

Chapter 9

Principal Components Analysis

Correlations among Bands

Principal Components Analysis

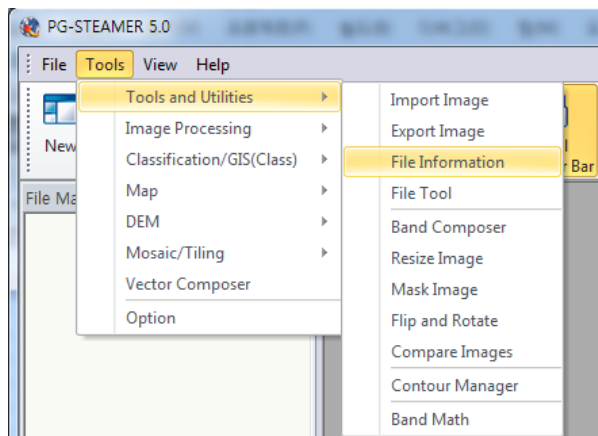
다중 밴드를 갖는 영상은 중복된 정보를 포함하고 있기 때문에 영상 파일의 크기가 커질 수 있습니다. 이번 장에서는 주성분 분석(Principal Components Analysis)을 이용하여 정보의 손실을 최소화하면서 영상 파일의 크기를 줄이는 방법에 대해 배웁니다.


Correlations among Bands


이번 실습에서는 밴드 간의 상관 관계를 다루는 방법에 대해 배웁니다.

[실습 9a]

1. PG-STEAMER를 실행합니다.
2. Menu → Tools and Utilities → File Information 순으로 실행합니다.




3. Open  툴 버튼을 클릭하고 Chapter09 경로에 있는 **landsat.xdm** 파일을 선택하여 엽니다.

4. Statistics  툴 버튼을 클릭합니다.

5. Real Time Processing  툴 버튼을 클릭합니다.

6. 6개의 Landsat 밴드들 간의 상관 관계를 확인합니다.


Correlation Matrix						
Band	1	2	3	4	5	6
1	1.0000	0.8667	0.8065	-0.2146	0.3533	0.6964
2	0.8667	1.0000	0.9104	0.1448	0.6203	0.8215
3	0.8065	0.9104	1.0000	-0.0053	0.4380	0.7035
4	-0.2146	0.1448	-0.0053	1.0000	0.6618	0.2533
5	0.3533	0.6203	0.4380	0.6618	1.0000	0.8507
6	0.6964	0.8215	0.7035	0.2533	0.8507	1.0000

7. Add  툴 버튼을 클릭하고 **SPOT.xdm** 파일을 추가합니다.

8. Real Time Processing  툴 버튼을 클릭합니다.

9. 밴드들 간의 상관 관계를 확인합니다.

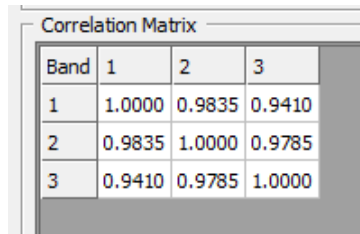
Correlation Matrix				
Band	1	2	3	4
1	1.0000	-0.2651	-0.2337	0.3709
2	-0.2651	1.0000	0.9785	0.5463
3	-0.2337	0.9785	1.0000	0.5330
4	0.3709	0.5463	0.5330	1.0000

10. Add  툴 버튼을 클릭하고 **IKONOS.xdm** 파일을 추가합니다.

11. Real Time Processing  툴 버튼을 클릭합니다.

Chapter 9: Principal Components Analysis

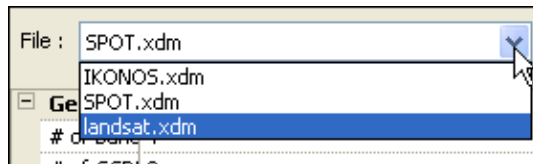
12. 밴드들 간의 상관 관계를 확인합니다.



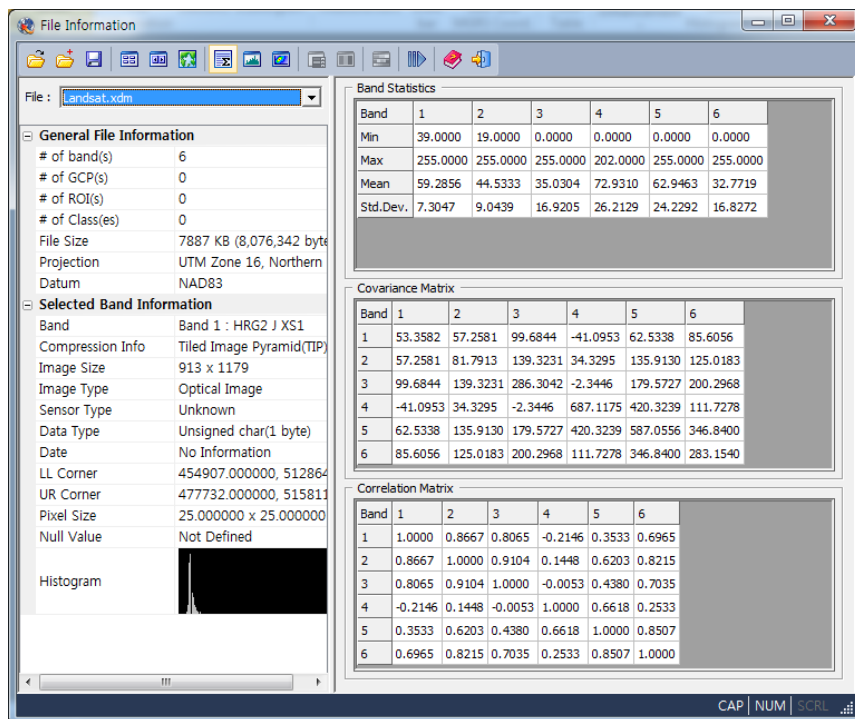
Band	1	2	3
1	1.0000	0.9835	0.9410
2	0.9835	1.0000	0.9785
3	0.9410	0.9785	1.0000

Add 톨을 사용할 때 파일 정보는 추후 검색을 위해 저장합니다.

13. File 항목의 리스트를 dropdown 시킵니다.



14. Landsat.xdm를 선택합니다.



Landsat.xdm 밴드들 사이의 상관 관계가 나타납니다.

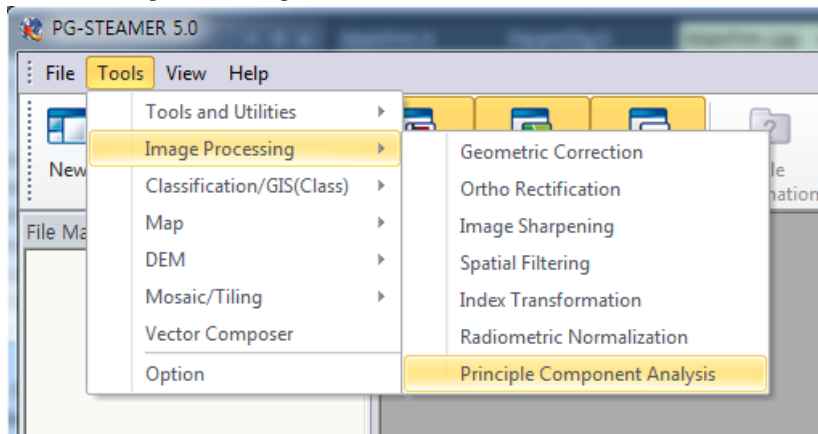
15. XDM File Information 창을 닫습니다.



Principal Components Analysis

이번 장에서는 principal components를 분석하는 방법에 대해 배웁니다.

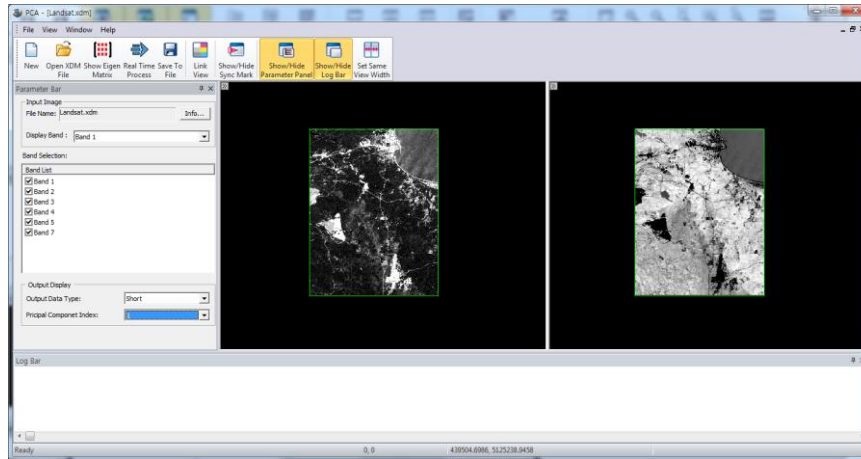
[실습 9b]

1. PG-STEAMER를 실행합니다.
2. Menu → Image Processing → PCA의 순으로 실행합니다.

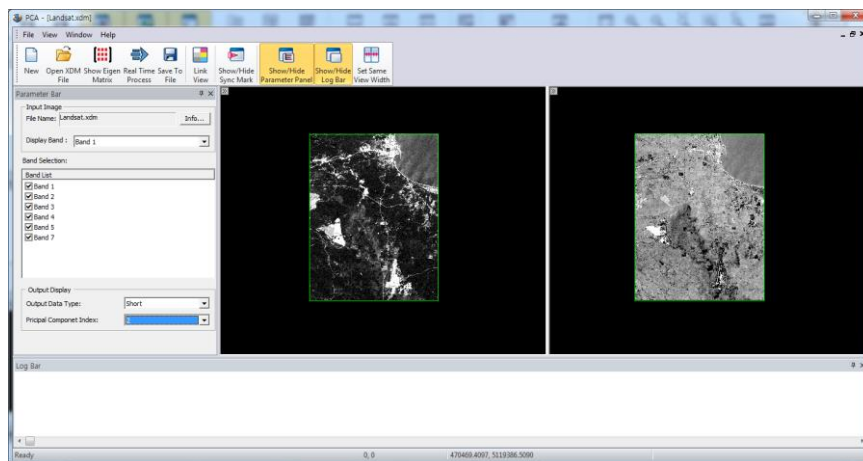


3. Open  툴 버튼을 클릭하고 Chapter09 경로에 있는 **landsat.xdm** 파일을 선택하여 엽니다.
4. RTP  툴 버튼을 클릭합니다.
5. Output Data Type을 Short로 변경하고 Principal Components Index 항목을 1로 지정하고 RTP 툴을 클릭합니다.

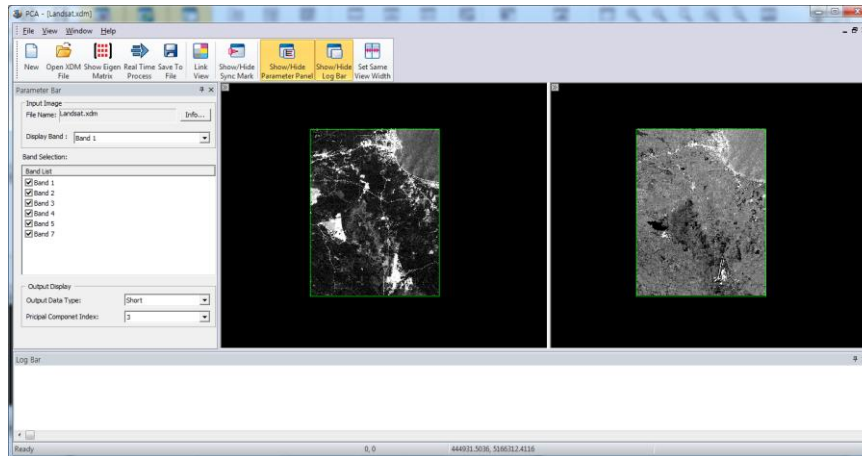
Chapter 9: Principal Components Analysis




6. Principal Components Index 항목을 2로 지정하고 RTP 톨을 클릭합니다.



7. Principal Components Index 항목을 3으로 지정하고 RTP 톨을 클릭합니다.



Eigen Matrix를 확인합니다

8. Show Eigen Matrix  툴 버튼을 클릭합니다.
9. 모든 데이터가 보이도록 PCA Matrix 대화상자의 크기를 조정합니다.

PCA Matrix of []

[Eigen Value]		PC1	PC2	PC3	PC4	PC5	PC6
Band	Value	1251.55	576.16	122.85	15.52	9.81	2.90
	Percent	63.25	29.12	6.21	0.78	0.50	0.15

[Eigen Vector]		PC1	PC2	PC3	PC4	PC5	PC6
Band							
1		0.07	-0.25	0.09	0.72	-0.24	0.59
2		0.17	-0.25	0.22	0.42	-0.29	-0.78
3		0.23	-0.53	0.67	-0.43	0.01	0.18
4		0.57	0.66	0.41	0.16	0.18	0.04
5		0.66	-0.08	-0.48	-0.26	-0.50	0.08
6		0.38	-0.38	-0.31	0.17	0.76	-0.07

[Degree of Correlation]		PC1	PC2	PC3	PC4	PC5	PC6
Band							
1		0.34	-0.83	0.13	0.39	-0.10	0.14
2		0.65	-0.67	0.27	0.18	-0.10	-0.15
3		0.48	-0.75	0.44	-0.10	0.00	0.02
4		0.77	0.61	0.17	0.02	0.02	0.00
5		0.97	-0.08	-0.22	-0.04	-0.06	0.01
6		0.80	-0.55	-0.20	0.04	0.14	-0.01

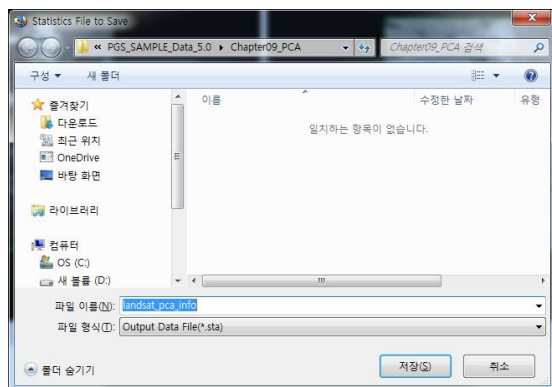
OK Cancel

eigenvalue는 해당 성분의 전체 분산도를 나타내며, 흔히 eigenvalue의 상대적 크기를 퍼센트로서 나타냄으로써 각 주성분이 포함하는 상대적 크기를 표현합니다. 본 실습에서 주성분 1(PC1)은 전체 분산의 63.25%정도이고 처음 세 개의 component가 전체 분산의 98.5 %정도를 차지합니다.


10. OK를 클릭합니다.

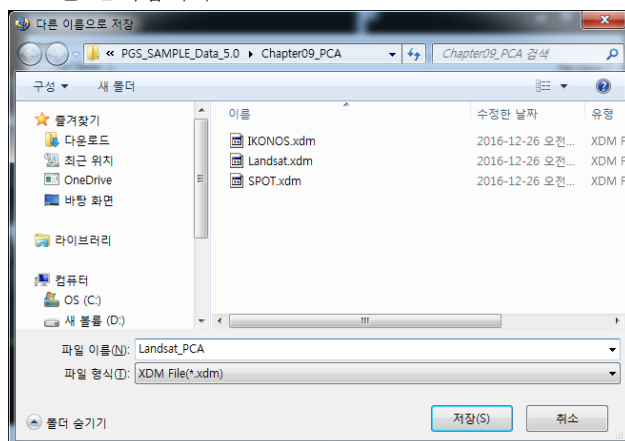
Chapter 9: Principal Components Analysis

11. Statistics File to Save 대화상자에서 file name으로 **landsat_pca_info**를 입력하고 Save합니다.






밴드의 principal components를 export 합니다

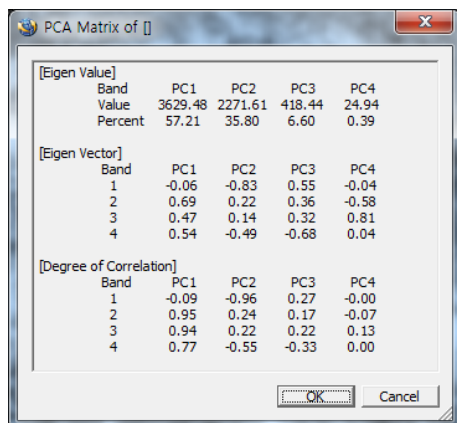
12. Save To File  툴 버튼을 클릭합니다.
13. File Name 버튼을 누르고 Save As 창이 나타나면 File name을 **Landsat_PCA**로 입력하고 Save를 클릭합니다.



14. Save를 클릭합니다.
15. File Information 창을 닫습니다.

SPOT data의 주성분 분석

16. Open  툴 버튼을 클릭하고 **SPOT.xdm** 파일을 선택하여 엽니다.
17. Output Data Type을 Short로 변경합니다.
18. RTP  툴 버튼을 클릭합니다.
19. Show Eigen Matrix  툴 버튼을 클릭합니다.
20. 모든 데이터가 보이도록 PCA Matrix 대화상자의 크기를 조정합니다.



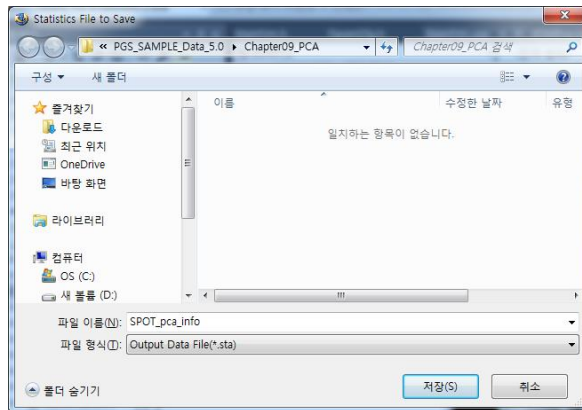
[Eigen Value]				
Band	PC1	PC2	PC3	PC4
Value	3629.48	2271.61	418.44	24.94
Percent	57.21	35.80	6.60	0.39

[Eigen Vector]				
Band	PC1	PC2	PC3	PC4
1	-0.06	-0.83	0.55	-0.04
2	0.69	0.22	0.36	-0.58
3	0.47	0.14	0.32	0.81
4	0.54	-0.49	-0.68	0.04

[Degree of Correlation]				
Band	PC1	PC2	PC3	PC4
1	-0.09	-0.96	0.27	-0.00
2	0.95	0.24	0.17	-0.07
3	0.94	0.22	0.22	0.13
4	0.77	-0.55	-0.33	0.00

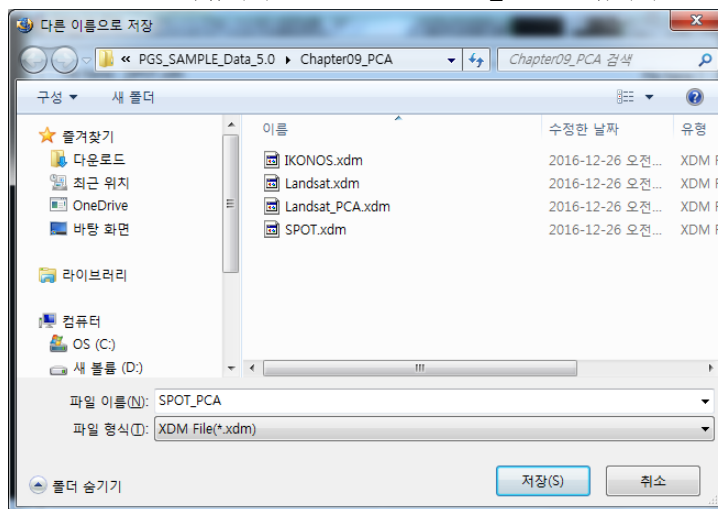
21. OK를 클릭합니다.
22. Statistics File to Save 대화상자에서 File name을 **SPOT_pca_info**로 입력하고 Save를 클릭합니다.

Chapter 9: Principal Components Analysis



23. Save to File  톨 버튼을 클릭합니다.




24. File Name을 클릭합니다. File name을 **SPOT_PCA**로 입력하고 Save를 클릭합니다.

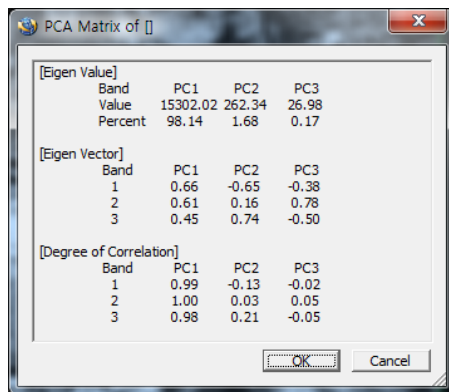


25. Save를 클릭합니다.

26. File Information 창을 닫습니다.

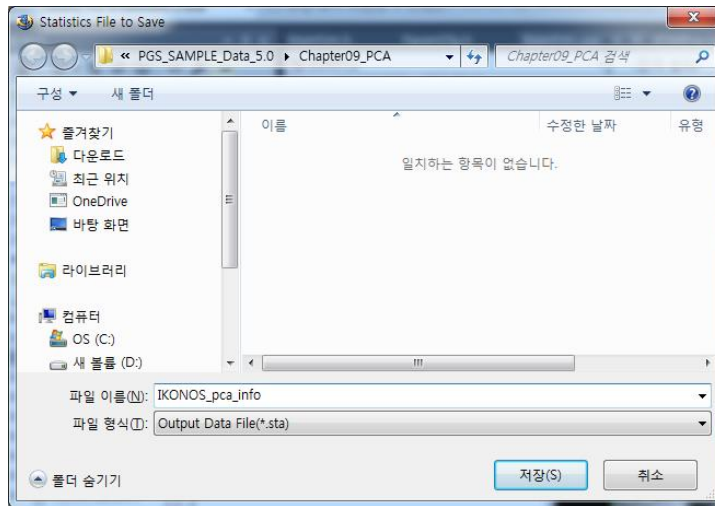
IKONOS data의 주성분 분석

27. Open  툴 버튼을 눌러 **IKONOS.xdm** 파일을 선택하여 엽니다.
- 28.
29. Output Data Type을 Short로 변경합니다.
30. RTP  툴 버튼을 클릭합니다.
31. Show Eigen Matrix  툴 버튼을 클릭합니다.
32. 모든 데이터가 보이도록 PCA Matrix 대화상자의 크기를 조정합니다.



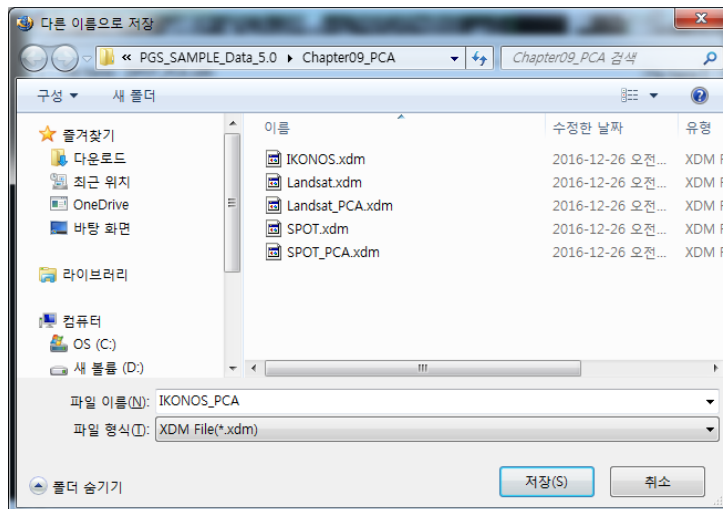
33. OK를 클릭합니다.
34. Statistics File to Save 대화상자에서 File Name을 **IKONOS_pca_info**로 입력하고 저장합니다.

Chapter 9: Principal Components Analysis



35. Save to File  툴 버튼을 클릭합니다.

36. File Name을 클릭하고 **IKONOS_PCA**로 입력한 후 Save를 클릭합니다.

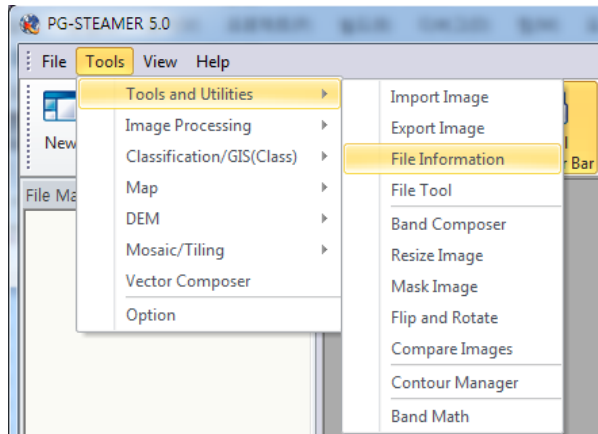



37. Save를 클릭합니다.


38. File Information 창을 닫습니다.

PCA 밴드 간의 상관 관계를 확인합니다

39. Menu → Tools and Utilities → File Information 순으로 실행합니다. .



40. Open  툴 버튼을 클릭하고 Chapter09에 있는 **landsat_PCA.xdm** 파일을 선택하여 엽니다.

41. Statistics  툴 버튼을 클릭합니다.

42. Real Time Processing  툴 버튼을 클릭합니다.

43. 밴드 간의 상관 관계를 확인합니다.

Correlation Matrix						
Band	1	2	3	4	5	6
1	1.0000	-0.0065	-0.0083	0.0086	0.0073	0.0053
2	-0.0065	1.0000	-0.0171	-0.0133	-0.0209	0.0032
3	-0.0083	-0.0171	1.0000	-0.0427	-0.0180	-0.0032
4	0.0086	-0.0133	-0.0427	1.0000	-0.0293	-0.0124
5	0.0073	-0.0209	-0.0180	-0.0293	1.0000	-0.0024
6	0.0053	0.0032	-0.0032	-0.0124	-0.0024	1.0000

44. Add  툴 버튼을 클릭하고 **SPOT_PCA.xdm** 파일을 선택하여 엽니다.

45. Real Time Processing  툴 버튼을 클릭합니다.

46. 밴드 간의 상관 관계를 확인합니다.

Chapter 9: Principal Components Analysis

Correlation Matrix				
Band	1	2	3	4
1	1.0000	-0.0004	0.0027	-0.0029
2	-0.0004	1.0000	0.0023	0.0008
3	0.0027	0.0023	1.0000	0.0012
4	-0.0029	0.0008	0.0012	1.0000

principal component 간의 상관 관계가 매우 약함을 알 수 있습니다.

47. Add  툴 버튼을 클릭하고 **IKONOS_PCA.xdm** 파일을 선택하여 엽니다.

48. Real Time Processing  툴 버튼을 클릭합니다.

49. 밴드 간의 상관 관계를 확인합니다.

Correlation Matrix			
Band	1	2	3
1	1.0000	-0.0021	0.0059
2	-0.0021	1.0000	0.0156
3	0.0059	0.0156	1.0000

50. File Information 창을 닫습니다.

Appendix

이론적 배경 : 주성분 분석

Chapter 9: Principal Components Analysis

주성분 분석은 특히 위성 영상 자료가 대부분 다중 분광(multi-spectral)자료라는 측면에서 정보의 손실을 최소화하면서 자료의 크기를 줄일 수 있다는 장점을 제공한다. 예를 들어 7개의 밴드를 가지고 있는 Landsat TM 영상 자료의 경우, 분석자가 영상으로서 동시에 분석할 수 있는 자료는 3개 밴드에 불과하다(RGB 색합성을 통하여).

그러나 이 영상 자료를 주성분 분석 변환하여 제1주성분, 제2주성분, 제3주성분을 각각 빨강, 초록, 파랑으로 할당하여 칼라 영상으로서 나타내면, 7개의 밴드 자료가 가지고 있는 거의 모든 정보를 동시에 육안으로서 확인하는 것이 가능하다.

우선 그림 1을 통해서 주성분 분석의 개념을 이해하도록 한다. 그림 1의 (a)와 (b)에는 두 개의 밴드를 가지고 있는 영상 자료의 첫 번째 밴드와 두 번째 밴드 자료를 영상으로서 나타냈다.

(c)는 두 밴드 자료의 스캐터그램을 보여주고 있으며, 영상과 스캐터그램을 통해서 두 영상이 매우 유사한 특징을 보이고 있음을 알 수 있다. 즉, 첫 번째 밴드 자료에서 큰 수치 화소값을 가지고 있는 화소는 두 번째 밴드 자료에서도 큰 수치 화소값을 가지고 있으며, 반대로 첫 번째 밴드 자료에서 작은 수치 화소값을 가지고 있는 화소는 두 번째 밴드에서도 작은 수치 화소값을 가지고 있다. 즉, 두 영상간의 상관 관계가 매우 크다는 것을 알 수 있으며, 이는 곧 두 영상이 거의 유사한 내용의 정보를 포함하고 있음을 의미한다.

그림 1의 (c)에 나타난 스캐터그램에서 가로축은 밴드 1 영상 자료를 나타내며, 세로축은 밴드 2 영상 자료를 나타내고 있다. 여기에서 제1주성분이라고 표시된 새로운 축을 가정하고 이를 제1주성분 축이라고 부른다.

이 때, 스캐터그램에 점으로서 표현되어 있는 모든 화소들을 이 제1주성분 축에 투영시키는 것이 가능하며, 모든 화소들의 수치 화소값이 이 제1주성분 축에 투영시킨 값으로 대체된 새로운 영상을 생성할 수 있다. 이렇게 하여 생성된 영상이 (d)에 나타난 제1주성분 영상이며, 이 영상은 밴드 1 영상 자료와 밴드 2 영상 자료가 포함하는 대부분의 정보를 포함하고 있다.

또한 제1주성분 축에 직교하는 새로운 축을 가정할 수 있으며, 이를 제2주성분 축이라고 부른다. 이 때 제1주성분 축과 제2주성분 축의 원점은 밴드 1 영상 자료와 밴드 2 영상 자료의 평균값이 만나는 지점이 된다.

앞서와 같은 원리로 제2주성분 축에 투영된 새로운 영상이 (e)에 나타나 있으며 이를 제2주성분 영상이라고 부른다. 그림을 통해서 입력 영상 자료의 대부분의 정보가 제1주성분 영상에 포함되어 있음을 알 수 있으며, 실제로 밴드 1 영상 자료와 밴드 2 영상 자료가 가지고 있는 정보의 95% 이상이 제1주성분 영상에 포함되고 있다. 따라서 제1주성분 영상만을 분석하여도 밴드 1 영상 자료와 밴드 2 영상 자료를 모두 분석하는 경우에 비교하여 약 5%이하의 정보만이 손실되었다고 말할 수 있다.

이제 실제 주성분 분석을 수행하는 과정에 대하여 설명하도록 한다. 주성분 분석은 밴드간 공분산 행렬(covariance matrix)이나 혹은 상관관계수 행렬(correlation matrix)의 고유벡터(eigenvector)와 고유치(eigenvalues)를 이용하여 수행된다.

주어진 정방 행렬 \mathbf{A} 에 대하여 다음의 식을 만족하는 벡터 \mathbf{x} 와 스칼라 λ 를 각각 고유벡터와 고유치라고 한다. 주어진 정방 행렬에 대한 고유벡터와 고유치를 계산하는 과정은 [Press등, 96] 등의 각종 수치 분석 관련 문헌을 참고한다.

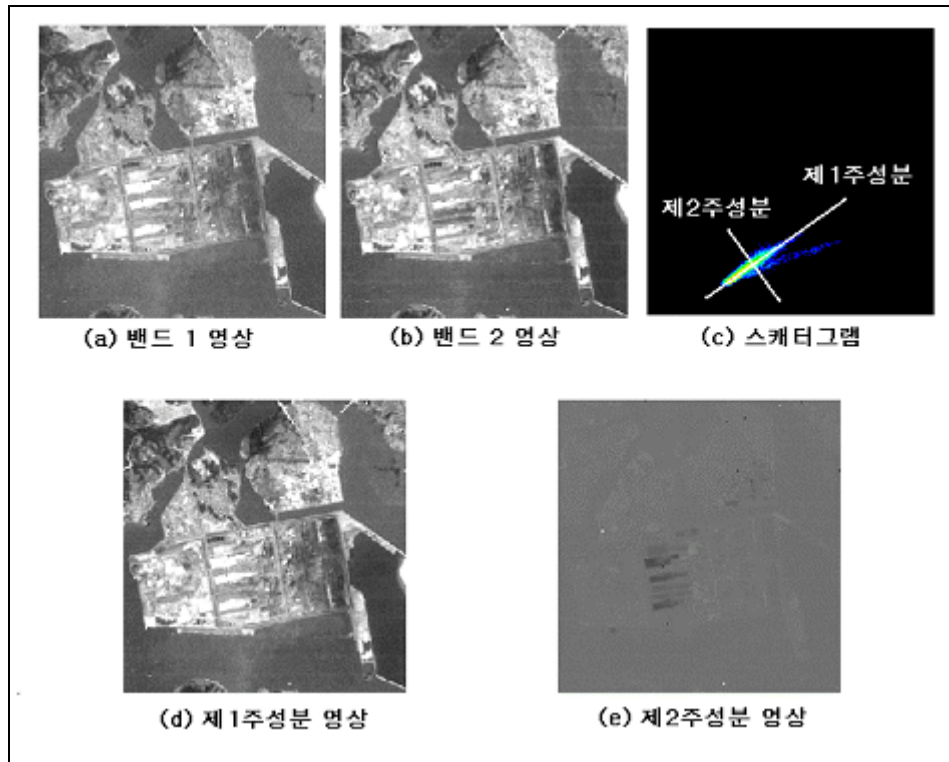


그림 1. (a)와 (b)에는 두 개의 밴드를 가지고 있는 영상 자료의 첫 번째 밴드와 두 번째 밴드 자료를 영상으로서 나타냈다. (c)는 두 밴드 자료의 스캐터그램을 보여주고 있으며, 영상과 스캐터그램을 통해서 두 영상이 매우 유사한 특징을 보이고 있음을 알 수 있다. (c)에 나타낸 스캐터그램에서 가로축은 밴드 1 영상 자료를 나타내며, 세로축은 밴드 2 영상 자료를 나타내고 있다. 여기에서 제1주성분이라고 표시된 새로운 축을 가정하고 이를 제1주성분 축이라고 부른다. 이 때, 모든 화소들을 이 제1주성분 축에 투영시키는 것이 가능하며, 모든 화소들의 수치 화소값이 이 제1주성분 축에 투영시킨 값으로 대체된 새로운 영상을 생성할 수 있다. 이렇게 하여 생성된 영상이 (d)에 나타낸 제1주성분 영상이며, 이 영상은 밴드 1 영상 자료와 밴드 2 영상 자료가 포함하는 대부분의 정보를 포함하고 있다. 또한 제1주성분 축에 직교하는 새로운 축을 가정할 수 있으며, 이를 제2주성분 축이라고 부른다. 이 때 제1주성분 축과 제2주성분 축의 원점은 밴드 1 영상 자료와 밴드 2 영상 자료의 평균값이 만나는 지점이 된다. 앞서와 같은 원리로 제2주성분 축에 투영된 새로운 영상이 (e)에 나타나 있으며 이를 제2주성분 영상이라고 부른다. 그림을 통해서 입력 영상 자료의 대부분의 정보가 제1주성분 영상에 포함되어 있음을 알 수 있으며, 실제로 밴드 1 영상 자료와 밴드 2 영상 자료가 가지고 있는 정보의 95% 이상이 제1주성분 영상에 포함되고 있다.

Chapter 9: Principal Components Analysis

그림 2에 나타난 영상 자료는 6개의 밴드를 가지고 있다. 이 영상 자료의 밴드간 공분산 행렬은 표 1에 나타난 바와 같다. 표 2에는 밴드간 공분산 행렬로부터 계산된 고유치와 고유벡터를 나타냈다.

공분산 행렬로부터 고유치가 계산되면, 고유치의 크기 순으로 제1주성분, 제2주성분 등으로 구분하며, 고유치는 해당 성분의 분산을 나타낸다. 따라서 고유치가 클수록 더 많은 영상 정보를 포함하고 있다고 말할 수 있으며, 흔히 고유치의 상대적 크기를 퍼센트로서 나타냄으로써 각 주성분이 포함하는 정보량의 상대적 크기를 표현한다.

표 2에 나타냈던 고유치와 고유벡터를 크기순으로 정렬하면 표 3과 같으며, 여기에는 각 고유치의 상대적 퍼센트도 함께 나타냈다.

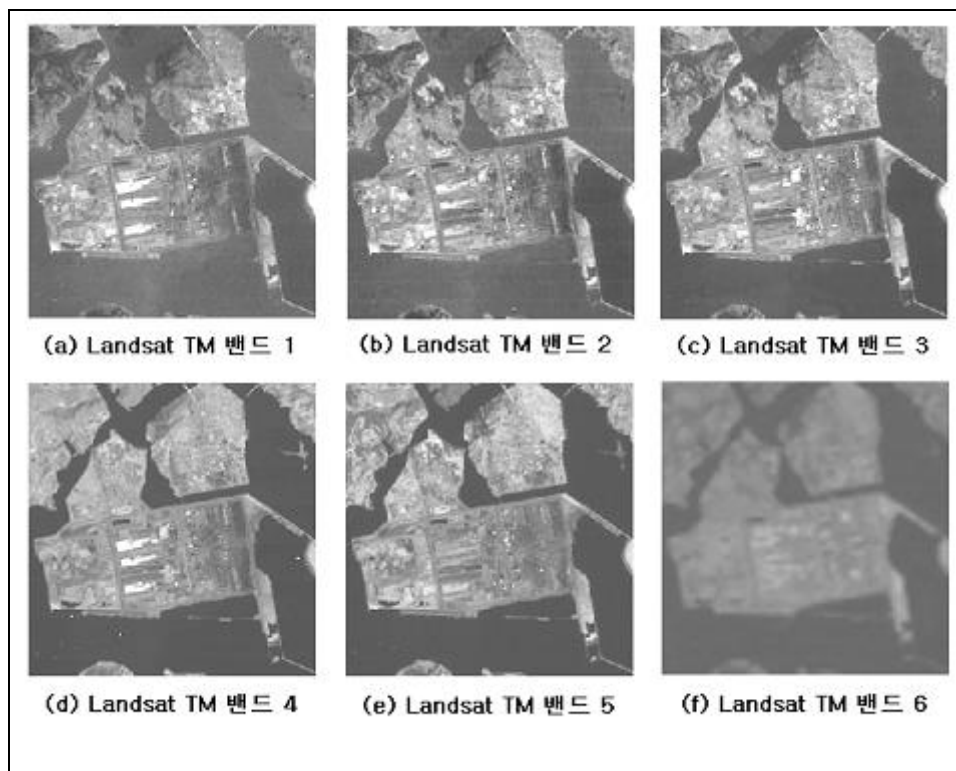


그림 2. 주성분 분석 변환을 위한 남해안 지역의 Landsat TM 영상 자료. 밴드 7 자료를 제외한 6개 밴드 자료가 사용되었다.

	밴드 1	밴드 2	밴드 3	밴드 4	밴드 5	밴드 6
밴드 1	138.90					
밴드 2	83.68	60.02				
밴드 3	156.62	117.72	260.16			
밴드 4	174.80	130.17	298.76	446.98		
밴드 5	324.46	256.45	588.17	816.15	1715.46	
밴드 6	155.50	109.03	260.46	351.85	680.98	415.00

표 1. 그림 4-26에 나타난 영상 자료의 밴드간 공분산 행렬

고유치		21.17	2.07	42.22	101.14	2736.51	133.43
고유벡터	밴드 1	0.63	-0.27	0.08	0.57	0.17	0.41
	밴드 2	-0.06	0.91	0.11	0.34	0.13	0.15
	밴드 3	-0.76	-0.30	0.13	0.41	0.28	0.25
	밴드 4	0.00	0.04	-0.92	0.03	0.39	0.02
	밴드 5	0.14	-0.02	0.32	-0.08	0.78	-0.51
	밴드 6	0.02	0.05	0.14	-0.61	0.34	0.70

표 2. 표1에 나타난 밴드간 공분산 행렬로부터 계산된 고유치와 고유벡터

표 3에 의하면 제1주성분이 전체 영상 자료의 90% 정도의 분산을 포함하고 있으며, 제1주성분, 제2주성분 및 제3주성분이 6개 밴드의 전체 영상 자료가 가지고 있던 정보량의 약 97% 정도를 가지고 있다고 할 수 있다.

실제 주성분 영상은 계산된 고유벡터를 이용한 간단한 선형 변환을 통하여 이루어질 수 있다. 예를 들어서 특정 화소의 밴드별 수치 화소값을 [DN1, DN2, DN3, DN4, DN5, DN6]이라고 할 때, 제1주성분 영상에서 해당 화소의 수치 화소값은 다음과 같이 계산된다.

성분		제 1 주성분	제 2 주성분	제 3 주성분	제 4 주성분	제 5 주성분	제 6 주성분
고유치		2736.51	133.43	101.14	42.22	21.17	2.07
퍼센트		90.12	4.39	3.33	1.39	0.70	0.07
고유 벡터	밴드 1	0.17	0.41	0.57	0.08	0.63	-0.27
	밴드 2	0.13	0.15	0.34	0.11	-0.06	0.91
	밴드 3	0.28	0.25	0.41	0.13	-0.76	-0.30
	밴드 4	0.39	0.02	0.03	-0.92	0.00	0.04
	밴드 5	0.78	-0.51	-0.08	0.32	0.14	-0.02
	밴드 6	0.34	0.70	-0.61	0.14	0.02	0.05

표 3. 고유치의 크기순으로 정렬된 고유벡터

$$DN_{PCA_1} = DN_1 \times 0.17 + DN_2 \times 0.13 + DN_3 \times 0.28 \\ + DN_4 \times 0.39 + DN_5 \times 0.78 + DN_6 \times 0.34$$

이와 같이 하여 6개의 밴드를 가지고 있는 영상 자료에 대하여 제1주성분 영상으로부터 제6주성분 영상까지 총 6개의 주성분 영상이 생성될 수 있으며, 그림 3 에는 그림 4-26 에 나타냈던 영상 자료의 주성분 분석 변환된 영상을 주성분별로 나타냈다.

한편, 주성분 영상들과 입력 영상 자료의 각각의 밴드들과의 상관 관계를 통해서 변환된 주성분 영상이 어떠한 종류의 정보들을 포함하고는 것인지를 파악할 수 있다. 특정 주성분 영상과 특정 입력 밴드 자료간의 상관 관계는 다음의 식으로서 표현될 수 있다.

$$\text{상관관계}_{ij} = \frac{\text{고유벡터}_{ij} \times \sqrt{\text{고유치}_j}}{\sqrt{\text{분산}_i}}$$

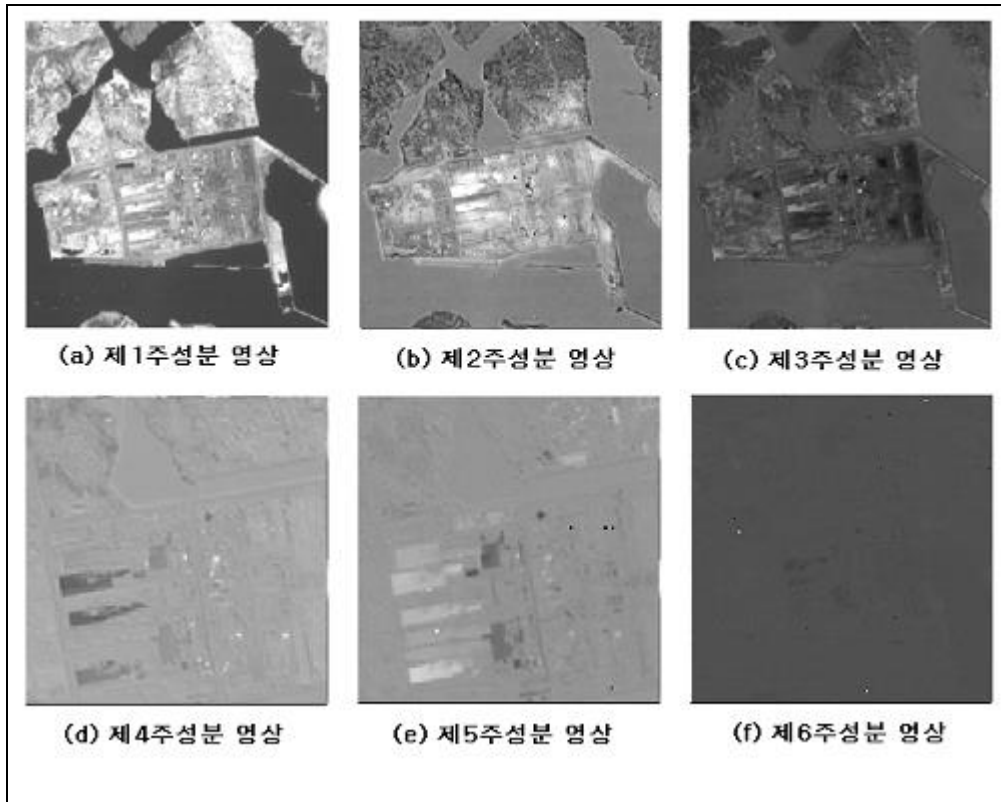


그림 3. 그림 2 에 나타냈던 Landsat TM 영상 자료에 대한 주성분 분석 변환 결과. 제1주성분 영상이 전체 영상이 가지고 있는 정보의 약 90% 정도를 포함하고 있으며, 제1주성분, 제2주성분, 그리고 제3주성분의 3개 영상 자료가 전체 6개 밴드의 영상 자료가 가지고 있는 정보의 약 97%를 포함한다.

여기에서 i 는 입력 영상 자료의 밴드를 나타내며, j 는 주성분을 나타낸다. 즉, 제1주성분 영상과 밴드 1 영상은 $0.17 \times \sqrt{2736.51} / \sqrt{138.90} = 0.754$ 의 상관 관계를 가지고 있다.

주성분 분석은 영상 자료의 정보를 압축하는데 뿐아니라, 영상 자료의 효율적인 대비신장(decorrelation stretch)이나 변화 탐지(change detection) 등을 위해서도 효과적으로 활용될 수 있다.

