

AUTOMATICALLY DETERMINING SCALE WITHIN UNSTRUCTURED POINT CLOUDS

Jayren Kadamen^a, George Sithole^b

^a Dept. of Geomatic Engineering, School of Architecture, Planning and Geomatics, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa, Email: george.sithole@uct.ac.za

^b Dept. of Geomatic Engineering, School of Architecture, Planning and Geomatics, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa, Email: kdmjun001@myuct.ac.za

Commission III, WG III/4

KEY WORDS: Unstructured Point Cloud, Monocular Camera, Real-world Scale, Smart-phones, Simplified Object-Recognition

ABSTRACT:

Three dimensional models obtained from imagery have an arbitrary scale and therefore have to be scaled. Automatically scaling these models requires the detection of objects in these models which can be computationally intensive. Real-time object detection may pose problems for applications such as indoor navigation. This investigation poses the idea that relational cues, specifically height ratios, within indoor environments may offer an easier means to obtain scales for models created using imagery. The investigation aimed to show two things, (a) that the size of objects, especially the height off ground is consistent within an environment, and (b) that based on this consistency, objects can be identified and their general size used to scale a model. To test the idea a hypothesis is first tested on a terrestrial lidar scan of an indoor environment. Later as a proof of concept the same test is applied to a model created using imagery. The most notable finding was that the detection of objects can be more readily done by studying the ratio between the dimensions of objects that have their dimensions defined by human physiology. For example the dimensions of desks and chairs are related to the height of an average person. In the test, the difference between generalised and actual dimensions of objects were assessed. A maximum difference of 3.96% (2.93cm) was observed from automated scaling. By analysing the ratio between the heights (distance from the floor) of the tops of objects in a room, identification was also achieved.

1. INTRODUCTION

It has become increasingly popular to acquire point clouds of real-world scenes by photogrammetric means. This can be attributed to advancements in multiple-image reconstruction algorithms, increased processing power of consumer-grade computers and the relatively inexpensive cost of cameras compared to laser scanners. The widespread penetration of smart-phones has equipped billions of users (Richter, 2012) with a high quality imaging platform at their fingertips. There are also new platforms entering the market that make use of a single camera to capture a scene; these include commercial drones which have a growing market-share. The field of computer vision has made use of unstructured point clouds for indoor mapping purposes. This is because cameras are often chosen over active systems such as laser scanners because they require less power and are significantly less expensive.

Point clouds that are created from images taken with a single camera have an arbitrary scale. Because of this such point clouds have to be scaled. Smart-phones are a good example of devices that are increasingly being used to take images for creating 3D models. It is envisaged that they will find greater usage for real-time indoor navigation in the future. Therefore, finding automatic solutions for scaling indoor models created from images captured by a single camera become attractive. Current systems rely on object detection within point clouds, e.g. Mostofi et al. (2014). In such systems a point cloud is segmented into objects. A database of known objects is then searched to identify the objects. Once identified the size of an object is retrieved from the database and used to scale the point cloud or model.

An alternative method to detect and identify objects is scene analysis. Scene analysis considers the spatial relationship of objects

in an environment. With scene analysis the arrangement of objects in relation to one another can be used as a descriptor (i.e. all computer screens are on desks in a classroom). A variety of descriptors based on the relationships between objects can be devised, such as relative distances, sizes and orientation. Scene analysis begins by segmenting a point cloud into objects based on geometric properties, e.g. proximity, curvature, etc. For example a point cloud of a room will be segmented into walls, ceiling, floor, desks and chairs. An scene analysis will then create a graph of the spatial relationship between the objects, e.g. desk next to a chair, desk on top of floor etc. Once the relationships are mapped, they can be studied to identify the different objects in the room based on the assumption that there is a defined relationship, or semantic map, between objects in the real world, e.g. Rusu et al. (2008). The work described here is premised on the simple idea that a height relationship between objects in a room is enough to detect and identify them. The idea stems from the observation that most furniture in a room have vertical dimensions that are very similar. For example, tables tops are at about the same height, doors are about the same height, and chair seats are about the same height.

Point clouds created from images captured by a single camera have an arbitrary scale. The objective of this investigation is to devise a simple method to automatically scale such a point cloud.

2. PREVIOUS WORK

Currently there exists three major methods to obtain a non-arbitrary scale for a scene represented by an unstructured point cloud. All of these methods require an absolute measurement to be made manually or automatic.

Manual Measurements: Physically measuring a dimension in the scene and propagating it through the generated 3D model is an established method of providing a real-world scale for point clouds. This can be accomplished by measuring an item in the scene or by using targets with known coordinates in a 3D reference frame. An example of this includes physically placing targets that can be uniquely identified in the scene before capturing the scene using photogrammetric methods. The target locations are determined by means of a survey such as a traverse which is often done beforehand. This is popular in the field of photogrammetry and surveying.

Use of Known Objects: Object detection is used to identify objects with known features stored within a database. Object detection attempts to solve the problem of obtaining a non-arbitrary scale for point clouds by placing objects which have at least one feature that describes a known dimension within a scene. This method is semi-automatic as a particular object needs to be identified within each scene but it only needs to be measured once. An example of this is illustrated in the work by Rashidi et al. (2014) where a pre-measured cube is placed within a scene that is captured using a monocular camera setup. Using Structure from Motion (SfM) the distance between subsequent frames of a video feed or collection of images can be recovered and used to propagate an arbitrary scale throughout the resulting point cloud. (Fleet et al., 2014).

Content-Based-Image-Retrieval (CBIR): One of the tools of CBIR is object recognition, an extremely important component of computer vision systems. The primary goal of this tool in relation to computer vision problems is to give platforms equipped with imaging sensors the ability to recognise unknown objects within a scene. Generalised features are stored within a database and a unique collection of them describe an object. This allows a computer vision system to ascertain which object is in the scene rather than searching for a particular object within a scene as in the case of object detection. The same criteria to obtain a real-world measurement using object detection is present here in object recognition (at least one descriptor of an identified object must be a measurement). The benefit of using object recognition is that it can identify many possible objects within a scene rather than having to place a known object such as a calibration pattern within it beforehand. This allows the scale of the point cloud, that has already been acquired, to be determined.

3. METHOD

3.1 Concept

Indoor scenes such as office spaces, lecture rooms and household environments contain common objects such as chairs, desks and tables. These objects have dimensions which are fairly similar in different environments which means they can be estimated without the need for physical measurement. Table 1 and Figure 1 illustrate the average dimensions for commonly found objects. The heights of the horizontal planes of chairs and desks can be used for scale determination. If these types of objects can be identified in point clouds then a scale can be readily obtained. The solution proposed is to determine those dimensions that show the least variation, e.g., the height of a chair. The benefits of this solution in contrast to existing methods mean that at no point is a manual measurement needed within the scene and object recognition can be simplified.

To test the concept scenes were captured using a laser scanner in order to determine the actual dimensions of objects within a scene and compare them to the generalised reference values. Heights

Typical Measurements for Furniture and Doors		
Item	Dimension	Label
Chair	45-53cm	Seat Height
	40cm	Seat Length
	81-107cm	Backrest Height
Desk	74cm	Tabletop Height
	122-152cm	Tabletop Length
	74cm	Tabletop Width
Door	200cm	Height
	96cm	Doorknob Height
	61-91cm	Width

Table 1: Table showing the average dimensions of furniture and doors, (Lefler, 2004) and (Griggs, 2001)

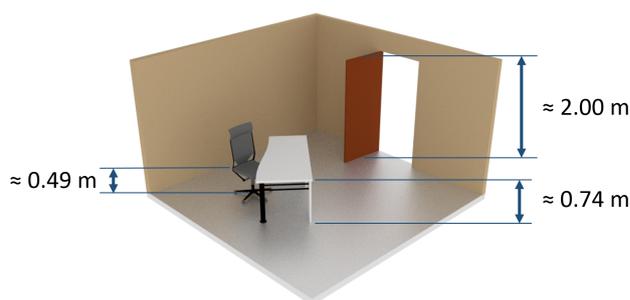


Figure 1: Schematic diagram of commonly found furniture (chair, desk, and a door)

from the ground for tabletops and chair-seats were taken from Table 1 and used as reference values. In the case of chair-seats where a range of possible values exists an average was taken. The reference heights can be seen in Figure 1. Scans were cleaned to remove points that had strayed outside of the indoor space (due to windows) and lastly the cloud was thinned to a 1cm resolution to reduce computation time. Results were expected to lie within 10% of their reference value which was 4.9cm and 7.4cm for chairs and desks respectively.

3.2 Processing Point Clouds

The segmentation process made use of the region growing algorithm from PCL. A Principal Components Analysis (PCA) was performed on each segment in order to determine whether it was a vertical or horizontal plane. If it was the latter then the segment was sorted into one of two groups based on the height of the segment from the ground (which was determined by the segment with the lowest height value). (a) A chair-seat has to lie between 45 & 53cm off the ground, (b) A tabletop has to lie between 72 & 80cm off the ground. (Slightly more than a 10% deviation from the reference value of 74cm quoted in Table 1 was used to derive this allowable height range.)

A weighted histogram of heights for the horizontal segments was used to illustrate the distribution of heights within a scene and to observe where spikes occurred which indicated a cluster of objects. In some scenes desks were placed very close together which resulted in strips of desks belonging to one segment. This under-segmentation was preferred as it erred on the side of identifying a group of similar objects rather than missing one that may be slightly occluded. As a result each segment within the histogram of heights was weighted using the number of points in said segment. If however a vertical segment was detected then the vertical width of the segment was used to determine whether it was a stair (the width would have to lie between 14 and 20cm).

The maximum value was determined by the South African Bureau of Standards (SABS) whilst the minimum value was decided based on observed width of stairs (as SABS does not set a minimum height for stairs) (SABS, 2011). The expected error for stairs was also 10% (using a median value of 17cm the maximum expected error would be 1.7cm). An F-test was performed to determine whether the variance of the height of an object type within each scene was equal. The hypothesis test is defined in equation 1. A large deviation from 1 of these variances suggest that they do not belong to the same population.

$$\begin{aligned}
 H_0 : \sigma_1^2 &= \sigma_2^2 \\
 H_a : \sigma_1^2 &\neq \sigma_2^2 \\
 \text{Test Statistic} : &\frac{s_1^2}{s_2^2}
 \end{aligned}
 \tag{1}$$

where H_0 = Null Hypothesis, H_a = Alternate Hypothesis, σ_n^2 = Sample Deviation, s_n^2 = Sample Variance, where $n \in \mathbb{R}$

4. RESULTS

4.1 Horizontal and Vertical Planes

The angle between the normal to the plane and the vertical was used to determine plane orientation. Figure 2 and 3 illustrate the segmented point cloud, vertical segments, horizontal segments and finally the classification of desks and chairs (blue and green respectively). Horizontal segments that have been recognised as desks or chair-seats are coloured in blue and green respectively. Multiple indoor scenes were scanned namely a classroom, (such as that found in Figure 2 and 3) a small computer lab and a seminar room. For these scenes the emphasis was on identifying chairs and desks. Two other indoor scenes were scanned, both of which were in open seating areas where the emphasis was on identifying stairs by the width of their vertical height. An example of this can be seen in Figure 4. Each scene was scanned using a single scan location in order to incorporate occlusions.

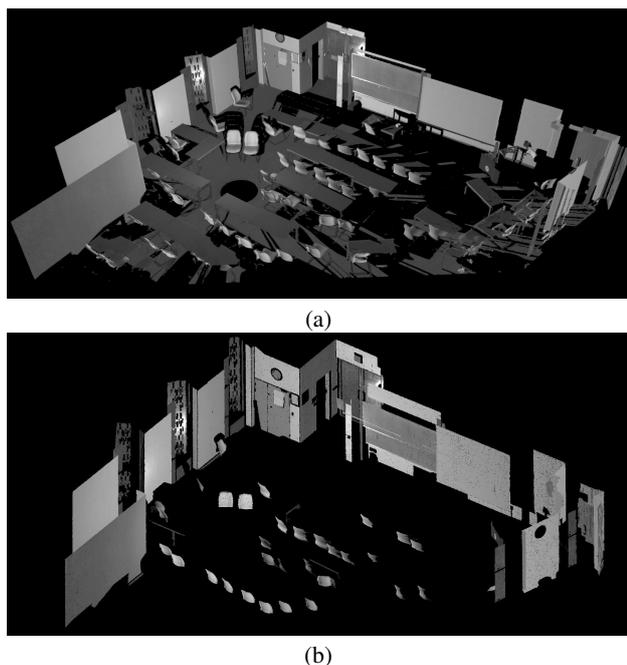


Figure 2: Point cloud of a classroom. (a) Original scene. (b) Vertical planes only.

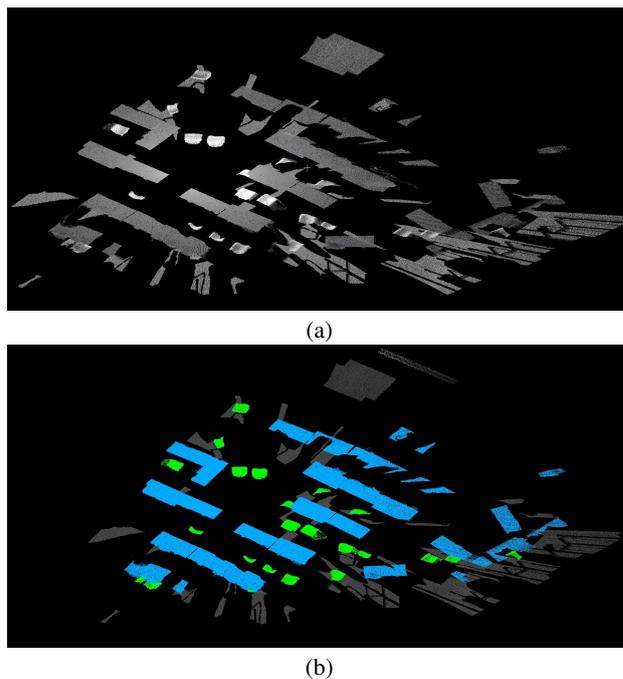


Figure 3: Point cloud of a classroom. (a) Horizontal planes. (b) Classification: Desks are coloured blue, chairs are green. They are identified by the height (Z-value) of the horizontal segments only.

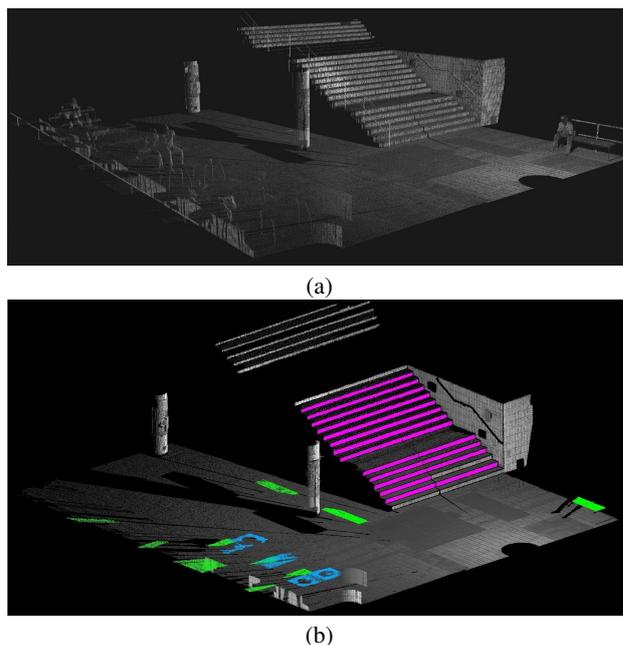


Figure 4: Point cloud of an auditorium. (a) Original scene. (b) Stairs were identified by vertical width of the segment and are highlighted in pink whilst tables and benches are coloured blue and green respectively.

4.2 Scene Analysis

Histograms of the horizontal height segments for each scene were combined into a single graph depicting heights of segments within the scenes. Each spike in the data depicts a cluster of objects at a respective height. Objects were weighted using the number of points per segment which compensated for under-segmentation (rows of desks formed a single segment rather than separate segments). By weighting the data each spike represents

the relative size of the cluster of segments at a given height illustrated in Figure 5. Table 2 illustrates the scale obtained when using horizontal segments (chairs and desks) and vertical segments where available (stairs). The results obtained with the stairs differed significantly from the horizontal segments as illustrated by their ratio. Column ‘Measured Height \bar{h} ’ represented the average of all measured values for a particular object type within a scene.

Weighted Histogram of Horizontal Height Segments per Room:

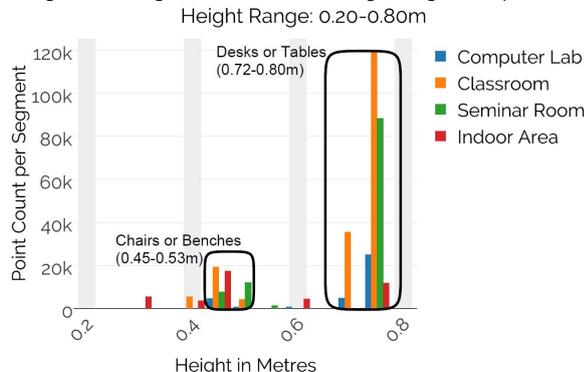


Figure 5: Histogram of height clusters per scene for the horizontal segments. These were weighted by point count per segment.

Deriving Scale from Recognised Objects					
Computer Lab					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	75.22cm	1.016	1.22cm	1.65%
Chair	49cm	47.94cm	0.978	-1.06cm	-2.17%
		Average	0.997	0.08cm	
Seminar Room					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	76.47cm	1.033	2.47cm	3.34%
Chair	49cm	50.01cm	1.021	1.01cm	2.06%
		Average	1.027	1.74cm	
Classroom					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	76.93cm	1.040	2.93cm	3.96%
Chair	49cm	48.54cm	0.991	-0.46cm	-0.94%
		Average	1.015	1.23cm	
Open Indoor Area 1					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	76.79cm	1.038	2.79cm	3.77%
Chair	49cm	46.94cm	0.958	-2.06cm	-4.21%
Stair	17cm	14.64cm	0.861	-2.36cm	-13.86%
		Average	0.952	0.54cm	
		Average w/o Stairs	0.998	0.36cm	
Open Indoor Area 2					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	76.06cm	1.028	2.06cm	2.79%
Chair	49cm	46.37cm	0.946	-2.63cm	-5.37%
Stair	17cm	14.87cm	0.874	-2.13cm	-12.55%
		Average	0.949	-0.90cm	
		Average w/o Stairs	0.987	-0.28cm	
Overall Scale for all scenes: 0.988					
Overall Scale for first three scenes: 1.013					
Overall Scale w/o Stairs: 1.005					

Table 2: Comparing measured and reference heights for identified objects per scene

Results of Applying the Average Scale of 1.013 to Scenes					
Computer Lab					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	74.24cm	1.003	0.24cm	0.33%
Chair	49cm	47.31cm	0.966	-1.69cm	-3.44%
		Average	0.984	-0.72cm	
Seminar Room					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	75.48cm	1.020	1.48cm	1.99%
Chair	49cm	49.36cm	1.007	0.36cm	0.73%
		Average	1.014	0.92cm	
Classroom					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	75.93cm	1.026	1.93cm	2.61%
Chair	49cm	47.91cm	0.978	-1.09cm	-2.22%
		Average	1.002	0.42cm	
New Overall Scale: 1.000					

Table 3: Table showing the effect of scaling each scene such that the objects are closer to their respective reference height.

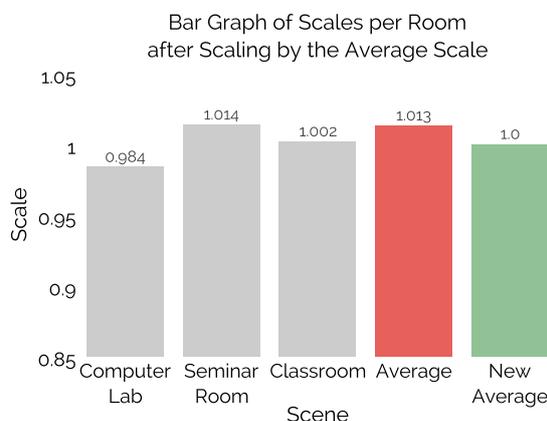


Figure 6: Bar graph illustrating the effect of applying the average scale to each scene.

Chairs in each scene were on average 1.04cm lower than their reference value of 49cm. Desks were 2.29cm higher than their reference value of 74cm. Both were calculated by taking the mean of the averages for object types in each scene illustrated in Table 2. This represents an average ratio of 2.13% and 3.10% for chairs and desks respectively between the measured and reference heights. Stairs yielded poorer results and in both indoor open area scenes breached the 10% expected error threshold. This can be attributed to significant occlusions based on the vantage point of the scan. As a result they were not used for further analysis.

Table 2 illustrated the discrepancy between the reference heights for chairs and desks. The scale obtained in Table 2 was then applied to the scenes in order to bring the heights of objects closer to that of the reference heights (the overall scale for the first three scenes which had no stairs were used as the average scale: 1.013). This was done to illustrate the difference between the heights of chairs and desks with respect to each scene rather than with respect to the reference heights. The results of this can be seen in Table 3 and Figure 6.

4.3 Object Analysis

The heights of objects were normalised by dividing the height of each recognised object by 80cm (this is the highest a desk could

be given the parameters laid out in Section 3.2). This was done to analyse the relationship between desks and chairs rather than to the reference heights as in Section 4.2.

Deriving Scale using Normalised Heights					
Computer Lab					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	0.92	0.94	1.016	0.015	1.65%
Chair	0.61	0.59	0.967	-0.020	-3.33%
		Average	0.992	0.003	
Seminar Room					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	0.92	0.96	1.033	0.031	3.34%
Chair	0.61	0.61	0.97	-0.002	-0.30%
		Average	1.015	0.015	
Classroom					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	0.92	0.96	1.038	0.035	3.77%
Chair	0.61	0.61	0.991	-0.006	-0.94%
		Average	1.014	0.015	
Overall Scale: 1.007					

Table 4: Comparison of measured and reference heights using normalised values for identified objects within each scene.

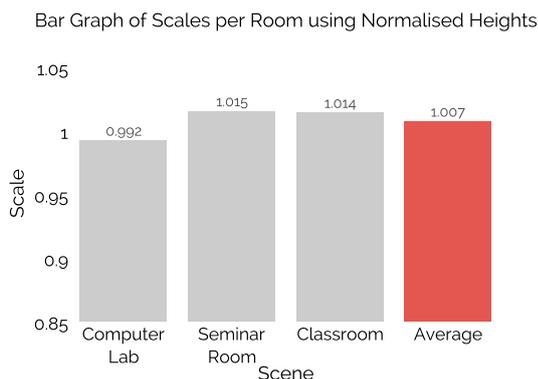


Figure 7: Bar graph of scales obtained per scene using normalised heights.

Results of Applying the Average Scale of 1.007 to Scenes					
Computer Lab					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	0.92	0.93	1.009	0.009	0.94%
Chair	0.61	0.60	0.971	-0.017	-2.85%
		Average	0.990	-0.004	
Seminar Room					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	0.92	0.95	1.026	0.024	2.62%
Chair	0.61	0.62	1.013	0.008	1.35%
		Average	1.020	0.016	
Classroom					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	0.92	0.95	1.032	0.030	3.24%
Chair	0.61	0.60	0.984	-0.010	-1.63%
		Average	1.008	0.010	
New Overall Scale: 1.006					

Table 5: Table showing the effect of scaling each scene using the scale obtained from the normalised height values.

The analysis of heights took place at the scene scale in the previous section which is why there is a larger difference in overall scene scales between Tables 2 and 3 compared to Tables 4 and 5 as these took place at the object scale by using normalised heights.

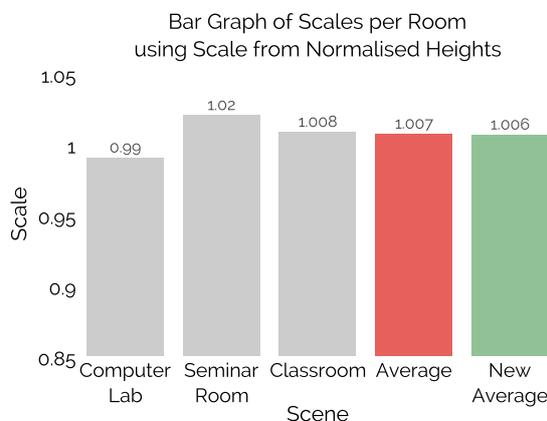


Figure 8: Bar graph illustrating the effect of applying a scale of 1.007 to each scene.

F-Test Two-Sample Variances at 95% Significance Level		
Chairs	Seminar Room	Classroom
	Variable 1	Variable 2
Mean	0.500069	0.485241
Variance	0.000327	0.000271
Observations	12	19
dF	11	18
F	1.205032	
P(F <= f) one-tail	0.350179	
F Critical one-tail	2.374155	
Verdict	PASS	

Table 6: Table showing the results of an F-Test performed in Microsoft® Office Excel 2013

5. F-TEST

An F-test was performed to determine if the heights for each object between scenes belonged to the same population. An example of this for chairs (using non-normalised heights) can be seen in Figure 6 whilst the results for all the scenes can be found in matrix form within Tables 7 and 8.

F-Test Two-Sample Variances at $\alpha = 0.05$			
Desks			
	Com-Lab	Sem-Room	Classroom
Com-Lab	-	PASS	PASS
Sem-Room	PASS	-	PASS
Classroom	PASS	PASS	-
Chair			
	Com-Lab	Sem-Room	Classroom
Com-Lab	-	PASS	PASS
Sem-Room	PASS	-	PASS
Classroom	PASS	PASS	-

Table 7: Matrix of F-Test results for non-normalised heights.

Table 7 illustrates that all the scenes pass the analysis of variance test for both desks and chairs. Using normalised heights however the F-test fails when comparing the variance of heights for

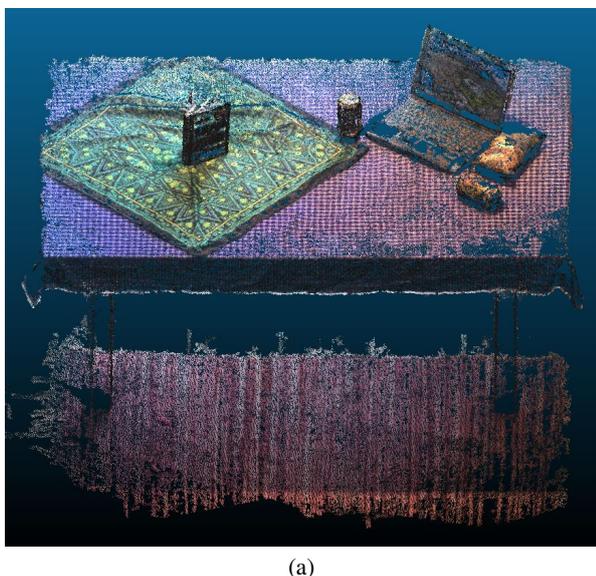
chairs between the computer lab and the classroom as seen in Table 8. This can be attributed to the large difference in degrees of freedom between the two scenes (4 and 19 respectively) and the fact that using normalised heights lowers the amount of allowable variation between samples. This is because the normalised heights allow the relationship between the desks and chairs to be analysed which is at a finer scale than analysing the objects within the scene as a whole.

F-Test Two-Sample Variances at $\alpha = 0.05$			
Desks			
	Com-Lab	Sem-Room	Classroom
Com-Lab	-	PASS	PASS
Sem-Room	PASS	-	PASS
Classroom	PASS	PASS	-
Chair			
	Com-Lab	Sem-Room	Classroom
Com-Lab	-	PASS	FAIL
Sem-Room	PASS	-	PASS
Classroom	FAIL	PASS	-

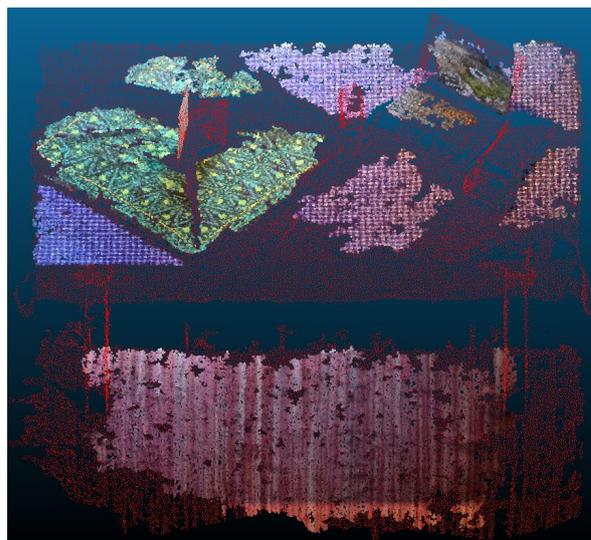
Table 8: Matrix of F-Test results for normalised heights.

6. TESTING THE THEORY ON PCD GENERATED FROM IMAGERY CAPTURED BY A SMART-PHONE

The next step was to capture a scene using a smart-phone camera. The device used was a two year old Huawei P6 which featured an 8 megapixel camera. This device was chosen in order to prove that a sufficiently dense point cloud can be captured with a mid-range smart-phone. Generating a point cloud using a series of images from a camera posed other problems outside of the scope of this research. For example monotonous surfaces would result in low to no feature matches between images thus creating an unusable cloud. To overcome this patterned tablecloths were used on those types of surfaces. The point clouds were generated using the software package VisualSFM developed by Wu Changchang which made use of SfM (Wu, 2013), Bundle Adjustment (Wu et al., 2011) and Feature Detection (Wu, n.d.). Once the point clouds were generated they were thinned to a resolution of 5mm to reduce the size of the point cloud and artificially scaled up by physically measuring objects within the scene. This was done in order to compare the results from the concept proposed within this paper and the real-world dimensions of the scene itself.

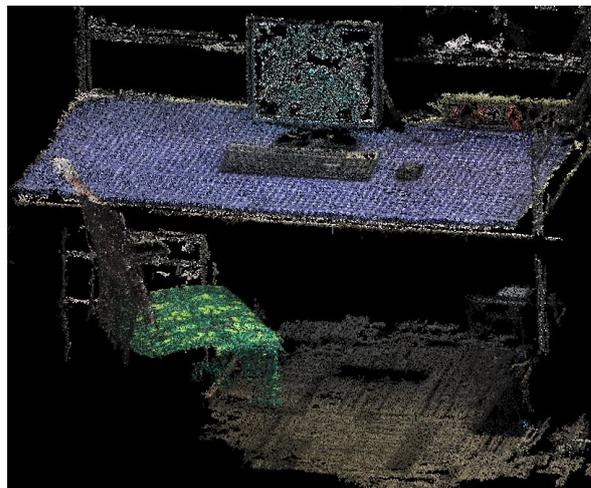


(a)

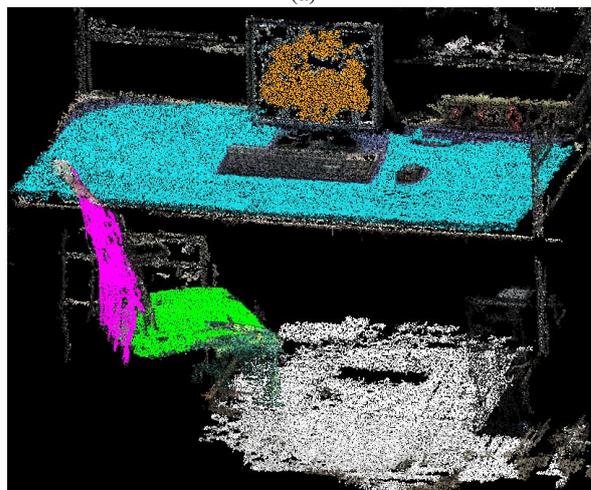


(b)

Figure 9: Quality of segmentation of a scene captured with a smartphone camera. (a) is the original cloud, (b) is the segmented one with segments in colour and unclassified points in red.



(a)



(b)

Figure 10: Image (a) illustrates the original cloud whilst (b) illustrates the identified segments - those in white represent the floor, green the chair seat, pink the backrest, blue the desk, and orange an unclassified vertical segment.

Figure 9 comprises of two images of a scene featuring a dining room table. The first image (a) is the original scene whilst (b) illustrates the difference between identified segments (those in colour such as the table, floor or vertical segment) and unidentified points (red). Unlike the laser scanner the points in these clouds are less uniform hence the fragmented nature of the segments. The office cubicle scene in Figure 10 was also supplemented by making use of a tablecloth on the desk, chair seat and backrest as these surfaces were monotonous in colour and texture. The segments that were successfully identified as represented by Figure 10, image (b) were the floor (white), chair seat (green), backrest (pink) and desk (blue) whilst the monitor (orange) was identified as a vertical segment but had no further classification. Even though the backrest was recognised it was not used further as only one instance of it in a single scene did not provide a statistically relevant sample.

Comparison of Dimensions					
Dining Room					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Table	74cm	76.46cm	1.033	2.46cm	3.32%
Office Cubicle					
Object	Reference Height	Measured Height \bar{h}	Scale	Difference	Ratio
Desk	74cm	75.04cm	1.014	1.04cm	1.40%
Chair	49cm	49.78cm	1.016	0.78cm	1.59%
		Average	1.015	0.87cm	

Table 9: Table showing the effect of scaling each scene such that the objects are closer to their respective reference height.

Due to the limited number of objects that could be used to scale the scene in Figures 9 and 10 the scale that was determined for both of them was less robust compared to the larger scenes captured with the laser scanner. Therefore instead of analysing the normalised heights, the ratio between the desk and chair was explored. Using the following ratio of reference chair-seat to desk: $49cm : 74cm$, a ratio of 0.6622 was calculated. Performing the same calculation on the measured values in the above table of $49.79cm : 75.04cm$ a ratio of 0.6635 was calculated. The two ratios differed by 1.35×10^{-3} . Further exhaustive testing is required to draw a significant conclusion from this result. It does however illustrate that using a ratio between these types of objects allows a check to be put in place for objects that may have strayed outside of the expected generalised dimension range (such as a coffee table) and therefore, even though they were detected, will not be used to scale the scene. The relationship between these objects still exist when the scene has an arbitrary scale and thus will aid in applying the correct scale to the scene.

7. SUMMARY OF RESULTS

Tables 2 and 9 illustrate the degree to which the average height of desks and chairs in a scene deviated from their respective reference height. The degree to which they differed is represented by the two columns labelled 'Difference' and 'Ratio' in either table. The values within these columns and the results that follow in the upcoming paragraph were calculated using the following equations:

$$Difference = Measured Height \bar{h} - Reference Height \quad (2)$$

$$Ratio = \frac{Difference}{Reference Height} \quad (3)$$

The maximum difference was $2.93cm$ for desks with a ratio of 3.96% which meant that the desks were $2.93cm$ taller than their expected (reference) height for that scene (as illustrated in Table 2). The smallest difference between measured and reference heights were for chairs at -0.94% which resulted in them being $0.46cm$ shorter than their reference height, (also from Table 2). Even though stairs differed from their reference value by similar magnitudes (under $3cm$), their ratio was outside the 10% threshold. This is attributed to the finer tolerance as stairs could only be a maximum of $20cm$ in width (along the Z -axis) with a median value of $17cm$, which meant they were only allowed to differ by a maximum of $1.7cm$, significantly less than desks or chairs ($7.4cm$ and $4.9cm$ respectively). Other reasons they were not included for further analysis include their sensitivity to occlusions and a lack of information concerning their possible range of values (no minimum stair height is set as mentioned in Section 3.2). There does however exist a ratio between the width of a stair and its allowed height (SABS, 2011), but in order to incorporate this in a way that would lead to meaningful results would have required significant analysis using edge detection which adds a layer of complexity to the proposed solution. This additional complexity would have defeated the primary purpose, which was to offer a solution that was simple and as computationally lightweight as possible. For the scenes captured by a smart-phone camera the biggest difference was between the dining room table and reference height with an overall discrepancy of $2.46cm$, still within the 10% threshold. In the office scene the height difference of the desk and chair-seat from the reference heights are $1.04cm$ for the desk and $0.78cm$ for the chair-seat height which are also well within the 10% threshold. Both scenes would have benefited from having more items to scale the scene with in order to derive a more robust scale.

Height values were normalised in order to analyse the relationship between desks and chairs. This yielded more accurate results. The values quoted here came from Tables 4 and 5. Where relevant they have been converted back to their non-normalised form by multiplying them out by $80cm$ as described in Section 4.3. The largest deviation from normalised reference heights was for desks at 3.77% which resulted in a $2.8cm$ discrepancy. The smallest deviation using normalised values was for chairs at 0.30% which meant they were $0.16cm$ shorter than their normalised reference height. By analysing the relationship between chairs and desks rather than the scene as a whole a more robust scale was determined. The scale obtained through normalised heights (1.007) was applied to the scenes again and the resulting scale (1.006) only varied by a factor of 0.001 from the original (as seen in Figure 8). Comparatively when this process was performed using non-normalised heights the difference in scale was 0.013 as seen in Figure 6. The results obtained using point clouds generated from images captured by a smart-phone were within the allowable tolerance (within 10% of the reference heights). This was incorporated to prove that the concept could work on unstructured point clouds.

8. CONCLUSION

The objective of this research was accomplished since the scale was obtained using a fully-automatic method. The proof-of-concept test yielded results below the expected 10% deviance from reference heights for desks and chairs but not for stairs which helped rule them out as potential candidate objects for automatic scale determination. The major find however was that the absolute measurement of objects yielded less accurate results compared to analysing the relationship between objects whose dimensions are based on human physiology. This has the potential

to change the approach taken when using object recognition to automatically derive scale: Rather than detecting an object in a scene whose dimensions are known and using that to derive scale for the scene, the relationship between different objects whose dimensions are based on human physiology can be used to derive scale by using typical or expected values for their dimensions.

The results obtained from the laser scans were from scenes on a university campus. There was variety between the scenes in terms of purpose and furniture. The university does not have a policy that determines the dimensions for furniture but they do have a list of preferred vendors which has been chosen based on having met quality standards. When researching furniture policies it became clear that the dimensions concerning furniture were not controlled to the exclusion of overall size so as to fit within an office or through the door for loading purposes. This allows one to rule out the possibility of obtaining misleadingly good results during the course of this research by having used scenes on a university campus as there exists no policy which determines furniture dimensions. In order to exhaustively test the proof-of-concept scenes in different contexts must be used. Other limits of the conclusions drawn from this research is that the typical or expected values for an object whose dimensions are based on human physiology can vary between different countries. This is because human physiology can vary between country populations, sometimes significantly enough to warrant differently scaled furniture for instance. In order to overcome this a greater variety of scenes must be tested.

9. RECOMMENDATIONS

Finding minimal descriptors for objects described herein will allow for a less computationally intensive form of object-recognition to be used in order to identify key items within an unstructured point cloud for the purposes of scaling it. There will always be objects that fall outside the expected range of dimensions and as such, checks should be put in place that explore the ratio between these objects in order to identify any that fall outside of the expected generalised dimensions (such as large doorways or coffee tables). Those objects should then not be used to scale the scene and provide a tainted result.

As this solution is geared towards being implemented on mobile devices due to their widespread use and available on-board technology, there are still a few limitations and refinements that should be discussed. The premise is that a smart-phone device will take a series of images and this will be used to reconstruct a scene represented by a series of points in 3D space. Segmentation of this point cloud is a computationally expensive procedure which can be mitigated by using object recognition (based on a series of generalised descriptors which can be stored locally) in order to identify objects such as desks and chairs within the images in order to segment only the corresponding area within the reconstructed 3D space. This offers a solution that mitigates the computational strain and eliminates the need for a server based solution. Another refinement would be to use a combination of 3D model-based object-detection with 2D appearance based imagery (Aubry et al., 2014). This will allow for objects with varying physical attributes such as chairs to be identified from various vantage points within a scene but it will require a large dataset of objects which makes it more appealing to a server-based solution. This will enable more accurate identification of objects within a greater variety of scenes. Once an object has been successfully recognised and the corresponding area in the 3D point cloud has been segmented, then the scalar attribute of the object can be used to scale the entire scene.

References

- Aubry, M., Maturana, D., Efros, A., Russell, B. and Sivic, J., 2014. Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of CAD models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3762–3769.
- Fleet, D., Pajdla, T., Schiele, B. and Tuytelaars, T., 2014. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings. Vol. 8695, Springer.
- Griggs, J. M., 2001. Typical Furniture Measurements. <http://www.fas.harvard.edu/~loebinfo/loebinfo/Proportions/furniture.html>, (22 Oct. 2015).
- Lefler, R. K., 2004. Choosing the Right Ergonomic Office Chair. <http://www.spine-health.com/wellness/ergonomics/office-chair-choosing-right-ergonomic-office-chair>, (25 Oct. 2015).
- Mostofi, N., Elhabiby, M. and El-Sheimy, N., 2014. Indoor localization and mapping using camera and inertial measurement unit (IMU). In: Position, Location and Navigation Symposium - PLANS 2014, 2014 IEEE/ION, pp. 1329–1335.
- Rashidi, A., Brilakis, I. and Vela, P., 2014. Generating absolute-scale point cloud data of built infrastructure scenes using a monocular camera setting. *Journal of Computing in Civil Engineering* 29(6), pp. 04014089.
- Richter, F., 2012. Infographic: 4.4 Billion Camera Phones... <http://www.statista.com/chart/653/prevalence-of-selected-features-in-the-global-installed-base-of-mobile-phones/>, (26 Oct. 2015).
- Rusu, R. B., Marton, Z. C., Blodow, N., Dolha, M. and Betsch, M., 2008. Towards 3d Point cloud based object maps for household environments. *Robotics and Autonomous Systems* 56(11), pp. 927–941.
- SABS, S. D., 2011. SANS 10400-M:2011 Edition 3: The application of the National Building Regulations Part M: Stairways. <https://law.resource.org/pub/za/ibr/za.sans.10400.m.2011.pdf>, (10 Oct. 2015).
- Wu, C., 2013. Towards linear-time incremental structure from motion. In: 3D Vision-3DV 2013, 2013 International Conference on, IEEE, pp. 127–134. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6599068, (10 April 2016).
- Wu, C., n.d. SiftGPU: A GPU implementation of scale invariant feature transform (SIFT). <http://cs.unc.edu/~ccwu/siftgpu/>, (28 March 2016).
- Wu, C., Agarwal, S., Curless, B. and Seitz, S. M., 2011. Multicore bundle adjustment. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, pp. 3057–3064. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5995552, (10 April 2016).