ARTIFICIAL NEURAL NETWORKS BASED AUTOMATIC LINEAR AND AREA FEATURE EXTRACTION FROM WORLDVIEW-02 SATELLITE IMAGES FOR CADASTRAL DATA COLLECTION; A CASE STUDY IN BELIHULOYA, SRI LANKA

Kaluarachchige Manjula Chathuranga[1], Adikari Mudiyanselage Janith Prasanna Adikari[1] and Thilantha Lakmal Dammalage[2]

[1]Survey Department of Sri Lanka, No.150, Kirula road, Narahenpita, Colombo 05, Sri Lanka,
Email: manjulakamc@gmail.com, adikari8711@gmail.com
[2]Sabaragamuwa University of Sri Lanka, P.O. Box 02, Belihuloya, 70140, Sri Lanka,
Email: thilantha@geo.sab.ac.lk

KEY WORDS: Cadaster, Geographical Information System, Learning Vector Quantization, Pattern Recognition

ABSTRACT: Cadastral Surveying deals with ownership, boundaries, extent, and value of land or property in a given area, mainly for the purpose of taxation. Traditional cadastral surveying approaches are always time consuming and require lot of efforts in field surveying especially in remote and mountainous areas. The photogrammetric technique is a possible solution for this. However, up-to-date aerial photographs are hard to obtain when compare to the high resolution satellite images. Therefore, in this study Worldview-02 satellite images were used for spatial data acquisition due to its resolution, coverage and continuous data availability. The implementation of automatic feature extraction from satellite images to collect some of the possible land features would definitely reduce the time and cost of completing cadastral mapping. There are a number of methods to extract the features from satellite images, but most of them are semi-automatic and cannot be applied for both linear and area feature extraction. Therefore, an Artificial Neural Network (ANN) based system has been developed to extract both liner and area features automatically from satellite images with providing up to date spatial data for cadastral data collection. The proposed system is developed based on two ANNs; Learning Vector Quantization (LVQ) and Pattern Recognition (PR) to extract stream, road, lake, and building features from worldview-02 satellite images automatically. The system is tested for worldview-02 satellite images in the surrounding area of Belihuloya, Sri Lanka. The system has achieved the performance for LVQ ANN as average completeness: 85.99 %, correctness: 92.29 % and quality: 80.68 % and PR ANN as average completeness: 79.57 %, correctness: 95.14 % and quality: 76.74 %. The results show the possibility of using the proposed approach for automatic liner and area feature extraction from worldview-02 satellite images for cadaster data collection.

1. INTRODUCTION

1.1 Background

Surveying may be defined as the science of determining the position, in three dimensions, of natural and man-made features on or beneath the surface of the Earth (Rusu & Musat, 2012). There are six different types of surveying (Seedat et al., 2012) and one of them is Cadastral Surveying. The name Cadaster is a Latin base term which refers to an official register of the ownership, extent, and value of real property in a given area, used as a basis for taxation (Kark, 2008 and Seedat et al., 2012). Cadastral Surveying is surveying of land so as to determine and define land ownership and boundaries (Seedat et al., 2012).

Traditionally, the cadastral map was prepared using plane tabling, sight rule, optical square, chain or steel measuring tapes as a ground survey (Rao et al., 2014). Then, high accurate cadastral surveys are performed by total station traverse (Onkalo, 2006). Traditional field surveying approaches were time consuming, very expensive and require lot of efforts as well as it was very difficult in remote areas especially in mountainous areas when the weather is very harsh (Onkalo, 2006 and Ali & Ahmed, 2013). Then, cadastral surveys were performed by aerial photographs and past photogrammetry techniques (e.g. aerial triangulation, stereo orientations, and photo-interpretation, stereo-digitizing on stereo- model) were commonly used to produce geo-information (Tuladhar, 2005). The photogrammetry was not sufficient in cadastral surveying because inapplicable for large area, difficult to maintain the equipment, there are numbers of steps involved in the processing, expert are needed and time consuming due to manual digitization (Tuladhar, 2005). In this case satellite images were used as an alternative approach for spatial data acquisition (Ali & Ahmed, 2013) and it is beneficial for cadastral surveying according to larger coverage, continuous capturing, and quick processing (Tuladhar, 2005).

Automatic feature extraction has been an active research topic in the field of digital photogrammetry and computer vision for many years (Lari & Ebadi, 2007). There is no method for fully automatic feature extraction from images, there may be at least little human interaction in all the current feature extraction methods (Wijesingha et al., 2013). There are a number of methods to extract the features from the satellite images for cadastral surveying, but most of them are semi-automatic and cannot be applied for both linear and area feature extraction. A method should be developed to extract both liner and area features automatically from satellite images. Feature extraction from satellite images using ANN is currently ongoing research for reducing human input from the feature extraction (Wijesingha et al., 2013). Therefore, ANN based system should be developed to extract both liner and area features automatically from satellite images supplying up to date spatial data for cadastral surveying.

The primary objective of the study was to develop an ANN based system for extracting both linear and area features automatically from worldview-02 satellite images for cadastral data collection. Specific objectives of the study were to identify the significant linear and area spatial features to collect automatically for cadastral surveying, detect the identified features automatically on satellite image, determine the best ANN and train it for extracting the linear and area features, analyses the performance of the system by validating the accuracy of extracted features and convert the extracted features into useable Geographical Information System (GIS) environment. The study area is surrounding area of Belihuloya village which is located in Rathnapura district of Sri Lanka. The upper left Universal Transverse Mercator (UTM) co-ordinates and lower right UTM coordinates of the study area are 742904 N, 466978 E and 733459 N, 477958 E respectively. The village area has both flat and mountainous terrain with some urban areas where some part of the area is built-up and very dense and covers an area of about 10370 ha. Belihuloya village area where most of the area is open and mostly covered with some forest, consists of all kind of topographic features such as streams, roads, lakes and buildings etc.

1.2 Artificial Neural Network

ANN is a mathematical model that tries to simulate the structure and functionalities of biological neural networks. The biological neural network is collection of biological neurons (Figure 1) and inspired by brains. An ANN is composed of many artificial neurons that are linked together according to specific network architecture (Figure 2). Basic features of an artificial neuron are similar to the real neuron. Inputs are alike to dendrites and output path alike to the axon. Processing element is similar to the cell body. ANNs can be trained to perform complex functions in various fields, pattern recognition, identification, classification, speech, vision, and control systems. ANNs can also be trained to solve problems that are difficult for conventional computers or human beings.

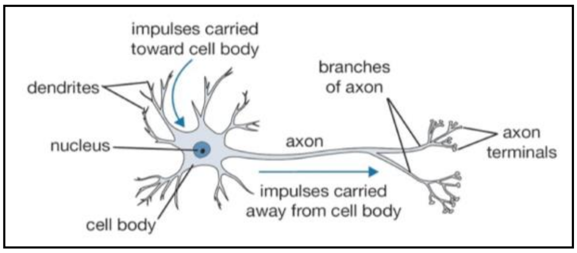

Figure 1: The biological Neuron
Source: http://www.embedded-vision.com

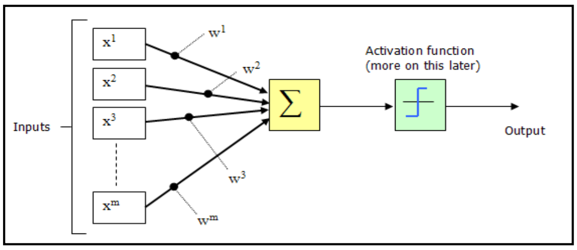

Figure 2: The Artificial Neuron
Source: http://blog.csdn.net

The system was developed based on two ANNs; LVQ and PR. An LVQ network has a first competitive layer and a second linear layer in network architecture (Figure 3). The linear layer transforms the competitive layer's classes into target classifications defined by the user. PR ANN that follow the feed forward network architecture (Figure 4) can be used for multi class classification.

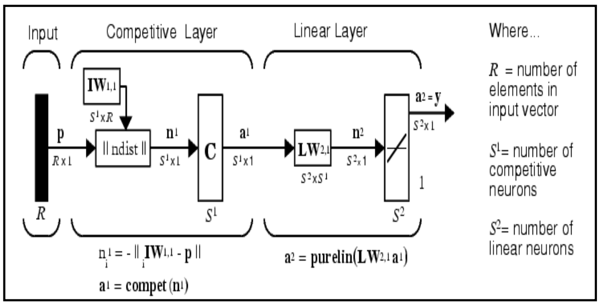

Figure 3: The LVQ network architecture
Source: User guide of NN toolbox in MATLAB

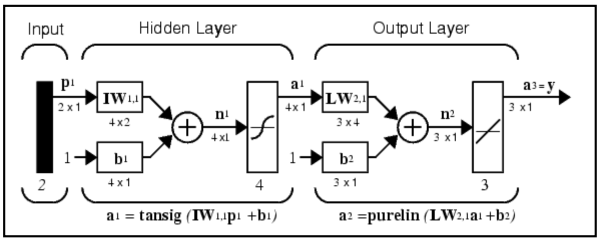

Figure 4: The PR network architecture
Source: User guide of NN toolbox in MATLAB

2. METHODOLOGY

With the help of literature, stream and road features were identified as the significant linear features as well as land parcels, lake and building features were identified as the significant area features in cadastral data collection. But, it is not try to develop the system for extracting land parcels automatically because it is difficult to separate land parcels on satellite images using Digital Numbers (DN) values.

First step of the methodology is the preprocessing of the satellite images. The preprocessing step includes conversion of DN of the satellite images to spectral radiance, conversion of spectral radiance to spectral reflectance, panchromatic sharpening for enhancing the resolution of the images and applying index ratios for enhancing the required features on the satellite images. Normalized Difference Water Index (NDWI) was used to detect areas of standing water and Non-Homogeneous Feature Difference (NHFD) index was used to detect the building features (Wolf, 2010). NDWI is calculated using Eqn (1) and NHFD is calculated using Eqn (2).

NDWI = (Coastal Band – NIR2 Band) / (Coastal Band + NIR2 Band) - Eqn (1)

NHFD = (Red Edge Band – Coastal Band) / (Red Edge Band – Coastal Band) - Eqn (2)

After preprocessing, two type of subsets (Training & Simulate) were selected for each type of features using ArcGIS software (Table 1). Ten training subsets and ten simulate subset for each feature types were used and the best training subset for each feature types were selected based on performance and confusion matrix.

Table 1: Selection of subset

Feature type	Type of Subset	Name of subset	Number of subsets
Streams	Training	S_T_(Number)	10
	Simulate	S_S_(Number)	10
Roads	Training	R_T_(Number)	10
	Simulate	R_S_(Number)	10
Lakes	Training	L_T_(Number)	10
	Simulate	L_S_(Number)	10
Buildings	Training	B_T_(Number)	10
	Simulate	B_S_(Number)	10

Then, LVQ and PR ANNs were defined by applying network parameters: number of hidden neurons, typical class percentage, learning rate and learning function. After defining the networks, the defined networks were trained by applying training subsets, target classes and training parameters such as number of epochs and training time. The trained networks were used to simulate ten simulate subset and relevant features were extracted automatically. As the next step, extracted features are vectorized automatically to get the vector layer.

Then, evaluation parameters: completeness, correctness and quality were used to check the accuracy of developed system (Sujatha & Selvathi, 2015). These parameters can be calculated using Eqn (3), Eqn (4) and Eqn (5) respectively. The features digitized manually from simulate subsets, were used as the reference data. Finally, the vector layers of extracted features automatically, were converted into shape file format automatically as the final output.

Completeness = (Area of matched extraction) / (Area of reference)
= (TP) / (TP + FN) - Eqn (3)

Correctness = (Area of matched extraction) / (Area of extraction)
= (TP) / (TP + FP) - Eqn (4)

Quality = (Area of matched extraction) / [Area of (extraction + unmatched reference)]
= (TP) / (TP + FN + FN) - Eqn (5)

Where: TP: True Positive (Matched extracted data), FN: False Negative (Unmatched reference data) and FP: False Positive (Unmatched extracted data)

3. RESULT AND DISCUSSION

3.1 Automatic stream features extraction

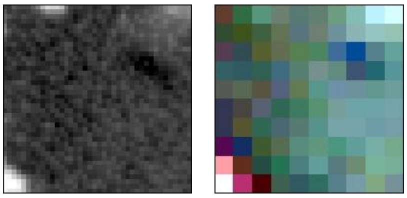

Figure 5: Best stream training subset (Panchromatic and Multispectral)

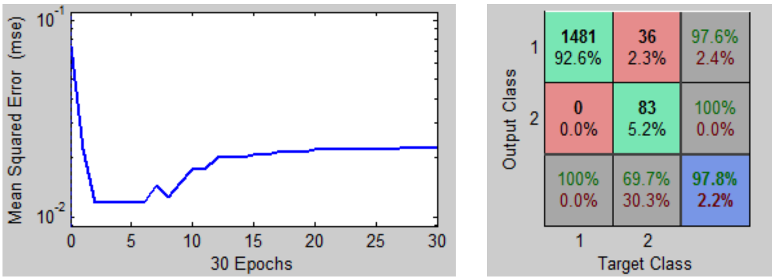

Figure 6: Performance and confusion matrix of best stream training subset

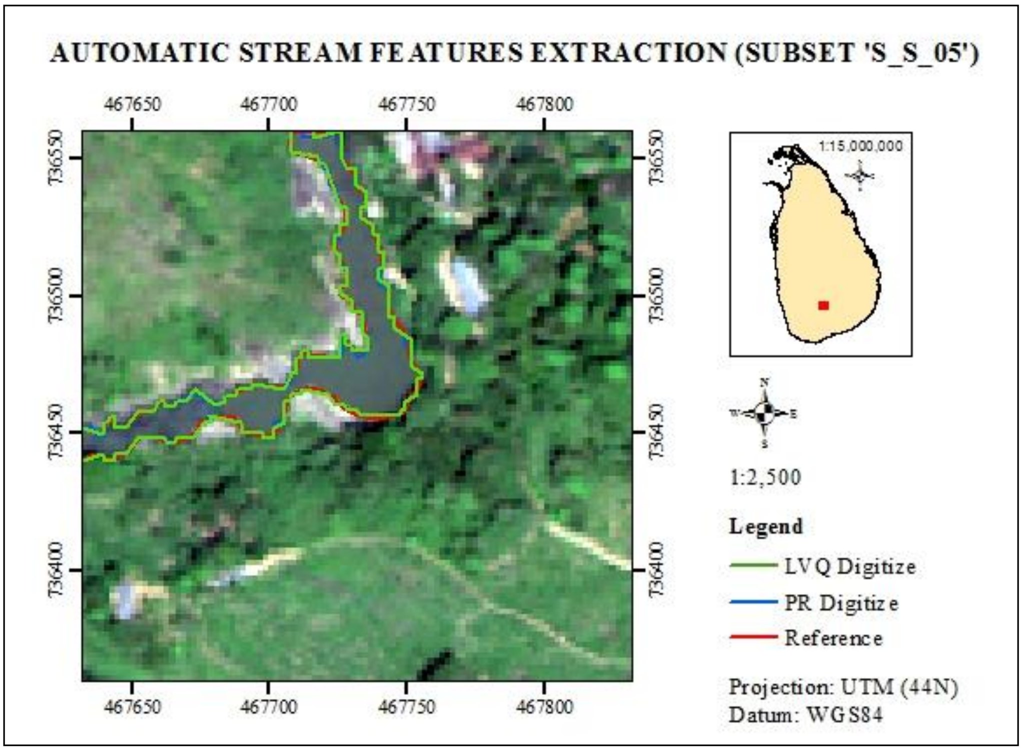

Figure 7: The extracted stream features from best simulate subset

Table 2: The accuracy of networks for stream simulate subsets

Subset	Completeness (%)		Correctness (%)		Quality (%)	
	LVQ	PR	LVQ	PR	LVQ	PR
S-S-01	73.75	52.59	93.70	95.59	70.27	51.35
S-S-02	80.96	72.27	96.57	97.73	78.69	71.08
S-S-03	75.72	70.86	89.60	92.25	69.60	66.88
S-S-04	85.08	80.88	96.36	97.28	82.43	79.09
S-S-05	96.30	95.12	95.52	96.34	92.14	91.80
S-S-06	86.66	83.38	91.86	92.77	80.48	78.30
S-S-07	94.25	92.41	95.69	96.77	90.41	89.64
S-S-08	68.74	66.00	98.04	98.61	67.81	65.40
S-S-09	95.45	93.35	94.60	96.46	90.52	90.26
S-S-10	85.80	84.87	95.68	96.80	82.60	82.56

3.2 Automatic road features extraction

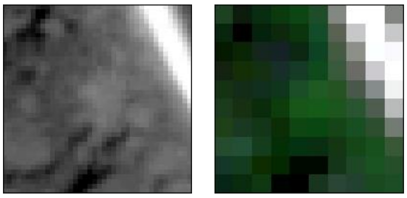

Figure 8: Best road training subset (Panchromatic and Multispectral)

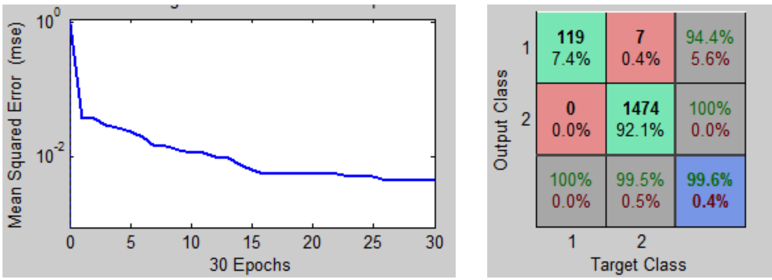

Figure 9: Performance and confusion matrix of best road training subset

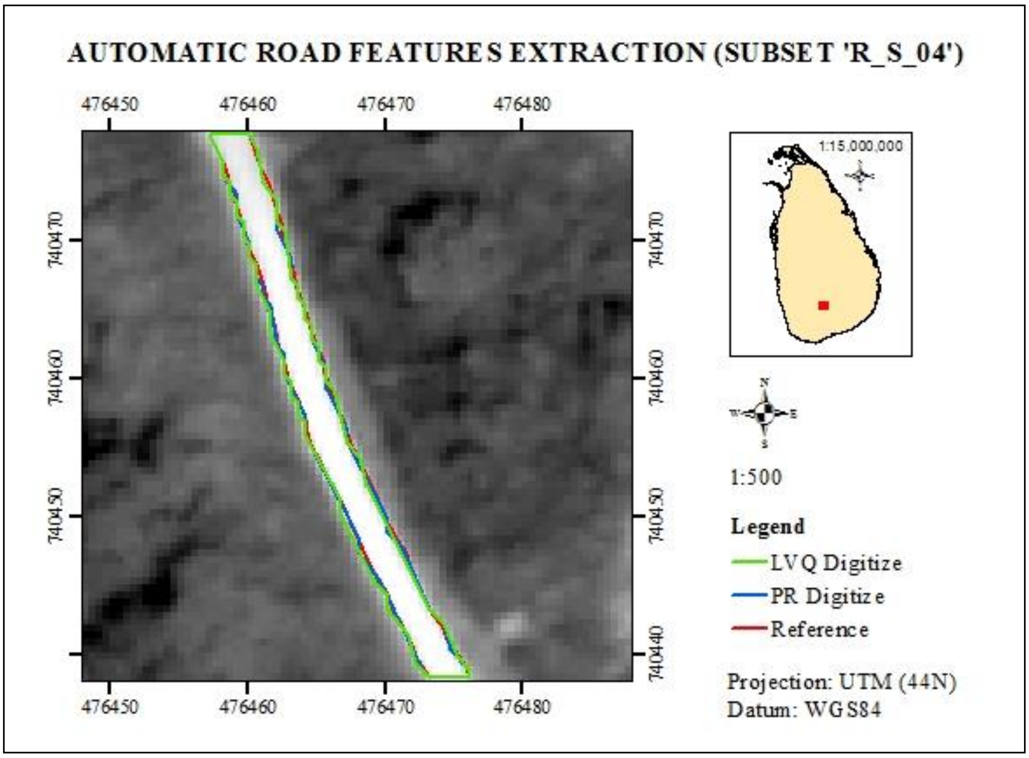

Figure 10: The extracted road features from best simulate subset

Table 3: The accuracy of networks for road simulate subsets

Subset	Completeness (%)		Correctness (%)		Quality (%)	
	LVQ	PR	LVQ	PR	LVQ	PR
R-S-01	80.44	73.59	96.30	98.72	78.02	72.89
R-S-02	96.90	93.50	92.00	94.55	89.37	88.73
R-S-03	82.13	80.35	95.68	97.12	79.19	78.48
R-S-04	96.81	95.69	93.03	95.98	90.26	92.00
R-S-05	78.04	74.66	88.77	90.03	71.03	68.96
R-S-06	92.73	87.18	93.91	97.25	87.46	85.08
R-S-07	89.48	86.97	73.26	77.14	67.45	69.15
R-S-08	27.43	20.27	95.82	98.53	27.11	20.21
R-S-09	88.31	83.50	86.25	88.31	77.41	75.19
R-S-10	84.23	77.60	93.81	96.90	79.80	75.72

3.3 Automatic lake features extraction

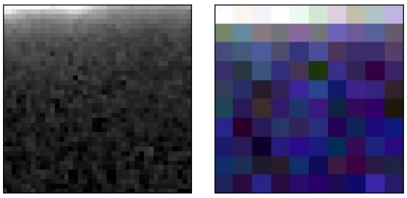

Figure 11: Best lake training subset (Panchromatic and Multispectral)

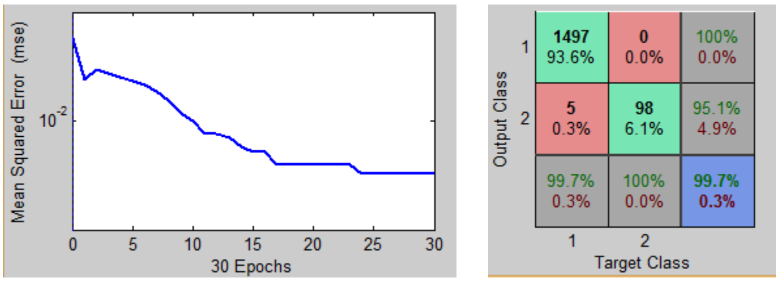

Figure 12: Performance and confusion matrix of best lake training subset

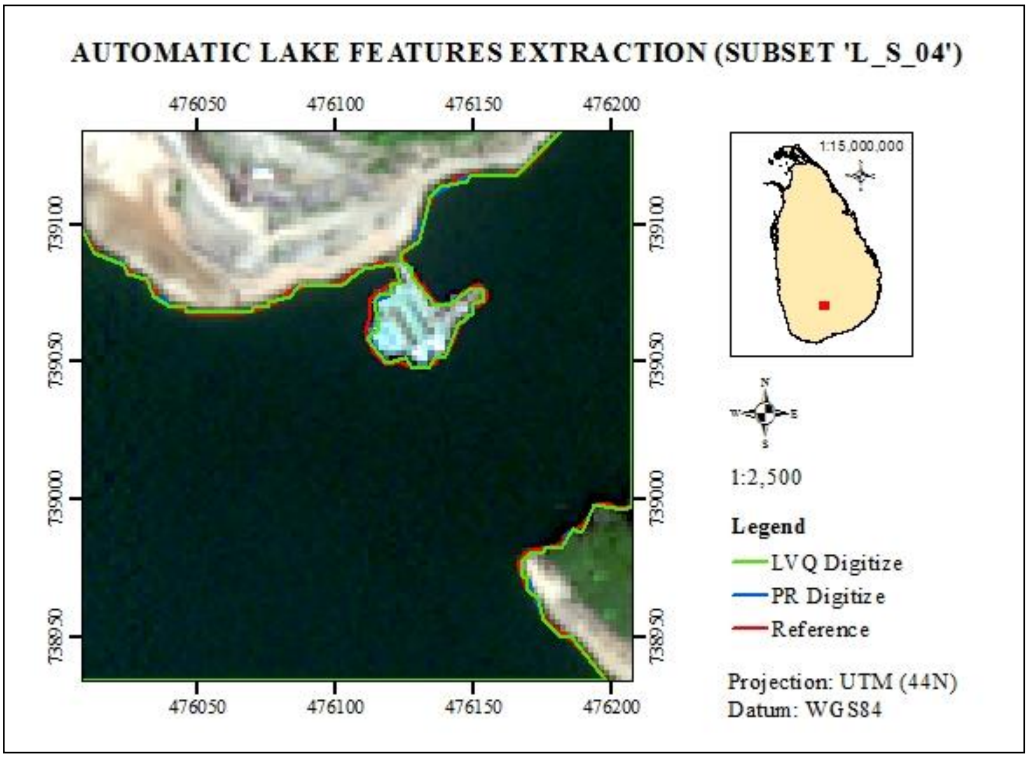

Figure 13: The extracted lake features from best simulate subset

Table 4: The accuracy of networks for lake simulate subsets

Subset	Completeness (%)		Correctness (%)		Quality (%)	
	LVQ	PR	LVQ	PR	LVQ	PR
L-S-01	99.76	99.76	99.06	99.01	98.83	98.78
L-S-02	99.79	99.80	99.72	99.67	99.52	99.47
L-S-03	99.76	99.87	99.62	99.55	99.39	99.42
L-S-04	99.72	99.75	99.32	99.28	99.05	99.03
L-S-05	99.89	99.92	99.83	99.82	99.72	99.75
L-S-06	99.83	99.91	99.67	99.62	99.50	99.54
L-S-07	99.28	99.34	99.83	99.85	99.11	99.20
L-S-08	99.69	99.76	99.72	99.69	99.41	99.45
L-S-09	99.85	99.84	99.77	99.74	99.62	99.58
L-S-10	99.75	99.80	99.87	99.83	99.62	99.63

3.4 Automatic building features extraction

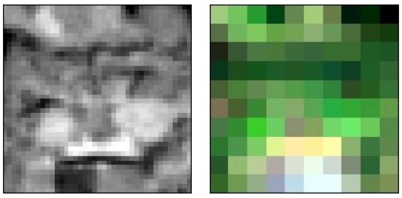

Figure 14: Best building training subset (Panchromatic and Multispectral)

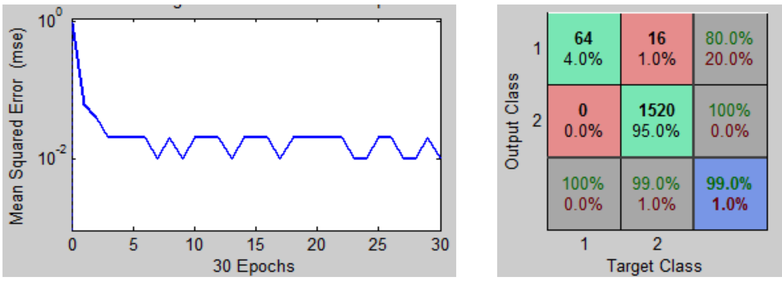

Figure 15: Performance & confusion matrix of best building training subset

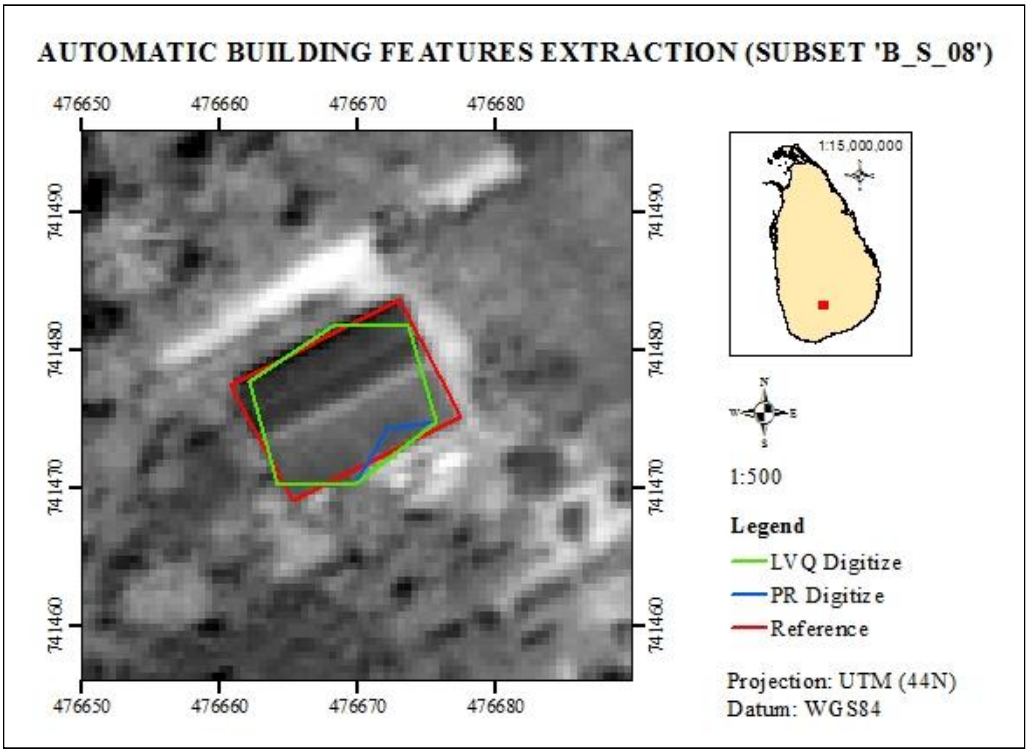

Figure 16: The extracted building features from best simulate subset

Table 5: The accuracy of networks for building simulate subsets

Subset	Completeness (%)		Correctness (%)		Quality (%)	
	LVQ	PR	LVQ	PR	LVQ	PR
B-S-01	82.03	72.95	86.70	91.79	72.86	68.48
B-S-02	78.10	64.87	89.89	95.73	71.79	63.05
B-S-03	82.81	52.51	72.18	82.26	62.77	47.17
B-S-04	71.84	57.05	57.82	68.02	47.14	44.98
B-S-05	73.53	51.69	94.33	100.0	70.42	51.69
B-S-06	47.19	18.07	81.02	100.0	42.49	18.07
B-S-07	86.76	70.88	88.65	97.35	78.09	69.54
B-S-08	84.08	80.57	95.78	97.15	81.08	78.71
B-S-09	87.56	76.23	83.12	90.66	74.34	70.68
B-S-10	89.20	75.35	89.05	91.52	80.38	70.43

3.5 The overall accuracy of features extraction

Table 6: The overall accuracy of networks for each feature types

Feature Type	Statistic	Completeness (%)		Correctness (%)		Quality (%)	
		LVQ	PR	LVQ	PR	LVQ	PR
Stream	AVG	84.271	79.173	94.762	96.060	80.495	76.636
	SD	9.510	13.665	2.474	2.045	9.006	12.997
	MIN	68.740	52.590	89.600	92.250	67.810	51.350
	MAX	96.300	95.120	98.040	98.610	92.140	91.800
	RANGE	27.560	42.530	8.440	6.360	24.330	40.450
Road	AVG	81.650	77.331	90.883	93.453	74.710	72.641
	SD	20.144	21.377	6.967	6.709	18.305	20.037
	MIN	27.430	20.270	73.260	77.140	27.110	20.210
	MAX	96.900	95.690	96.300	98.720	90.260	92.000
	RANGE	69.470	75.420	23.040	21.580	63.150	71.790
Lake	AVG	99.732	99.775	99.641	99.606	99.377	99.385
	SD	0.170	0.165	0.257	0.269	0.289	0.297
	MIN	99.280	99.340	99.060	99.010	98.830	98.780
	MAX	99.890	99.920	99.870	99.850	99.720	99.750
	RANGE	0.610	0.580	0.810	0.840	0.890	0.970
Building	AVG	78.310	62.017	83.854	91.448	68.136	58.280
	SD	12.365	18.509	11.398	9.809	13.430	18.052
	MIN	47.190	18.070	57.820	68.020	42.490	18.070
	MAX	89.200	80.570	95.780	100.000	81.080	78.710
	RANGE	42.010	62.500	37.960	31.980	38.590	60.640

Completeness and quality of automatic feature extraction using LVQ ANN is better than PR ANN for stream, road and building features but PR ANN is better than LVQ ANN for lake features extraction. Correctness of automatic feature extraction using PR ANN is better than LVQ ANN for streams, roads and buildings features but LVQ ANN is better than PR ANN for lake features extraction. However, automatic building feature extraction is poor in accuracy relative to the other features extraction.

4. CONCLUSIONS AND RECOMMENDATIONS

The study was intended to develop an ANN based system for extracting both significant linear and area features automatically from worldview-02 satellite images for cadastral data collection. Therefore, LVQ and PR ANNs based system was developed to extract both linear and area features automatically from worldview-02 satellite images providing up to date cadastral data. Especially for streams, road and lake boundaries, the result suggested it worked well.

According to the obtained results, it is recommended that ANNs are appropriate for automation as well as LVQ and PR ANNs are suitable for automatic feature extraction from worldview-02 satellite images. In addition, worldview-02 satellite images are respectable data source for collecting cadastral data. But available resolution of the satellite images are not enough. Therefore, panchromatic sharpening method was used to create high spatial and spectral resolution images and it is recommended that panchromatic sharpening method is significant for enhancing the resolution of worldview-02 satellite images.

First, DN values of the satellite images were converted into reflectance value. It is recommended that reflectance value should be used to calculate index ratios. In this study, two index ratios: NDWI and NHFD were applied. NDWI is very significant index for enhancing water area on worldview-02 satellite images as well as NHFD is significant index for enhancing man-made features on worldview-02 satellite images. Further, the used generalization technique: Douglas-Peucker line simplification algorithm, is suitable for line simplification. In addition, the system was not developed to extract land parcels from the satellite images. Finally, it is concluded that combination of the develop system and conventional methods facilitate more in cadastral data collection.

It is recommended following developments further;

- ✓ The system should be developed to extract the land parcels automatically from satellite images
- ✓ Suitable generalization techniques: filtering and smoothing, should be applied further
- ✓ The system can be developed for multi class classifications
- ✓ Different ANNs should be analyzed further
- ✓ Common MATLAB interface should be developed for extracting all type of features

5. REFERENCES

Ali, Z., & Ahmed, S. (2013). Extracting parcel boundaries from satellite imagery for a land information system. In RAST 2013 - Proceedings of 6th International Conference on Recent Advances in Space Technologies (pp. 79–81). http://doi.org/10.1109/RAST.2013.6581319

Kark, R. (2008). Mamluk and Ottoman Cadastral Surveys and Early Mapping of Landed Properties in Palestine, 71(1), 46–70.

Lari, Z., & Ebadi, H. (2007). Automatic Extraction of Building Features from High Resolution Satellite Images Using Artificial Neural Networks. In Dept. of photogrammetry & remote sensing, Geodesy and Geomatics Engineering Faculty.

Onkalo, P. (2006). Cadastral Survey Methodologies and Techniques in Developing Countries ; Case Cambodia and Kosovo (pp. 1–10). XXIII FIG Congress, Munich, Germany.

Rao, S. S., Sharma, J. R., Rajasekhar, S. S., Rao, D. S. P., Arepalli, A., Arora, V., … Kanaparthi. (2014). Assessing Usefulness of High-Resolution Satellite Imagery (HRSI) for Re-Survey of Cadastral Maps. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-8.

Rusu, G., & Musat, C. C. (2012). Modern Concepts of 3d Modeling in Geodetic Works (pp. 137–146). University of Alba Iulia.

Seedat, M., Thusi, S., Zaca, M., Mnembe, L., & Ndaba, N. (2012). Different Methods of Surveying. Retrieved from http://www.indabook.org/d/Different-methods-of-Surveying-Final-Durban.pdf

Sujatha, C., & Selvathi, D. (2015). Connected component-based technique for automatic extraction of road centerline in high resolution satellite images. EURASIP Journal on Image and Video Processing (2015). http://doi.org/10.1186/s13640-015-0062-9

Tuladhar, A. M. (2005). Innovative use of Remote Sensing Images for Pro Poor Land Management. In UNESCAP-in Bangkok, Thailand.

Wijesingha, J. S. ., Kumara, R. W. D. M., Kajanthan, P., Koswatte, R. M. K. G. S. P. B., & Bandara, K. R. M. . (2013). An Automatic Road Feature Extraction from High Resolution Satellite Images using LVQ Neural Network. ASIAN JOURNAL OF GEOINFORMATICS, VOL13(No1).

Wolf, A. (2010). Using WorldView 2 Vis-NIR MSI Imagery to Support Land Mapping and Feature Extraction Using Normalized Difference Index Ratios . Digital Globe.