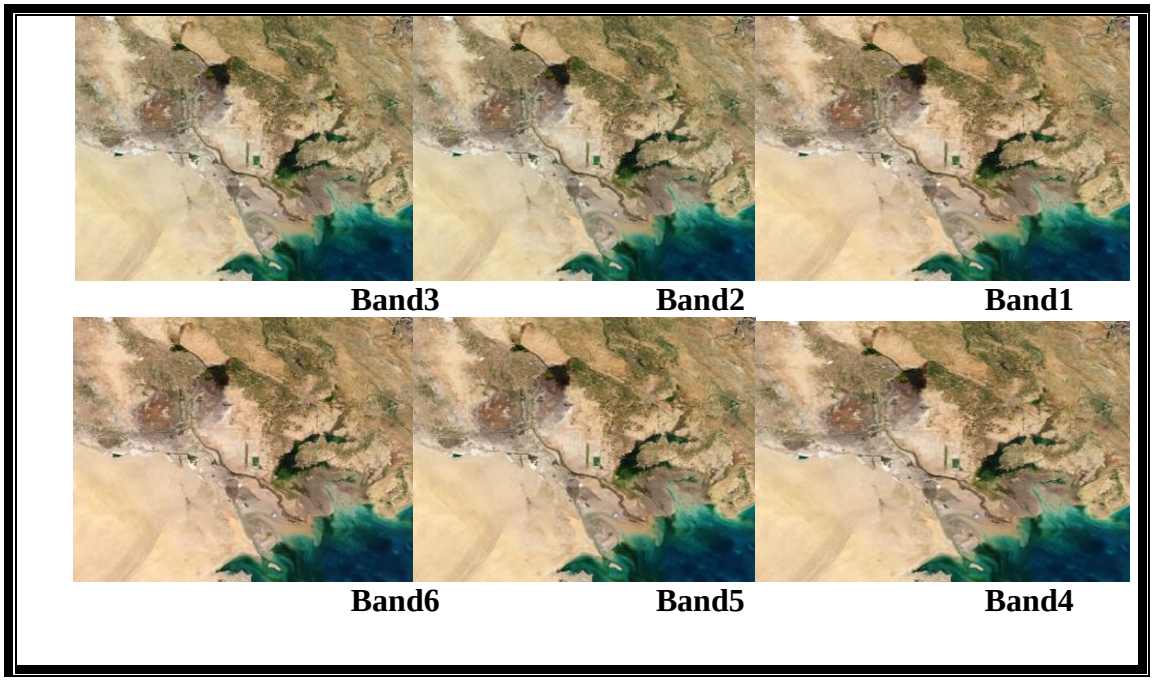
وفيما يلي المخطط الانسيابي للخوارزمية والذي تم اعداده من قبل الباحثة .



شكل رقم (1)
يمثل المخطط الانسيابي لخوارزمية طريقة PCA

3- الجانب العملي

لغرض تطبيق تقنية المركبات الرئيسية في الجانب النظري لتقليل الابعاد الصورية وبقياسات (256×312) بكسل لكل صوره تم اختيار صور لمصب نهري دجلة والفرات في شط العرب في محافظة البصرة والتي تم الحصول عليها من وكالة ناسا الفضائية عبر شبكة الانترنيت. تم اعتماد حزم النظام TM وكما موضح بالشكل (4) والتي تمثل الحزم الصورية الاصلية للقمر الصناعي(LANDSAT-5)وكالاتي :



شكل(4)

يمثل الحزم الطيفيه الاصلية لمشهد مصب نهري دجلة والفرات في شط العرب في محافظة البصرة

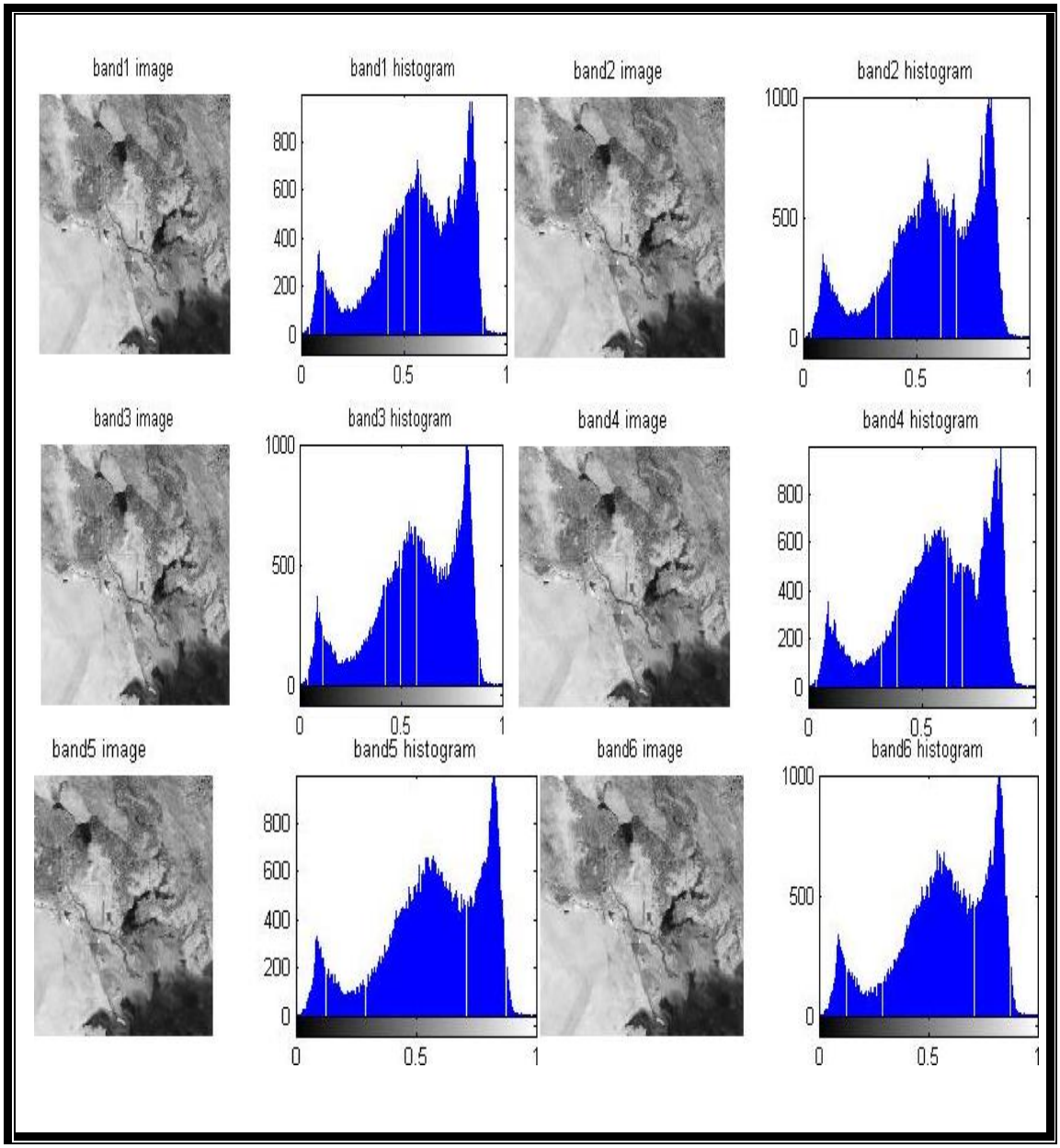
وبعد تلوين الحزم الصورية الست بالون الاخضر تم استخراج قيم الوسط لحسابي والتباين والانحراف المعياري لكل حزمة (Band) والتي تم كتابة بلغة برنامج matlab وكانت النتائج الاتية:

جدول (1)

المقاييس الاحصائية للصور الرقمية

Maximum	Minimum	Variance	Std. Deviation	Mean	
1	0	0.0879	0.2964	0.9027	Band 1
1	0	0.1623	0.4028	0.7962	Band 2
1	0	0.1965	0.4432	0.7314	Band 3
1	0	0.1153	0.3395	0.8671	Band 4
1	0	0.1777	0.4216	0.7689	Band 5
1	0	0.1968	0.4436	0.7307	Band 6

نلاحظ ان متوسط الحزم متقاربة وهذا يدل على ان شدة التدرج الرمادي للحزم متقاربة وان اقل شدة كانت للحزمة السادسة اي ان الحزمة مائلة للون الاسود اما اعلى قيمه للمتوسط هي الحزمة الاولى وهذا يدل ان شدة التدرج الرمادي عالية اي ان الحزمة مائلة للبياض,اما في ما يتعلق في التباين فان الحزمة الاولى كان تباينها قليل اي ان التغيرات بالكثافة قليله وبهذا فان مستويات التدرج الرمادي قليلة ,اما الحزمة السادسة فان التغيرات في الكثافة تكون اعلى جميع الحزم وبهذا فان مستويات التدرج الرمادي عالية لذلك الكثافة تكون عالية .والإشكال اداه تبين التوزيعات التكرارات.



شكل (2)
يبين التوزيع التكراري للحزم الصورية الاصل

جدول (2)
يبين الارتباط بين الحزم الطيفية

Band 6	Band 5	Band 4	Band 3	Band 2	Band1	
0.5408	0.5989	0.8387	0.5419	0.6491	1	Band 1
0.8333	0.9227	0.7739	0.8384	1	0.6491	Band 2
0.9981	0.9047	0.6461	1	0.8384	0.5419	Band 3
0.6449	0.7141	1	0.6461	0.7739	0.8387	Band 4
0.9031	1	0.7141	0.9047	0.9227	0.5989	Band 5
1	0.9031	0.6449	0.9981	0.8333	0.5408	Band 6

اما في ما يتعلق بالجدول رقم (2) فانه يمثل مصفوفة الارتباطات بين الحزم الطيفية والذي له اهمية كبيره في معالجة الصور الرقمية ,يقيس الارتباط قوة العلاقة بين الحزم وكذلك يوضح تكرار البيانات في الحزم.

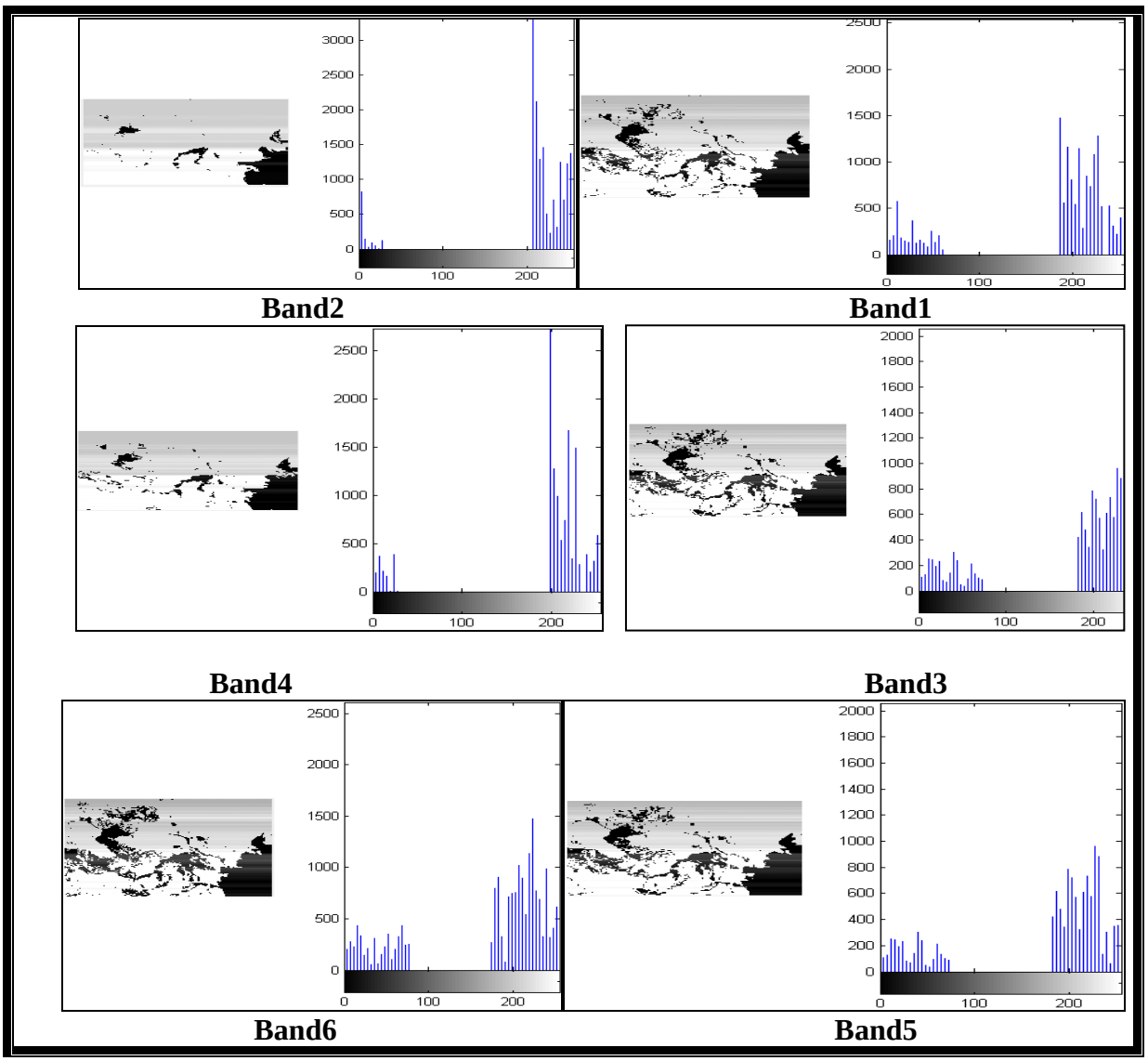
كما نلاحظ ان هنالك ارتباط بين الحزم الصورية وهذا يدل على ان البيانات مكرر في الحزم,وان قيمة الارتباط موجبة عالية بين الحزمة الصورية السادسة والحزمتان الثالثة والخامسة يليه الارتباط موجب بين الحزمة الصورية الخامسة والحزمتان الثانية و الثالثة .اما الحزمة الصورية الاولى فان ارتباطها قليل مع جميع الحزم الصورية اي ان تكرار البيانات قليل .

و بعد تطبيق تقنية المركبات الرئيسية على الحزم الصورية تم الحصول على القيم الكامنة لطريقة المركبات الرئيسية في جدول (3)مع اخذ اول خمس مركبات فقط بنظر الاعتبار تبين ان الحزمة الصورية الرابعة هي الافضل وذلك لكونها اخذت اعلى قيمة كامنة حيث بلغت قيمتها (26.3199) .

جدول (3)
يبين قيم الجذور الكامنة لكل طريقة

Eigen Values For P.C					
Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
24.6403	23.8814	21.0981	26.3199	22.5182	21.0706
3.5959	8.0245	9.8062	5.1488	8.9254	9.8228
2.4153	5.5065	7.4579	3.7615	6.1324	7.4966
1.6604	3.8433	5.6619	1.7334	4.8575	5.6810
0.8347	2.3036	3.1727	1.2329	2.6910	3.1856

والصور في الشكل (3) يوضح الحزم الصورية والمدرج التكراري الناتجة بعد اجراء تحليل المركبات الرئيسية.



شكل (3)

نتائج تطبيق خوارزميات تحليل المركبات الرئيسية على الحزم الطيفيه الستة

4-**الاستنتاجات**(Conclusions)

وفيما يلي الاستنتاجات التي تم التوصل اليها في متن هذا البحث :

1- من خلال تنفيذ تقنيات تقليل الابعاد المعتمدة اتضح ان الاعتماد على تقليل الابعاد للصورة اعطت نتائج جيدة وأكثر وضوحاً من اساليب اخرى ، وان الآلية التي اعتمدت لتنفيذ الخوارزمية مكنتنا من دراسة صور طيفية ذات احجام عالية.

2- تبين من تطبيق تحليل المركبات الرئيسية والذي يعتمد على القيم الكامنة ان تطبيق PCA اعطت نتائج عاليه ودقيقه في تحديد افضل حزمة صوريه من بين الحزم الصورية الستة .فقد تم التوصل الى ان الحزمة الرابعة والتي تمتلك اعلى قيمة كامنة هي الافضل وهذا يدل على انها تحتوي على اهم الخواص المستقلة في الصور والتي تستخدم للتحليل كعزل المناطق المائية والمناطق الزراعيه ونوع التربه وغيرها .

5- التوصيات(Recommendations)

1-تطبيق اساليب تقليل ابعاد مختلفة اخرى مثل تقنية تقليص واختيار القيم المطلقة الصغرى (Least Absolute Shrinkage and Selection Operator(LASSO)) ومقارنة النتائج مع تقنية تحليل المركبات الرئيسية كأساليب خطية لتقليل الابعاد.

2-اعتماد اساليب غير خطية لتقليل الابعاد الصورية مثل (Kernel Principal Component analysis ومقارنة النتائج مع تقنية تحليل المركبات الرئيسية كأساليب خطية وأخرى غير خطية لتقليل الابعاد .

المصادر

1- حسو ,مهى عبد الرحمن و صديق ,منى جعفر (2010)" اختيار أفضل حزمة باستخدام خوارزمية تحليل المركبات الاساسية على بيانات الاستشعار عن بعد" مجلة الرافدين لعلوم الحاسوب والرياضيات وقائع المؤتمر العلمي الثالث في تقانة المعلومات ص[39-29_].

2- حمزة، عبد المجيد الناصر و الراوي، أسماء غالب (2006)" استخدام بعض أساليب متعدد المتغيرات لتقليل الابعاد الصورية في تطبيقات علم الحياة ". المجلة العراقية للعلوم الاحصائية، المجلد 1427،العدد10،ص[1 -29_].

3- صالح, وصفي طاهر و احمد , زركار مغديد (2012)"استخدام تحليل المركبات الرئيسية لاستخلاص المعلومات وتحسين النسجة في معالجة الصور الرقميه" المجلد2،العدد 1.

4- محمد ,محمد عبد الودود, (2018),"استعمال بعض طرائق متعدد المتغيرات في التحليل الصوري مع تطبيق عملي"" ,رسالة ماجستير في الإحصاء ,كلية الإدارة والاقتصاد , جامعة بغداد .

5-Bajcsy, Peter; Groves, Peter(2004)" Methodology for Hyperspectral Band Selection", Photogrammetric Engineering & Remote Sensing, Number 7 , pp. [793-802] .

6-Changshui Zhang, Jun Wang, Nanyuan Zhao, David Zhang, (2004) , " Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction", Pattern Recognition 37,pp[325 – 336] .

7-Kwang In Kim, Matthias O. Franz, and Bernhard Scholkopf ,(2005), " Iterative Kernel Principal Component Analysis for Image Modeling" Transactions on Pattern Analysis and Machine Intelligence , 27,P[66- 1351].

8-Santo, Rafael do Espírito ,(2012)," Principal Component Analysis applied to digital image compression" .

9-L.J.P. van der Maaten, (2007)," An Introduction to Dimensionality Reduction Using Matlab", Universiteit Maastricht .

Use Principal Component Analysis Technique to Dimensionality Reduction to Multi Source

Abstract:

This paper tackles with principal component analysis method (PCA) to dimensionality reduction in the case of linear combinations to digital image processing and analysis. The PCA is statistical technique that shrinkages a multivariate data set consisting of inter-correlated variables into a data set consisting of variables that are uncorrelated linear combination, while ensuring the least possible loss of useful information. This method was applied to a group of satellite images of a certain area in the province of Basra, which represents the mouth of the Tigris and Euphrates rivers in the Shatt al-Arab in the province of Basra. In this research, when selected the best imaging band in terms of taking the highest eigen value, it is shown that the fourth image band is best when using the PCA method . The application principal component analysis, which depends on the eigen values, showed that the application of PCA gave high and accurate results in determining the best image among the six image beams. It was found that the fourth images which has the highest eigen value is the best and this indicates that it contains the most important Independent properties in images, which are used for analysis, such as isolating water areas, agricultural areas, soil type, etc.

Keywords: images , images processing ,digital image ,dimensionality reduction , principal component analysis .