

## Best Band Selection using Principle Component Analysis Algorithm from Remote Sensing Data

Maha A. Hasso

Mona J. Siddiq

College of Computer and Mathematics

University of Mosul

Received on: 24/10/2010

Accepted on: 10/11/2010

### ABSTRACT

The best band selection from remote sensing image plays an important roles in multispectral and hyperspectral remote sensing image processing due to the intercorrelation that inherent in the multispectral images taken by remote sensing sensors.

In this paper we use principle component analysis algorithm applied on remote sensing data and find covariance matrix for bands that should be processed then find eigen vector using Jacobi methods. The algorithm was applied on multispectral images of Thematic Mapper sensor, it concluded that the six band was the best band, the value of it's eigen value was the biggest one and the value of signal to noise ratio equals to 74.7217. This algorithm is constructed using Visual C# 2008 that is characterized by efficient and high speed implementation.

Keywords: Band, Principle Component Analysis, Remote Sensing Image.

اختيار أفضل حزمة باستخدام خوارزمية تحليل المركبات الأساسية على بيانات الاستشعار عن بعد

مهي عبد الرحمن حسو منى جعفر صديق

كلية علوم الحاسبات والرياضيات

جامعة الموصل

تاريخ قبول البحث: 2010/11/10

تاريخ استلام البحث: 2010/10/24

### المخلص

ان عملية اختيار افضل حزمة من بين حزم بيانات الاستشعار عن بعد تلعب دور مهماً و اساسياً في عملية معالجة صور متعددة الاطيف (hyperspectral and Multispectral images) وذلك بسبب الترابط المتواجد بين صور المتعددة الاطيف الماخوذة باجهزة الاستشعار عن بعد.

تم في هذا البحث استخدام خوارزمية تحليل المركبات الاساسية على صور الاستشعار عن بعد لتحديد افضل حزمة، حيث يتم ايجاد مصفوفة Covariance للحزم المراد معالجتها ثم ايجاد متجة ايكن باستخدام نظرية جاكوبي وقد تم تطبيق الخوارزمية على صور متعددة الاطيف لمتحسس الخرائط الموضوعي (Thematic Mapper, TM) تبين ان الحزمة السادسة هي افضل حزمة حيث كانت قيمة ايكن لها اكبر قيمة وكانت قيمة نسبة الاشارة الى الضوضاء (Signal to noise ratio, SNR) تساوي 74.7217. و قد تم بناء الخوارزمية باستخدام لغة فيجول سي شارب 2008 (Visual C# 2008) لما تمتاز به هذه اللغة من كفاءة وسرعة عالية في التنفيذ.

الكلمات المفتاحية: حزمة، تحليل المركبات الأساسية، الاستشعار عن بعد.

## 1- مقدمة

في السنوات الاخيرة اصبحت عملية استخلاص الخواص وبالاخص عملية اختيار افضل حزمة من الامور المهمة في معالجة الصور المتعددة الاطيايف (hyperspectral and Multispectral images) لبيانات الاستشعار عن بعد .واخذت اهتمام العديد من الباحثين في التطبيقات التي تمتاز باحتواءها على كميات هائلة من البيانات .

في الصور المتعددة الاطيايف اصبح من الضروري التقليل من عدد الحزم (Bands number) والتقليل من الابعاد العالية (dimensionality reduction) ، ان عملية اختيار افضل حزمة تعني كبس او تقليل عدد الحزم بدون فقدان المعلومات المهمة. وتوجد طريقتان للتقليل من عدد الحزم الاولى هي اختيار الحزم (Band selection) ونقصد بها اختيار الحزمة التي تحوي المعلومات الاكثر والطريقة الثانية هي استخلاص الخواص (Feature Extraction) او بمفهوم اخر كبس الحزم باستخدام التحويلات الرياضية [5] .

## 2- مفهوم الاستشعار عن بعد

من اجل التعمق في معرفتنا عن الكرة الارضية ، كان لابد من الحصول على معلومات اكثر شمولية واكثر دقة عنها وعن بعض الظواهر التي تحدث عليها .وحتى يتم ذلك كان لابد من ايجاد وسائل للرصد والمراقبة عن بعد وخاصة من الفضاء لاماكن على سطح الكرة الارضية يصعب الوصول اليها. تعتبر صور الاقمار الصناعية من مصادر البيانات الهامة للكثير من مجالات العلوم التطبيقية وعلوم الارض المختلفة ولكل ما له صلة بالبحث عن مصادر الثروة التي تحتضنها الكرة الارضية، حيث تغطي صور الاقمار الصناعية منطقة واسعة على سطح الارض لايتسنى لعين الانسان او أي تقنية اخرى الاحاطة بها [3] .

لقد استخدم مصطلح الاستشعار عن بعد لأول مرة في الستينات من القرن الماضي حيث ارتبط ارتباطاً عضوياً بدراسة سطح الارض ومحيطها الحيوي، وثرواتها الدفينة بواسطة الاقمار الصناعية ،التي اطلقها الانسان لتدور حول الكرة الأرضية ،على ارتفاعات مختلفة وتستشعر الارض طبعا للاجهزة التي تحملها ورغم تنوع وتعدد التعريفات التي صاغها الباحثون لتعريف هذا العلم الا انهم اتفقوا فيما بينهم على وصفه بأنه علم وفن وتقنية للحصول على معلومات عن جسم او ظاهرة من مسافات او ارتفاعات مختلفة باستخدام اجهزة تحسس واستشعار متنوعة ودقيقة تكون محمولة على الطائرات او الاقمار الصناعية وذلك من خلال دراسة وتحليل الاشعة او الطاقة الكهرومغناطيسية التي تنعكس او تبت من تلك الاهداف والتي تحمل خواص الهدف الذي تحت الدراسة [3].

## 3-الدراسات السابقة

في عام 2005 قام Xiaoling واخرون [15] باختيار حزمة باعتماد على خوارزمية تحليل المكونات الغير المعتمدة Independent Component Analysis حيث تم حساب Weight Matrix لكل الحزم ثم ايجاد معدل القيمة المطلقة لمصفوفة الوزن وترتيبها تنازليا واختيار الحزمة التي تحتوي على اعلى قيمة من قيم معدل القيمة المطلقة. لاجل حساب Weigh matrix لكل الحزم نحتاج الى وقت تنفيذ كبير .

عام 2008 قام Hongjun واخرون [5] باستخدام خوارزمية مؤشر الحزمة المثالية Optimal Band Index حيث في يتم حساب البعد الكسري كمقياس لازالة الضوضاء من الحزم وخزن الحزم التي تحوي على اهم

الخصائص المكانية بعد ذلك يتم حساب covariance matrix ومعامل الارتباط Correlation Factor بين الحزم والذي يستخدم لحساب افضل تسلسل لحزمة ثم اختيار الحزمة حسب تسلسلها. ولكن معامل الارتباط لا يعتمد عليه في صور التحسس النائي

في عام 2009 قام Sanaeinejad و اخرون [14] بدراسة اختيار افضل مجموعة من الحزم لدراسة ملوحة التربة باستخدام صور المتحسس ETM على المنطقة الدراسية نيشابور الواقعة في الجهة الشمالية الشرقية من ايران حيث يتم اجراء التصنيف عليها ثم ادماج الحزم Color Composite.

#### 4- تحليل المركبات الأساسية Principle Component Analysis

تحويل المركبات الرئيسية يسمى كذلك تحليل المركبات الرئيسية، تحويل متجهه أيكن (eigen)، تحويل هوتلينك أو مايسمى بتحويل كارهونين لوف في مجال الاستشعار عن بعد وتمييز الانماط. قدم هذا التحويل لأول مرة من قبل بيرسن (1901) وتمت عليه تحويلات عدة الى ان تم تعميمه من قبل لوي (1963) ولاتزال هذه الطريقة اليوم إحدى اهم طرق المتغيرات المتعددة (Multivariate techniques) [8].

يعتبر هذا التحويل طريقة خطية قياسية لاختزال الاتجاهية (dimensionality reduction) واستخلاص الصفات (feature extraction). تستخدم هذه الطريقة في عملية اختيار حزمة (Band) واحدة (افضل حزمة تحتوي على عدد كبير من الصفات) من بين مجموعة من الحزم bands، وكذلك في الاحصاء المتعدد المتغيرات (Multivariate statistics)، وفي التوجيه الآلي (Machine learning) والمعالجة الرقمية للصور والرؤيا الحاسوبية (ComputerVision) حيث يمكن من خلاله حساب المئات من الخصائص [9][11].

وبسبب الترابط المتواجد أصلاً بين الصور المتعدد الاطراف المأخوذة بأجهزة الاستشعار عن بعد فقد استخدم هذا التحويل على نطاق واسع في تطبيقات الاستشعار عن بعد وذلك باختزال الاتجاهية (dimensionality reduction) للصور متعددة الاطراف الناتجة عن متحسسات الاقمار الصناعية [11].

ان مجموعة من الصور بحزم مختلفة يمكن اعادة ترتيبها بواسطة تحويل المركبات الاساسية (PCA) كمجموعة جديدة من الصور تدعى بالمركبات الغير المترابطة احداها مع الاخرى وهي مرتبة نسبة الى مقدار التباين (Variance) [5]. أن الحزم الجديدة لمركبات (PCA) غير مترابطة وغير معتمدة على بعضها وهي عادة قابلة للتفسير اكثر من الصور الاصلية .

تحتوي حزمة PCA الاولى على اكبر نسبة تباين (Variance) بين المعلومات في حين أن حزمة PCA الثانية تحتوي على ثاني اكبر نسبة للتباين Variance بين المعلومات وهكذا. اما الحزمة الاخيرة فتحتوي على ضوضاء (noise) لانها تحوي على اقل نسبة تباين (Variance).

ان الفائدة المهمة لتحويل المركبات الاساسية هي ان معظم المعلومات ضمن الحزم المتمثلة بالتباين يمكن حفظها أو كبسها الى عدد أصغر من الحزم بدون خسارة تذكر للمعلومات [10].

يتطلب تطبيق تحويل المركبات الاساسية حساب مصفوفة (Variance/ Covariance matrix) للخصائص وبعدها يتم حساب المركبات الاساسية والتي يمكن حسابها بطريقتين اساسيتين هما طريقة الاس (Power Method) وطريقة جاكوبي (Jacobi Methods). تعتمد طريقة الاس على حساب قيم أيكن (Eigen values) الواحدة تلو الاخرى ابتداء من القيمة الاعلى المرتبطة بالمركبات الرئيسية (Principal

components) التي تحتوي على معظم المعلومات ثم يتم حساب القيمة الاصغر فالاصغر وهكذا. لهذا توفر طريقة الاس الفرصة لحساب قيم الايكن والمركبات الرئيسية المرتبطة مع بعضها الى مستوى مطلوب من قبل الباحث عن محتوى المعلومات في حين أن طريقة جاكوبي تحسب كل قيم الايكن دفعة واحدة ولهذا قد تتطلب زمن تنفيذ اكبر على الحاسبة ، ولكن عند التنفيذ المتزامن لكليهما (طريقة الاس وطريقة جاكوبي) فان طريقة جاكوبي تعمل بمقياس أفضل من طريقة الاس عندما يكون عدد المركبات الرئيسية كبيراً وهكذا فان الحالات التي تتطلب كمية اصغر نسبياً من المعلومات فان طريقة الاس قد تعمل بشكل اسرع في حين عندما يكون المطلوب مستوى اعلى من المعلومات يتم استخدام طريقة جاكوبي [7] .

ان الهدف من تحويل المركبات الاساسية هو تحويل كمية البيانات العالية والمبعثرة الخصائص الى بيانات تحتوي على الخصائص الاساسية وبصورة مرتبة. وكذلك يستخدم لاستخلاص مركبات الخصائص ( feature components) القريبة من بعضها [8] .

#### 4-1 الوصف الرياضي لتحليل المركبات الاساسية

يتطلب تحليل المركبات الاساسية رياضياً ايجاد مصفوفة covariance للصور المتوفرة .يمكن حساب مصفوفة covariance وفقاً للمعادلة التالية :-

$$Cov_{ij} = \frac{1}{MN} \sum_{k=1}^M \sum_{l=1}^N (X_i(k,l) - M_i) (X_j(k,l) - M_j) \quad \dots(1)$$

حيث أن  $M, N$  عدد الوحدات الصورية في الصورة و  $i, j$  تمثل عدد الحزم و  $k, l$  تمثل موقع النقطة في الصورة و  $M_i$  و  $M_j$  هي المتوسط الحسابي للحزم  $i$  و  $j$  على الترتيب ويمكن حسابها وفقاً للمعادلة التالية

$$M_i = \frac{1}{MN} \sum_{k=1}^M \sum_{l=1}^N X_i(k,l) \quad \dots(2)$$

نلاحظ من معادلة رقم (1) بان قيم المصفوفة تكون عبارة عن تباين variance لعناصر القطر الرئيسي و covariance لبقية العناصر وكذلك فان مصفوفة covariance مصفوفة متناظرة أي ان العناصر فوق القطر الرئيسي مساوية بالقيم للعناصر تحت القطر المتناظرة لها.

وهكذا فان مصفوفة Covariance  $n \times n$  عدد من الصور هي مصفوفة  $n \times n$  ويمكن ترتيبها كما يلي :

$$\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix} \quad \dots(3)$$

وحيث أن  $c_{ij} = c_{ji}$  لكل  $j = i$  فان المصفوفة  $C$  هي متناظرة وتكون ايضاً Semi\_definite لان المحدد لها لا يساوي صفر .

مصفوفة متجة الايكن  $T$  تشتق من مصفوفة Covariance باستخدام طريقة جاكوبي الموضحة في الفقرة اللاحقة [2] والتي نحصل منها على  $D$  ، حيث أن  $D$  هو covariance القطري للمركبات الاساسية وتمثل بالمعادلة التالية:

$$\dots D = \begin{bmatrix} \lambda_{11} & 0 & \dots & 0 \\ 0 & \lambda_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \lambda_{nn} \end{bmatrix} \dots(4)$$

أن العناصر القطرية  $D$  تسمى قيمة ايكن Eigen value لمصفوفة Variance/Covariance حيث  $\lambda_{ii}$  لكل  $i = 1, 2, \dots, n$  هي التباينات لمحاور المركبات الرئيسية ان العناصر خارج القطر (Off\_diagonal) للمصفوفة  $D$  هي صفر مما يدل على أن المركبات الخارجة غير مترابطة أي أنها مستقلة [6][13].  
تتم عملية تحويل الصورة باستخدام المركبات الاساسية وفقا للمعادلة التالية:

$$\bar{Y} = T\bar{X} \dots(5)$$

حيث أن

$\bar{X}$  متجهه يحوي قيم مدخلات الصور

$\bar{Y}$  متجهه يحوي قيم المركبات الرئيسية

$T$  مصفوفة تمثل الصفوف لها قيم متجهه ايكن المعياري .

#### 4-1-1 نظرية جاكوبي Jacobi's Method

1. يتم ايجاد اكبر عنصر في مصفوفة معينة مربعة ولتكن  $C$  بحيث لا يكون من عناصر القطر الرئيسي أي  $Max\_element = a_{ik}$  و  $i, k$  هي الصف والعمود في مصفوفة Covariance و  $i \neq k$  .
2. يتم ايجاد الزاوية  $\theta$  وذلك عن طريق مايلي :

$$\theta = \frac{1}{2} \arctan(2a_{ik} / (a_{ii} - a_{kk})), \text{ if } a_{kk} \neq a_{ii}$$

$$\theta = \begin{cases} \pi / 4 & \text{when } a_{ik} > 0 \\ -\pi / 4 & \text{when } a_{ik} < 0 \end{cases} \text{ if } a_{kk} = a_{ii}$$

3. اجراء عملية التدوير (rotation) على المصفوفة  $C$  وارجاع الناتج في مصفوفة وكما يلي

$$d_{ii} = a_{ii} \cos^2 \theta + 2a_{ik} \sin \theta \cos \theta + a_{kk} \sin^2 \theta$$

$$d_{kk} = a_{ii} \sin^2 \theta + 2a_{ik} \sin \theta \cos \theta + a_{kk} \cos^2 \theta$$

$$d_{ik} = d_{ki} = -(a_{ii} - a_{kk}) \sin \theta \cos \theta + a_{ik} (\cos^2 \theta + \sin^2 \theta)$$

او عن طريق المعادلات التالية

$$d_{ii} = \frac{1}{2}(a_{ii} + a_{kk} + \sigma R)$$

$$d_{kk} = \frac{1}{2}(a_{ii} + a_{kk} - \sigma R)$$

$$d_{ik} = d_{ki} = 0$$

حيث يتم ايجاد قيمة  $R$  وفقا للمعادلة التالية  $R = \sqrt{(a_{ii} - a_{kk})^2 + 4a_{ik}^2}$

ويتم ايجاد قيمة  $\sigma$  وفقا للمعادلة التالية

$$\sigma = \begin{cases} 1 & \text{if } a_{ii} \geq a_{kk} \\ -1 & \text{otherwise} \end{cases}$$

اما بقية عناصر مصفوفة التدوير فنجدها باستخدام المعادلتين التاليتين:

$$d_{ir} = a_{ir} \cos \theta + a_{kr} \sin \theta$$

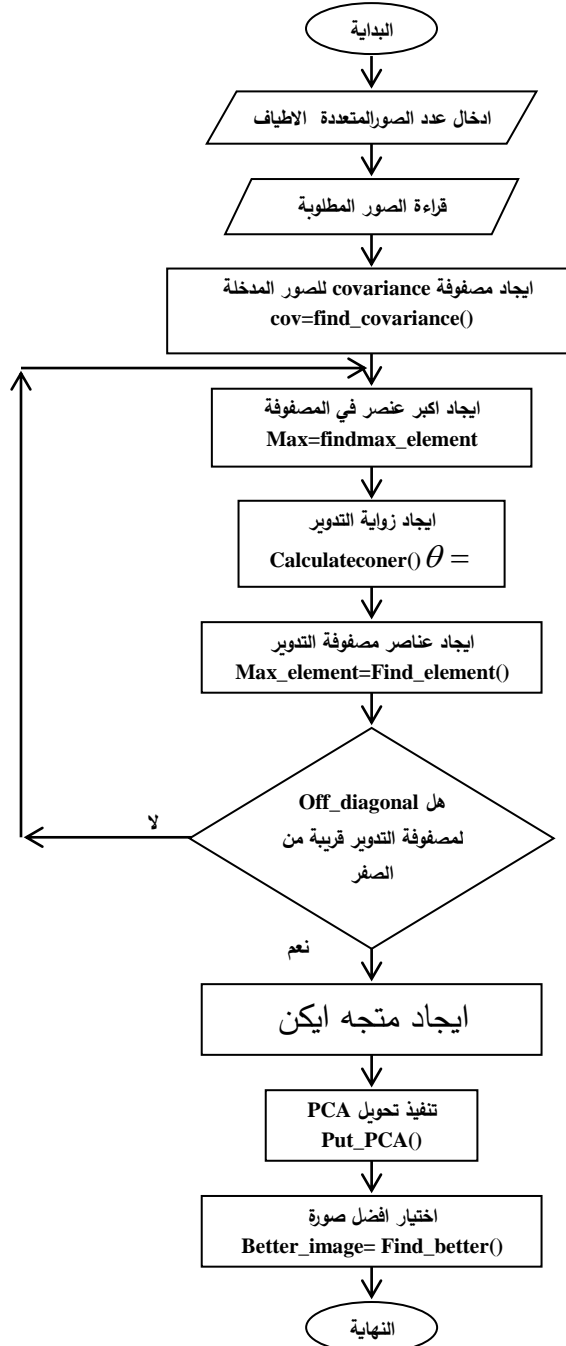
$$d_{kr} = -a_{ir} \sin \theta + a_{kr} \cos \theta$$

حيث  $r \neq i$  و  $r \neq k$  و  $i, k$  تمثل قيمة الصف والعمود داخل المصفوفة .

4. اعادة الخطوات من 1 الى 3 على المصفوفة الناتجة الى ان يتم الحصول على قيم عناصر القطر غير الرئيسي

(off\_diagonal) قريبة من الصفر [4] .

لشكل (1) يوضح المخطط الانسيابي لخوارزمية تحليل المركبات الاساسية (PCA).



## الشكل (1) المخطط الانسيابي لخوارزمية PCA

## 5- التطبيق العملي للخوارزمية

عند تطبيق خوارزمية تحليل المركبات الاساسية على صور متعددة الاطياف لمتحسس الخرائط الموضوعي (TM) حيث كانت مصفوفة Variance /Covariance للصور هي كما يلي:

Variance /Covariance Matrix of TM

2847.8108	2964.0857	3055.1168	2974.3864	3076.7410	3180.0915
2964.0857	3176.5428	3301.7963	3230.6633	3399.4561	3509.3644
3055.1168	3301.7963	3479.5078	3397.4943	3617.4120	3741.7662
2974.3864	3230.6633	3397.4943	3420.9303	3583.5001	3674.6240
3076.7410	3399.4560	3617.4120	3583.5001	4434.7075	4514.1952
3180.0915	3509.3644	3741.7662	3674.6240	4514.1952	4650.3946

وكانت قيم متجة ايكن للصور الستة هي كما يلي:

Eigen Vector Matrix of TM

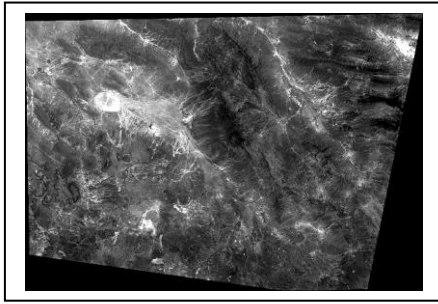
20896.8049	0.0000	-0.0000	0.0000	0.0	0.0000
0.0000	908.1359	-0.0000	0.0	0.0000	-0.0000
-0.0000	-0.0000	127.1215	0.0000	-0.0000	0.0
0.0000	0.0	0.0000	48.0547	-0.0000	-7.5530
0.0	0.0000	-0.0000	-0.0000	19.1278	0.00000
0.0000	-0.0000	0.0	-7.5530	0.00000	10.6486

والجدول التالي يوضح قيمة SNR وقيمة Eigen value لكل حزمة من الحزم

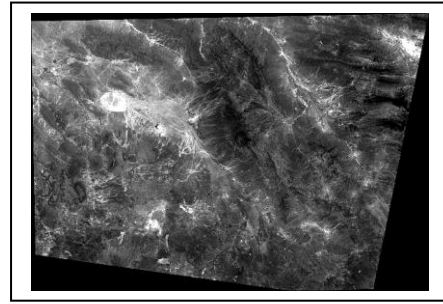
SNR	Eigen value	Band
73.3707	127.1215	TM1
73.6854	10.64886	TM2
73.6788	908.1358	TM3
73.745	48.0547	TM4
74.714	19.1278	TM5
74.7217	20896.8049	TM6

والصورة (1) توضح الصور متعددة الاطياف الاصلية والصور بعد اجراء تحليل المركبات الاساسية عليها

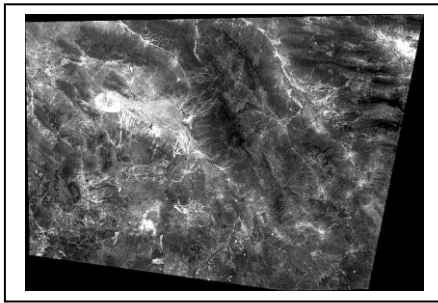




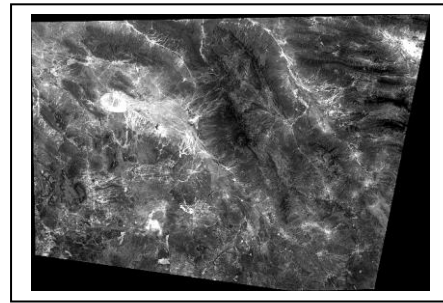
TM1



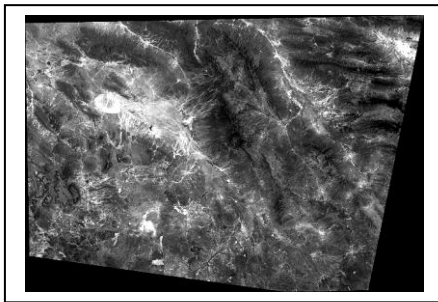
PCA1



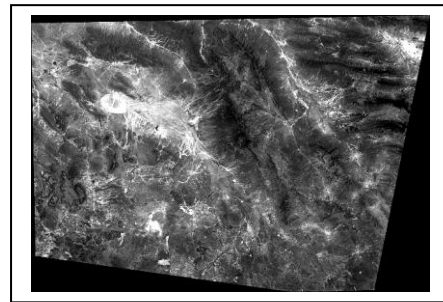
TM2



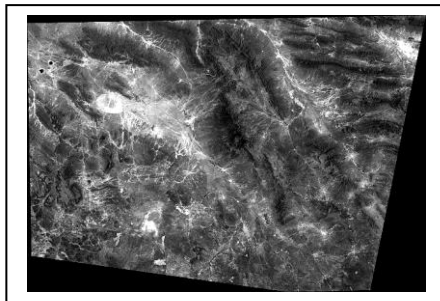
PCA2



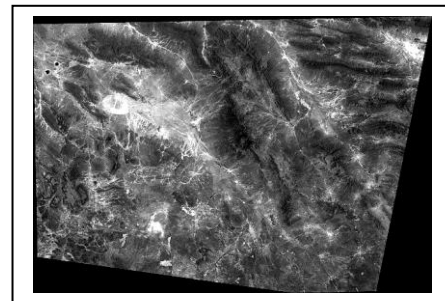
TM3



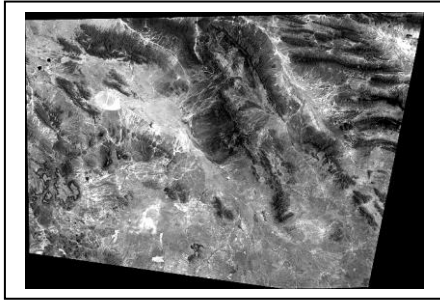
PCA3



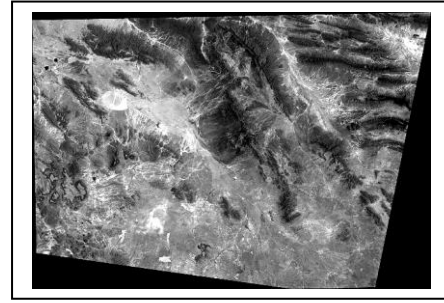
TM4



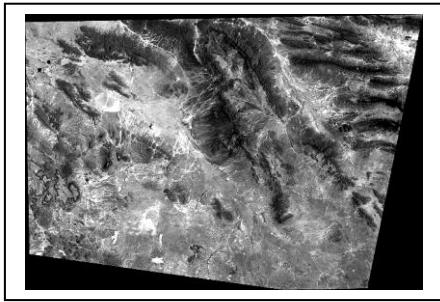
PCA4



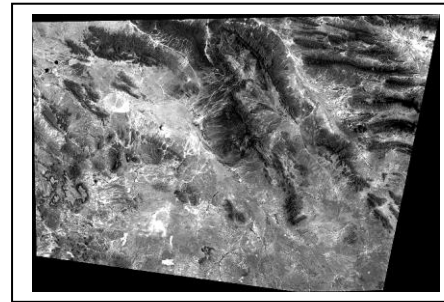
TM5



PCA5



TM6



PCA6

الصورة (1): صورة متعددة الأطياف لمتحسس TM، الجهة اليسرى تمثل الصورة الاصلية والجهة اليمنى تمثل الصورة بعد اجراء تحليل المركبات الاساسية عليها.

#### 6- النتائج والمناقشة

تم في هذا البحث تطبيق خوارزمية تحليل المركبات الاساسية على صور الاستشعار عن بعد .تبين من تطبيق تحليل المركبات الاساسية والذي يعتمد على طريقة جاكوبي في ايجاد متجه الايكن Eigen vector ، ان تطبيق PCA اظهرت كفاءه عالية ودقة في تحديد الحزمة الافضل من بين الحزم الست المدخلة واعتماداً على قيم الايكن ونسبة الاشارة الى الضوضاء والتي بلغت لافضل حزمة 20896.8049 و 74.7217 على التوالي تبين ان الحزمة السادسة هي افضل حزمة . وهذا يدل على ان الحزمة السادسة تحتوي على اهم الخواص المستقلة في الصور وانها تفيد في عملية تحليل صور الاستشعار عن بعد كعزل مناطق المياه وانواع التربة وغيرها.

#### المصادر

- [1]. Dinesh Maidasani,2009,"C# .Net 2005", Firewall Media.
- [2]. Dreamtech Press,19-A, Ansari Road ,Daryaganj,New Delhi ,2009,"C#2008 Programming Covers.NET 3.5 ",Print Well Offset ,New Delhi.
- [3]. El Hassan Ismat M. , 2007 , "*Digital Image Processing in Remote Sensing*" ,College of Engineering , King Saud University.
- [4]. Froberg C.,1969,"*Introduction to Numerical Analysis*" 2<sup>nd</sup> edition, Addison – Wesley.
- [5]. Hongjun , Sheng and Peijun ,2008,"*A New Band Selection Algorithm For Hyperspectral Data Based on Fractal Dimension*", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. pp. 279- 284
- [6]. H. Eesa, 1998, " *A PC\_Based Multispectral Image Processing Software for Remote Sensing Applications* " , M. SC. Thesis, university of Mosul, Collage of Science, Computer Sc. Dept.
- [7]. IDRISI Guide to GIS and Image Processing Volume 1, "*Introduction to Remote Sensing and Image Processing* ", Chapter 3, pp. 17-34
- [8]. Jacqueline Le Moigne, 2003, " *Multi-Sensor Image Registration, Fusion and Dimension Reduction* " .
- [9]. M. Pal, 2002, " *Factors Influencing The Accuracy Of Remote Sensing Classifications* " , Ph. D. Thesis, University of Nottingham.
- [10]. P. Bajcsy , P. Groves,2004, "*Methodology For Hyperspectral Band Selection* " , Published in Photogrammetric Engineering and Remote Sensing journal, Vol. 70, Number 7, pp. 793-802.
- [11]. R. Amer, T. Kusky, P. Reinert and A. Ghulam, 2009," *Image Processing And Analysis Using Landsat ETM+ Imagery For Lithological Mapping At Fawakhr, Central Eastern Desert Of Egypt* " , ASPRS Annual Conference Baltimore, Maryland.
- [12]. Rajeev A .Parida ,2009,"The Power of C#" ,Firewall Media.
- [13]. T. Shumam,2001,"*An Integrated Software for Remotely Sensed Data Classification* " , M. Sc. Thesis, University of Mosul, College of Science , Computer Sc. Dept.
- [14]. Sanaeinejad, S. H.; A. Astaraei,. P. Mirhoseini.Mousavi and M. Ghaemi, 2009 "*Selection of Best Band Combination for Soil Salinity Studies using ETM+ Satellite Images (A Case study: Nyshaboor Region,Iran)*", World Academy of Science, Engineering and Technology , pp. 519 – 521.
- [15]. Xiaoling Wang, Hongtao Du, Hairong Qi, Rajeev Ramanath, Wesley E. Snyder, 2005, "*Band Selection Using Independent Component Analysis for Hyperspectral Image Processing*" , [rajeev.ramanath@ieee.org](mailto:rajeev.ramanath@ieee.org) , [wes@eos.ncsu.edu](mailto:wes@eos.ncsu.edu).