CHAPTER 7

HYBRID CLASSIFIER RESULTS AND APPLICATIONS

In this chapter, a new powerful Multiple Classifier System (MCS) has been developed to combine the results obtained in Chapter 6, for the three classifiers (SOM, CT, and SVM), based on the D-S theory of evidence. The three Dempster-Shafer models, *D-S1, D-S2 and D-S3*, described in Chapter 4 were run in this combination process. For the *D-S1* approach, the class memberships for each classifier were directly used as probability masses. For the *D-S2* approach, the class memberships were first weighted according to the classifiers' performances for each class, and then used as probability masses. For *D-S3* approach, beside the weighting process, the amount of belief from a given classifier output that must be discarded i.e. assigned to the ignorance hypothesis was taken into account. After combining all evidences by means of the *D-S* algorithm, results were obtained in the form of layers that defined the degree of belief or probability of each pixel belonging to each of the hypotheses or classification categories (buildings, trees, roads and ground). A land cover classification map was then obtained by assigning each pixel to the category that was the most probable after the three classifiers had been combined. Figure 7.1 summarizes the steps of *D-S*-based decision fusion and land cover map production. The resulting images for the UNSW, Bathurst, Fairfield and Memmingen test areas are given in Figure 7.2 (*a*) – (*d*).

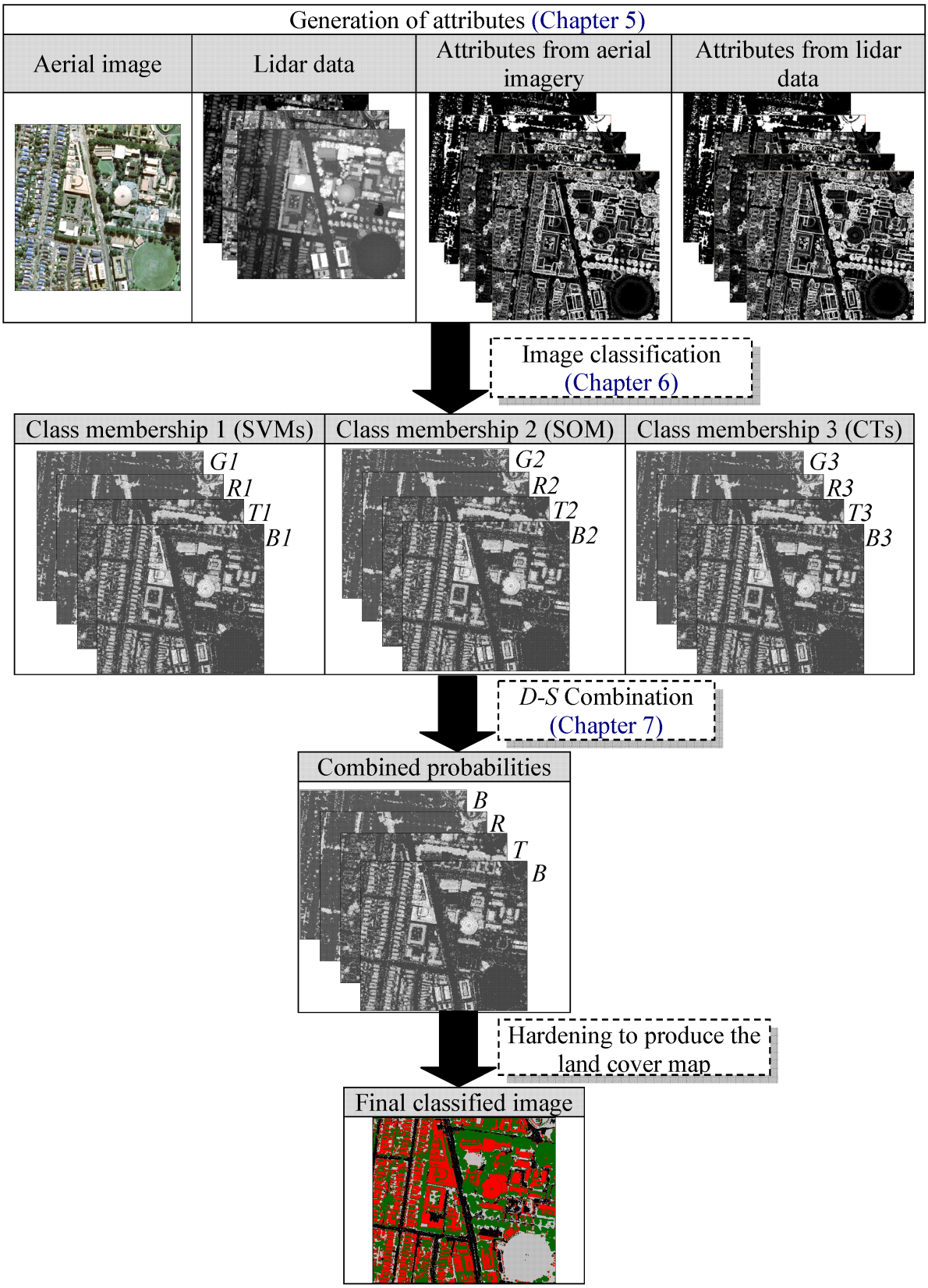

Figure 7.1: Summary of *D-S*-based decision fusion and land cover map production.

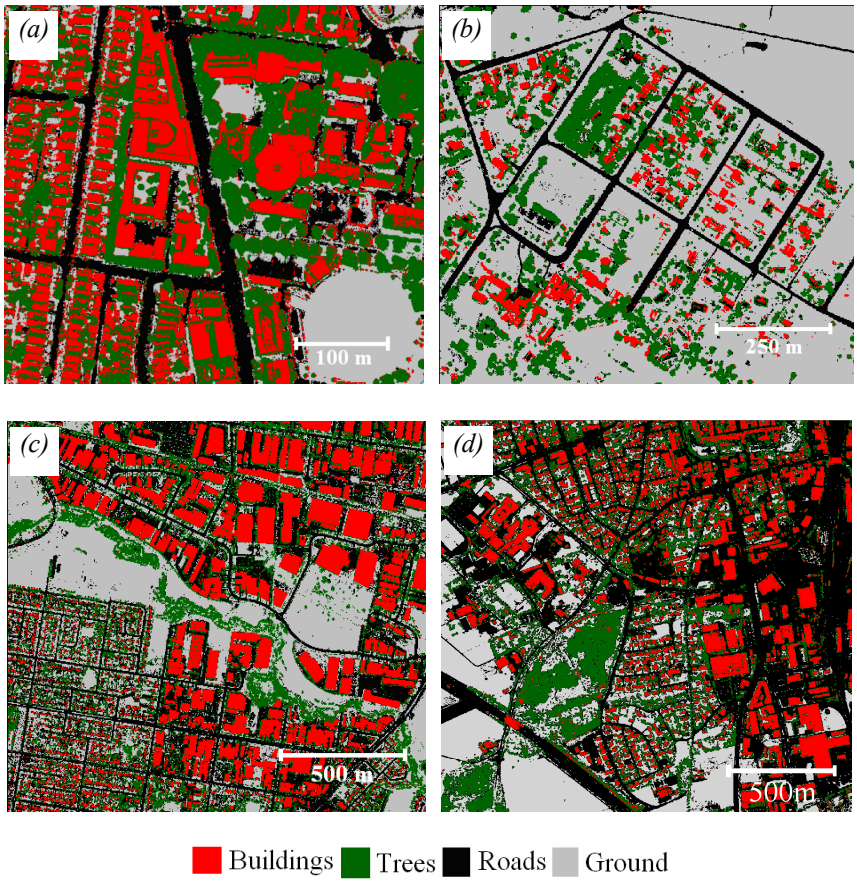

Figure 7.2: *D-S*-based classifier combination results for: (a) UNSW; and (b) Bathurst; (c) Fairfield; and (d) Memmingen.

The results were evaluated and compared to: (1) the results of the individual classifiers; (2) the results with some of the existing fusion rules. Finally and as an application of the combined classification results, the classified images have then been processed through a series of image processing techniques to produce the digital vector map. Building and roads as the most important components of large scale maps were intensively evaluated. In order to evaluate the results of extracting buildings, an accurate estimation of extracted buildings is carried out by comparing the reference data against the corresponding buildings derived from the classified image after the process of raster to vector conversion. In this case, the comparison is based on completeness and correctness. Also, estimates of the planimetric accuracy have been determined of the vectorized building outlines, for the UNSW test area, based on field-surveyed GPS positions. On the other hand, the final produced vector roads have been compared to the reference data in terms of completeness, correctness, mean, SD of the extracted centerlines and the widths.

7.1 COMPARISON WITH EXISTING FUSION ALGORITHMS

The overall classification accuracies of individual classifiers, based on the reference data, were evaluated first with the overall accuracy of the best classifier serving as a reference. The three probability combination strategies, discussed in Chapter 4, were also tested and compared to the proposed method with its three alternatives (*D-S1, D-S2 and D-S3*). These strategies include: Maximum Rule (*MR*); Fuzzy Majority Voting (*FMV*); and Weighted Sum (*WS*).

The overall accuracy for each of the classifications can be assessed using the reference data, but since it is a global measure of the performance of the classifications, the two additional measures of commission and omission errors will reveal locations of specific errors arising from the classifications. Unlike the overall classification accuracy, commission and omission errors clearly reveal how the different combined classifiers improve or deteriorate for each individual class. Commission errors are the percent of incorrectly identified pixels associated with a class, and omission errors are the percent of unrecognized pixels that should have identified as belonging to a particular class.

The overall classification accuracies of individual classifiers, based on the reference data, are given in Table 7.1. SVM performs the best with 96.8 percent average overall classification accuracy, followed by SOM and CTs with average overall classification accuracies of 95.5 percent and 93.7 percent respectively. CTs are the worst performing classifier for complex class distributions, since they use only one feature attribute at each node in the classification tree. The overall accuracy of the best classifier served as a reference in the following.

Table 7.1: Performance evaluation of single classifiers for the four test areas.

	Classification accuracy (%)		
Test area	SVM	SOM	CT
UNSW	96.9	96.8	95.05
Bathurst	96.5	95	92.85
Fairfield	97	96.8	96.15
Memmingen	96.6	95	90.75
Mean	96.75	95.9	93.7
SD	0.24	1.04	2.40

The improvement in overall classification accuracies achieved by each combination method compared with the best individual classifier, SVM, was determined as shown in

Figure 7.3. For the four test areas, it is clear that the overall performances of *D-S3* are better than those of the other combination methods. *D-S2* performs slightly better than *FMV*, and both outperform *D-S1*. *MR* resulted in the worst performance followed by the *WS*. It is worth mentioning that even though the *MR* resulted in the worst performance, it still performed better than the best single classifier.

The question still remains as to whether these improvements are statistically significant. In order to answer this question, first, the standard deviation (SD) of the classification accuracies produced by each classifier for the four test areas is determined to express the variability in classification accuracies from the mean as shown in Table 7.1. With only four test areas the estimate of the SD is limited. However, the low standard deviation of 0.24 percent for the SVM results indicates that the spread of the accuracies for the four tests areas is small and hence accuracies tend to be very close to the mean. In the case of SOM and CTs, the higher SD values, 1.04 percent and 2.40 percent respectively, indicate a larger spread of accuracies for the four test areas. This is another reason why SVMs are used as the reference. The SD was then used as a confidence measure in the conclusions on the quality of the accuracies derived by the three classifiers and the combined classifiers. It can be assumed that the reported margin of error is typically about plus/minus twice the standard deviation (a range for an approximately 95 percent confidence interval). For this work a margin of accuracy of 0.72 percent was used, which is three times the standard deviation of the SVMs results, to define improvements in accuracy that are considered statistically significant, as shown by the dashed line in Figure 7.3. Any improvements in classification accuracy more than this dashed horizontal line are deemed to be significant. It can be concluded that the application of *D-S3* results in the most significant improvement in classification accuracy. The improvements achieved by other techniques are either extremely close to the significance value, and therefore considered to be marginally significant, or below the level of significance.

Taking into account the limited room for improvement beyond the 96.9 percent accuracy achieved by the SVM method, since the image and lidar data are subject to errors in the acquisition and image to lidar geographic registration, the best average improvement in classification accuracy of 1.3 percent is obtained from *D-S3* algorithm, followed by 1.1 percent average improvement from *D-S2* algorithm. *FMV* and *D-S1* algorithms give an average improvement of 1.1 percent and 0.9 percent respectively. *MR* resulted in the

worst performance and only improved the results by 0.7 percent followed by *WS* with 0.8 percent average improvement. The results clearly indicate that the performance of *D-S* fusion has been improved twice: firstly through the weighting process; and secondly by taking into account the amount of belief from the output of a given classifier that must be discarded i.e. assigned to the ignorance hypothesis. Before applying these two processes, the traditional *D-S* resulted in lower classification accuracies and comparable with *FMV*.

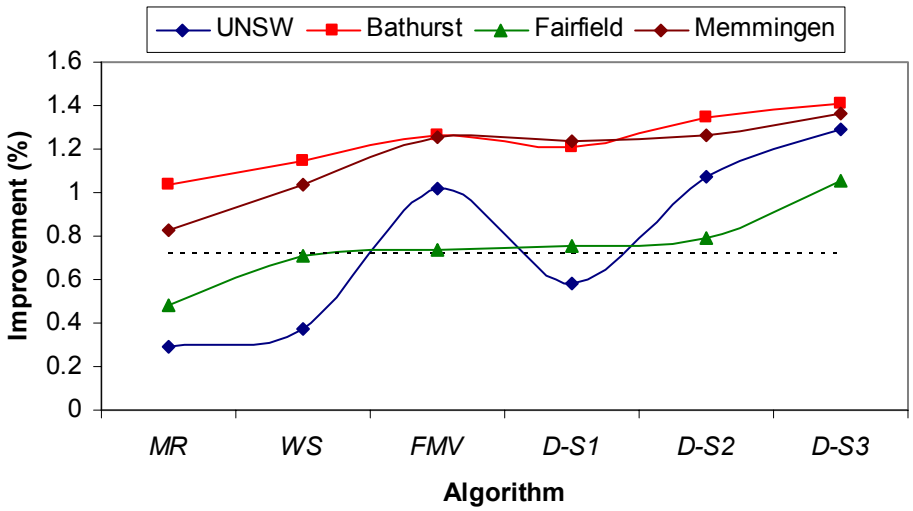

Figure 7.3: Performance comparison of the proposed fusion algorithms with existing fusion algorithms, compared with the performance of the best individual classifier, SVM. Improvements exceeding the dashed horizontal line are considered to be significant.

7.2 ADJUSTING THE CORRECTION FACTOR δ FOR *D-S3* APPROACH

In order to study the relation between the produced overall classification accuracy and the correction factor δ, first, δ in Equation 4.11/Chapter 4 was set to zero, then it was decreased gradually by 0.1 down to -0.9 and the improvements in the overall classification accuracies, compared with the best classifier, were computed and plotted against the corresponding values of δ as shown in Figure 7.4. It was found that adjusting the values of δ can be achieved during the first few iterations, after which no further improvement can be achieved.

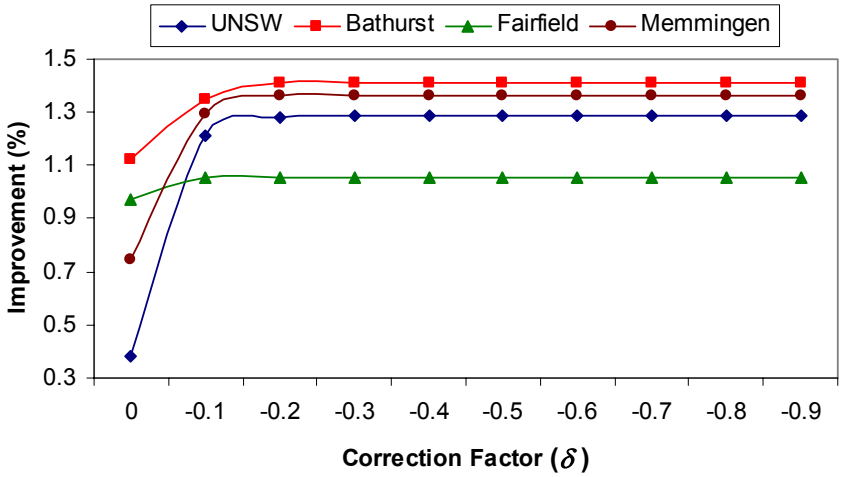

Figure 7.4: Performance of the D-S3 as a function of the correction factor (δ).

7.3 COMPLEMENTARITIES OF CLASSIFIERS

For each data set, classifiers were sorted according to their performance, then results obtained by combining the best two classifiers (SVMs, SOM), were considered, then best and worst ones (SVMs, CTs), the worst two classifiers (SOM, CTs) and finally the whole set of three classifiers (SVMs, SOM and CTs). Since *D-S3* has outperformed other combination methods, as observed in Figure 7.5, it was used as the method for this test. Figure 7.5 shows the improvement in the overall classification accuracy obtained by these combinations with reference to the results of the best single classifier, SVM.

For Fairfield and Memmingen test areas, the combination of the best two classifiers (SVMs, SOM) resulted in the highest accuracy improvement. They were followed by the combinations of the best and worst classifiers (SVMs, CTs), then followed by the combination of all classifiers (SVM, SOM, CTs), and finally the combination of the worst two classifiers (SOM, CTs). For UNSW and Bathurst test areas, except for the combination of the worst two classifiers (SOM, CTs) which resulted in the worst performance, the results of all combination methods are similar. The superiority of the combination of SVMs + SOM classifiers reveals the advantages of combining the best classifiers, or any combination of classifiers which includes the best classifier, SVMs in this case. The SVM classifier brings more complementary information into the solution and results in significant improvements in classification accuracies compared with other combinations. On the other hand, in the case of two-classifier problem, adding more classifiers to the combined system does not guarantee improvements in the performance. This conforms to results obtained in Yu-Chang and Kun-Shan, (2009).

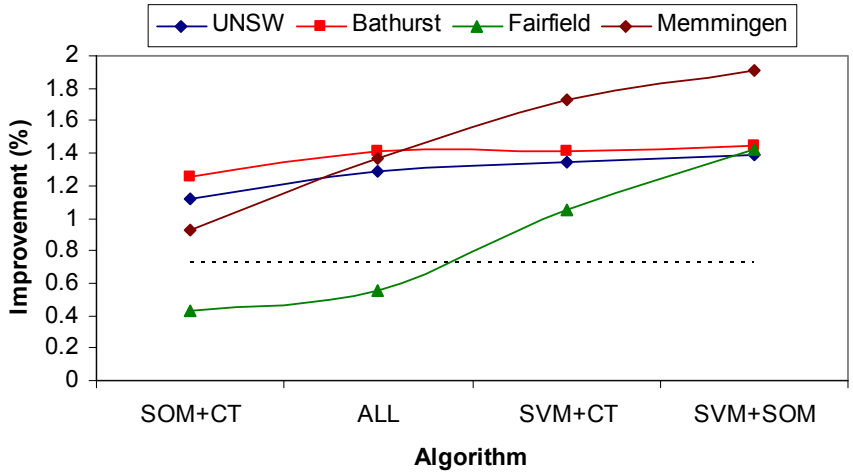

Figure 7.5: Improvements in performance of combinations of classifiers using the fusion algorithm *D-S3*. Results show comparisons with the best performing individual classifier, SVM. Improvements exceeding the dashed horizontal line are considered to be significant.

7.4 PERFORMANCE ANALYSIS OF THE PROPOSED METHOD FOR INDIVIDUAL CLASSES

Table 7.2 shows the commission and omission errors based on the algorithm *D-S3* and the combination of the best two classifiers, compared to the commission and omission errors of the best individual classifier. It can be seen that a considerable amount of the misclassified pixels have been recovered by the combined classification process. For the best classifier, the commission errors vary from 0.06 percent to 10.3 percent, while the omission errors range from 0.06 percent to 21.28 percent respectively. For the combined classifier based on the best two classifiers and *D-S3* algorithm, the commission errors range from 0.11 percent to 5.98 percent, while the omission errors range from 0.08 percent to 4.07 percent respectively. In general, the combination process has significantly reduced the commission and omission errors compared to the best single classifier.

An assessment of the mean commission and omission errors for the four test areas confirms that *DS-3* outperformed SVMs in terms of commission and omission errors. One advantage of *DS-3* combination over SVMs is that the achieved commission and omission errors are less variable as shown in Table 7.2. Whereas the application of SVMs resulted in standard deviation of 3.2% and 5.3% for commission and omission errors respectively, the application of *DS-3* reduces these values to 1.52 % and 1.22 %. In general fusing different classifiers improve commission and omission errors. Moreover the errors are more balanced. This demonstrates the benefit of combining different sensor sources at different classification levels.

Table 7.2: Comparison of errors using the best classifier, SVM, with the classification resulting from the best fusion results, SVM + SOM, for the four test areas. B, T, R and G refer to buildings, trees, roads and grass respectively.

		Best classifier (SVMs)		Best fusion (*DS-3*)	
		Commission (%)	Omission (%)	Commission (%)	Omission (%)
UNSW	B	4.65	2.77	1.50	0.95
	T	3.18	1.97	2.15	1.37
	R	4.81	0.06	0.12	1.51
	G	0.06	5.10	1.55	0.12
Bathurst	B	9.79	7.80	5.98	3.57
	T	0.35	6.12	0.90	0.08
	R	4.36	0.98	1.62	4.07
	G	10.30	4.06	0.40	0.57
Fairfield	B	8.23	11.11	0.99	2.79
	T	0.89	3.36	4.06	1.46
	R	4.08	0.76	0.11	1.61
	G	3.69	7.04	1.65	0.11
Memming.	B	4.04	21.28	1.50	0.95
	T	0.63	3.94	1.37	2.15
	R	4.10	0.42	1.51	0.12
	G	7.96	5.30	0.12	1.55
Mean		4.45	5.13	1.60	1.44
SD		3.22	5.25	1.52	1.22

The visual interpretation of the final classification result clearly shows a relatively high degree of noise in the SVM-based classification results. In contrast to this, the classification that is based on the *DS-3* appears more homogenous. Within the main land cover classes, some pixels in the aerial image are misclassified, whereas these pixels are correctly classified by *DS-3*. Figure 7.6, which is an enlarged portion from Figure 7.2(a), is a typical example showing the results from the SVM classifier in Figure 7.6(*b*), and in Figure 7.6(*c*) the *D-S3* algorithm used to combine classifiers SVMs and SOM for UNSW test area. Sharp building edges in aerial images, appear blurred in SVMs classification results; consequently they are hard to identify. The noise is significantly reduced by combining the two classifiers based on the *DS-3* method and edges appear sharp and clear. This clearly illustrates the different and complementary information provided by the three classifiers.

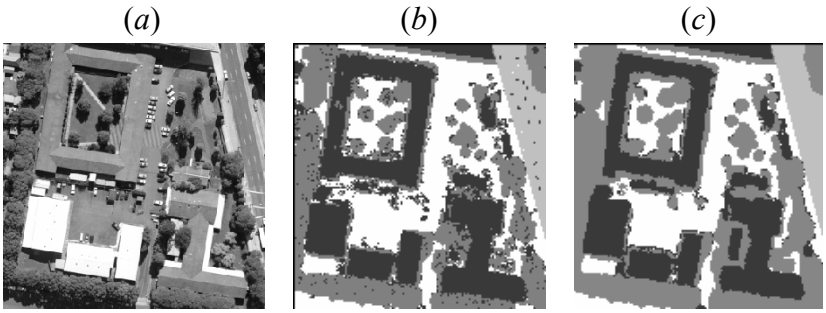

Figure 7.6: (a) Aerial image; (b) classification results of the best classifier (SVMs); (c) Error correction after applying the D-S fusion algorithm. Black: buildings; dark grey: trees; light grey: roads; white: ground.

7.5 COMPUTATIONAL COST

The computational cost involved in implementing the combined classifiers based on the Dempster-Shafer method is much higher than that of other combination methods as shown in Figure 7.7. Most of the time is used to accurately estimate the evidence of each classifier from its corresponding class membership. However, the processing time could be reduced by splitting large test areas into smaller parts, processing each part separately and combining the results later. For example, dividing the Fairfield test area, which is 4km^2 in area, into four equal parts can saved more than 85 percent of processing time required for processing the whole image in the case of using the *D-S3* algorithm. It is also worth mentioning that, using two classifiers instead of three in case of the *D-S3* algorithm reduced the processing time by about 95 percent, while still maintaining the classification accuracy.

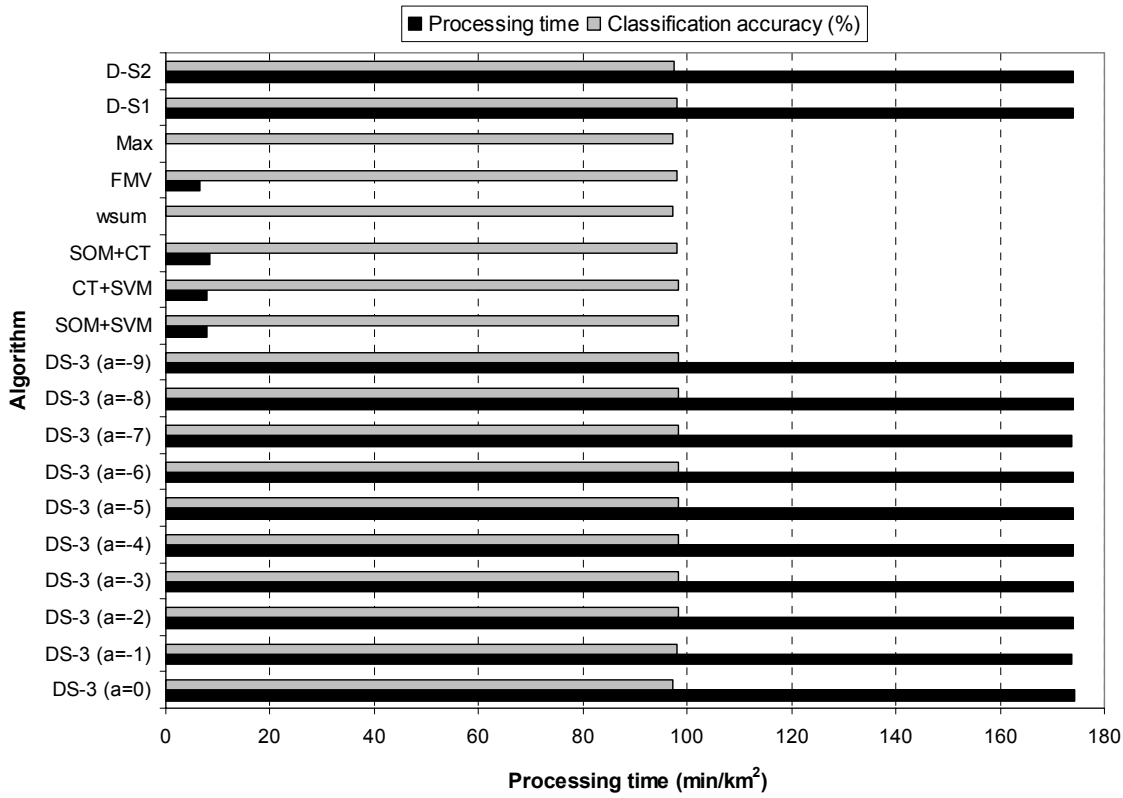

Figure 7.7: Computational cost comparison of D-S based fusion with existing algorithms.

7.6 LINEAR FEATURE EXTRACTION AND GRAPHIC REPRESENTATION

As a result of D-S combination, the hybrid classifier yields land cover maps in the form of pixels. Although these data sometimes are used directly, standard practice is to convert this form of data into vector format that are considerably more useful for the display of digital spatial information and in the computation of spatial information such as coordinates, lengths, directions, perimeters and areas. As an application of the results of the D-S combination classifications, this chapter describes a method of extracting the final vector maps. While the process have been evaluated and presented for the four test areas, only the resulting images of processing buildings and roads for the UNSW test area are presented as a typical example in this chapter. The results of processing buildings and roads in for the Bathurst, Fairfield and Memmingen test areas can be found in Appendix C.

7.6.1 Building Extraction

7.6.1.1 Post Classification Smoothing

In order to extract buildings from the classified data, the classified image is converted to a binary image, with pixels representing buildings displayed with one, and non-building pixels (roads, trees and ground) as zeros. Then, the smaller raster homogeneous building regions were merged into the larger neighboring homogeneous regions or deleted according to an arbitrary 1m distance and 30m^2 area threshold respectively. The area threshold represents the expected minimum building area that can be reliably extracted, while the distance threshold was set to 1m to fill in any gaps produced through the classification process. Building regions were retained if at least one of two conditions was satisfied: (1) the building region was larger than the predefined area threshold; and/or (2) more than 1m distance from a larger homogeneous region. As a last step, building borders were cleaned by removing structures that were smaller than 8 pixels and that were connected to building borders. There was a compromise between cleaning thresholds less than 8 pixels, which may leave the original buildings unclean, and thresholds greater than 8 pixels which may remove parts of the buildings. The result was a black and white image that represents the detected buildings without noisy features and also without holes as shown in Figure 7.8. Finally, the 2-D binary image was labeled, with the background pixels labeled as 0, and the pixels representing each building labeled consecutively from one onwards.

Figure 7.8: the final labeled 2-D binary image for UNSW test area.

7.6.1.2 Boundary extraction

Starting from the upper left corner of the image and using a 3 X 3 moving window, the exterior boundaries of buildings are extracted in the labeled binary image. The central pixel of the window is considered to be part of the perimeter if it is nonzero and it is connected to at least one zero-valued pixel, as shown in Figure 7.9. This returns a two column vector containing only the rows and columns of the perimeter pixels of each building in the binary image.

0	0	0
1		1
1	1	1

1	1	1
1		1
1	1	1

Figure 7.9: Perimeter pixel (left), non perimeter pixel (right).

7.6.1.3 Simplification of the extracted boundary

The extracted boundaries of buildings are still noisy vectors because of the pixel structure of the raster image. Therefore, the extracted building boundaries are processed again through a line simplification routine. This process simplifies the shapes of buildings to remove unnecessary or unwanted details, while maintaining their essential shape and size. There are several algorithms to produce a simplified polyline that approximates the original one within a specified tolerance. The Douglas–Peucker algorithm is the

commonly used algorithm for this purpose, since it is fast and works well for 2D planar polylines. The initial form of the algorithm was suggested by Ramer (1972) and by Douglas and Peucker (1973). Initially, the algorithm automatically marks the first and last point to be kept. It then finds the point that is furthest from the line segment defined by the first and last points. If this point is greater than the reduction tolerance ε from the line segment then that point must be kept. All points not currently marked to be kept are discarded. Using recursion, this process continues for each edge until all vertices of the original polyline are within ε. When the recursion process is completed a new output is generated consisting of only those points that have been marked as kept. Figure 7.10 is a typical example of this process. Based on the experiments in this thesis it can be concluded that the tolerance can be initially specified equal to the pixel size of the data. If the output still contains too much detail, then the tolerance can be doubled and so on. Similarly, if the output lines do not have enough detail, the tolerance can be halved.

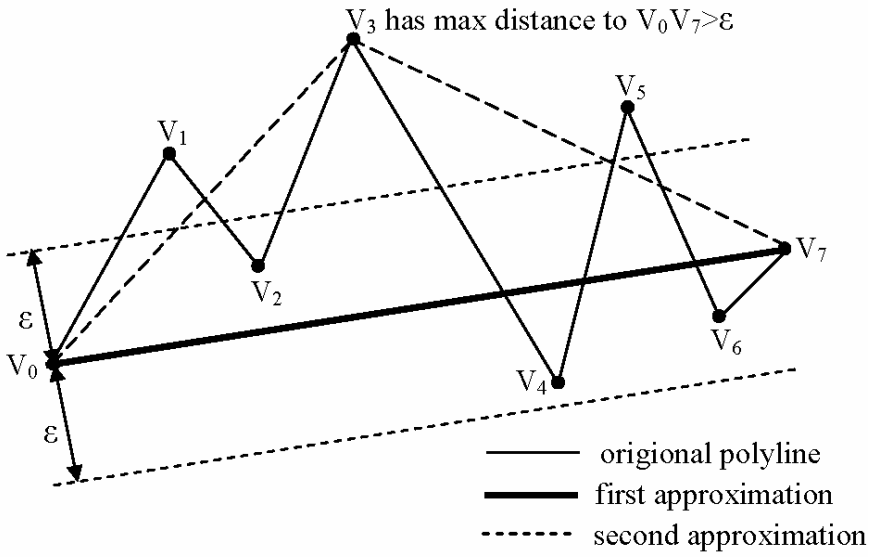

Figure 7.10: Smoothing a line segment with the Douglas–Peucker algorithm (Douglas and Peucker, 1973).

7.6.1.4 Adjusting the parallelism and rectangularity of buildings

A method based on *Freeman chain codes* (Bribiesca, 1981) has been proposed to enhance the parallelism and rectangularity of buildings. The method starts with: aligning one of the coordinate axes, x-axes in this case, with the major axis of the building (length of the building); then the building covered with a grid of an arbitrary resolution, 1% of the building width in this case. This small resolution provides accurate shapes with minimum deviations from the origional shapes. Finally, the grid points closest to the boundary were

recorded. Based on this method, the boundaries of buildings are enhanced so that all near-90-degree angles become exactly 90 degrees. Figure 7.11 is a typical example obtained in these experiments to demonstrate this process in which the grid resolution has been reduced to 10% for clarity. The horizontal and vertical axes refer to rows and columns respectively.

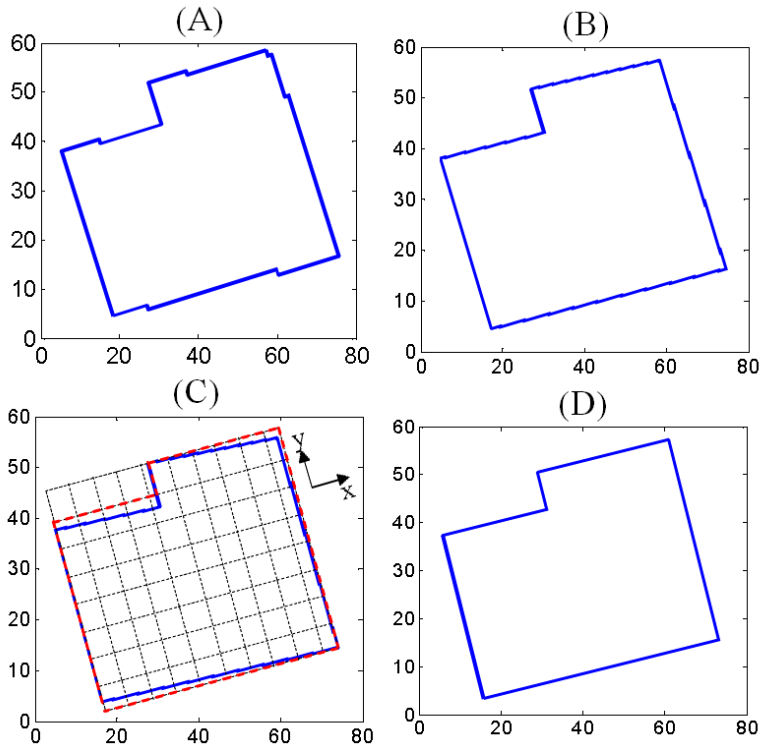

Figure 7.11: (a) the extracted building after raster to vector conversion, (b) building after Douglas–Peucker simplification, (c) coordinate transformation and grid construction, (d) bulding after adjusting the rectangularity.

Usually buildings have regular shapes, either squares or rectangles with few exceptions, such as buildings with curved boundaries. A common shape measure called the circularity ratio has been applied in these experiments to test the circularity of each building before adjusting the rectangularity. Circularity is the ratio of the area of the shape to the area of a circle having the same perimeter. That ratio is expressed mathematically as $M = 4*\pi*(\text{area})/ (\text{perimeter})^2$. For a circle, the ratio is one; for a square, it is $\pi/4$; for an infinitely long and narrow shape, it is zero (Bottema, 2000). The area is considered as the number pixels of each building multiplied by the pixel size. The circularity ratio is close to one for buildings with curved boundaries or with circular shape. Based on this fact, these buildings are not processed by the rectangularity test, but still considered in the simplification test. Figures 7.12 and 7.13 show the extracted buildings, before and after the simplification and adjustment of the rectangularity.

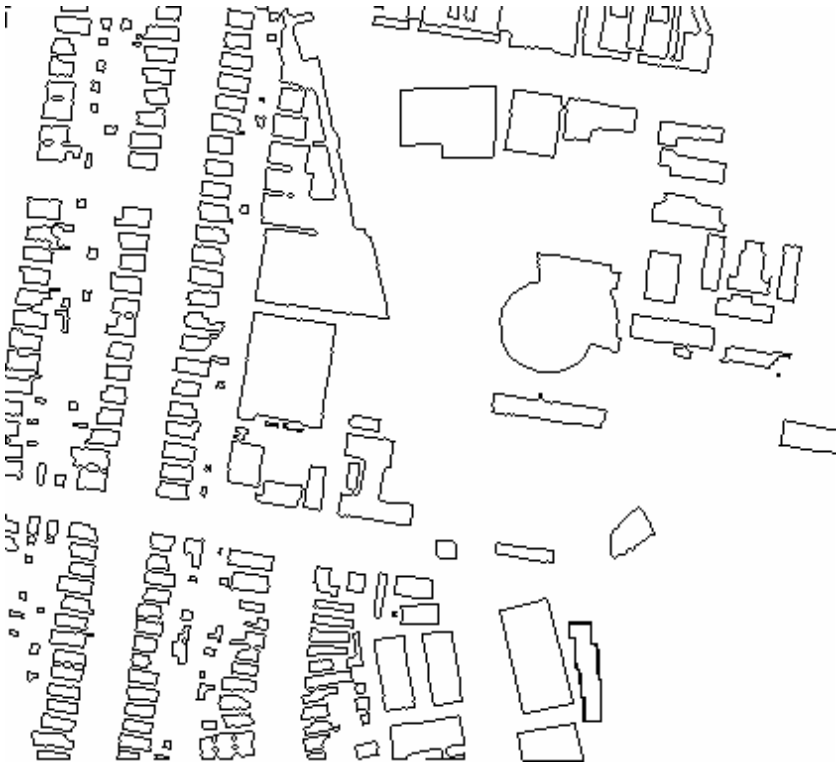

Figure 7.12: Buildings map after raster to vector conversion in case of UNSW test area.

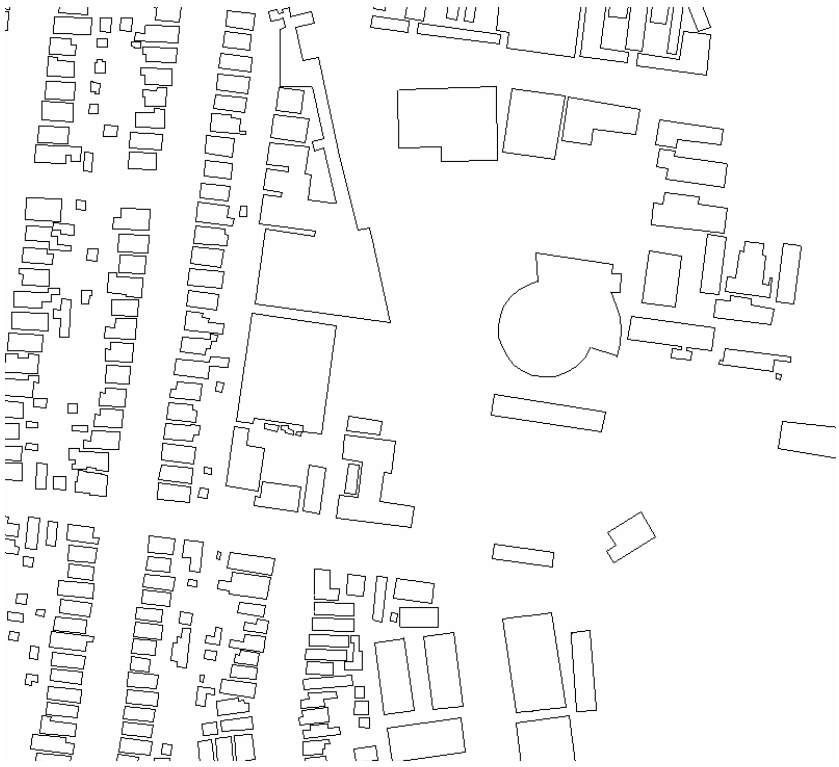

Figure 7.13: Buildings map after adjusting the rectangularity in case of UNSW test area.

7.6.1.5 Completeness and correctness of the extracted buildings

In order to evaluate the performance of the building extraction process after raster to vector conversion, the *completeness* and the *correctness*, described in Chapter 3, of the detected buildings were investigated based a per-building level. For the evaluation on a per-building basis, a building in the reference data set is counted as a true positive if at least 80 % of its area is covered by buildings detected automatically and vice versa. The results by the best two classifiers and the *D-S3* algorithm, which achieved the most accurate classification results, were used to evaluate the ability of *D-S* fusion to extract buildings. Figure 7.14 shows that false positives and false negatives mostly occurred at the building outlines. Possible reasons for that can be:

- In the resulting binary image of buildings, small elongated areas of buildings were eliminated by the post classification smoothing process.
- The effect of between-class variance on the edge pixels which caused many of these pixels to be placed in an incorrect category during the classification phase of the work.
- The errors in image to lidar geographic registration, as well as in filtering of lidar point clouds could have a negative impact on the classification results especially at class edges.

Also it is clear that fences with a height of more than 3m, the height threshold which was applied to the nDSM, were still detected as buildings as shown by blue rectangles at the lower right corner of the Figure 7.14 since the three classifiers have already detected them as buildings.

Figure 7.14: The spatial distribution of errors for the UNSW dataset. Blue: false positives; red: false negatives.

Figure 7.15 shows the completeness and correctness against the building size for the: UNSW; Bathurst; Fairfield; and Memmingen case studies respectively for SOM, CTs, SVMs classifiers as well as the combined classifier (SVMs+SOM) using the DS3 algorithm. For the UNSW case study, buildings around 30m^2 were detected with average improvements in completeness and correctness over the best classifier of around 0.3% and 0.7% respectively. For Bathurst, Fairfield and Memmingen case studies, buildings around 30m^2 were detected with average improvements in completeness and correctness over the best classifier of around: (1.4%,1.5%); (0.75%, 0.7%); and (1.2%, 1.1%) respectively. For all cases, all buildings larger than 60m^2 were detected with average improvements in completeness and correctness over the best classifier of around 1.9%. The difference between completeness and correctness is a matter of 0.02–0.1 compared to 1–2% in case of results for the individual best classifier. For buildings smaller than 60m^2, the difference between completeness and correctness is up to 2.7% compared to 10% in case of the best classifier. This further confirms the higher reliability of *D-S* combination for improving the detection of buildings smaller than 60m^2. The overall results are very consistent for all test areas, which are derived by 3 different lidar systems in terms of scanning pattern and resolution, in different urban environments in Australia and Germany and different vegetation types. It can therefore be concluded that these tests strongly represent achievable accuracies for detection of buildings by the proposed method using a combination of lidar data and images.

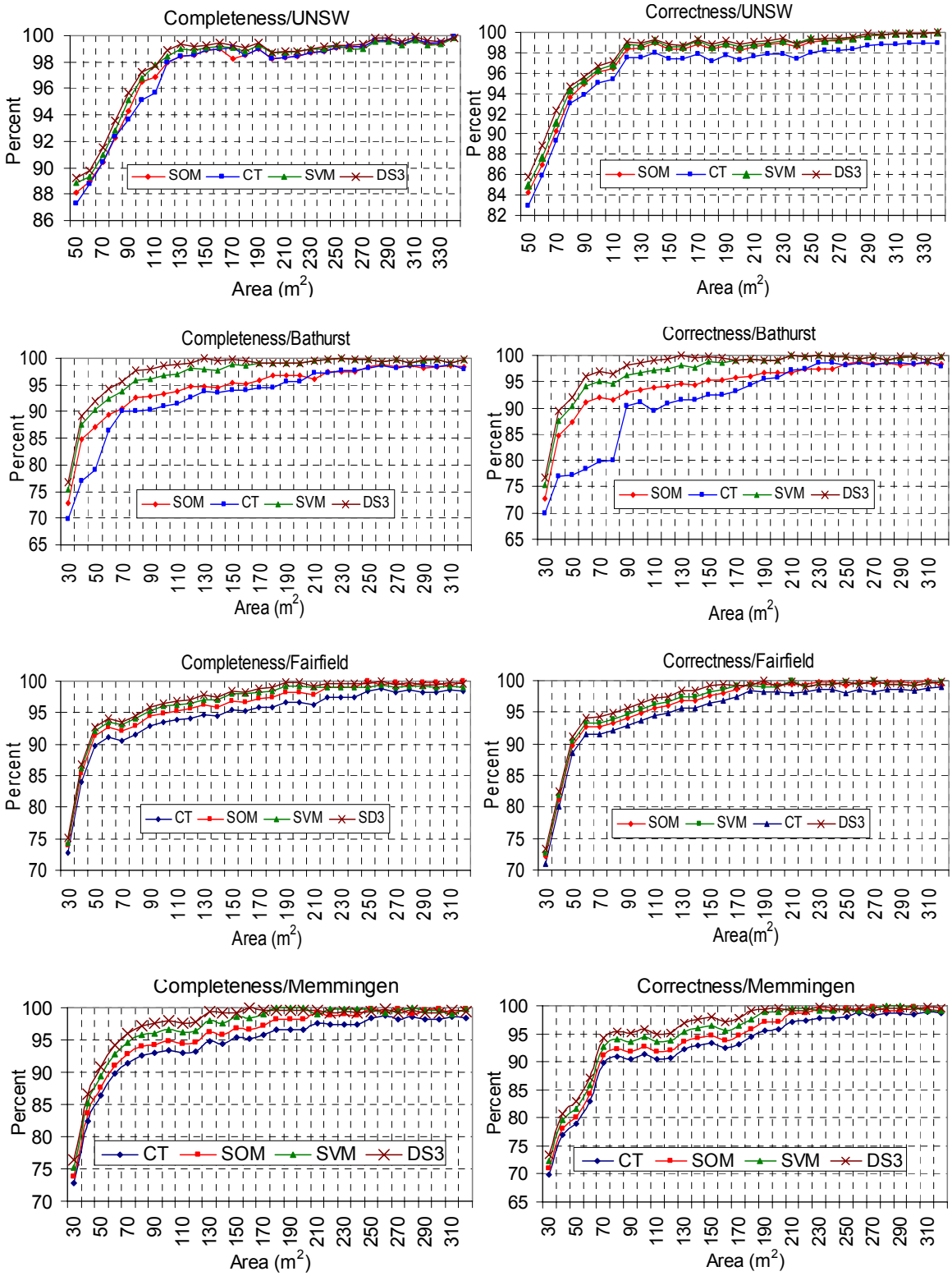

Figure 7.15: Completeness and correctness derived for SOM, CTs, SVMs and combined classifier (SVMs+SOM) using the DS3 algorithm plotted against building areas in cases of: *(a)* UNSW; *(b)* Bathurst; *(c)* Fairfield; and *(d)* Memmingen.

7.6.1.6 Evaluation of the planimetric accuracy of the final vector buildings

Due to limitations of available ground observations, only UNSW results were evaluated for planimetric accuracy based on ten GCPs determined by field surveys. The GCPs were selected to be evenly distributed throughout the study area and well defined on both the extracted building data and ground as shown in Figure 7.16. GPS observations were carried out and compared against the extracted building data coordinates as shown in Table 7.4.

Figure 7.16: Spatial distribution of GPS check points.

Table 7.3: The accuracy estimate of the building vectorization process for UNSW test area.

E_{map}	N_{map}	E_{GPS}	N_{GPS}	ΔE	ΔN	$\sqrt{\Delta E^2 + \Delta N^2}$
336012.26	6245794.85	336011.44	6245794.54	-0.82	-0.31	0.88
336048.23	6245505.13	336048.85	6245505.86	+0.62	+0.73	0.96
336196.30	6245778.30	336196.64	6245777.47	+0.34	-0.83	0.90
336163.63	6245700.83	336163.01	6245701.74	-0.62	+0.91	1.10
336192.51	6245662.65	336193.18	6245663.48	+0.67	+0.83	1.07
335933.13	6245618.86	335932.46	6245618.45	-0.67	-0.41	0.79
335891.40	6245458.21	335892.31	6245458.96	+0.91	+0.75	1.18
335768.24	6245548.42	335768.95	6245547.99	+0.71	-0.43	0.83
335763.63	6245756.31	335763.35	6245755.29	-0.28	-1.02	1.06
335838.84	6245866.81	335838.36	6245867.58	-0.48	+0.77	0.91
Mean				+0.04	+0.10	0.97
RMS				0.64	0.73	0.97

The accuracy of the final derived products were evaluated according to the three major map accuracy standards, namely, the National Map Accuracy Standard (NMAS), the American Society of Photogrammetry and Remote Sensing (ASPRS), and the National Standard for Spatial Data Accuracy (NSSDA). According to American National Map Accuracy Standards (NMAS; 1947) for maps at publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30inch (0.8mm), measured on the publication scale. These limits of accuracy shall apply to positions of well-defined points that are easily plotted on the scale of the map within 1/100inch (0.25mm). According to the three map standards, the horizontal accuracies for large scale maps are given in Table 7.4:

Table 7.4: Horizontal map accuracy according to NMAS, ASPRS, and NSSDA map standards.

	Map Standard		
Map Scale	NMAS	ASPRS	NSSDA
1:1,000	0.90	0.27	0.45
1:2,500	2.11	0.63	1.055
1:5,000	4.20	1.26	2.1
1:10,000	8.5	2.55	4.25

For the sake of clarity, the way the horizontal map accuracies are determined for 1:10,000 map scale, as a typical example, is outlined as follows (NMAS; 1947):

- For NMAS: 0.03333 x 10,000 x 2.54 / 100 = 8.5 meters.
- For ASPRS: (0.03333 x 10,000 x 2.54 / 100)/3.3333 = 2.55meters.
- For NSSDA: (0.03333 x 10,000 x 2.54 / 100)/2 = 4.25 meters.

The field survey measurements by GPS with accuracies of the order of 0.05m represent an estimate of the absolute accuracy of the extraction of buildings. They are shown as the first line in Table 7.5. The RMS of the results obtained for UNSW test area is 0.97m. From the accuracy figures presented in Tables 7.4, we can conclude that the horizontal accuracy defined by an RMS < 1m is well within the required map accuracy standard for a map scale of 1:5,000 according to the three map accuracy standards NMAS, ASPRS, and NSSDA. On the other hand, the results meet the accuracy requirements for map scale of 1:2,500 according to the NMAS, and NSSDA standards.

Table 7.5: The accuracy of the building vectorization process in cases of: Bathurst; Fairfield; and Memmingen test areas.

Test area	Pixel size (m)	RMS (m)	Standard map scale by NMAS 90%	Standard map scale by ASPRS 68%	Standard map scale by NSSDA 95%
UNSW	0.1	0.97	2,500	5,000	2,500

The same comparison has been repeated for Bathurst, Fairfield and Memmingen test areas, but since the sites could not be visited to make GPS measurements, 20 well defined and evenly distributed control points were derived from the manually digitized reference data, and compared against the extracted building data coordinates as shown in Table 7.6.

The accuracy of the final derived products for Bathurst, Fairfield and Memmingen test areas is well within the required map accuracy for map scale of 1:10,000 according to the three map accuracy standards NMAS, ASPRS, and NSSDA. It is worth mentioning that the achieved horizontal accuracies met the horizontal accuracy requirement for:

- Scale 1:5,000 for Bathurst test area according to NMAS map standards.
- Scale 1:2,500 and 5,000 for Fairfield test area according to NMAS and NSSDA map standards respectively. The same conclusion applies for Memmingen test area.

Table 7.6: The accuracy of the building vectorization process in cases of: Bathurst; Fairfield; and Memmingen test areas.

Test area	Pixel size (m)	Mean ΔE (m)	Mean ΔN (m)	RMS (m)	Standard map scale by NMAS 90%	Standard map scale by ASPRS 68%	Standard map scale by NSSDA 95%
Bathurst	0.5	-0.19	0.14	2.33	5,000	10,000	10,000
Fairfield	0.15	0.11	-0.09	1.60	2,500	10,000	5,000
Memmingen	0.5	-0.12	0.17	1.86	2,500	10,000	5,000

On the other hand, although both Bathurst and Memmingen images have the same pixel size, 0.5m, the positional accuracy in case of Memmingen test area is much higher than that in case of Bathurst test area. A possible reason for this may be that for Memmingen test area, sensors and scene characteristics are significantly different from those for Bathurst test area. Memmingen dataset include an IR channel and also the density of lidar points is higher than Bathurst dataset. As well, Bathurst is a largely rural area that contains small residential buildings, while Memmingen features industrial areas.

Finally, the accuracy of the feature extraction process only was estimated by subtracting the estimated errors in the reference data from those derived for the tests shown in Tables 7.5 and 7.6. In case of UNSW test area, the accuracy of GPS observations was estimated to be 0.05m. For Buthurst, Fairfield and Memmingen test areas, the accuracy of the digitized points derived from orthophotos as well as the digitizing process is estimated to be about 1 pixel (Abdullah, 2007). From the accuracy figures presented in the last column of Table 7.7, we can conclude that the proposed system has performed well for building extraction by aerial images and lidar data fusion.

$$\sigma_{Feature\ Extraction} = \sqrt{\sigma^2_{map} - \sigma^2_{reference\ data}} \tag{7.1}$$

Table 7.7: The accuracy of the building extraction process in cases of: UNSW; Bathurst; Fairfield; and Memmingen test areas.

Test area	Pixel size (m)	σ^2_{map} (m)	$\sigma^2_{reference\ data}$ (m)	$\sigma^2_{Feature\ Extraction}$ (m)
UNSW	0.1	0.97	0.1	0.93
Bathurst	0.5	2.33	0.5	1.35
Fairfield	0.15	1.60	0.15	1.20
Memmingen	0.5	1.86	0.5	1.17

7.6.2 Road Extraction

Roads are the second important component required for acquisition and/or updating of digital map data and Geographical Information Systems. Automatic and semi-automatic road recognition is an active area of research (Mena, 2003). Again, the classified images using the combination of the best two classifiers and the *D-S3* algorithm, which achieved the most accurate classification results, were used to evaluate the ability to extract roads. The roads were initially extracted by converting the classified image into a binary image, with the roads represented by one, and non-road pixels (building, trees and ground) by zeros. As roads form long thin connected objects, the area ratio of each individual road segment and the corresponding minimum bounding rectangle (MBR) are considerably small compared with the area ratio of any car parks and its corresponding MBR. Using a minimum ratio of 0.5, car parks (all road segments with MBR>0.5) which were classified as roads were detected and removed from the final binary image as shown in Figure 7.17. However, depending on the definition of the road model, car parks may or may not be considered to be included in the road model.

Another problem is that some parts of road sides were not properly detected because of the occlusion of trees as shown in Figure 7.17. In order to overcome this problem and to avoid its negative impact on the extracted centerlines as well as road edges, the binary image was smoothed by a Gaussian low pass filter (Nixon and Aguado, 2008) to remove sharp changes and inconsistencies along edges by the blending intensities with those of neighboring pixels. The smoothed image obtained after filtering is shown in Figure 7.18. As described by Clode et al. (2004) the road centerline, orientation φ, and width w are three critical parameters that need to be determined in order to model any road network. For this research, two experiments have been conducted to provide the centerline and width of the road. No information was provided about the orientation, since it can be defined easily from the centerline.

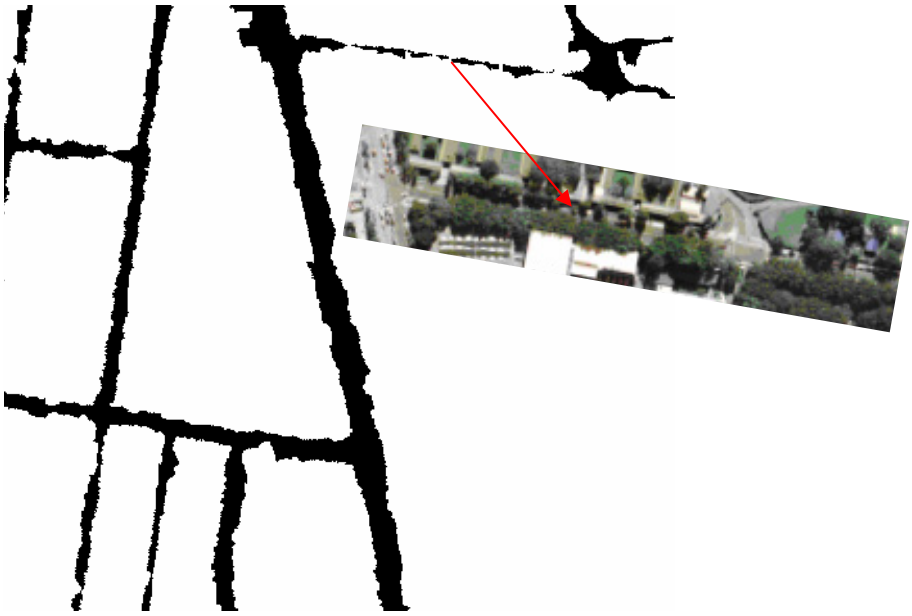

Figure 7.17: The resultant binary road network in the case of UNSW data set. The arrow highlights a problem that occurs due to tree occlusion.

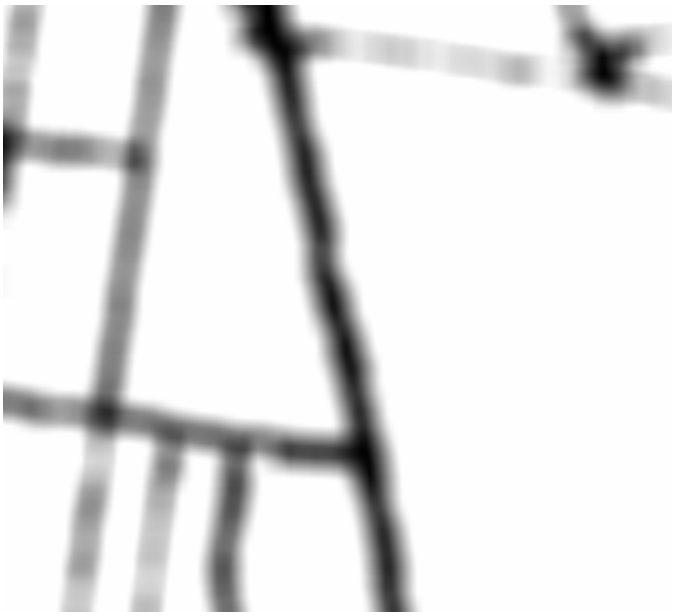

Figure 7.18: Smoothed road network using Gaussian low pass filter in the case of UNSW data set.

7.6.2.1 Centreline extraction

In order to extract the centerline of the road, the smoothed image was processed through an image skeletonizsation algorithm to produce a single pixel thickness version of each road that is equidistant from its boundaries. This new image emphasizes geometrical and topological properties of the road in the binary image, such as its connectivity, topology, length, direction, and width, as well as the distance of centreline points to the road boundary. There are many different algorithms for image skeletonization (Serra, 1988). Since morphological skeletons have well-defined mathematical characterizations that allow exact reconstructions of the centerline, as well as lower computational cost, it has been adopted for this work. Figure 7.19 shows the resulted skeleton in the case of UNSW road network. After that, the resulted centerlines were simplified using Douglas–Peucker algorithm to remove unnecessary details, while maintaining the essential shape of the centerline. The final centerlines are smooth and continuous as shown in Figure 7.20. Finally, the x- and y-coordinates of every pixel along the centerlines of the roads was determined and recorded.

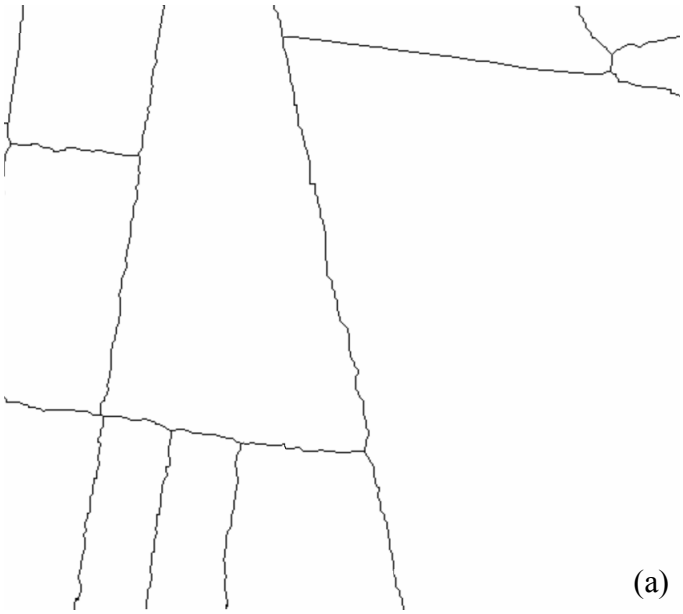

Figure 7.19: Road centerlines as extracted by morphological skeletonization for UNSW data set.

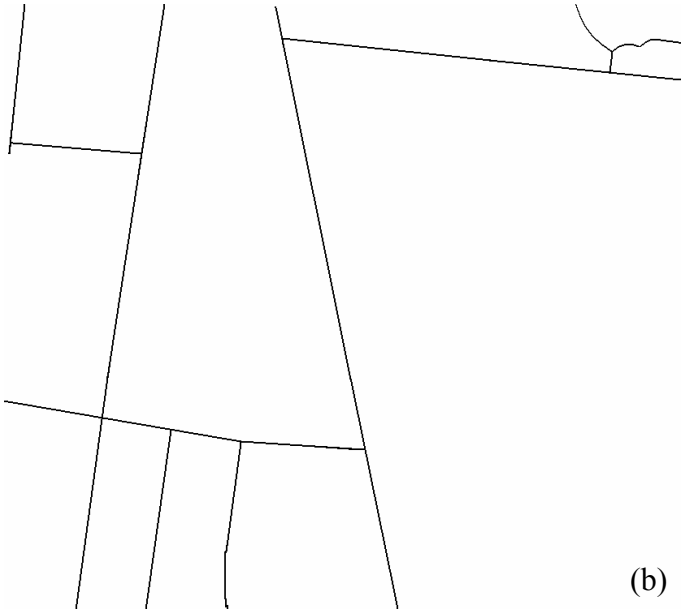

Figure 7.20: The extracted centerlines after Douglas–Peucker simplification for UNSW data set.

7.6.2.2 Width calculation

At each point along the centerline of each road segment, the distance between the centerline point and the corresponding left and right points on the edges of the binary image were calculated. The two calculated half widths are referred to as left half width (wh_l) and right half width (wh_r) as shown in Figure 7.21 (a). If the difference between wh_l and wh_r is within five pixels, the road width was calculated as the mean value of wh_l and

wh_r, and kept, other wise the width was neglected. Finally, the width of the road segment was determined as the mean value of the recorded widths as shown in Equation 7.2.

$$wh = \frac{\sum_{i=1}^{n} (wh_l(i) + wh_r(i)) / 2}{m}, | wh_l(i) - wh_r(i) | \leq 5 \qquad (7.2)$$

wh_l and wh_r: the left and right half width at a given point on the centerline.

n: the total number of points along the centerline

m: the number of points which satisfy the condition in Equation (7.2).

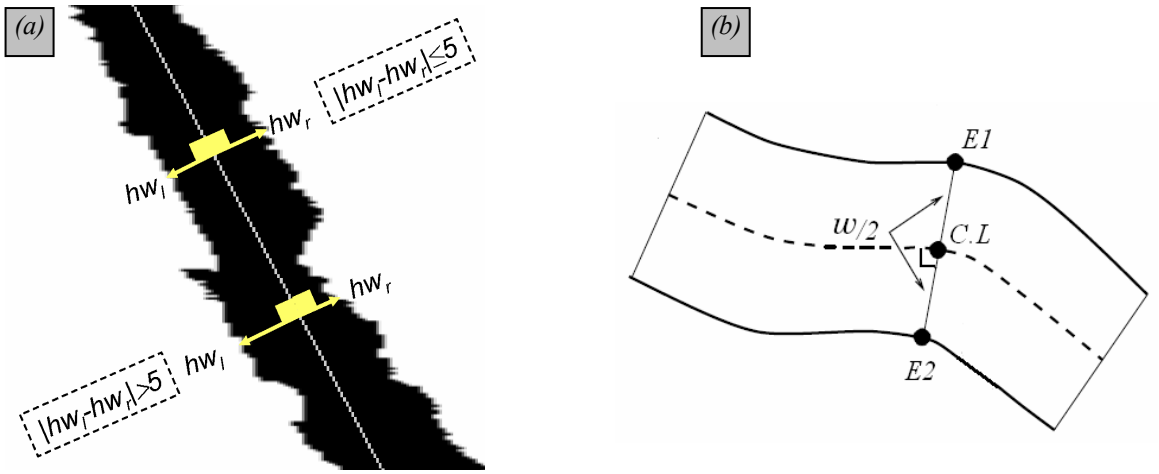

Figure 7.21: Construction of edges: (a) calculating the width; (b) shifting the centerline pints by the calculated width for UNSW data set.

Finally, edges were constructed as being parallel to the centerline by applying half of the predetermined width orthogonally to the road centerline as shown in Figure 7.21 (b). Figure 7.23 shows the vectorized results overlaid on an aerial image for UNSW test area. The results show a good approximation of the original network.

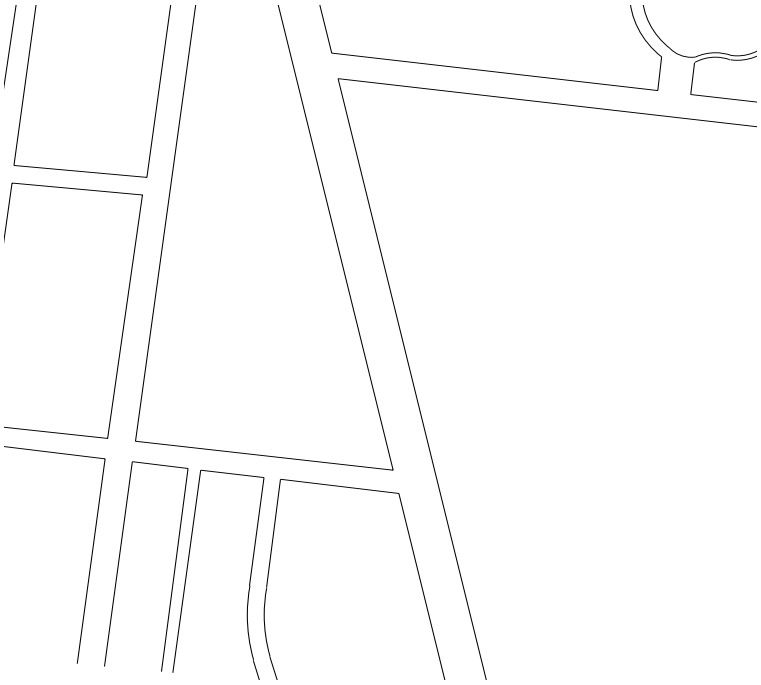

Figure 7.22: The final vectorized roads for UNSW data set.

Figure 7.23: The road edges are overlaid on the aerial image for UNSW data set.

7.6.2.3 Completeness and correctness of the extracted roads

In order to analyze the results, the final vectorized map was converted from vector to raster and compared with the manually digitized reference data. For comparison purposes, completeness and correctness were evaluated. The spatial distributions of the errors in case of UNSW are plotted in Figure 7.24 displaying the FN and FP in blue and red

respectively. The automatically extraction process results missed one road segment that is displayed in blue in Figure 7.24 and also reflected in the low completeness numbers in Figure 7.25. Although the upper part of this segment appears in the classified image, it disappeared in the final labeled binary image because it satisfies the minimum bounding rectangle (MBR) ratio condition, and hence has been removed from the binary image. The reason why this section is missing is similarity of properties of this lower intensity segment and the surrounding terrain surface, especially when they have approximately the same height values from the lidar data. This meant that this segment was classed as ground and hence were removed from the detected road network. The majority of the false positives and negatives are located at road edges. The quality of the detected road width appears good.

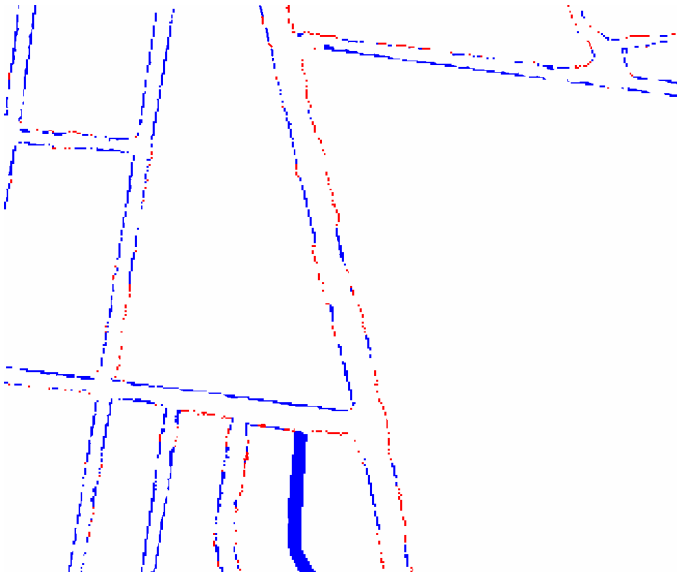

Figure 7.24: Spatial distribution of the errors in case of UNSW test area (FP in red, FN in blue).

Figure 7.25 shows the completeness and correctness for the: UNSW; Bathurst; Fairfield; and Memmingen case studies. Detailed analysis of the completeness for the four test areas shows that 90-98% of the reference road network was recognized. The difference between completeness of the best classifier and the best combination of classifiers amounts to 4-8%. The high completeness values are very encouraging. However, it clearly appears that *DS-3* consistently outperforms the others in terms of completeness and correctness. Also, SVM classifiers outperform both SOM and CTs for the four test areas, with correctness with lower values than completeness. A possible reason for these results is the similar properties of the lower intensity roads and the surrounding terrain surface. This means

that some parts of the ground are classed as roads and some parts of roads as ground. On the other hand, the lower intensity of the UNSW road network as compared to Bathurst, Fairfield and Memmingen roads gives an explanation why the completeness and correctness of the UNSW results are lower than for the other test areas as shown in Figure 7.25. This clearly indicates that the quality of automatically extracted features does not depend only on the pixel size but on other image quality factors such as: illumination conditions and image sharpness, even if the ground pixel size is similar or slightly better. It is worth mentioning that high completeness is more important than high correctness. However, future research should be directed towards distinguishing roads from ground under lower height and intensity differences in order to also improve the correctness.

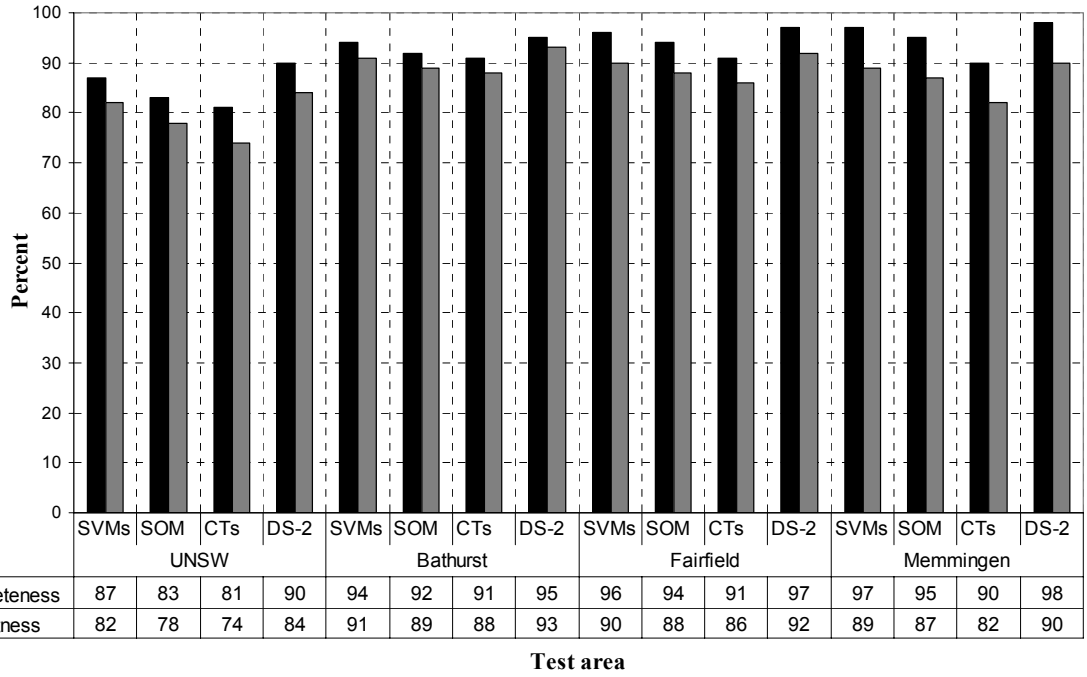

Figure 7.25: Completeness and correctness of roads derived for SOM, CTs, SVMs and combined classifier (SVMs+SOM) using the D-S3 algorithm in cases of: UNSW; Bathurst; Fairfield; and Memmingen.

7.6.2.4 Evaluation of the planimetric accuracy of the final extracted roads

The geometrical accuracy of the detected centerlines and edges were assessed. The manually vectorized roads were compared to the automatically extracted roads except for Bathurst test area, where reference data for roads centerlines only was provided by NSW Land and Property Authority (LPMA) in vector format. The edges in the reference data were used to calculate a centerline positions and widths at each extracted vertex. The positional errors, also known as displacements, or distortions which represent the difference between the reference and extracted coordinates were calculated for all

centerline points and road widths. The RMS of centerlines and widths were used as measures of quality of the results of all individual segments. The huge number of check points used, as shown in Tables 7.8 and 7.9, justifies a high level of confidence that the RMS is an accurate measure of the performance of the process of road extraction.

For clarity, the way the statistical figures are determined is outlined by Figure 7.26 and Equations 7.3, 7.4 and 7.5. It is worth mentioning that Equations 7.4 and 7.5 are applicable for width calculations.

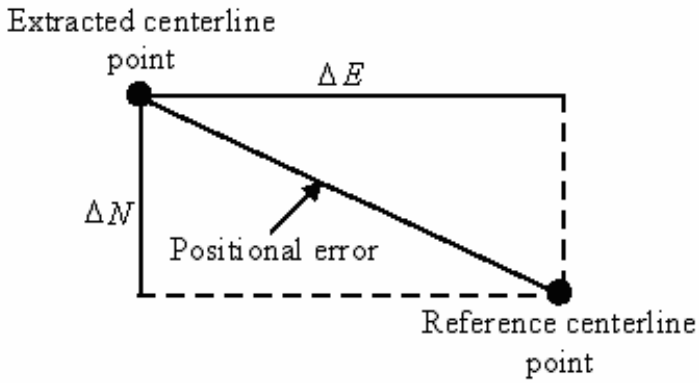

Figure 7.26: Positional error at centreline points.

$$PE = \sqrt{\Delta E^2 + \Delta N^2} \tag{7.3}$$

$$Mean\ PE = \frac{1}{n}\sum_{i=1}^{n} PE_i \tag{7.4}$$

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(PE_i - Mean\ PE)^2} \tag{7.5}$$

ΔE: The difference in *E* coordinates ($E_{reference}$- $E_{extracted}$)

ΔN: The difference in *N* coordinates ($N_{reference}$- $N_{extracted}$)

Mean PE: Mean positional error

PE: Positional error

n: No. of points on the centerline (No. of pixels)

The accuracy of the road vectorization process is summarized in Tables 7.8 and 7.9, which reveal that the accuracy of the final derived products for all test areas is well within the required map accuracy for map scale of 1:10,000 according to the three map accuracy standards NMAS, ASPRS, and NSSDA. It is worth mentioning that the achieved horizontal accuracy met the horizontal accuracy requirement for:

- Scale 1:2,500 and 1:5,000 for UNSW test area according to NMAS and NSSDA map standards respectively. The same conclusion applies for Fairfield test area
- Scale 1:5,000 for Bathurst test area according to NMAS map standards. The same conclusion applies for Memmingen test area.

The small mean values for both centerlines and widths indicate that there are no systematic errors in the extraction process. The lower accuracy of the extracted centerlines and widths in the case of UNSW results, as compared to Fairfield results, clearly confirm the fact that the quality of automatically extracted roads does not depend only on the pixel size but also on other image quality factors as aforementioned.

The accuracy of road extraction process only was estimated by subtracting the estimated errors in the reference data from the accuracies shown in Table 7.8. Again, for the accuracy of the digitized reference data derived from the estimated accuracy of the orthophotos as well as the digitizing process is estimated to be about 1 pixel (Abdullah, 2007). The accuracies of centerline and width extraction process only are shown in Tables 7.10 and 7.11 respectively. From the accuracy figures presented in the last column of Tables 7.10 and 7.11, we can conclude that the proposed system has performed well for centreline as well as width extraction by aerial images and lidar data fusion.

Another aspect of interest is to visually compare the extracted roads with the reference data, provided by NSW LPMA, in case of Bathurst test area. This comparison showed that two major problems exist with the reference data. On the one hand, some roads have been missed in the reference data. This is because these roads are very similar to its surrounding areas, especially roads with lower intensity, which makes it difficult for the operator to detect and digitize this category of roads. Figure 7.28 is a good example where roads a, b and c in the automatically detected results have been missed in the reference data. On the other hand, some roads which fall into categories such as driveways were picked up by the operator. Figure 7.29 shows a good example of this problem. The reference roads 1, 2 and 3 are driveways and not-roads. Also, the curvilinear extracted roads are quite smooth and represent the original image well in the extracted network. This confirms the potential of using different sources of information as well as combined classifiers as a base for automatic feature extraction.

Table 7.8: RMS of positional errors of the extracted centerlines.

Test area	No. of points	Mean ΔE(m)	Mean ΔN(m)	RMS(m)	Standard map scale by NMAS 90%	Standard map scale by ASPRS 68%	Standard map scale by NSSDA 95%
UNSW	28152	0.13	-0.23	1.54	1:2,500	1:10,000	1:5,000
Bathurst	9566	0.3 4	-0.2 9	2.48	1:5,000	1:10,000	1:10,000
Fairfield	103728	-0.11	0.26	1.32	1:2,500	1:10,000	1:5,000
Memmingen	39201	0.2 1	0.25	2.16	1:5,000	1:10,000	1:10,000

Table 7.9: Differences between extracted and digitized widths.

	No. of points	Mean Difference (m)	RMS (m)
UNSW	28152	-0.25	1.46
Bathurst	9566	+0.34	1.82
Fairfield	103728	+0.18	1.24
Memmingen	39201	-0.49	1.89

Table 7.10: The accuracy of the centerline extraction process in cases of: UNSW; Bathurst; Fairfield; and Memmingen test areas.

Test area	Pixel size (m)	σ^2_{map} (m)	$\sigma^2_{reference\ data}$ (m)	$\sigma^2_{Feature\ Extraction}$ (m)
UNSW	0.1	1.54	0.1	1.44
Bathurst	0.5	2.48	0.5	1.98
Fairfield	0.15	1.32	0.15	1.17
Memmingen	0.5	2.16	0.5	1.66

Table 7.11: The accuracy of the width extraction process in cases of: UNSW; Bathurst; Fairfield; and Memmingen test areas.

Test area	Pixel size (m)	σ^2_{map} (m)	$\sigma^2_{reference\ data}$ (m)	$\sigma^2_{Feature\ Extraction}$ (m)
UNSW	0.1	1.46	0.05	1.41
Bathurst	0.5	1.82	0.5	1.32
Fairfield	0.15	1.24	0.15	1.09
Memmingen	0.5	1.89	0.5	1.39

Figure 7.27: Bathurst aerial image.

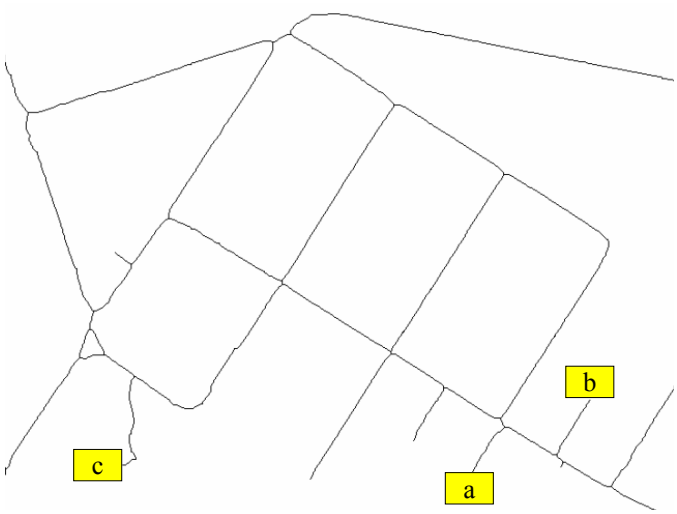

Figure 7.28: Extracted centerlines in case of Bathurst test area.

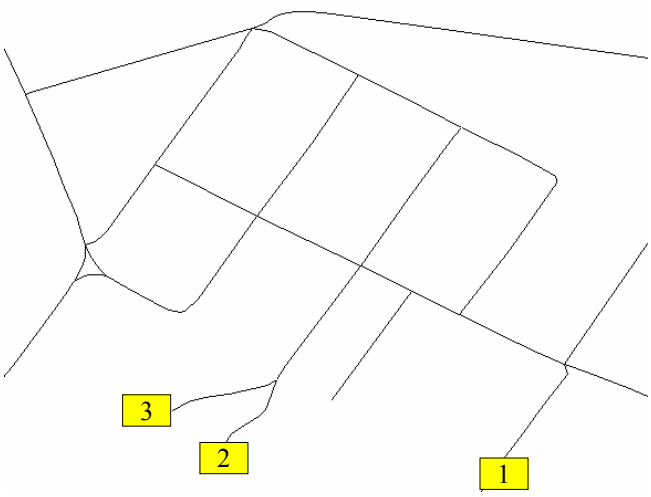

Figure 7.29: Reference centerlines in case of Bathurst test area.

7.7 SUMMARY

In this chapter, a strategy for combining results from multiple classifiers (SOM, CTs and SVMs) with different characteristics, based on *D-S* theory of evidence, was introduced. The investigation clearly shows that results from *D-S* combination are superior to those from individual classifiers, as well as to those from other hybrid systems (*MR*, *WS* and *FMV*), in terms of overall accuracy and commission and omission errors.

The assessment of commission and omission errors shows that the proposed system achieves the best performance for almost every class. Furthermore, results from D-S combination show a positive balance between commission and omission errors of individual classes. The main reason for this can be the underlying assumption of the system. As aforementioned, the system first weights the class membership at each classifier output based on the reliability of that classifier, then it determines the amount of belief from a given classifier output that must be discarded i.e. assigned to the ignorance hypothesis. In addition the individual classifiers do not perform equally. One classifier might be more efficient at detecting one class, while another classifier is more efficient for another class. A visual interpretation of the produced image shows the positive impact of the proposed system for significantly reducing the degree of noise in the combined classification results, even compared with the image produced by the best classifier (SVMs).

The proposed system is computationally expensive compare to *MR*, *WS* or *FMV*. However, the processing time could be reduced either by splitting large test areas into smaller parts, processing each part separately and combining the results later, or by using two classifiers instead of three. It is worth mentioning that in the case of two-classifier problem, adding more classifiers to the system does not guarantee improvements in the performance.

The produced digital vector maps for buildings and roads based on the proposed system were intensively evaluated. An assessment of the completeness and correctness, as well as positional accuracy of the vectorized buildings and roads clearly demonstrated the reliability of the proposed system for large scale mapping.