A Combined GIS and Stereo Vision Approach to Identify Building Pixels in Images and Determine Appropriate Colour Terms

Bartie, P 1, Reitsma, F 1, Mills, S 2

1 Department of Geography, University of Canterbury, Christchurch, NZ
2 Areograph Ltd (NZ), 90 Crawford St, Dunedin, New Zealand

Received: January 1, 2011

Abstract: Colour information is a useful attribute to include in a building’s description to assist the listener in identifying the intended target. Often this information is only available as image data, and not readily accessible for use in constructing referring expressions for verbal communication. The method presented uses a building polygon GIS layer in conjunction with street level captured imagery to provide a method to automatically filter foreground objects and select pixels which correspond to building façades. These selected pixels are then used to define the most appropriate colour term for the building, and corresponding fuzzy colour term histogram. The technique uses a single camera capturing images at a high frame rate, with the baseline distance between frames calculated from a GPS speed log. The expected distance from the camera to the building is measured from the GIS layer and refined from the calculated depth map, after which building pixels are selected. In addition significant foreground planar surfaces between the known road edge and building façade are identified as possible boundary walls and hedges. The output is a dataset of the most appropriate colour terms for both the building and boundary walls. Initial trails demonstrate the usefulness of the technique in automatically capturing colour terms for buildings in urban regions.

Keywords: GIS; Computer Vision; Stereo Depth Mapping; Colour Terms; Referring Expressions

© by the authorsLicensed under Creative Commons Attribution 3.0 License
1 Introduction

When talking about a place people like to include descriptive words to conjure up a pictorial representation in the listener’s imagination. Such feature descriptions are also often included in way finding instructions, such as the details of a building façade material or colour, as in “we’re the red brick house with the long white fence”. Capturing this level of detail has received a lot of attention in recent years, including initiatives such as Google’s Street View [56] and Microsoft’s Street Slide [33]. However texture information is only presented as imagery and not transformed into values suitable for use in cartographic symbology, or for inclusion in way finding feature descriptions. This paper presents a method whereby depth mapping is used to automatically filter foreground object from scenes, allowing the automatic extraction of colour terms to describe buildings.

Object descriptions are known in natural language research as “referring expressions” and are used, for example, to draw someone’s attention to a particular building in a cityscape [16]. They include visual clues which the speaker considers to be useful aids for the listener to determine which item in view is the intended target. The most useful terms are those which the listener can identify quickly, and limit the number of candidates rapidly without leading to any confusion. In some ways this is similar to determining landmark saliency, which focuses on ways to measure the prominence or distinctiveness of a building according to a number of factors, including visual, semantic, or structural attraction [52].

There are two main methods for extracting landmark candidates, by assigning a saliency score based on various attributes. Elias (2003) uses characteristics such as the building area, number of corners, density of buildings in the district, orientation to north and so on, while an alternative definition for saliency measurement was proposed by Raubal and Winter [44] which scores buildings according to Sorrows and Hirtle’s (1999) visual, semantic and structural characteristics. The visual factors include façade area, shape, colour, and visibility, translating well to the egocentric projective view experienced by street observers. These visual variables closely reflect Bertin’s [8] set of seven visual variables (position, orientation, size, colour, value, texture, and form) which should be considered when displaying graphical information. While traditional GIS datasets store position, orientation, and planimetric size of buildings, they fail to show information relating to building height, colour or texture. One of the challenges therefore is sourcing this information and making it accessible in a format suitable for use in constructing referring expressions. While LiDAR now offers a viable solution for capturing building height and form [42, 47], colour and texture details are either unavailable or stored in an inaccessible form, such as street level images.

The purpose of the research presented here is to offer a way using low-tech equipment available to most communities, to collect images from street level from which the foreground objects may be automatically filtered, enabling the remaining building pixels to be classified with a colour term. Information in this format may

www.josis.org
verbalised, and will be an important component for the future of speech interfaces whereby a user operating both hands-free and eyes-free may request information on a building in view, selected by its description. In addition a Location Based Service (LBS), such as a virtual city guide application [4], may direct the user’s attention to a specific building using a narrative which singles it out in the current view, simulating natural language descriptions. Emergency services would also benefit from technology to assist in locating people based on a description of their surroundings [18]. The paper begins with a discussion of colour as a descriptor in Section Error! Reference source not found., followed by examination of two methods to select the relevant building pixels from an image by ignoring foreground items in Section 2. These pixel values may be turned into relevant colour terms using fuzzy colour sets as discussed in Section 3. A trial of the proposed method is demonstrated in Section 4, as well as a brief outline of how this information may be used, including possible uses in the cartographic context.

2 Stereo Depth Mapping for Façade Colour Retrieval

There are a number of methods which may be used to recover depth information from images, including the Structure from Motion approach [32, 54], and the Stereo Vision method [9, 30, 37, 49]. The objective is to generate information on the distance from the camera to the real world object in each pixel, such that when combined with a GIS building layer those pixels which corresponding to buildings may be retrieved, and those from foreground objects ignored.

2.1 Structure from Motion Approach

Through combining many views of the same real world object from different distances and angles, it is possible to reconstruct the object’s structure. This is achieved through a processing pipeline which begins with matching pixels in at least 3 of the images. Pixel matching is an automated procedure, whereby points with high contrast gradients are found, such as SURF-based features [5]. These features are tracked between images and the 3D structure of the real world object is recovered, along with camera pose estimates.

The following images (Figure 1) were rendered from a model built using this Structure from Motion approach, whereby 26 images collected while walking in an arc around a property were processed using PhotoScan software [1].
There are a number of considerations for using this technique in urban regions. Firstly, a large number of images are required from a variety of viewing angles; a linear set of images from a single drive-by will not be sufficient to produce a detailed model. It can be rather difficult to obtain enough viewing angles in confined streets or when there are many moving pedestrians and cars. Additionally, processing times are fairly high, as feature matching, bundle adjustment [55], and geometry reconstruction are computationally intensive tasks. Finally, regions of low texture lack features for the matching process, resulting in poor depth estimates.

There have been a number of attempts to overcome these shortcomings, such as sourcing images from popular image sharing websites such as FlickR to build community volunteered virtual models [51], and improvements in depth reconstruction by imposing model restrictions such as enforcing planar surfaces [38].

www.josis.org
Currently this approach is best suited to capturing information in the more open urban spaces, a good example of which exists for Cathedral Square, in New Zealand [46]. The colour details in the 3D point dataset are maintained, so for instance the green top of the Cathedral Spire can be identified. The data may be rectified to fit the corresponding GIS layer as shown in Figure 2, allowing the retrieval of texture colour information for any part of a building. However for the purpose of collecting a colour classification this process is rather cumbersome, and an alternative simpler stereo vision technique may be used to capture colour information more easily.

![Green Pixels on Cathedral Spire](image)

*Figure 2: a) 3D Render of Cathedral Square, Christchurch captured by RedPaw (2008) b) Data rectified and reference for display on Google Earth (TM)*

### 2.2 Stereo Baseline Vision

Stereo baseline vision works by comparing two images taken parallel to each other but a short distance apart, known as the baseline distance. The technique requires the same feature to be identified in both images, so that the horizontal disparity may be measured. Larger disparities indicate objects are closer to the camera, as a result of distance parallax.

Computer vision techniques are used to identify stereo correspondences by locating interesting features (e.g. corners, edges) in one image, for example using Harris [25] or SURF [5] definitions, and search the paired image for the most similar matching template. The automatic process produces a disparity map, which can be
transformed into a depth map with knowledge of the camera’s focal length, pixel size, and the baseline distance between the image pair [11].

A recent addition to Google Streetview is a 3D mode whereby depth information is rendered as an anaglyph giving the viewer the opportunity to perceive depth when wearing red/cyan glasses [3]. Here the red channel is used to convey the secondary camera view. However Google’s façade depth details, collected using either optical or laser range finding techniques, are not publically available, nor is the stereo view accessible via the Google Maps API for automated local reconstruction.

In order to collect a sample dataset to trial colour extraction without foreground objects in a suburban region, a simple stereo rig was built from two low cost webcams set a baseline distance of 60cm apart. It is necessary to perform a one off calibration of the stereo camera setup to calculate various intrinsic and extrinsic details, allowing lens distortions to be removed from future images. We did this by capturing multiple views of a flat chessboard pattern using Emgu.CV [13], a C# implementation of OpenCV [27]. Initial static trials of the rig showed that depth could be recovered successfully up to a distance of around 30 metres; however the webcams suffered from various distortions when moving at 50km/h due to their slow shutter speeds, rendering the images useless for depth mapping.

Instead a single high quality video camera with optical stabilisation was used to capture 25 images per second with a shutter speed of 1/500th second to ensure sharp images without motion blur. The camera was fixed perpendicular to the direction of travel, and a GPS device was used to log speed, orientation, and location at 1Hz. This allowed the baseline distance between subsequent video frames to be calculated, which for a car travelling at 50km/h would be 55cm. The advantage of this setup is that it is extremely simple to implement, requiring only intrinsic details for a single lens, and uses technology available to a wide audience, potentially allowing it to be implemented on public transport vehicles such that urban colour datasets could be regularly updated.

GPS speed is considered to be accurate to within 0.2m/s (0.72km/h), and is calculated using Doppler shift making it more robust in multipath environments than full location triangulation [57]. At a frame rate of 25 images per second the speed accuracy would constitute a baseline discrepancy in the region of 0.8cm, resulting in negligible depth inaccuracies. To ensure the most accurate baseline estimate for any frame pair, the speed was interpolated from the surrounding GPS measurements, as shown in Figure 3. Here a captured frame (801) falls between two GPS readings, resulting in an interpolated speed, from which the baseline distance may be calculated (ground distance between frames 800 and 801) so that disparity information may be translated into real world depth units (metres).
The relationship between disparity and depth is inversely proportional, and being non-linear high depth resolution is only available for objects near the camera. The equation is given as,

\[ Z = \frac{F \times B}{D} \]

where \( Z \) is the distance from the camera to the house, \( F \) is a constant for the camera lens setup, \( B \) is the baseline distance moved between frames, and \( D \) is the disparity measured between features in the image [11].

An example of this relationship for a camera travelling at 50km/h is shown in Figure 4. Notice that for this setup an object at a distance of 26 metres from the camera would have a disparity of 28.6 pixels between left and right images, while an object 27 metres away would be 27.6 pixels. This is the limit at which a 1 metre change in distance can be measured at the pixel level for this camera. The working depth range can be increased by using a larger baseline which can be accommodated with this technique by using a 2 frame offset between left and right images, extending the 1m depth resolution limit to a distance of 38 metres. A 3 frame offset extends the search resolution to 46m from the camera, however greater offsets result in less similar foreground views, rendering it more difficult to find stereo correspondences for closer items [49].
A slower moving car will result in a shorter baseline between images, dropping depth resolution for distant items. This can be overcome automatically by calculating the current forward speed for each frame and the expected distance to the target building from a GIS layer, and then choosing the lowest frame offset suitable for the required depth, thereby balancing the highest level of stereo correspondence matching and depth resolution.

![Figure 4: Pixel Disparity at Various Distances for a Camera Travelling at 50kmh](image)

**2.3 Sampling Strategy**

The sampling strategy for this project was to consider views from directly in front of buildings to be the most suitable starting point, with no attempt to model the building’s visibility from the road. The distance to the building was measured from the known camera location and a GIS building layer, allowing image pixels at matching depths to be selected. However it proved necessary to estimate also the horizontal position of the building within the image frame, by projection the building’s edges into the image based on the camera’s known Field of View (FOV) and object distance, so neighbouring houses and other objects at similar distances could be filtered out. The camera’s FOV information was calculated during the initial camera calibration stage and remains constant throughout data capture, as the lens is fixed at infinite focus on
the widest zoom setting. The current viewing direction is fixed at a 90 degree offset from the GPS orientation collected from the direction of travel.

The first sample for each building was selected from the video stream at the road location aligned to the building midpoint along the road axis, with subsequent samples radiating outwards from this point offset a frame in either direction, forming a sampling sequence as shown in Figure 5. Each sample location was only considered valid if the number of pixels successfully placed within the expected house location was above a given threshold. For our trials this threshold was set to 100 pixels, deemed the minimum required to give a fair reflection of the house colour. To reduce the sampling time a limit of ten successful sample locations was set per target, after which the next target would be selected, thereby ensuring at least 1000 pixels were sampled per building. In Figure 5 the locations A and B would be expected to give the highest pixel counts for each building, however where high fences or vegetation restricted the view of the target the most successful samples were collected across driveways, such as location C between houses.

*Figure 5: Estimating House Position in Image*
2.4 Improving System Robustness

To accommodate any system noise, from errors introduced in the GPS, GIS, or baseline calculations a further step was added to the processing pipeline. As buildings generally have planar walls it is possible to search the depth map for planar surfaces in the expected location. This may be done by passing a 3 by 3 kernel over the depth map to compare the depth gradients between cells, identifying regions of constant value, as depicted in Figure 6. By using a recursive function it was possible to identify significant planar candidates with similar gradient properties, merging results to define the larger planar surfaces, which in turn were filtered to leave only those vertical and facing the camera.

![Figure 6: Detecting Planar Surfaces using a 3 by 3 Kernel](image)

The depth map may also be summarised vertically (against the x-axis) to generate a frequency graph showing the most commonly occurring depths. Vertical planar surfaces result in high frequency depth counts, as a high proportion of the column exhibits the same depth value. A linear stretch of similar high frequency values across the graph (y-axis) indicates a wide planar surface, such as a building or wall. Figure 7 shows an example of the histogram, where only significant values are displayed by applying a low pass filter to remove background noise, superimposed and aligned onto the corresponding image. This depth summary was used to refine the expected depth value for the target building, in addition a depth tolerance of 1.5 metres was permitted in all cases, judged to be shallow enough to exclude cars parked in front of target buildings.
Figure 7: Using Depth Frequency Histogram to Refine Selection of Target Building Search

The process also allows foreground boundary wall information to be gathered from the images by searching for planar surfaces at distances between the road edge and target building. The search logic is that boundary walls would be located in the lower part of the image, as shown in Figure 3, following the order road, pavement, boundary wall, house wall, roof, sky. By comparing the presence of any linear features identified in the depth histogram (Figure 7), and GIS road boundary polygons, it was possible to limit the image search to find large flat surfaces likely to be fences, walls, and hedges at the property boundary. An example of the output from each stage of the process from generating the disparity map is shown in Figure 8, with the most likely boundary planar surface highlighted in red. The topic of identifying house and boundary material (e.g. vegetation, brick, wood, fence) is discussed, with the results, in Section 4.
2.5 Windows

One of the issues when automatically recovering building colour histograms is the presence of windows, which have no colour of their own but either allow light to pass into the interior, or reflect the surroundings. There are methods which can be used to
automatically determine window locations based on gradient projection approaches [45], however these require the extent of the façade to be pre-defined in each image. A more simplistic approach is to filter the image for sky hues in HSL colour space, thereby identifying those surfaces reflecting the sky. The pixels selected with similar colour may be removed from those considered as façade, leaving a smaller set of candidates for the colour analysis, but with a higher likelihood that the pixels being considered were wall pixels. It was found that on both bright sunny days and bright overcast days the procedure worked fairly well, although future research should look to implement more sophisticated procedures to identify windows under all lighting conditions.

![Image of house with sky hue filter applied]

Figure 9: Avoiding window regions from selection using sky hue filter

2.6 Shadows

The colour summary is also complicated by shadow regions, which are darker patches resulting from changes in lighting as a result of surrounding features [19]. A Retinex filter [35] may be applied to images to reduce the effect of illumination variation, as shown in Figure 10. Here an image is reduced to use the closest of only 11 colours before and after the Retinex filter is applied, showing an improvement in colour matching for regions in shadow (such as trees) after processing.

Although the Retinex process improves the colour classification, some illumination artefacts remain in the image (e.g. strong shadows show up as sky blue), and further shadow reduction techniques may prove beneficial [21, 29, 48]. The issue of strong shadows is reduced by limiting data capture to bright overcast days, when the lighting source is more diffuse.
2.7 Processing Pipeline

The processing pipeline is outlined in Figure 11, and begins by identifying a target building and finding the nearest GPS point from the data collected. Next the corresponding image frame is retrieved and the expected distance to the house measured from the GIS building layer. The current speed is interpolated from nearby GPS points, and the most appropriate baseline distance between frames determined allowing the disparity map to be transformed into a depth map (Section 2.2). The expected building depth value is then refined using the depth histogram approach (Section 2.4), before building pixels are selected. The image is then filtered for sky hue reflections, and those pixels dropped from the selection (Section 2.5). If the pixel count remains above the threshold the image is considered suitable for inclusion in the classification, otherwise the next image in the sequence is processed. Before classifying the Retinex filter is applied, and the image blurred and dilated slightly to remove pixel colour noise. The selected pixel colour values are saved as an array with the target building identification number, and the process repeated until ten locations are collected for each target building. In some cases it was impossible to collect enough good data for a target and so these were marked as irretrievable.
At this point in the processing pipeline two pixel groups were identified, one for the most likely house pixels and the other for any likely boundary wall. The final stage of the process was to determine the most appropriate colour term for each of these groups, as explained in the next section.

![Processing Pipeline](image)

**Figure 11: Processing Pipeline**

### 3 Colour terms

The mapping of pixel values to colour terms is rather complex, as the perception of colour is related to many factors [34]. There are issues of chromatic induction [28], which is when similar hues are judged to be different due to the contrast with surrounding colours, and also effects of lighting where similar colours appear very differently depending on shadows cast onto the surface. Furthermore colour terms describe regions in colour space which are only vaguely defined, and vary depending upon the viewer.

There are many colour terms which are highly specialised and not in general use by the public (e.g. chartreuse). It is therefore necessary to determine first a list of the most popular terms and rate each building against these to generate a fuzzy colour classification [6]. The results from such a fuzzy classification allow an object to be
identified using a range of terms, such that one user may calls it ‘red’ while another calls it ‘orange’, with both terms having fairly strong membership values.

Previous research highlights 11 main perceptual colour foci [7, 53], which are black, white, red, green, yellow, blue, brown, purple, pink, orange, and grey. Boynton and Olson [10] found that 424 subjects could repeatedly consistently identify the red, green, yellow, blue, orange, purple, brown and pink foci without any confusion. These 11 foci were adopted for this research, with each term defined in colour space according to the HP colour thesaurus, an online databank of defined colour centres constructed from people around the world [39]. For example the most widely used red definition uses Red Green Blue values 216,35 and 44 respectively.

The comparison between a selected building pixel and each foci was conducted by calculating the Euclidean distance in HSL (i.e. Hue, Saturation, Lightness) colour space. This is a transformation of colour space recognised to more closely match the human eye’s perception of colour similarity than the more common RGB colour space. The process was repeated for all 11 colours giving a fuzzy classification for the building sample against the colour terms. This was repeated for the 10 samples for each target building, and the classifications were summed to produce a single building fuzzy colour set.

### 3.1 Colour Entropy

Colour entropy is a measure of the ambiguity of the assigned colour[15], which may be measured by counting the number of fuzzy classes with significant values (Figure 12). In cases where a building has a single strong classification the colour entropy is low, and the generated natural description may include a single colour term. However where colour entropy is high a number of colour terms may be required, such as “the red-ish orange building”. This fuzzy classification accommodates variation in user-formed descriptions, such that inclusion of “red” or “orange” would give the same search results. Where many buildings in view match a given selection criteria other classifications may be required to narrow the results, such as building size, or roof colour.
Determining Building Pixels using GIS and Stereo Vision

4 Trials and Results

The system was trialled in a number of streets within Christchurch (NZ) during the summertime. The central city is grid based with a large suburban expanse, much vegetation and many gardens, providing a suitable test environment for identification and filtering of house façades from behind foreground objects such as parked cars, trees, bushes and people.

Figure 13 shows a selection of the identified façades, with corresponding fuzzy colour classifications. Notice that despite fairly limited views house colour could still be recovered from the views across driveways (A), and also where foreground vegetation was present (B, C). Roofs were not included in selections as a result of being non-vertical planar surfaces, and could more easily be captured from aerial imagery. Garage doors proved to be an issue however (C, D) and were generally ignored for the classification as stereo depth mapping failed to produce stable results on the very similar textured surfaces. Therefore these regions tended to contain rather patchy irregular distance values, and were filtered out from selection as non-vertical planar surfaces.

Despite efforts to reduce selections on window surfaces there were occasions where pixels were selected (D). However on most occasions the non-window façade pixels dominated the selection, rendering the window pixels insignificant in the colour classification.

To evaluate the system’s performance a comparison was carried out with a sample collected from a walk along a number of streets noting house colour information. It was found that the automated procedure was able to correctly identify the most prominent colour of 33 out of 43 houses (77%) when compared to these manual values. The occasions it failed were mostly a result of incorrect colour term classification due to shadows, rather than incorrect pixel selection.
Figure 13: Example Façade Determination and Colour Classification

In addition boundary wall information was gathered by selecting linear planar features. Boundary wall locations were identified correctly in 36 of the 43 targets (84%). An attempt was made to classify the type of boundary (brick wall, fence, vegetation etc) which proved to be a much more difficult task.

Vegetation could be detected due to the dense high counts of Harris corners [25], while slat fences could be found by applying a Gaussian blur to the image before running the Canny edge detector [14]. The part of the image corresponding to the boundary was then scanned horizontally tallying the number of intersections with detected Canny edges. Regions with a low standard deviation were deemed to be likely slat fence candidates, as shown in Figure 14.
Material recovery is a difficult task [41], and during trials vegetation and slat fence were the only reliable classifications which could be made. Therefore the boundary material types were divided into wall/fence, vegetation and slat fence. Vegetation shadows sometimes caused incorrect classifications, particular when cast onto plain painted fences. A summary of the target houses processed are shown in Figure 15, including the graph of depth used to first identify the presence and location of vertical linear features in front of the house, the main planar regions, and the results of the Harris corner detection and Canny filter process.
4.1 Mapping Symbology

Colour information for roof, wall and boundary could be presented on large scale city maps, or GIS layers, to assist in navigation tasks. An example of a 1:2500 and a 1:5000 scale map are shown in Figure 16, demonstrating how this may be achieved. Although perhaps initially counter-intuitive the 1:2500 scale map shows wall colour as the polygon fill colour, and roof colour on the border. This was considered more appropriate than the alternative as wall colour is the predominant colour viewed from the street, and most relevant in building descriptions and locating tasks.

As an example of how this information may be used, an ambulance searching for a particular address would be able to see very clearly that they are searching for a blue
house, or perhaps the yellow house next door to the blue house, or the white house opposite. Way-finding tasks should therefore benefit from displaying this additional contextual information for a neighbourhood.

Figure 16: Cartographic Design Example for Showing House Colour Information

5 Conclusion and Future Work

Human formed descriptions of buildings often include references to attributes which are lacking in our spatial datasets, such as façade colour and texture information. While in recent times there has been a dramatic rise in the capturing of this texture and colour information from street level, the information is locked up in 2D image formats restricting its usefulness beyond visualisation. To extract building colour terms, for use in forming natural language descriptions and verbalisation, requires the ability to identify which pixels correspond to buildings by filtering out foreground objects. These pixels may then be summarised to generate the most appropriate colour term for use in forming the building’s description.

The work presented here demonstrates how this may be achieved using a low-tech hardware solution, available to most communities, in conjunction with spatial
analysis on a building polygon layer. The processing pipeline begins by determining the expected distance from the camera to the building by analysing the GIS dataset. Computer vision techniques are used to build a disparity map from images captured from a moving camera, and converted to a depth map once the baseline distance between frames is determined from GPS speed information. From these inputs it is possible to filter the image for pixels at the expected building distance and generate a building colour summary. Although a full evaluation of colour perception was beyond the scope of this work, the method of using a single moving camera and GPS unit to build stereo views was able to generate meaningful colour descriptions, and also to identify boundary walls and fences in images. There were however a number of issues which require attention in future versions of the system.

The use of a single camera limits the system’s operation to straight road sections, as when navigating corners the rotation of the camera means parallel image views are not available. This was only a minor problem for our tests in a grid based city, but could be a significant issue in other regions. To overcome this restriction the Structure from Motion approach may be used to recover an estimate of the camera pose and orientation from which pixel depth information may then be calculated, although at additional processing cost. Alternatively a stereo camera could be used, but the operating range would be limited by the fixed baseline.

Colour term retrieval was influenced by the presence of strong shadows in images, and although the Retinex filter stage of the processing pipeline improved the situation it became evident that more sophisticated approaches would beneficial in some cases. New techniques such as entropy minimisation [20] may be worth considering for future versions of the system.

The ability to retrieve material types for both boundary walls and house facades would be beneficial when generating natural descriptions. However this proved to be difficult, for example although bricks are uniform shapes they occur in a range of colours. In theory shape detection methods would be suitable to identify them, but unfortunately Canny edge detection proved unreliable at the operating distances required. Higher quality cameras and lenses would improve this, but shadows are often cast on to the solid surfaces and may still impede these edge detection methods. Slat fences were one of the most easily recognised boundary types, as the edges were clearly defined even under strong shadow.

Currently only a single colour definition is collected per building facade, or fence, however some buildings have different coloured side walls, or a multiple boundary types (eg low wall with a fence above). By exploiting the planar surface details it would be possible to divide target buildings into sections based on the direction the wall faces, and map these colour values to the GIS layer, thereby creating colour summaries for subsections of the target building. In addition the vertical boundary regions could be subdivided into those sections which appear most wall-like, and most vegetation-like to derive more complete descriptions.
Garage doors are often different colours to the main house, and would be useful additions to the descriptors. Currently they tended to be regions ignored from the summary, due to difficulties in retrieving stable depth value for such similar textured regions. In addition window frames and doors may be segmented in the image for separate colour analysis, giving rise to very detailed house description possibilities. It is possible that windows may be identifiable as regions which change appearance with direction across views, therefore tracking a lack of consistency could be used to detect their locations.

By automating the process it is possible to imagine a future whereby such implementations could be installed on public transport (e.g. buses, taxis), or domestic car volunteers, to ensure regular updates to a building colour database.

6 Acknowledgements

The authors would like to thank Dr Cynthia Brewer and Dr William Mackaness for their contributions during the concept phase of this paper.

7 References

1. Agisoft. PhotoScan, 2010.
2. Amri Musliman, I., Alizadehashrafi, B., Chen, T.K. and Abdul-Rahman, A. Modeling Visibility through Visual Landmarks in 3D Navigation using Geo-DBMS. Developments in 3D Geo-Information Sciences(2010). 157-180.
3. Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L. and Weaver, J. Google Street View: Capturing the World at Street Level. Computer, 43 6 (2010). 32-38.
4. Bartie, P.J. and Mackaness, W.A. Development of a speech-based augmented reality system to support exploration of cityscape. Transactions in GIS, 10 1 (2006). 63-86.
5. Bay, H., Ess, A., Tuytelaars, T. and Gool, L.V. Speeded-up robust features (SURF). Computer Vision and Image Understanding, 110(3) 3 (2008). 346–359.
6. Benavente, R., Vanrell, M. and Baldrich, R. A data set for fuzzy colour naming. Color Research & Application, 31 1 (2006). 48-56.
7. Berlin, B. and Kay, P. Basic Color Terms: Their Universality and Evolution. Univ. of California Press, Berkeley, 1969.
8. Bertin, J. Semiology of graphics. Madison: University of Wisconsin Press, 1983.
9. Birchfield, S. and Tomasi, C. Depth discontinuities by pixel-to-pixel stereo. International Journal of Computer Vision, 35 3 (1999). 269-293.
10. Boynton, R.M. and Olson, C.X. Salience of chromatic basic color terms confirmed by three measures. Vision Research, 30 9 (1990). 1311-1317.
11. Bradski, G. and Kaehler, A. Learning OpenCV. O’Reilly, Sebastopol, CA, 2008.
12. Brown, P.J., Bovey, J.D. and Xian, C. Context-aware applications: from the laboratory to the marketplace. Personal Communications, IEEE [see also IEEE Wireless Communications], 4 5 (1997). 58-64.
13. Canning. Emgu.CV, 2010.
14. Canny, J. A Computational Approach To Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligenc, 8 6 (1986). 679-698.
15. Chuang, J., Stone, M. and Hanrahan, P. A Probabilistic Model of the Categorical Association between Colors Color Imaging Conference, 2008.
16. Dale, R. and Reiter, E. Computational interpretations of the Gricean maxims in the generation of referring expressions. Cognitive Science, 19 2 (1995). 233-263.
17. El-Hakim, S.F., Brenner, C. & Roth, G. A multi-sensor approach to creating accurate virtual environments. Journal of Photogrammetry and Remote Sensing, 53(1998). 379-391.
18. Fabrikant, S., Reichenbacher, T., van Kreveld, M., Schlieder, C., Le Yaouanc, J.-M., Saux, É. and Claramunt, C. A Visibility and Spatial Constraint-Based Approach for Geopositioning, in Geographic Information Science, Springer Berlin / Heidelberg, 2010, 145-159.
19. Finlayson, G., Hordley, S. and Drew, M., Removing Shadows from Images. in ECCV 2002: European Conference on Computer Vision. Lecture Notes in Computer Science (2002), 823-836.
20. Finlayson, G.D., Drew, M.S. and Lu, C. Entropy Minimization for Shadow Removal. International Journal of Computer Vision, 85 1 (2009). 35-57.
21. Finlayson, G.D., Hordley, S.D. and Drew, M.S., Removing shadows from images using retina. in Color Imaging Conference, (2002), 73-79.
22. Forsyth, D., Torr, P., Zisserman, A., Korah, T. and Rasmussen, C. Analysis of Building Textures for Reconstructing Partially Occluded Facades. in Computer Vision – ECCV 2008, Springer Berlin / Heidelberg, 2008, 359-372.
23. Frueh, C., Sammon, R. and Zakhor, A. Automated texture mapping of 3D city models with oblique aerial imagery. (2004).
24. Han, J. and Ma, K.K. Fuzzy color histogram and its use in color image retrieval. IEEE Transactions on Image Processing, 11 8 (2002).
25. Harris, C. and Stephens, M., A combined corner and edge detector. in Proceedings of the 4th Alvey Vision Conference, (1988), 147–151.
26. Helvacioglu, E. and Olguntürk, N. Colour contribution to children’s wayfinding in school environments. Optics and Laser Technology, 43 2 (2009). 410-419.
27. Intel. OpenCV, 2010.
28. Jameson, D. and Hurvich, L.M. Essay concerning color constancy. Annual Review of Psychology, 40 1 (1989). 1-24.

www.josis.org
29. Jianhong, G., Lu, L. and Peng, G. Removing shadows from Google Earth images. *Int. Journal of Remote Sensing*, 31 6 (2010). 1379-1389.
30. Kanade, T., Yoshida, A., Oda, K., Kano, H. and Tanaka, M., A stereo machine for video-rate dense depth mapping and its new applications. in, (2002), IEEE, 196-202.
31. Kataoka, K.a.O., T. and Ozawa, S. and Wakabayashi, K. and Arakawa, K., Building Façade Model Reconstruction Using Parallel Images Acquired by Line Scan Cameras. in *Image Processing*, (2005), I-1009-1012.
32. Koenderink, J.J. and Van Doorn, A.J. Affine structure from motion. *Journal of the Optical Society of America A*, 8 2 (1991). 377-385.
33. Kopf, J., Chen, B., Szeliski, R. and Cohen, M.F., Street Slide: Browsing Street Level Imagery. in *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010)*, (2010), 1-8.
34. Lammens, J.M. and Shapiro, S.C. Learning symbolic names for perceived colors. *Machine Learning in Computer Vision: What, Why and How*(1993).
35. Land, E.H. The Retinex Theory of Colour Vision. *Scientific American*(1977). 108-129.
36. Lensch, H.P.A., Heidrich, W. and Seidel, H.P., Automated texture registration and stitching for real world models. in, (2000), 317-452.
37. Lucas, B.D. and Kanade, T., An iterative image registration technique with an application to stereo vision. in, (1981), Citeseer, 674-679.
38. Micusik, B. and Kosecka, J., Piecewise planar city 3D modeling from street view panoramic sequences. in, (2009), IEEE, 2906-2912.
39. Moroney, N. HP’s On-line Color Thesaurus, 2010.
40. ODPM. Final report for signage and wayfinding for people with learning difficulties, 2006.
41. Pajdla, T., Matas, J., Hayman, E., Caputo, B., Fritz, M. and Eklundh, J.-O. On the Significance of Real-World Conditions for Material Classification. in *Computer Vision - ECCV 2004*, Springer Berlin / Heidelberg, 2004, 253-266.
42. Palmer, T.C. and Shan, J.A. Comparative study on urban visualization using LIDAR data in GIS. *URISA Journal*, 14(2002). 19-25.
43. Pylvänäinen, T., Roimela, K., Vedantham, R., Itäranta, J., Wang, R. and Grzeszczuk, R., Automatic Alignment and Multi-View Segmentation of Street View Data using 3D Shape Priors. in *3DPTV 2010*, (Paris, France, 2010).
44. Raubal, M. and Winter, S., Enriching Wayfinding Instructions with Local Landmarks in *Geographic Information Science: Second International Conference, GIScience 2002, Boulder, CO, USA, September 25-28, 2002. Proceedings*, (2002), Springer Berlin / Heidelberg, 243-259.
45. Recky, M. and Leberl, F., Windows Detection Using K-means in CIE-Lab Color Space. in *ICPR*, (Istanbul, Turkey, 2010), 356-360.
46. RedPaw. Cathedral Square (12mm), 2008.
47. Rottensteiner, F. and Briese, C. A New Method for Building Extraction in Urban Areas from High-Resolution Lidar Data. International Archives of Photogrammetry and Remote Sensing, 34 3 A (2002). 295-301.

48. Scanlan, J.M., Chabries, D.M. and Christiansen, R.W., A shadow detection and removal algorithm for 2-d images. in, (2002), IEEE, 2057-2060.

49. Scharstein, D. and Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision, 47 1 (2002). 7-42.

50. Seaborn, M., Hepplewhite, L. and Stonham, J., Fuzzy colour category map for content based image retrieval. in, (1999).

51. Snavely, N., Seitz, S.M. and Szeliski, R., Photo tourism: exploring photo collections in 3D. in, (2006), ACM, 835-846.

52. Sorrows, M. and Hirtle, S., The nature of landmarks for real and electronic spaces. in Spatial information theory. Lecture notes in Computer Science, (1999), Springer, Berlin, 37-50.

53. Sturges, J. and Whitfield, T.W.A. Locating basic colours in the Munsell space. Color Research and Application, 20 6 (1995). 364-376.

54. Sturm, P. and Triggs, B. A factorization based algorithm for multi-image projective structure and motion. Computer Vision—ECCV’96(1996). 709-720.

55. Triggs, B., Zisserman, A., Szeliski, R., Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A. Bundle Adjustment — A Modern Synthesis. in Vision Algorithms: Theory and Practice, Springer Berlin / Heidelberg, 2000, 153-177.

56. Vincent, L. Taking online maps down to street level. Computer 40 12 (2007). 118–120.

57. Witte, T.H. and Wilson, A.M. Accuracy of non-differential GPS for the determination of speed over ground. Journal of biomechanics, 37 12 (2004). 1891-1898.

58. Wurm, L.H., Legge, G.E., Isenberg, L.M. and Luebker, A. Color improves object recognition in normal and low vision. J Exp Psychol, 19(1993). 899-911.

59. Yi-Leh, W., Cheng-Yuan, T., Maw-Kae, H. and Chi-Tsung, L. Automatic Image Interpolation Using Homography. EURASIP Journal on Advances in Signal Processing(2010).

www.josis.org