Building on Past Achievements: Generalising OS MasterMap® Rural Buildings to 1:50 000

Patrick Revell
Ordnance Survey, Research and Innovation
Romsey Road, Southampton, UK
patrick.revell@ordnancesurvey.co.uk

Abstract

This paper deals with vector-based automated cartographic generalisation and focuses on generalising of OS MasterMap® rural buildings to 1:50 000 scale. The aim of the research is to automate production of these rural buildings, but still conform to the current Ordnance Survey raster product specification. Firstly, urban areas and rural building clusters are detected, then amalgamation candidates are identified within the rural clusters. New algorithms are proposed for computing a squared amalgam for a group of buildings, and simplifying the resulting geometry. The paper concludes by discussion how the rural building amalgams can be placed to conform with the proximity requirements in the specification.

KEYWORDS: Generalisation, OS MasterMap®, Agent, Clarity, Buildings, Amalgamation.

1 Introduction

1.1 Generalising OS MasterMap®

In 2001 the vector-based Ordnance Survey (OS) MasterMap® Topography layer was launched, followed by the Integrated Transport Network™ (ITN) layer in 2003 [Ordnance Survey 2004]. This dataset has a high level of detail, being captured at 1:1250 scale for urban areas and 1:2500 scale for rural areas. Figure 1 shows an extract of OS MasterMap® displayed using the standard base scale representation.

Figure 1

The creation of OS MasterMap® introduced the possibility of automatically deriving smaller scale vector and raster products from a single base dataset, turning OS’s “capture once, use many times” philosophy into a reality. The logical starting point for achieving this goal is to develop automated tools for deriving the traditional OS map scales from OS MasterMap®.

Products created in this way would have the advantage of precision and consistency with the base scale, coupled with vastly reduced production costs. Once the traditional products have been derived, the aim would then be to extend the tools for generalising to arbitrary scales and specifications.

1.2 Research strategy

One of the key traditional products of OS is the raster coverage of Great Britain at 1:50 000 scale (1:50K). This is sold as both the Landranger® paper map series and as 20x20km image tiles. The raster data is currently revised and updated manually. A choice has been made to focus generalisation research on developing tools to achieve OS specification 1:50K data automatically from OS MasterMap®.

The task of deriving 1:50K maps has been subdivided into rural, urban and road generalisation. This paper concentrates on the generalisation of OS MasterMap® rural buildings to 1:50K. A summary is provided of the work that has been completed so far and the paper concludes with an overview of the planned next steps in the process.

The software platform chosen for developing generalisation tools is Clarity from Laser-Scan. Clarity includes a re-implementation of the Agent core [Regnauld 2002], constraints, measures and algorithms from the AGENT project [AGENT 2000], XML support for map specifications and data models, plus a new Java API for customising the system [Laser-Scan 2003].

The AGENT core allows a hierarchical generalisation strategy to be developed, where the *meso agents* in the upper levels are used to control groups of features which have a geographic or cartographic meaning. Meso agents are able to create and control other meso agents, while at the bottom of the hierarchy they control *micro agents*. Micro agents are responsible for the generalisation of individual cartographic features.

2 Map specifications for buildings

2.1 Existing specifications

The OS 1:50K specification document needed some interpretation to bring it in terms of dataset units for vector generalisation [Ordnance Survey 2000]. For vector-based representation the specification requires that building area geometries are represented with a black line of a constant width and an orange stipple fill. Isolated buildings, if large enough, are shown to scale, subject to a minimum width. Minimum size buildings are shown by a square with a specific side length. If necessary, buildings are enlarged to comply with this dimension.

Juts on buildings are shown subject to a specific minimum width. If space permits, small juts are enlarged to comply with this dimension. Detached buildings or groups of buildings closer than a certain distance are generalised together with a single outline. In the generalisation of such buildings, the general shape must be maintained. Building representation must be separated from water and forest boundaries by a minimum distance, or share geometry.

2.2 Deduced specifications

The specification for 1:50K maps is not sufficient to generalise buildings according to the current standard. It is clear that much of the specification is embedded in the current product, with cartographers presumably checking the consistency of their updates by eye. Hence additional generalisation criteria needed to be inferred through empirical analysis of existing 1:50K mapping. A key technique for this was to overlay OS MasterMap® buildings data on top of 1:50K raster images, to understand the basis for the desired result.

Empirical analysis revealed that isolated buildings smaller than a certain size can be deleted. Buildings must be separated from road/track casings by a minimum distance, or share an edge or corner with the road/track. Buildings must be separated from each other by a minimum distance, although they are permitted to touch corner to corner. Rural building geometries tend to be comprised of straight segments and right angles, unless influenced by their surrounding features.

3 Identifying rural buildings

Observations from existing Ordnance Survey 1:50K maps reveal that the generalisation of buildings differs quite dramatically between urban and rural areas. Figure 2 illustrates that the urban buildings on the left are represented as solid fill, adhering to the road curvature, whereas in rural areas the buildings are represented by isolated angular amalgams. Hence it is clear that distinct agents are required to deal with urban and rural building generalisation.

Figure 2

There is no differentiation in OS MasterMap® between urban and rural buildings, so some criteria were required to automatically detect rural buildings. The principal reason for urban buildings being generalised to solid fill is that they are too close to be represented individually, hence it is sensible to base the criteria on building proximity. An urban area detection algorithm developed during the AGENT project seemed appropriate for this purpose. The algorithm clusters the buildings by 50 metre proximity, then any clusters over 200000 metres² are taken to be urban areas, while the remaining are classed as *rural clusters*.

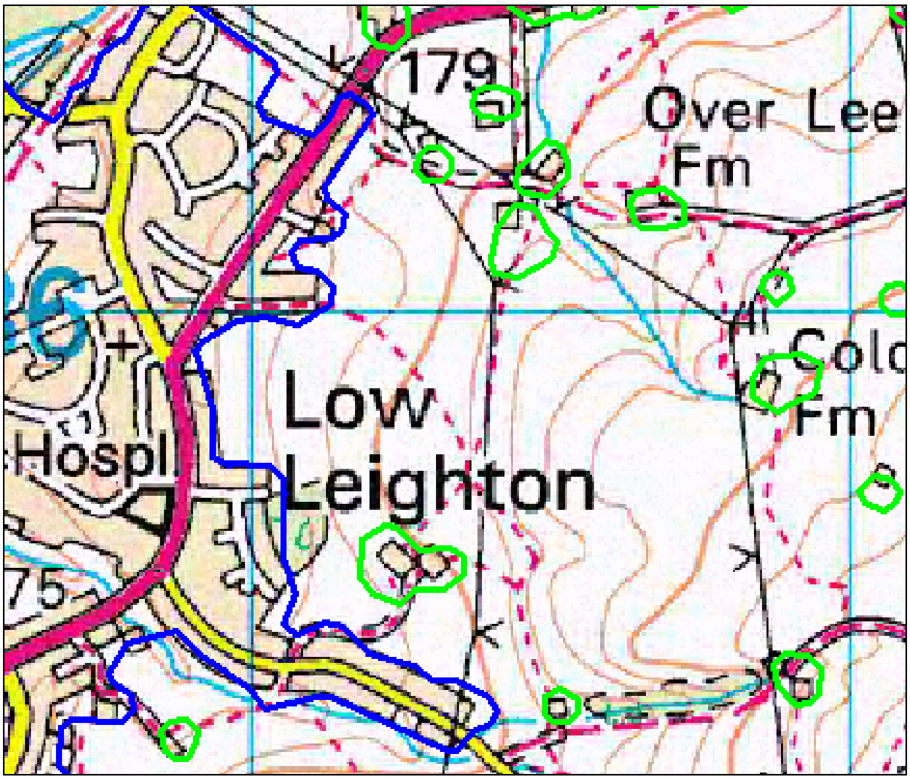

Figure 3

A visual check of the results of the algorithm against existing 1:50K images confirmed that the parameters were appropriate. Figure 3 shows an extract of 1:50K mapping with an urban area geometry overlaid on the left in dark blue and the rural clusters displayed on the right in green.

There are cases where rural-like buildings appear inside urban areas, and places where buildings are given a urban-style treatment in rural clusters. Such cases are comparatively rare, and at this stage the urban-rural distinction is a useful way of ordering the processing. If an agent detects that it has been classified unsuitably, the option is always available for it to create a new agent of a more appropriate class and delete itself.

4 Selecting amalgamation candidates

The rural clusters are a convenient means of processing the rural buildings since they are small, distinct and independent. A specific type of meso agent has been designed to control the generalisation of a single rural cluster. The rural cluster meso agent then selects candidates for amalgamation, and for each group creates a building amalgam micro agent responsible for representing a single building on the generalised map.

Inside a rural cluster candidates for amalgamation are selected on the basis of the minimum building separation distance in the specification. These proximity groups are split further according to separating features, such as roads, tracks, paths and rivers. This process is illustrated by Figure 4, which shows a rural cluster in green, OS MasterMap® roads/track polygons in red and buildings in brown.

Figure 4

A shrink-wrap hull is used as a temporary means of representing the building amalgam agents as shown in Figure 4. The shrink wrap hull is calculated from the convex hull by snapping points inwards to building co-ordinates according to a tolerance of 7 metres. The shrink-wrapping algorithm is part of Clarity, originating from Gothic Generaliser.

Each building amalgam micro agent has a reference to the buildings from which it is comprised. This allows the agent to quickly refer to its origins during the processing and provides the traceability needed for incremental generalisation.

Once the rural cluster meso agent has created the building amalgam agents, they are activated and processed in sequence. Each building amalgam agent then becomes responsible for evaluating itself against the map specification using the constraints which will be detailed in the following sections. The constraints are able to propose suitable algorithms to improve the building amalgam agent's adherence to the map specification.

5 Building size constraint

The first constraint that is applied to a building amalgam agent is the building size constraint. This constraint is triggered first, since the remaining constraints are irrelevant if the building is below the minimum size. Isolated amalgams below the specification deletion threshold can simply be removed and play no further part in the processing.

The specification requires that minimum size buildings are shown by a square of a specific side length. Therefore, if the amalgam area is below the squared value of this side length, an algorithm is triggered which sets a rectangular geometry at least as large as the minimum size square, on the agent. This rectangle is defined by its orientation, its two side lengths and its position.

5.1 Building group orientation

The orientation of the smallest minimum bounding rectangle (SMBR) of the shrink-wrap hull was evaluated as a possible measure of building group orientation. The SMBR algorithm is part of Clarity, originally proposed during the AGENT project [AGENT DC1 1999]. This measure was found to be unrepresentative of the orientations of the original buildings.

Hence a new measure was developed, based on previous work by IGN France on the *wall statistical weighting* orientation of a single building [Duchêne et al 2003]. Exactly the same procedure is used, except instead of considering the walls of an individual building, all external walls in the building group participate in the calculations. This produces a building group orientation modulo π/2, which is accurate to π/180 radians.

5.2 Building group rectangle

The centre of gravity of the shrink-wrap hull is then calculated, and the building group is rotated about this point by the group orientation. This orientates the walls as closely as possible with the X-Y axis. The X-Y oriented bounding rectangle of the reoriented building group is then calculated. The sides of this rectangle are compared against the specification minimum side length and are extended to conform, if necessary.

The final step is to restore the group orientation by rotating to the bounding rectangle about the centre of gravity, and setting the new geometry on the agent. This approach ensures that all of the original buildings are contained within the amalgam rectangle geometry, and the rectangle is at least as large as the minimum size square. The centre of gravity algorithm is part of Clarity, originally proposed during the AGENT project [AGENT DC1 1999].

6 Building squareness constraint

Of the remaining amalgams, the required result is usually comprised of straight segments and right angles (squared) as in Figure 5, so a logical starting point is to obtain a squared geometry for the building amalgam agent. Many building generalisation algorithms operate on the assumption that the geometry is squared, so ensuring a squared amalgam geometry is an appropriate operation to apply first. The squaring algorithm in Clarity was not designed to square up highly angular shrink-wrap hulls, so a new algorithm needed to be developed.

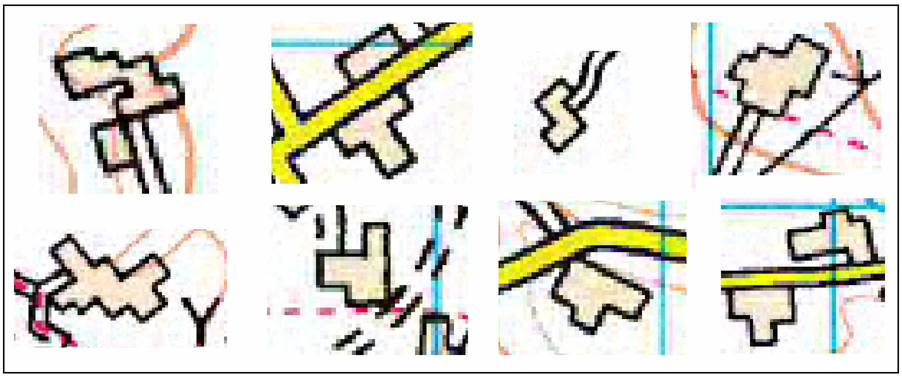

Figure 5

6.1 Compute the orientation of the group

The first step is to decide on the orientation for the amalgam geometry. For this, the wall statistical weighting measure of building group orientation is employed, as described in section 5.1. Figure 6 illustrates a building group contained by a bounding rectangle with sides oriented according to the group orientation. The initial shrink-wrap hull and centre of gravity are also shown in the picture.

Figure 6

The method outlined here is clearly not the only approach for computing group orientation. The algorithm could be easily extended, allowing an arbitrary orientation to be specified, so that an amalgam could be calculated which is oriented parallel to a road, for example.

6.2 Simplify the buildings to rectangles

All of the buildings in the group are then rotated about the centre of gravity by the building group orientation. The objective of this is to temporarily orient everything according to the direction of the X-Y axis, simplifying calculations considerably. The amalgam can be easily restored to the correct orientation at the end of the calculations.

The next step is to take the re-oriented buildings and dissolve any shared boundaries. Then for each remaining geometry the X-Y oriented bounding rectangle is calculated. This ensures that the amalgam geometry will completely contain the original buildings, fitting them as tightly as possible. Figure 7 shows the buildings reoriented from Figure 6, with boundaries dissolved and each building contained in its X-Y oriented bounding rectangle.

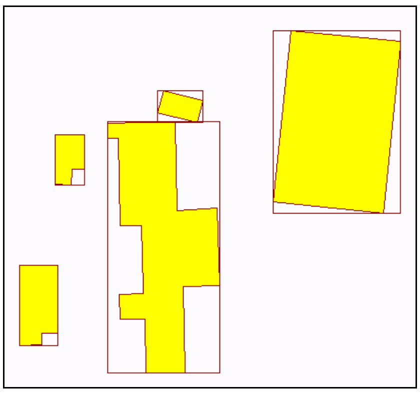

Figure 7

6.3 Bridge the rectangles

In order to create the final amalgam, the X-Y oriented rectangles need to be joined together using simple bridges comprised of X-Y oriented line segments. The minimum spanning tree is computed for the rectangles to determine the pairs of rectangles which need to be bridged. A process then begins which considers each link in the minimum spanning tree, evaluates the relative positions of the two rectangles, then builds a suitable bridge.

The simplest case is when the pair of rectangles intersect, by virtue of the original building positions. These circumstances can be observed in the centre of Figure 7. In this case the two geometries can be combined, with no need for a bridge. In evaluating the remaining cases, it is helpful to list all of the possibilities.

6.3.1 Total overhang

Figure 8 shows a set of straightforward cases, where there is only one choice of bridge, shown by the dotted lines. All that is required is to extend the smaller rectangle to join the larger rectangle.

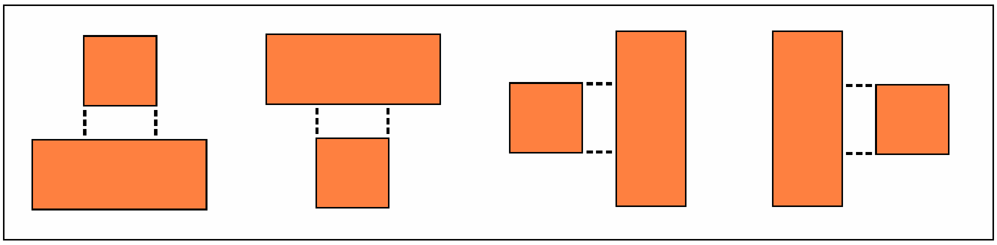

Figure 8

6.3.2 Partial overhang

Figure 9 shows that the choice of bridge is more complex when there is a partial overhang between the two rectangles. The first choice of bridge is the area labelled A. However there is a danger of creating narrow corridors between the two rectangles. Therefore the algorithm takes a threshold value which forces bridges to be over a specified width and height. The next choice of bridge is the bar comprised of areas {B, A, C}. If this still does not comply with the threshold, then two choices remain.

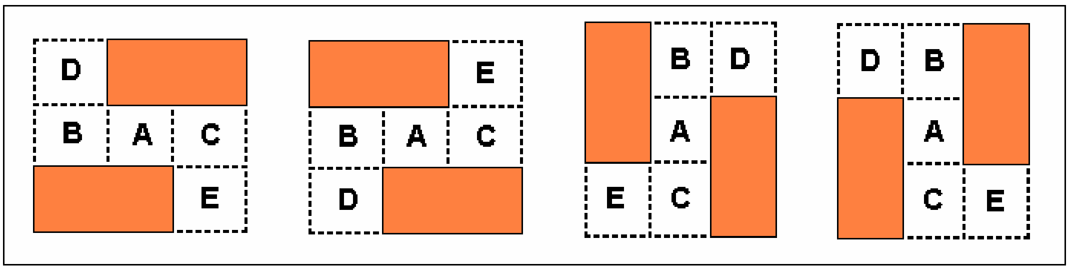

Figure 9

Either {D, B, A} could be used, {A, C, E} could be used. Note that A is included in the bridge, since by virtue of the threshold it is guaranteed to have a small area. When deciding between these remaining two bridge choices, the one with the smallest area is preferable. This maintains the requirement that the amalgam should fit the original buildings as tightly as possible.

6.3.3 Almost overhanging

If the previous case is taken to the extreme, the overhang becomes of zero width or height, and bridge A vanishes completely. These cases are comparatively rare, although it is necessary to ensure the algorithm can cope with them. Fortunately the logic for dealing with these circumstances is similar to when there is a partial overhang.

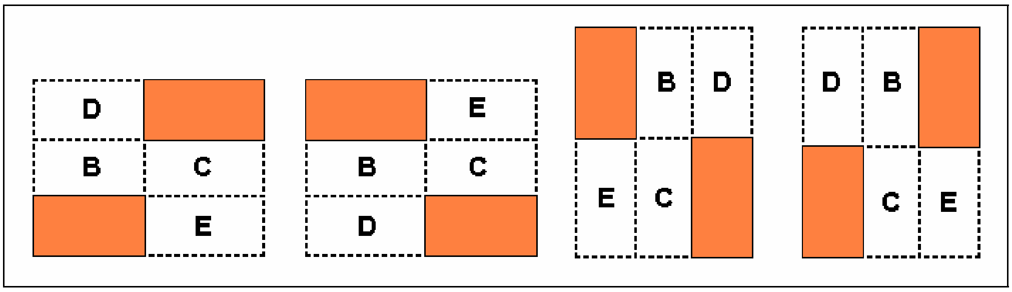

Figure 10

The first choice of bridge is the area {B, C}, as shown in Figure 10. If this bridge is too narrow, when compared against the threshold, then two choices remain. The bridges {B, D} or {C, E} are then evaluated and the bridge with the smallest area is selected.

6.3.4 Corners touching

The two cases shown in Figure 11 are even more rare than the almost overhanging case. Here there are two possible bridge choices shown by the dotted lines. The algorithm calculates the area of both possible bridges, and chooses the bridge which has the smallest area.

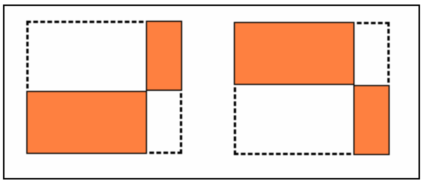

Figure 11

6.3.5 Total separation

The most complex case for bridging is when the two rectangles are separated with no overhang in either direction, as in Figure 12. First the width and height of area A are evaluated. If both width and height are greater than the threshold, then the smallest bridge of {B, A, C} or {D, A, E} is chosen. If only the width is less than the threshold, then {B, A, C} must be used. Similarly if only the height is less than the threshold, then {D, A, E} is the appropriate choice.

When both the width and height are less than the threshold, the bridge must be an L-shape either comprised {B, F, D} or {C, G, E}.

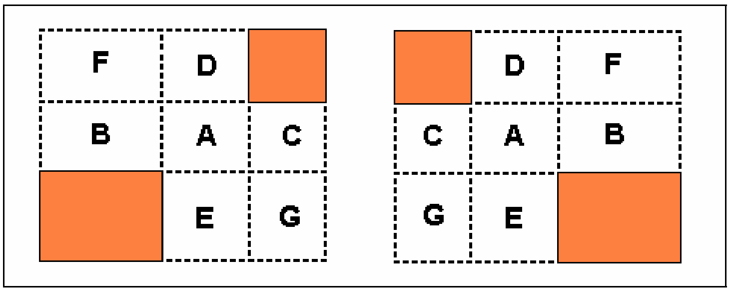

Figure 12

6.4 Re-orientate the amalgam

Once all the gaps between the rectangles have been bridged, the final step is to reorient the amalgam geometry back to the original orientation of the building group. The results of applying the algorithm to a four selections of OS MasterMap® buildings are shown in Figure 13. The final amalgam for the example in Figure 6 is shown in the top left corner. Note that the amalgams are near-optimal, although in some cases there is some empty space due to the positions of the original buildings.

Figure 13

The algorithm has been designed to work on any group of buildings, so it could be employed for generalising to other scales as well. For scales smaller than 1:50K, the rural building amalgams would become larger, up until the point where build-up regions can only be shown with solid fill. Hence there may be some use for it between 1:50K and 1:100K, although OS does not currently produce in this scale range. For the larger OS traditional scales such as 1:25K and 1:10K there is clear potential for using this algorithm in both rural and urban contexts.

7 Building granularity constraint

The granularity of the building amalgam is computed using the Clarity function to determine the shortest edge of the geometry. This measure originates from the AGENT project [AGENT DC1 1999]. This value is compared against the specification value for the side length of the minimum size square and a simplification algorithm is triggered to improve the compliance. The Clarity building simplification algorithm has been evaluated in this context and it has been found to have some shortcomings.

The main problem is that the simplified geometry no longer follows the positions of the original buildings, which is required by the OS specification. A constraint could be used to prevent this, by measuring the proportion of the original building areas contained inside the amalgam geometry and rejecting amalgams which are unrepresentative of the original buildings. However, an algorithm is still needed to produce an acceptable result.

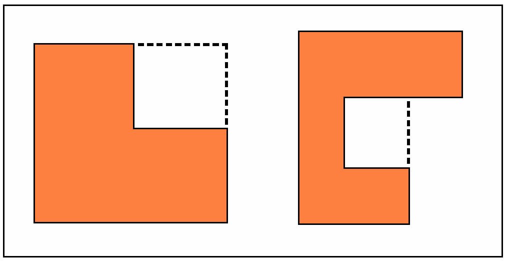

Figure 14

A new building simplification algorithm is currently being developed which will simplify according to OS requirements. The basic idea behind the algorithm is to expand concave corners and U-shapes on the perimeter by applying a rectangular patch to fill the hole, as shown in Figure 14.

The algorithm will iterate around the perimeter applying the smallest expansions first, until all edges are longer than the minimum side length, or there are no more concave corners and U-shapes.

It is quite possible that using such an algorithm will result in everything becoming a rectangle. A monitoring constraint could be used to check the concavity of the result, which is the ratio of its area to the area of its convex hull. A more elaborate constraint could measure how well the amalgam follows the original buildings, by calculating the portion of the original building areas inside the amalgam as a proportion of the total amalgam area.

In response to the monitoring constraints, the granularity constraint would need a second plan, which starts by enlarging the amalgam geometry slightly (about its centre of gravity), and then applies the new simplification algorithm. A lot of experimentation is required to obtain the desired result, and visual validation can be performed in Clarity by using existing OS 1:50K mapping as a backdrop.

8 Building proximity constraint

Once a suitable geometry has been found for the building amalgam, the last step is to ensure the building is positioned in an appropriate place on the map. The goal is to represent as best as possible the relationship that the original building group had with the surrounding features. Building amalgams can either be snapped to share geometry or displaced to satisfy the minimum separation specifications, depending on the position and context of the original buildings.

A building can interact with any other feature class, but for the moment analysis has concentrated on the inter-building proximity, the road/track proximity, the water proximity and the forest proximity. These are the main sources of proximity conflict for buildings at 1:50K and resolving these will pave the way for future work. The rural cluster meso agent will assume responsibility for its building amalgams, controlling the triggering of algorithms to improve the conformance to the specification.

8.1 Road/track proximity

The first building amalgams which should be placed are those where a connectivity is required with the surrounding roads/tracks. Examples of such buildings from existing 1:50K mapping are shown in Figure 15.

Buildings can be at the end of a cul-de-sac (sometimes wrapped around the end), at the side of a road/track, between two roads/tracks at a junction or entirely surrounded by two or more roads/tracks. Disjoint sections of the boundary can be shared with roads/tracks, and it is also possible for a building corner to touch the road/track casing. When part of a building boundary is a road/track, then the corner angles and shared geometry reflect the curvature of the track.

Figure 15

An algorithm is required which will move each building towards the roads/tracks, and clip its geometry so that it conforms with the shape of the road/track. It is likely that the algorithm will have to first decide on the context of the building, from the possibilities listed above, then apply a suitable operation depending on the situation. The granularity constraint should monitor the parts of the building which do not share the road/track casing, to ensure they are not reduced below the specification threshold.

8.2 Water and forest proximity

At the same time as connecting the building amalgams to the roads/tracks, the buildings may also need to share geometry with forest boundaries or with rivers/lakes. In this the forest boundary or water’s edge tend to be modified to fit snugly against the building shape. Examples of this can be seen in Figure 16.

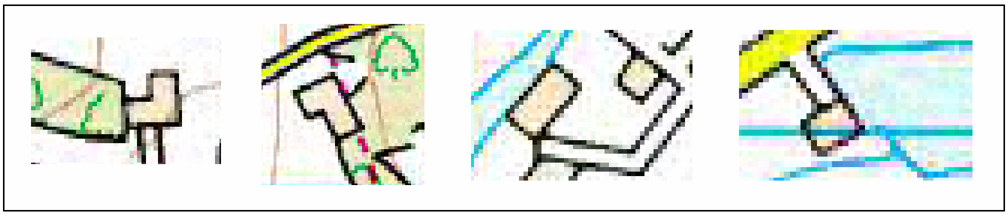

Figure 16

8.3 Inter-building proximity

Of the remaining building amalgams which have not been connected to anything, there are two possibilities. One choice is for the building to stand independent on the map, separated according to the minimum separation distances, from the surrounding features. The alternative is for the building to touch another building corner-to-corner. Examples of chains of buildings touching corner-to-corner can be seen in Figure 17.

Figure 17

9 Conclusion

This paper has given an overview of the progress made with using the Clarity environment for generalising rural buildings to 1:50K scale. In particular, rural buildings have been identified and broken down into manageable clusters. Within each cluster buildings have been grouped, with small groups either being deleted or enlarged to rectangles. For the remaining groups, a new building amalgamation algorithm has been developed.

In addition, plans have been presented for continuing the rural building 1:50K generalisation research. Future tasks include the development of a building simplification algorithm, plus algorithms which will connect or separate the buildings from their surrounding features.

A lot of progress has been made, and a familiarity has been gained with the new Agent core and with the cartographic style of OS 1:50K maps. The results from the work so far have been promising and a clear plan has been presented for the continuation this research. When complete, the work will provide an impressive demonstration of the benefits and feasibility of generalising OS specification data products automatically from OS MasterMap®.

10 References

AGENT (2000), *Project AGENT, ESPRIT/LTR/24939*, website.
http://agent.ign.fr/summary.html

AGENT DA2 (1998), *Constraint analysis, Report DA2 of the AGENT project*, ESPRIT/LTR/24939.
http://agent.ign.fr/deliverable/DA2.pdf

AGENT DC1 (1999), *Selection of basic measures, Report DC1 of the AGENT project*, ESPRIT/LTR/24939.
http://agent.ign.fr/deliverable/DC1.pdf

AGENT DD2 (1999), *Selection of basic algorithms*, Report DD2 of the AGENT project, ESPRIT/LTR/24939.
http://agent.ign.fr/deliverable/DD2.pdf

DUCHENE, C., BARD, S., BARILLOT, X., RUAS, A., TREVISAN, J., HOLZAPFEL, F. (2003), *Quantitative and qualitative description of building orientation*, paper presented at the 5th ICA workshop on progress in automated map generalization in Paris, France.
http://www.geo.unizh.ch/ICA/docs/paris2003/papers/duchene_et_al_v1.pdf

LASER-SCAN (2003), *Clarity technical information*.
http://www.laser-scan.com/pdf/Clarity_technical_english.pdf

ORDNANCE SURVEY (2000), *1:50 000 specification*, Ordnance Survey internal document.

ORDNANCE SURVEY (2004), *OS MasterMap®: definitive digital map of Great Britain designed by Ordnance Survey*, Ordnance Survey website
http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/

REGNAULD, N. (2002), *Agent redesign*, commissioned by IGN France, implemented by Laser-Scan.