Principal Component Analysis of Mul ti-Temporal Image Pairs

Laith A. Al-Ani, Ayad A. Al-Ani and Alyaa H. Ali
Department of physics, College of Science, University of Al-Nahrain
Received: 28/5/2001 Accepted: 22/9/2001

Abstract

The PCA is statistical technique that transforms a multivariate data set consisting of inter-correlated variables into a data set consisting of variables that are uncorrelated linear combination. In our project principal component analysis “PCA” was applied for two set of original bands in two dates (bands 1, 5, and 7 in 1988 and bands 1, 5, and 7 in 1990).

In this method the PCA of six channel data sets consisting of multi-temporal LANDSAT TM image pairs often generates higher order principal components that are related to the changes in brightness.

Although the image produced by the first component summarizes the information’s that are common to all channels, we can see that the first principal component is dominated by the contribution of the infrared band (band 7) in1988. Our result also, show that over 73.5% and 83.7% of the variability lies in the direction defined by the first and second principal component images respectively.

الخلاصة

تحليل المركبات الاساسية تقنية احصائية تقوم بتحويل مجموعة المتغيرات المترابطة الى مجموعة خطية غير مترابطة. في بحثنا هذا تم تطبيق تقنية المركبات الاساسية باستخدام مجموعتين من الحزم وبتوقيتين مختلفين (الحزم 1،5،7 في عام 1988 والحزم 1،5،7 في عام 1990).

ان القنوات الست الناتجة من تحليل المركبات الاساسية للحزم الاصلية نوع Land Sat- TM تصف مقدار التغيرات الحاصلة في الحزم الاصلية عن طريق التباين في التدرجات اللونية وفي قنوات المركبات الاساسية ذات الرتب العليا.

بالرغم من ان صورة المركبة الاساسية الاولى تمثل المعدل لجميع الحزم الاصلية, كذلك يمكن ملاحظة ان المركبة الاساسية الاولى تمثل اعلى مساهمة للحزمة التحت الحمراء (حزمة 7). النتائج تشير ايضاً الى أن 73.5% ، 83.7% من المتغيرات تعطى بالمركبة الاساسية الاولى والثانية على التوالي.

Introduction

The term remote sensing is a method to use the electromagnetic energy to detect and measure the characteristics of the object from the low frequencies to high frequencies bands through the microwave, far infrared, near infrared, visible and ultraviolet bands [1].

Remotely sensed image can be used to monitor changes in land surface condition. There are a large number of methods used for change detection, which have been proposed and applied.

A principal component analysis (PCA) is an important data transformation technique used in remote sensing work with multi-spectral data.

Donker and Mulder 1n 1976 derived the principal components from a variance- co-variance matrix and used it change detection [2]. Lodwick in 1979 found that changes could be measured using principal component and he found that the first two PC”s contained the differencing in the image [3]. Byrne et.al in 1980 superimposed two Land-sat images of the same area and they found that PCA of this data contained useful information about temporal change [4]. Richards in 1984 used the application of PCA in remote sensing. He used eight bands multi-temporal data, he found that the first of the eight components was an overall brightness factor while temporal change was associated with variation in the values of principal component 2, 3 and 4 [5]. Fung and Ledrew in 1987 use the standardized PCA to study changes because the PC's found to be better aligned along the object of interest [6]. Collins and Wood Cock in 1994, apply several methods of transformation to detect that changes, one of them is K-L transformation, They made a comparison between the K-L and the other transformation, they found that the K-L is the best method for studying the changes [7]. Muchoney and Haack in 1994 examined several approaches to detect defoliation including the principal component [8].

Mathematical Representation

Let us consider two-dimensional image represented by an array

$$X_{NxN} = \begin{bmatrix} x_{11} & x_{12} & . & x_{1N} \\ x_{21} & x_{22} & . & x_{2N} \\ . & . & . & . \\ x_{N1} & x_{N2} & x_{N3} & x_{NN} \end{bmatrix} \text{ ...... 1}$$

The mean of the columns is defined by

$$M_x = E\{X\} \text{ ...... 2}$$

Where, E is the expectation value. So the covariance matrix C_x could be defined as follows

$$C_x = E((X_{NN-}M_x)(X_{NN-}M_x)^T) \text{ ...... 3}$$

Where, T indicates to the transpose of the matrix. The covariance matrix takes the form of the correlation matrix whose elements represented the covariance between the images. The diagonal elements are the variance of each image (i.e. $C_{ii} = \sigma_i^{\ 2}$). Where, σ_I is the standard deviation corresponds to the contrast of the i^{th} image. In order to compute the principal component the eigen-vector should be extracted from the covariance matrix. The covariance matrix should be diagonalized to compute the eigen value. The diagonal elements of the diagonalized covariance matrix correspond to the variance of a new PC image, the variance being related to the amount of contrast [9].

The correlation coefficient R_{np} of each n^{th} can be computed by the following relationship [10].

$$R_{np} = \frac{e_{np}(\lambda_p)^{1/2}}{\delta_n^2} \text{ ...... 4}$$

Where, e_{np} is the eigen vector of band (n) and associated with component p, λ_p is the p^{th} eigen value component and σ_n^2 is the variance of band (n) in the covariance matrix.

This transformation minimized the mean square approximation errors. The process has been viewed as information compression into smaller number of components from the large number of features by discarding redundant information into higher order component [9].

The Study Area

The study area"SAMARA" lies in the northwest of Baghdad. This region generally consist of residential areas, agricultural areas depends on the rain and artificial irrigation. Also, the area has a rocky region and channels used for irrigation. One of these channels is the old "NEHRAWAN" channel. The agricultural areas that depend on the rain lies below the level of the Tigris directly, this region is called the flood plane.

Moreover, the study area has a main way, which clearly appears in the six bands [11] where we have in our study. Generally, our study area is an agriculture area which Tigris passes through it and AL- Othaim River which flow into the Tigris passes through this region also. For more details see [11].

In 1988 AL-Othaim River has been flooded on the agriculture area which lies near the river and this is clearly seen in the three bands of 1988. In 1990 the area become completely dry, and changes has been happened.

Methodology

Traditional Use of PCA

The PCA has been traditionally applied to a multi-spectral data set with TM images of this study. The variance matrix obtained from the entire image was used to determine eigen structure (principal component axes) of the feature space after applying PCA to the two data TM images set. The first set represents the images in the year 1988 and the other represents the images in the year 1990, each set consists of three bands. Sets of six principal component images are shown in the figure (1). Fig (2) shows the histogram of the PC image (probability density function (PDF)). The result shows that

PC1, PC2, and PC3 contain approximately all the information contained in the original images, since the commutative variance is 92.9%.

Table (1) show some of the multivariate statistics for the original bands in Samara scene, band (1, 5, and 7 for 1988 and band 1, 5, and 7 for 1990). Band 5 in 1988 has the largest mean and the smallest standard deviation. This means that the information in this band is small while band 7 in 1988 has the largest standard deviation and the smallest mean value. This indicates that the variance is large and the information in this band is very high. Table (2) shows the covariance-variance matrix. The result of this table show that the largest value is in band 7 in 1988, that smallest variance is in band 5 in 1988. This is obvious from table (1) for the standard deviation.

Table (3) shows the correlation between bands. The largest correlation is between band 1 in 1990 and band 5 in 1990, this means that there is common information in these bands i.e. these correlation values suggest that there are a substantial amount of redundancy in the information contents among these bands. The information that is common to both bands is mapped to the first and second component. The lowest correlation is between band 5 in 1988 and 7 in 1988. This means that the relation between them is very small i.e. there is some information in band 7 that not founded in band 5.

Table (4) shows the eigen value for each band and its commutative variance. Band 1 in 1988 has the largest eigen value and about 73.5% of variability in the data defined by the first principal component

Table (5) shows the eigen vector, although the image produced by the first component summarizes the information that are common to all channels, we can see that the first principal component is dominated by the contribution of the infrared band (band 7) in1988.

The first three principal component contains most of the information (changes and unchanged) in the original bands at different years. Since the flooded happen about the AL-Athaim River in 1988 has been flooded around the near by area, which is an agricultural area, this can be seen clearly in the three bands of 1988. In 1990 the area becomes completely dry and this obvious clear from the three bands for 1990, so that the principal component for these six bands could show these changes.

From figure (2) one can see that the histogram of PC1 have wide range of brightness value, this means that PC1 image contain most the information in the region.

Table (6) shows the correlation between each band and component. The first principal component has highest degree of correlation with the six bands. The first PC represents the average of all the six PC, so it has higher correlation with the all the six bands. The second PC has less degree of correlation than the first PC and the last PC has the lowest correlation with the six bands.

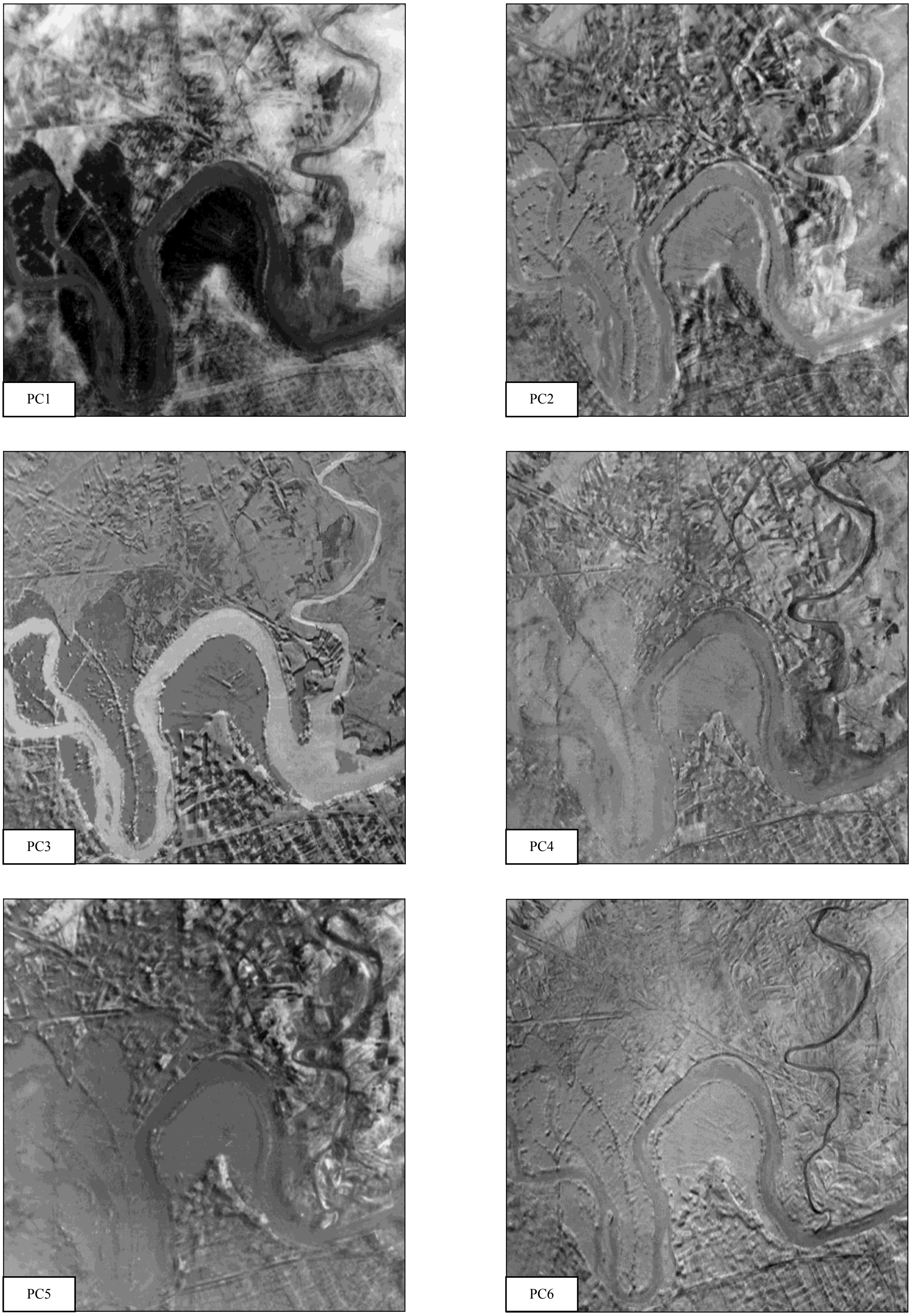

Figure (1): Principal Component Images for Samara Scene.

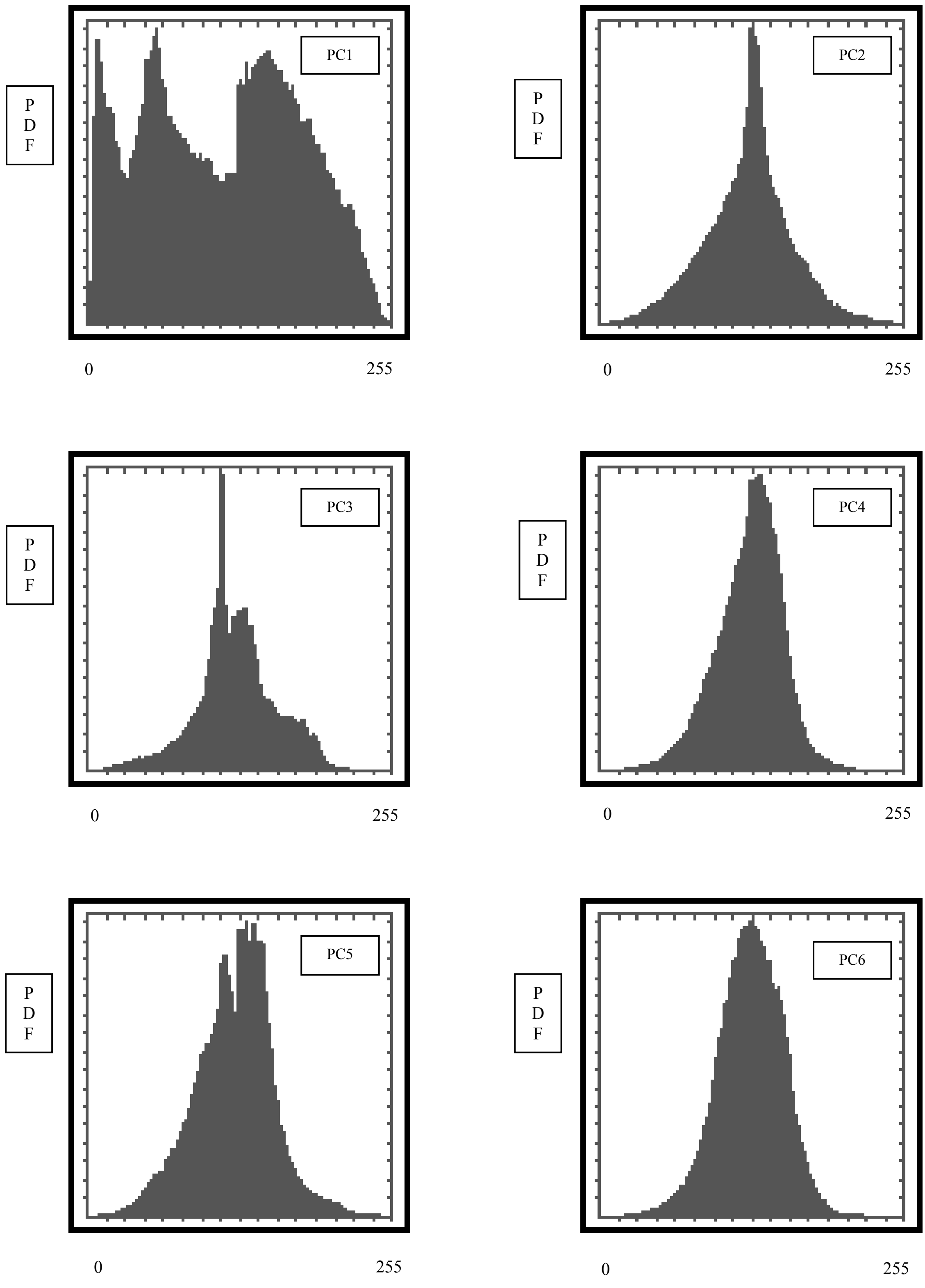

Figure (2): The Histogram of PC Images

Table (1): Multivariate Statistics for the Original Bands in Samara Scene.

Year	Band	Mean	Min	Max	Std
1988	1	87.3	0	192	60.3
1988	5	107.3	76	218	39.3
1988	7	88.5	0	213	68.2
1990	1	90.0	26	255	39.8
1990	5	101.9	0	213	55.9
1990	7	97	53	213	23.4

Table (2): Covariance-Variance Matrix of the spectral Bands.

Bands	1	5	7	1	5	7
1	3636					
5	254	87				
7	2809	269	4657			
1	1504	201	1910	1588		
5	2149	299	2438	1603	3126	
7	531	76	673	371	664	546

Table (3): Correlation between Bands.

Bands	1	5	7	1	5	7
1	1					
5	0.45	1				
7	0.68	0.42	1			
1	0.63	0.54	0.7	1		
5	0.64	0.57	0.64	0.72	1	
7	0.38	0.35	0.42	0.40	0.50	1

Table (4) Eigenvalues of the Covariance matrix

Principal axis	Eigen value	Commutative variance %
1	10030	73.52
2	1384	83.66
3	1270	92.97
4	526	96.84
5	375	99.59
6	54	100

Table (5): Eigen vectors Computed for the covariance matrix

0.5196	0.1211	-0.8454	-0.0155	0.011	-0.0146
0.0517	0.0584	0.0226	0.0355	0.0271	0.9957
0.6195	-0.7043	0.2819	-0.182	-0.0861	0.0115
0.3285	0.1165	0.2059	0.8919	0.1906	-0.6560
0.4705	0.679	0.3893	0.2976	-0.2728	-0.0550
0.1193	0.106	0.1064	0.2852	0.9386	-0.0302

Table (6): Degree of Correlation between Each Band and Component.

N/p	1	2	3	4	5	6
1	0.863	0.750	-0.500	-0.006	0.004	-0.002
5	0.555	0.233	0.086	0.087	0.056	0.787
7	0.909	-0.384	0.147	-0.061	-0.024	0.001
1	0.826	0.109	0.184	0.514	0.093	-0.012
5	0.843	0.452	0.248	-0.122	-0.095	-0.007
7	0.511	0.169	0.162	0.28	0.778	-0.001

References

1. Elachi, C., 1987, Introduction to the Physical and Techniques of Remote Sensing., John Wiley and Sons.
2. Lodwick, G D., 1979, Measuring Ecological Changes in Multi Temporal Land Sat Data Using Principal Components. *Proceedings of the International Symposium on Remote Sensing of the Environment, Ann Arbor, Michigan,* pp. 1131-1141.
3. Donker, N. H. W , and Meuerink, A. M. J., 1977, Digital Processing of Land Sat Imagery to Produce a Maximum Impression of Terrain Ruggedness. *ITC Journal* 4, 683-704
4. Byrne, G, F., Crapper, P. F., and Mayo, K. K., 1980, Monitoring Land Cover Change by Principal Component of Multi Temporal Land Sat Data. *Remote Sensing Environ.,* 10, 175-184
5. Richards, J. A., 1984, Thematic Mapping from Multi Temporal Image Data Using The Principal Component Transformation. *Remote Sensing of Environment.,* 16, 35-46.
6. Fung, T., and Ledrew, E., 1987, Application of Principal Components Analysis to Change Detection. *Photogramm. Eng. Remote Sens.,* 53, 1649-1658.
7. Collins, J. B. and Woodcock, C. E., 1994, Change Detection Using the Gramm-Schmidt Transformation Applied to Mapping Forest Mortality. *Remote Sensing. Environ*., 50, 267-279.
8. Muchoney, D. M., and Haack, B. N., 1994, Change Detection for Monitoring Forest Defoliation. *Photogramm. Eng. Remote Sens.,* 1243-1251
9. Gonzaliez, R., and Wintz, P., 1987, Digital Image Processing. *Addison-Wesley Publishing Company, England*
10. Jensen, J. R., 1986, Introductory Digital Image *Processing. Perspective-Hall., Englewood Cliffs, New Jersey.*
11. Alyaa, H. A., 1998, Studying the Flooded Area by Principal Component Analysis of Multi-Temporal Land-Sat Thematic Mapper Data. *M. Sc Thesis, Saddam University, Baghdad- Iraq.*