Towards the Automatic Integration of Ground Level Images into a Virtual Urban Environment

R. G. Laycock
School of Computing Sciences
University of East Anglia
Norwich, UK, NR4 7TJ
robert.laycock@uea.ac.uk

A. M. Day
School of Computing Sciences
University of East Anglia
Norwich, UK, NR4 7TJ
amd@cmp.uea.ac.uk

ABSTRACT

In this paper we present a method which progresses towards automatically generating the texture maps for the building façades of an existing coarse three dimensional urban model. Rather than focusing on a method for achieving the exact texture map the algorithm creates two texture maps. These represent the building façade's wall material and building features, such as windows and doors. The paper describes two algorithms. The first defines an area in the image to be used as a repeatable wall pattern. It undertakes this procedure without prior knowledge of the wall material. The second algorithm presented creates an object map from the building image. The object map is an image containing the spatial arrangement of building features on a transparent background. These features are extracted from the building image automatically using a variant on an active contour model. The approach focuses on residential scenes where occlusions cause problems with existing automated techniques.

Keywords

Urban Modelling, Virtual Environments, Ground Level Images, Texture Mapping.

1. INTRODUCTION

Large urban environments are fundamental to many applications such as town planning, virtual tourism and environment simulation. When modelling small urban environments interactive modelling packages such as 3D Studio Max, [3dsa], are employed. These enable a user to generate photorealistic models albeit via a slow labour intensive modelling process. The need for the reconstruction of large virtual urban environments has prompted researches in the fields of remote sensing, computer vision and computer graphics to focus on techniques to aid this interactive process.

Overlapping aerial images may be used to automatically construct the geometry for a large urban environment, [Bai99a]. Alternatively the height map for a location may be obtained directly from the environment using LIDAR data, [Fru01a]. To progress towards a photorealistic render of the scene texture maps are applied to the building's façades. Overlaying an aerial image onto the existing geometric model is one method of undertaking this texture mapping procedure. However the result is less than desirable. Errors are prominent on the building façades where the aerial image is distorted. Furthermore due to the insufficient resolution of the aerial image, the rendered building façades will become pixilated when viewed in close proximity.

WSCG POSTERS proceedings
WSCG'2004, February 2-6, 2004, Plzen, Czech Republic.

To reduce these artifacts ground level images can be incorporated. Ground level images are cheaper to capture and have a higher resolution since they are focused particularly on the building façade. A system is required which when provided with an initial coarse three dimensional urban model has the ability to apply the textural detail to the building façades. This operation should be undertaken with minimal user interaction. Two vehicle-borne systems, documented in [Fru01a] and [Man01a], discuss this problem. Their solutions involve fitting laser scanners and CCD cameras to a vehicle. The vehicle is driven around the physical environment capturing depth and texture information simultaneously. The initial three-dimensional model is modified by the recorded depth maps and the texture maps are applied automatically. The disadvantage with this approach is that in order to maintain the speed of data capture the camera is fixed. Consequently texture information is available for what lies in the camera's field of view. This means that tall buildings will have parts left untextured. Furthermore whilst occluding objects, such as trees and fences, are removed via the scanned depth information occluded areas remain untextured.

A similar problem is present with multiple view geometry techniques. For example in [Fau98a], a three dimensional model plus its texture are captured automatically. By using this method only parts of the model visible in the images can be recovered.

In this paper the key ideas towards the development of a system for recovering the texture detail of an urban environment is presented. In particular focusing on residential scenes where occlusions cause problems with existing techniques.

2. PROCEDURE OUTLINE

The two data sources required by the system are an initial coarse three dimensional urban model and a set of ground level images. In this paper the method in [Lay03a] has been employed to automatically generate a three dimensional urban model. This method combines Ordnance Survey's Landline.Plus product with LIDAR data. The former provides building features, line segments, which are processed automatically to form building footprints. The LIDAR elevation data allows these building footprints to be extruded into three dimensional polyhedral models representing the urban environment. Figure 1 illustrates the urban model generated using this technique. The roof models are synthesized for visualisation purposes.

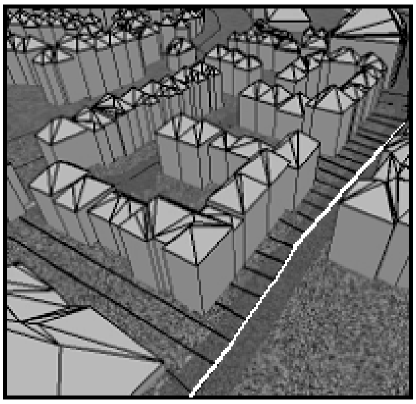

Figure 1: Mapping ground level images to buildings in the Virtual Urban Environment.

Included in the Landline.Plus data is a set of line segments representing the road centre lines. These can be automatically grouped to form a graph of the road network. The road network provides a mapping for a set of ground level images to the building façades. In figure 1 the road centre line is indicated in white for a selected street. As the street is traversed rays are sent perpendicular the tangent of the road centre line. They intersect the building façades indicating the order building façades appear as the street is traversed. A user may obtain an image of each building façade as they walk down the street. These are subsequently processed and applied to the building façade in the virtual urban environment. There are many important issues to resolve to enable this process to occur automatically. These include the identification of the building image in each of the ground level images, the registration of the building image to the geometry in the scene and the removal of occluding objects.

Currently the image registration problem is undertaken by selecting four points on the building façade and four corresponding points in the ground level image. The transformation matrix which warps the image onto the building façade can then be calculated. This process will resolve issues with perspective distortion and images taken at different scales. However it introduces significant user interaction and consequently is an area of future work.

It was stated previously that in residential scenes occlusions cause significant problems. Therefore to apply a single image to the building façade will introduce undesirable effects. These include the mapping of trees to the sides of buildings. In the UK a residential building can be partitioned into two main components. The first is the wall material such as brick, flint, concrete etc. The second component is the set of objects which reside in or on the wall material including doors, windows etc. Therefore an approach is suggested which partitions the texture map into two layers. The first called the wall map contains a repeatable wall pattern and the second is the object map. The object map contains the spatial arrangement of objects on a transparent background. These two texture maps are combined in hardware using multitexturing. The first texture unit on the graphics card contains the wall map and the second texture unit contains the object map. By conducting the process of image registration the texture coordinates for the object map can be determined. The wall map's texture coordinates are calculated automatically so that the scale of the repeating pattern corresponds to the scale of the object map.

The next two sections discuss the initial methods undertaken for the construction of the object and wall maps.

3. WALL MAP GENERATION

The following outlines the method that takes as input a ground level image and returns a wall pattern image.

Input: High Resolution Ground Level Image.

Step1: Scale the image into a 256*256 image maintaining the image's aspect ratio. Perform interpolation using the super-sampling interpolation scheme.

Step2: Construct an edge map using the Canny Edge Detect algorithm, [Can86a]. The result selects the important edge detail such as window, doors and the boundary of the building.

Step3: Process the edge map to obtain a set of non overlapping rectangles. Each rectangle must not contain any edge detail.

Step4: Group rectangles into sets using a greedy clustering algorithm. The similarity metric is the Euclidean distance between the hue channel of each of the rectangle's histograms.

Step5: Select the set of rectangles that has the most evidence of it containing a wall pattern.

Output: The largest rectangle in the wall pattern rectangle set identified in step 5.

Figure 2: The original image and the corresponding edge map.

Figure 2 illustrates the results of completing step 2 of the proposed algorithm. Steps 3 to 5 are conducted on the edge map. The aim being to identify an image sub region which can be used as the wall map. The assumption used is that a wall pattern on a building façade will contain very little edge detail after the image has been reduce in size. Consequently step 3 identifies a set of non overlapping rectangles which do not contain any edge detail. These rectangles form a candidate set which potentially contains wall pattern. These rectangles are separated into groups in step 4. This is based on the histograms of the image sub regions contained within the rectangles. Figure 3 illustrates the set of rectangles and the grey scale denotes the groups each rectangle is assigned to.

From the candidate sets the set is chosen in step 5 which represents the wall pattern. This is achieved using the structure of a building façade. A wall pattern is assumed to be a moderately large area containing many sufficiently large and convex holes or concavities. These holes and concavities being due to an object on or in the wall pattern surface, such as windows and doors.

Figure 3: The image illustrates the candidate rectangles.

For each rectangle a seed fill algorithm is performed, initialised from the rectangle's centroid. This process generates a boolean mask. The area can be tested to check for sufficiently large holes or concavities. Each set is tested to determine which rectangles are classified as containing wall pattern. The set with over 70% of the rectangles returning that they contain wall pattern is selected. The result of this procedure is shown in figure 4.

Figure 4: The final selected set of rectangles plus the extracted wall pattern.

4. OBJECT MAP GENERATION

In this section the outline for the object map generation procedure is presented.

Input: High resolution image plus the set of rectangles containing the wall pattern.

Step1: Construct a boolean wall mask by performing a seed growing algorithm. A seed is initialised at all the centres of the rectangles in the input set. The mask is created by adding pixels to the seeds if they are connected and within a threshold of 30 from the seed pixel's R, G and B components.

Step2: Initialise an active contour model, [Kas87a], surrounding pixels not included in the wall mask.

Step3: Minimise the energy of the active contour models. This will identify the features contained within the wall pattern.

Step4: Insert the objects identified in step 3 into the object map.

Output: An object map which is then superimposed onto the wall map in hardware.

Figure 5 illustrates a sub image region of one of the ground level images. Its illustrates the wall mask which has been generated by performing step 1 of the algorithm. The contour in step 2 is initialised at the holes or concavities of this wall mask. To complete the contour, at concavities, edges are added from the convex hull of the wall mask. Figure 5 illustrates these additional edges with dashed black line segments. As the internal energy of the contour is minimised the nodes move inwards around the object to be extracted. The external energy of the contour is defined such that it will be set to zero when any node intersects edge detail in the edge map.

Figure 5: The identification of the initial object polygons. The wall mask is illustrated in light grey and the initial object polygons in black. The dashed black lines represent the connected convex hull edges.

Once the features have been extracted they are positioned into the object map. The object map is subsequently superimposed onto the wall map. This is shown in figure 6.

Figure 6. The wall pattern and object map are merged using multitexturing on the graphics card.

5. CONCLUSION

In this paper a framework has been presented which enables the building façades in an initial coarse model of an urban environment to be textured. The approach requires a user to obtain ground level images which are automatically processed. Each building image is split into a wall pattern texture containing a tileable wall pattern and an object map containing the spatial arrangement of the building features on a transparent background. These are combined at run time using multitexturing to provide a fast method to texture mapping an urban environment enabling the appearance of the environment to be realised. By performing texture mapping in this way memory is reduced and many varied textures can be synthesised by combining the resulting object and wall texture maps. Thus providing variety to the scene in areas which have yet to have their ground level images captured from the real world location.

6. ACKNOWLEDGMENTS

The authors wish to thank EPSRC for funding the Rapid World Modelling Project, [Rap00a].

7. REFERENCES

[Lay03a] R. G. Laycock, A. M. Day, Automatically Generating Large Urban Environments based on the Footprint Data of Buildings, ACM Solid Modelling 2003, Seattle, June, 2003.

[3Dsa] 3D Studio Max, http://www.discreet.com/index-nf.html

[Kas87a] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active Contour Models, International Journal of Computer Vision 1, pp 321-331, 1987.

[Can86a] J. F. Canny, A Computational Approach to Edge Detection, IEEE Transaction on Pattern Analysis and Machine Intelligence, pp 679-698, 1986.

[Fau98a] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, I. Zoghlami, 3-D Reconstruction of Urban Scenes from Image Sequences, Computer Vision and Image Understanding, Volume 69, Number 3, March, pp 292-309, 1998.

[Bai99a] C. Baillard, A. Zisserman, Automatic Reconstruction of Piecewise Planar Models from Multiple Views, Proc. IEEE Computer Vision and Pattern Recognition, pg 559-565, 1999.

[Fru01a] C. Fruh, A. Zakhor, Fast 3D Model Generation in Urban Environments, International Conference on Multisensor Fusion and Integration for Intelligent Systems, Baden-Baden, Germany, August 2001.

[Man01a] D. Manandhar, R. Shibasaki, Feature Extraction from Range Data, Proceedings of ACRS 2001 – 22nd Asian Conference on Remote Sensing, 5-9 November 2001, Singapore, Vol 2, pp 1113-1118.

[Rap00a] Rapid World Modelling Project, http://www.cmp.uea.ac.uk/research/rwm/rwm1.html.