The combinatorial explosion: defining procedures to reduce data redundancy and to validate the results of processed hyperspectral images.

A. Traviglia

Department of Archaeology – Archaeological Computing Laboratory, University of Sydney, 2006, NSW, AUSTRALIA, atraviglia@usyd.edu.au

Keywords: MIVIS, hyperspectral images, image processing, image evaluation, GIS, Aquileia (Italy)

ABSTRACT:

This paper will focus on ways to manage large numbers of remotely sensed data images as results of image processing, a primary problem especially for those dealing with hyperspectral images that pose considerable issues due to the elevated number of channels. While briefly introducing the results of the application of several common image processing techniques in the target area of Aquileia (NE Italy), the current paper will discuss the necessity to define a set of procedures to reduce the number of final images to be used for visual inspection, selecting the ones that do not carry redundant information. Consequently, cross process coverage, detected traces evaluation and process validity tables will be presented and the results of their application discussed in order to provide information about how to reduce the images to a small number and be able to insure the complete coverage as regards to the detectable traces.

1. INTRODUCTION

The application of hyperspectral images in archaeological research in the last years has demonstrated that this kind of data can represent a valuable resource as a complementary source of information for archaeological goals. In the many projects using hyperspectral data, various processes are normally applied to part or to all of the image bands used for detection of archaeological features. When dealing with one or more of these images formed by a number of bands, which can be over a hundred, even considering realistically to use a small part of the bands for multiple processes, inevitably the result is an extremely large number of images that must be subjected to visual analysis and interpretation; at the same time, as a side problem, since many of the different processes produce similar results that provide the same traces, the number of repeatedly identified and recorded features can became too large.

The primary goal of the research project introduced here was to ascertain the usefulness of the MIVIS (Multispectral Infrared and Visible Imaging Spectrometer) images in the identification of unknown archaeological features in the case study area of Aquileia (Italy) and to establish how much these images could produce in terms of new archeological information. The process of exploitation of the spectral content of the MVIS data led to an unexpected issue to be faced, that is to say the overproduction of images that must be subjected to visual analysis and interpretation and the consequent elevated number of detected features. MIVIS is in fact a 102 bands image: even utilizing a minor number of the bands for simple TC/FC composites and for the several applied image processes, the combinatorial explosion created an almost unmanageable situation. Consequently it became critical to define a set of procedures to reduce the number of final images to be used for a visually performed interpretation of the imagery, selecting the ones that did not carry redundant information. Correlation and selection matrixes (Traviglia 2007: 295), comparison tables, cross process coverage tables were adopted on a test MIVIS run in order to define procedures to reduce the images to a smaller number and be able to insure the complete coverage as regards to the detectable traces. Only the most successful processes or composites were then adopted for the other MIVIS runs.

The set of chosen images has been managed through a GIS that provided also the archaeological and topographical data necessary to eventually recognize the detected surface features as ancient origin traces, attributing each of them a value of visibility and archaeological reliability as result of the interpretation process.

2. HYPERSPECTRAL IMAGES AND ARCHAEOLOGICAL INVESTIGATION

2.1 *The research project*

In the present research, the spectral content of the MIVIS images was used to reveal the presence of ground features related to archeological sites and structures under the ground surface on the basis of the different spectral characteristics of the terrain and of the vegetation. MIVIS is a simultaneous multispectral imaging system that operates in the wide range of wavelengths from visible to Thermal-IR regions of the spectrum, with a high spectral resolution and elevated number of channels (102). The runs are normally taken from a distance of around 1500-2500m from the Earth surface: based on flight altitude their pixel resolution is usually around 3x3m. The MIVIS images used in this research have been captured on two different days during October 1998 in two shots: a daytime shot (about h. 12 pm) and a so-called night shot (about h. 9-10 am). There are 10 runs that cover the target area, divided in 5 runs for each day or shot. Various processes have been applied to these images and their results compared in order to identify the ones that better match the different research targets. The goal of the enhancement techniques is to increase and improve the optic distinction between traces recorded in the scene and the surrounding areas by generating a new image where the useful information is more easily identifiable.

The case study area on which the MIVIS data have been tested included the Roman foundation town of Aquileia, located in NE Italy, and its surrounding countryside, comprising the neighboring Communes of Terzo d'Aquileia and Fiumicello.

2.2 Methodological approach to spectral analyses

Among the several processing methods the MIVIS runs have been subjected to in this project, the vegetation indexes (VI) have found prominence. As well known these indices can enable the identification of underground archaeological deposits that enhance or, in opposition, limit the growth of the vegetation. Heterogeneity of the texture of the subsoil has in fact a strong reflection on the growth of the vegetation, determining the appearance of the so called "crop-, grass-, or weed- marks". Various VI (such as DVI, NDVI, MSAVI2) have been calculated, tested, and compared to determine the best method for evaluating vegetation since different vegetation indices can work better for archaeological research, in accordance with certain environmental situations.

A specific Soil Index (SI) has been developed based on the study of the different degrees of water absorption to support the investigation of the numerous traces over non-vegetated zones: emphasizing the wetness or the dryness of a portion of the ground, it allowed the identification of feature on bare soil possibly related to underground structures (Traviglia 2005).

Due to the highly correlated nature of hyperspectral data, the Principal Component Analysis (PCA) was adopted to reduce the information previously contained in the original n-band data set into a smaller number of new bands that could be used in place of the original ones. However, in applying the most common procedure of visual inspection in PCA, consisting of generating a RGB false color composite of some of the principal components of a scene and increasing to 6 the number of used principal components, which have proven to convey significant information, 120 different combinations were generated losing the advantage of computing an "image reducing" process (Traviglia 2006: 126). In addition to the traditional PCA performance, a number of principal component images have been created through spectral subsets in order to convey only the information of spectral regions of interest.

A Selective Principal Component Analysis (SPCA) has been computed for groups of homogeneous bands belonging to different spectral regions or to different sensors' spectrometers. A selected number of the resulting images (originally 140 for each MIVIS run) has been then visually inspected using the chosen images both singularly and through their composites. Their selection was based on the intrinsic dimensionality (ID) of the data that led to a reduced number of 14 images and subsequently a correlation matrix was calculated in order to automatically identify in that group the best possible combinations that could carry the whole of the information (Traviglia 2007: 294-295).

Among the numerous forms of linear data transformations that have been developed for vegetation monitoring the Tasseled Cup Transformation, one of the most applied in archaeological studies, has been computed. The absence of specific coefficients for the calculation of a TCT for MIVIS data led to test a flow of operations in order to perform a modification of the MIVIS bands to make them applicable the same coefficients used to perform a Tasseled Cap Transformation over a Landsat ETM image. To do that, the MIVIS bands had to be converted into the same spectral coverage of a Landsat image, where possible. The first 3 axes of the resulting images from each run have been singularly visually inspected as well as their combinations (6 different composites for each run).

Finally, some of the most successful images delivering an elevated number of features have been obtained through the combination and display in RGB of different bands selected from different processes. The most interesting appeared to be the ones involving Tasseled Cap components, like for example the composite in RGB display of Tasseled Cap Brightness, NDVI and Tasseled Cap Wetness.

It is easy to understand that at the end of the processing activity and before starting the phase of inspection and interpretation of the images, the number of the images to submit to visual analysis, which included the single original bands, their simple combinations in true and false color composites, as well as all the described processes and the consequent derived composites, was extremely high.

3. EVALUATING, COMPARING, SELECTING: THE IDENTIFICATION OF SUCCESFUL PROCESSES

Among the many possibilities that the integration of the MIVIS images in the GIS allowed, there is the opportunity to compare in quantitative and statistic ways the number of detected and digitized features in order to evaluate the potential of each performed process in terms of capability of emphasizing the traces. This leads to the identification of which process is more successful for reaching the set goals. Attributes have been assigned to every features digitized in the GIS in order support this comparison.

The processes can be evaluated from two points of view: from the point of view of the visibility of the traces they can provide and the point of view of the number of traces they allow to identify. Moreover they can be compared through the reciprocal coverage of traces they exhibit with one another.

3.1 Evaluation through "Visibility Index Table"

The visibility of a trace can be defined as the degree to which an object can be distinguished from the surrounding areas through the contrast that is realized in the digital images by significant difference of DN values of the pixels. In order to define which are the MIVIS images (image processes or simple visualizations) able to better emphasize or display the traces over the landscape, an index of visibility was created on the base of a visibility attribute conferred to each detected trace. The "Index of Visibility" of the anomalies is an index of the quality of visual appearance of the traces; it is evaluated through a scale of values from 1 to 5, where 1 indicates slight possibility to identify the trace and 5 means that the trace is extremely visible, independently of the type of surface materials surrounding the trace (bare soil or vegetation). In the GIS, one scale value is attributed to each identified trace for a given process. The Visibility Index Table (see below Table 1) lists, in order of highest score of average visibility, the MIVIS images that showed highest visibility of traces.

Process name	[illegible]	[illegible]	[illegible]	Anomalies per process
MIVIS- PCA (TotRun) 2-4-6	2,368	1	5	28
MIVIS- TIR b96	2,250	1	4	32
MIVIS- TCAP (PCA) Br-NDVI-Wet	2,193	1	5	124
MIVIS- False Color	2,162	1	5	65
MIVIS- NDVI	2,138	1	4	54
MIVIS- TCAP Br-SLI-Wet	2,120	1	5	132
MIVIS- TCAP (PCA) Wet	2,088	1	4	60
MIVIS- TIR PCA	2,083	1	4	20
MIVIS- PCA S1P2-S1P1-BluP1	2,079	1	5	114
MIVIS- TCAP Br-Wet-Gr	2,030	1	4	99
MIVIS- MSAVI2	2,000	1	4	43
MIVIS- DVI	1,960	1	4	47
MIVIS- PCA (TotRun) 1-2-5	1,898	1	4	40
MIVIS- S4P1-b18-b13	1,882	1	4	55
MIVIS- TCAP Wet	1,873	1	4	45
MIVIS- PCA S1P1-S1P2-S4P1	1,871	1	4	52
MIVIS- MIR b57	1,857	1	4	20
Ortofoto	1,858	1	5	100
MIVIS- TCAP Green	1,814	1	4	36
MIVIS- PCA R1-G1-B1	1,803	1	4	51
MIVIS- True Color	1,790	1	5	253
MIVIS- Red b13	1,783	1	3	59
MIVIS- PCA S1P1-S1P2-BluP1	1,693	1	4	68
MIVIS- Red PCA	1,636	1	4	45
MIVIS- Green PCA	1,574	1	4	37
MIVIS- SLI	1,542	1	4	68
MIVIS- Green b7	1,524	1	4	40
MIVIS- NIR b18	1,471	1	4	50
MIVIS- Blue b3	1,429	1	3	14
MIVIS- Blue PCA	1,385	1	3	13
MIVIS- NIR PCA	1,344	1	4	48

Process	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	True Color	TCAP PCA [illegible]	TCAP [illegible]	PCA (TotRun [illegible]	PCA TotRun [illegible]	PCA S1P2- [illegible]	PCA S1P1- [illegible]	PCA S1P1- [illegible]	PCA R1-G1-B1	DVI	NDVI	MSAVI 2	[illegible]	TCAP Wet	TCAP PCA Wet
Contrast	3	3	3	0	2	4	2	2	3	2	2	1	1	1	0
Clarity	2	3	3	0	3	3	3	3	2	2	1	1	1	0	0
Definition	1	2	2	0	2	3	3	2	3	3	2	1	1	1	0

Process	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	TCAP Green	TIR b96	MIR b57	NIR b18	Red b13	Green b7	Blue b3	NIR PCA	Red PCA	Green PCA	Blue PCA	False Color	TIR PCA	S4P1-b18-b13	TCAP Br-Wet-Gr
Contrast	0	1	0	3	0	0	0	2	2	0	0	3	1	2	2
Clarity	1	2	0	2	0	0	0	3	2	1	0	3	1	3	2
Definition	1	1	0	3	0	1	0	3	2	1	0	3	1	2	3

Table 1 (leftt). Average Visibility Index by process with min. and max. range included Table (right) 2. Contrast-clarity-definition table example: the values go from 0 up to 4, where 4 is the better level.

In the proposed example, the composite of the Principal Components 2, 4 and 6 (RGB display) of the PCA obtained using all the bands appears in that table to have the best absolute results in terms of the visibility. However, it must be made clear that this does not correspond to having a "clear", easily readable image; in actuality, this PC composite is very "difficult" to decode and one needs to have a very good knowledge of the area and of the traces that are there. In front of situations like this one, it became obvious that a more rigorous index was necessary. Another two attributes took part in the redefinition of the visibility: clarity and definition, where "clarity" refers to the sharpness of the boundaries of the traces and the "definition" refers to the homogeneity of pixel DN values within the trace. Thus, in order to really evaluate the success or failure of a process, meaning its readability, it is fundamental to have for each trace a related table of information about the compared qualities of contrast, clarity and definition (see Table 2), to give a realistic judgment even in situations where the contrast is very good, but the process clarity and definition are very poor.

Back to PCA 2-4-6, this process demonstrated to be able to create a strong contrast between the feature itself and what is surrounding it, but it was indispensable to know in advance where the specific trace was located. Consequently, PCA 2-4-6 does not allow the identification of a large number of surface features just by itself. For this reason, in interpreting the data, it is extremely important to compare the average of visibility allowed from a process with the total number of features identifiable through the process itself (see Table 2 and Table 3). In the case of the PCA 2-4-6, it is a low number, a clear demonstration that this image cannot be used by itself in a visual analysis for the research of surface features.

3.2 Evaluation through Per Process Trace Count Table

The second way to compare the processes is from the point of view of the number of traces they allow to identify. In Table 3 (left) it is shown how the image processes or simple displays of images work in terms of identifying surface features. In this case, there is not a judgment on the quality of the visibility of the trace, so, even if True Colors images have shown to allow the identification of the highest number of features, the quality of definition, contrast and clarity of features is not always the best and the display of the same trace in other processes is needed to better identify the contours of the trace itself. True Color images, in any case, have demonstrated to be necessary to guide the process of identification of anomalies, because the comparison of the processed images with a True Color display image allows to determine the possible nature of the feature.

MIVIS- True Color	253	88
MIVIS- TCAP Br-WWI-Wet	132	5
MIVIS- TCAP (PCA) Br-NDVI-Wet	124	2
MIVIS- PCA S1P2-S1P1-BluP1	114	1
Ortofoto	100	55
MIVIS- TCAP Br-Wet-Gr	99	2
MIVIS- PCA S1P1-S1P2-BluP1	68	0
MIVIS- SLI	66	0
MIVIS- False Color	66	0
MIVIS- TCAP (PCA) Wet	60	0
MIVIS- Red b13	58	3
MIVIS- S4P1-b18-b13	55	0
MIVIS- NDVI	54	1
MIVIS- PCA S1P1-S1P2-S4P1	52	1
MIVIS- PCA R1-G1-B1	51	0
MIVIS- NIR b18	50	0
MIVIS- NIR PCA	48	0
MIVIS- DVI	47	0
MIVIS- TCAP Wet	46	1
MIVIS- Red PCA	45	0
MIVIS- MSAVI2	43	1
MIVIS- PCA (TotRun) 1-2-5	40	0
MIVIS- Green b7	40	0
MIVIS- Green PCA	37	0
MIVIS- TCAP Green	36	0
MIVIS- TIR b96	32	0
MIVIS- PCA (TotRun) 2-4-6	28	1
MIVIS- MIR b57	20	0
MIVIS- TIR PCA	20	0
MIVIS- Blue b3	14	0
MIVIS- Blue PCA	13	0

Process #	Name of process
[illegible]	Ortofoto
[illegible]	MIVIS- True Color
[illegible]	MIVIS- TCAP (PCA) Br-NDVI-Wet
[illegible]	MIVIS- TCAP Br-WWI-Wet
4	MIVIS- PCA (TotRun) 2-4-6
[illegible]	MIVIS- PCA (TotRun) 1-2-5
[illegible]	MIVIS- PCA S1P2-S1P1-BluP1
[illegible]	MIVIS- PCA S1P1-S1P2-S4P1
[illegible]	MIVIS- PCA S1P1-S1P2-BluP1
[illegible]	MIVIS- PCA R1-G1-B1
[illegible]	MIVIS- DVI
[illegible]	MIVIS- NDVI
[illegible]	MIVIS- MSAVI2
[illegible]	MIVIS- SLI
[illegible]	MIVIS- TCAP Wet
[illegible]	MIVIS- TCAP (PCA) Wet
[illegible]	MIVIS- TCAP Green
[illegible]	MIVIS- TIR b96
[illegible]	MIVIS- MIR b57
[illegible]	MIVIS- NIR b18
20	MIVIS- Red b13
21	MIVIS- Green b7
22	MIVIS- Blue b3
[illegible]	MIVIS- NIR PCA
[illegible]	MIVIS- Red PCA
[illegible]	MIVIS- Green PCA
[illegible]	MIVIS- Blue PCA
[illegible]	MIVIS- False Color
[illegible]	MIVIS- TIR PCA
[illegible]	MIVIS- S4P1-b18-b13
[illegible]	MIVIS- TCAP Br-Wet-Gr

Cross Process Coverage Table

253	103	100	98	41	77	54	53	64	50	55	45	38	40	38	37	33	24	36	39	28	31	39	36	22	24	17	16	14	14	12
	132	63	81	16	78	64	59	63	47	43	41	28	42	42	41	37	27	38	39	25	31	34	33	21	24	23	15	14	13	11
		124	80	16	67	47	43	46	47	48	42	44	39	38	37	36	31	34	32	31	35	30	26	25	23	23	18	14	14	13
			114	18	70	62	55	65	48	52	46	37	46	48	43	39	32	31	40	28	37	36	34	29	23	19	17	12	14	13
				100	10	11	10	13	10	14	8	7	11	8	5	5	6	7	7	6	8	7	7	6	7	5	6	4	3	4
					99	46	51	50	46	42	44	28	37	40	40	38	27	34	38	23	28	31	29	19	23	18	14	17	12	10
						69	50	60	33	38	28	22	51	48	35	34	28	30	35	10	34	32	32	21	18	18	18	12	13	12
							68	51	32	37	30	17	30	45	39	36	25	33	43	18	29	36	34	20	18	17	15	13	14	13
								65	33	38	30	20	42	45	41	36	27	33	42	21	32	35	32	20	19	21	17	14	14	13
									60	32	29	17	28	28	25	22	13	39	25	13	22	25	23	16	15	15	13	12	12	10
										59	27	20	31	30	25	25	15	23	32	18	26	32	31	16	16	16	16	13	13	12
											55	18	28	28	28	24	16	19	28	14	19	22	19	12	19	13	11	17	9	8
												54	21	17	20	18	30	10	15	34	21	12	12	23	9	13	12	4	8	8
													52	38	30	27	23	23	31	18	28	27	28	16	15	16	16	12	12	12
														51	37	32	26	26	34	18	30	27	27	29	15	18	14	10	13	12
															50	43	30	24	32	22	20	23	21	24	12	19	13	10	12	10
																48	26	22	29	18	25	21	19	20	14	18	11	11	10	9
																	47	11	18	31	22	12	10	30	8	14	10	4	6	6
																		45	28	10	20	25	22	12	13	16	12	12	12	10
																			46	12	24	32	30	15	18	14	14	12	12	11
																				43	18	12	10	25	10	14	9	5	5	4
																					40	21	20	23	12	18	14	10	14	13
																						40	35	12	14	10	13	12	13	11
																							37	11	14	10	14	11	13	12
																								36	10	13	11	5	6	5
																									32	9	9	17	6	5
																										28	11	8	8	8
																											20	5	9	9
																												20	5	4
																													14	12
																														13

Table 3 – (left) Total number of traces for each process including unique recognitions. -Cross process coverage table (right). ID process number (center).

The Tasseled Cap appears to be one of the most useful processes to identify features, if used for producing composites of different images: the Tasseled Cap processes employing combinations of images in RGB display are in fact localizable in the upper part of the index, while the single components of the same process are located in lower part of the index. Also, in this index the Near Infrared band or the processing involving its use are placed in the middle-upper level, as it was in the Visibility Index Table, confirming its good efficiency. Low efficacy is shown from Green bands and the corresponding SPCA; the thermal band 96, seen at the second place in the Visibility Index Table, is here in the lowest levels, showing how, even if being good for highlighting the shape of traces, it is not for detecting the traces themselves on the image. The least effective images are the simple displays of single original bands or the relative SPCA: it is not surprising that in the worst levels the Blue band and the SPCA of Blue can be found, since the blue band has already demonstrated the scarce usefulness, being blurred from atmospheric scattering.

3.3 Coverage between image processes

Another way to compare the results obtainable from the processes is to look at the coverage between processes, comparing the number of traces common to two given processes against the total of one of the processes. This is possible through a table listing the traces found in a process versus the number of those specific traces also found in other processes. For example, 39 of the 253 traces visible in True Colors are also seen in SPC of Red bands: this gives SPC Red about 15% coverage over True Colors. Reversing it, this means that 39 of the 45 traces found in SPC Red are also seen in True Colors, giving True Colors about 87 % coverage over SPC Red (Table 3, center and right).

4. CONCLUSIONS

At the beginning of the research problem the issue of a combinatorial explosion of the produced images and obtained detected traces was not sufficiently estimated. This led, in the progress of the work, to the definition of a complex GIS architecture able not only to manage archaeological and topographical data through overlay but also to support the selection of valid image processes delivering useful and original information. The GIS has played a critical role in reducing the processing time by indicating in the first stage the successful techniques that should be adopted. It is evident however that the obtained results can only be considered valid in relation to the used data and that new tests must be performed when using different data and examining environments with different characteristics from the ones here presented.

5. REFERENCES

Traviglia, A. 2005. A semi-empirical index for estimating soil moisture from MIVIS data to identify subsurface archaeological sites, in *9° Conferenza Nazionale ASITA, Federazione delle Associazioni Scientifiche per le Informazioni Territoriali ed Ambientali, 15-18 Novembre 2005, Catania, Atti Volume II.* Catania, ASITA, 1969-197.

Traviglia, A. 2006. Archaeological usability of Hyperspectral images: successes and failures of image processing techniques, in Campana S., Forte M. (Eds.), *From Space to Place. Proceedings of the 2nd International conference on Remote Sensing in Archaeology*, *Rome, Dec. 4-7 2006*. Oxford, Archaeopress, 123-130.

Traviglia, A. 2007. MIVIS Hyperspectral Sensors for the Detection and GIS Supported Interpretation of Subsoil Archaeological Sites, in Clark J.T., Hagemeister E. (Eds.), *Digital Discovery: Exploring New Frontiers in Human Heritage. CAA 2006. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 34th Conference, Fargo, United States, April 2006,* Budapest, Archaeolingua, 287-299.