

18 extracted corner points are then matched between the consecutive frames. Finally, the
19 corresponding corner points are reconstructed along with other features of the scenes to
20 determine the real world scale. To evaluate the performance of the method, ten indoor and ten
21 outdoor cases were selected and the absolute-scale PCD for each case was computed. Results
22 illustrated the proposed algorithm is able to reconstruct the predefined objects with a high
23 success rate while the generated absolute scale PCD is sufficiently accurate.

24 **Keywords:** Absolute scale; Monocular videogrammetry; Point Cloud Data; 3D reconstruction

25

26 **Introduction**

27 According to the results of current studies conducted by Golparvar-Fard et al. (2013) and
28 Becerik-Gerber et al. (2013), monitoring the health of infrastructure is one of the most
29 imposing challenges faced by civil engineers in the 21st century. Lack of viable methods to
30 map and label existing built infrastructure is an important component of this challenge. As-
31 built 3D geometry comprises a significant portion of the total as-built information and any
32 efforts towards automating its acquisition will translate to cost savings and improved quality
33 assurance in the delivery and maintenance of the built environment.

34 The current state-of-the-art approach to collecting spatial data and converting it to as-built
35 geometry of built environment scenes is through active sensors (total stations and laser
36 scanners) and surveying methods. This approach encapsulates the 3D geometry in a set/cloud
37 of 3D points. Although as-built geometry generation is assisted by recent technological
38 advancements both in hardware and software, most of its steps are costly, both in terms of

39 equipment and labor, and time consuming. As a result, there is increasing demand for
40 automated, cost effective methods for collecting spatial data of built infrastructure scenes and
41 converting the data to as-built models (Brilakis et al. 2011).

42 Within the last two decades, advances in high resolution digital photography and
43 increased computing capacity, have made it possible for image/video-based 3D
44 reconstruction methods to produce promising results. Over the past few years, researchers in
45 the fields of computer vision and civil engineering have heavily focused on developing
46 algorithms to improve the performance of this technology.

47 Based on the number of cameras, photo/videogrammetric-based algorithms are divided
48 into two major categories: a) monocular, defined as using a single camera; and b) binocular,
49 defined as using a stereo set of cameras. Additional cameras can also be used if needed in
50 multi camera systems. For binocular, the relative position and orientation of one camera in
51 relation to the other camera is measured in advance and considered as a known parameter,
52 thus making it directly possible to obtain 3D measurements in Euclidian space. However,
53 stereo cameras are specialized equipment, and far less feasible hardware solutions than
54 monocular setups, such as the cameras in most smart phones that on-site personnel carry. In
55 general, a single camera (monocular setting) is a much more practical way to capture
56 images/video data since most individuals on a jobsite has access to a single digital camera or
57 smart phone. However, implementing a monocular camera setup only generates unknown
58 global scale PCD (Scaramuzza et al. 2009). In order to compute the absolute scale, the
59 operator needs to know the base line of the camera motion or at least one dimension of the
60 scene. The traditional way of solving the problem is measuring the distances between a set of

61 predominant points in the scene before or after the data collection. The corresponding 3D
62 locations of these predominant points should be manually identified by the operator from the
63 generated PCD. The ratio of the real Euclidian distance between the predominant points
64 compared to the computed distance in the PCD is the absolute scale of the scene.

65 Measuring such dimensions in a job site is a manual task that increases the time and effort
66 needed to collect the geometry and induces human error in one of the most sensitive parts of
67 the 3D reconstruction process; consequently, the results can be inaccurate. Furthermore, there
68 is no guarantee that the corresponding measured points are successfully reconstructed and
69 already exist in the PCD. As explained in section 6 of this paper, the authors conducted
70 experiments and measured a number of dimensions in outdoor built environments using a
71 total station. These experiments indicate that it takes an average of 15 minutes to manually
72 measure one dimension of the scene, find the corresponding points in PCD and calculate the
73 scale factor within a reasonable error tolerance.

74 Several new methods have been proposed for automatically retrieving the absolute scale
75 of a scene using a monocular setup. These methods, however, either lose the practicality of
76 the monocular setup by adding extra sensors or are limited to explicit scenes and are not
77 general enough to be useful by Architecture/Engineering/Construction (A/E/C) practitioners
78 in their daily tasks (Scaramuzza et al. 2009). In this paper, we propose a general method for
79 automatically computing the absolute scale of PCD from monocular video, without the use of
80 additional sensors. The proposed method is based on using pre-measured, simple
81 standardized objects that are commonly available or easily obtained; in particular, a letter-
82 size sheet of paper for indoor settings (up to approximately 7 meters distance from target),

83 and a simple colored cube made of plywood material for outdoor environments (up to 25
84 meters distance from target). The vertices of these predefined objects are detected in video
85 frames using a novel algorithm. The detected vertices in 2D frames are then reconstructed
86 along with the other feature points extracted from the scene. Knowing the distance between
87 the vertices, the entire PCD is then scaled up using an existing method. The paper is
88 organized as follows: the background section summarizes the existing states of
89 practice/research on absolute scale calculation for monocular photo/videogrammetry. Our
90 method for automating the absolute scale calculation is presented in the next section. In the
91 experiments section, tests are conducted to test the validity of the proposed algorithms and
92 the entire pipeline. Finally, conclusions are drawn in the last section.

93 **State of practice: recovering absolute measurements in photo/videogrammetry**

94 In computer vision, 3D reconstruction of different scenes is achievable in different levels and
95 based on the priori available knowledge about the scene/camera (Table 1).

96 Insert Table 1 here

97 Many of the available commercial software packages (Photosynth, Photo-Modoler and
98 Photofly) fall into the second category, i.e. the intrinsic camera parameters can be achieved
99 by calibration; however, the camera motion is unknown. As the result, the obtained PCD is
100 up to an unknown global scale. Nowadays, applications of commercial 3D reconstruction
101 software packages (Photosynth, Photo-Modoler and Photofly), which work by processing
102 taken images/captured videos, vary from accident reconstruction and forensics to archeology,
103 geology and surveying (Overview of applications for Photo-Modeler 2013, Fathi and
104 Brilakis, 2014). However, all of these packages suffer from one issue: it is not possible to

105 directly extract real measurements since the global scale is unknown. This limitation is of
106 great significance since almost all measurements take place in Euclidean space with real
107 values in both civil and infrastructure engineering applications.

108 In manufacturing practices, the entire measurement procedure takes place in indoor,
109 controlled settings so it is feasible to arrange specific settings for directly extracting real
110 dimensions of objects. One popular approach is using specific target projectors called PRO-
111 SPOT. This structured-light system works like an ordinary slide projector. A light source
112 illuminates a target slide. As the next step, the illuminated pattern (usually a dot pattern)
113 passes through a number of lenses which magnify the slide and project it onto the object's
114 surface. By knowing the dimensions of the pattern, it is possible to extract the actual
115 dimensions of the objects (Figure 1).

116 Insert Figure 1 here

117 The proposed solution is feasible for indoor, controlled manufacturing environments;
118 though, it does not practically fit the random, uncontrolled built infrastructure scenes.
119 Theoretically, for built infrastructure scenes, it is possible to compute the global scale of the
120 PCD by measuring only one dimension in the scene. However, in practice, a number of
121 issues would occur:

- 122 - The common practice to precisely measuring dimensions in a built infrastructure
123 jobsite is using a total station (Coaker 2009). Using total stations for measurement
124 purposes leads to very accurate results (average error = ± 1 mm); yet, the entire
125 procedure is not straight forward and requires certain levels of training. A surveyor
126 should carefully setup the equipment in a proper location of the job site and conduct

127 the measurements (Coaker 2009). The surveyor then goes back to the office and
128 implements relevant software for post processing steps including visualization of PCD,
129 extracting corresponding measured dimensions from it and scaling up the entire PCD.
130 Obviously this procedure is time consuming and labor-intensive.

131 - Unlike scanning senses using laser scanners, in some cases, processing images and
132 video frames does not result in generating PCD that are uniformly dense enough
133 (Rashidi et al, 2013). There might be poorly reconstructed areas (due to several
134 reasons, e.g. insufficient coverage during sensing, reconstruction errors and texture-
135 less areas), and there is no guarantee that the corresponding points used for actual
136 measurements already exist in the PCD.

137 - The devices used for measuring dimensions of the scene are either expensive, e.g. laser
138 measurer and total stations, or inaccurate, e.g. tape measurer (Dai et al. 2013).

139

140 **State of research: absolute scale PCD for monocular settings**

141 As stated before, manually measuring dimensions of a scene or implementing a stereo
142 camera setup are two feasible solutions for calculating the absolute scale of a scene. For
143 monocular camera settings, two major approaches are suggested to automatically recovering
144 the absolute scale:

145 The first approach relies on the application of supplemental electronic sensors for
146 acquiring extra information about the scene or motion of the camera. Global Positioning
147 System (GPS), inertial measurement units (accelerometers, gyroscopes, magnetometers), and
148 odometry measurements are examples of the applied sensors for providing supplemental
149 measurements for absolute scale computation purposes (Tribou 2009). Nutzi et al. (2011)

150 fused inertial measurement unit (IMU) and visual data for absolute scale estimation in
151 monocular SLAM (Simultaneous Localization and Mapping). Eudes et al. (2010) solved the
152 scale drift problem observed in long monocular video sequence using a standard odometer
153 installed on a car. Kneip et al. (2011) combined accelerometer and attitude measurements
154 with feature observations in order to compute the metric velocity estimation of a single
155 camera. Supplemental sensors can also be applied in the form of range measurement devices
156 or additional monocular cameras (Gutierrez-Gomez and Guerrero, 2012). Jung et al. (2008)
157 implemented a range finding device for use in a SLAM context by projecting a structured
158 light on the environment and measuring the resulting distortions with a monocular camera.
159 2D laser range finder (LRF) is another popular sensor used by the robotics and computer
160 vision community to address the global scale issue (Castellanos et al. 2000).

161 Applying additional sensors is not always a cost effective solution, so other researchers
162 have tried to use prior knowledge about the scene obtained through predefined existing
163 objects and visual fiducials (Tribou 2009). In the SLAM area, different classes of objects and
164 artificial landmarks are utilized to acquire necessary information about the environment and
165 therefore solve the robot positioning or localization problem. Olson (2011) proposed a visual
166 fiducially system based on 2D planar targets with specific bar code patterns for accurate
167 localization of robots. Obtained results for localizing groups of robots in indoor and outdoor
168 settings have been promising. Botterill et al. (2012) proposed an innovative solution to the
169 problem of scale drift in single camera SLAM based on recognizing and measuring different
170 classes of objects. Anati et al. (2012) developed a robot which can localize itself by
171 recognizing specific groups of objects (bins, clocks, ticket machines) on a simple map of a

172 train station. Li et al. (2011) incorporated the structure of instances of known objects into the
173 3D reconstruction of a scene. Specific poles have been used for 3D reconstruction of large
174 scale, cultural heritage in absolute scales (Pavlidis, et al., 2007)

175 Acquiring extra information from existing objects in the scene or visual fiducials is a
176 feasible solution. However, the selected objects are not simple enough (from points of
177 material, shape and pattern) to be commonly found (built) in regular jobsites. Furthermore,
178 the success rate of the suggested algorithms for reconstructing the predefined object(s) should
179 be high enough to be reliably used in various conditions and environments.

180 Other than the two major approaches, there have been attempts to mathematically solve
181 the problem for explicit settings by imposing extra constraints/assumptions. Kuhl et al.
182 (2006) proposed a method based on a Depth-from-Defocus approach to calculate the absolute
183 scale of monocular settings by combination of geometric and real-aperture methods. The
184 proposed method does not require any prior knowledge about the scene; however, it is based
185 on tracking objects and, hence, is not a feasible solution for large scale civil infrastructure
186 scenes. Scaramuzza et al. (2009) mounted a single camera on a specific wheeled vehicle to
187 automatically recover the absolute scale of the scene. The method is applicable for large scale
188 scenes; though, mounting the camera on a wheeled vehicle is not feasible in common
189 construction job sites.

190 In the area of A/E/C, specific settings might be applied to solve particular problems.
191 Golparvar-Fard et al. (2012) used 3D coordinates of predominant benchmarks, e.g. corners of
192 walls and columns, and the building information modeling (BIM) of the built infrastructure to
193 solve the absolute scale calculation and registration problems. Later on, Golparvar-Fard et al.

194 (2012) proposed a solution based on placing specific registration targets on rebar meshes to
195 compute the absolute scale and 3D locations of rebars and embedments. In a NIST report,
196 Saidi et.al, (2011), introduced the application of fiduciary markers combined with specific
197 elaborated patterns to extract the absolute scale of built infrastructure PCD. The proposed
198 solutions are all practical, yet limited to specific settings and are not general enough to be
199 considered for a vast range of indoor and outdoor built infrastructure scenes, e.g. fiduciary
200 markers with specific elaborated patterns cannot easily be found at job sites. In addition,
201 there is no guarantee that the corners of walls and columns are reconstructed properly.

202 In the area of structural health monitoring, Jahanshahi et al. (2011) proposed an
203 innovative approach for measuring dimensions of cracks on concrete surfaces. They assumed
204 that the working distance (the distance between camera and the object) is known. This extra
205 known dimension was implied to calculate the Euclidian dimensions of cracks. Zhang et al.
206 (2012) utilized an unmanned aerial vehicle-based imaging system, equipped with GPS and
207 INS for 3D measurement of unpaved road surface distresses. Carozza et al. (2012) proposed a
208 mark-less monocular vision based approach for localization within an urban scene based on
209 an offline map of the environment. Their method requires a manual learning stage and
210 manually matching several 3D model points with their corresponding image points.

211 As observed, most of the proposed solutions either required specific extra electronic
212 sensors/equipment or are limited to particular settings/scenarios and are not generic enough to
213 immensely be applied by practitioners in the areas of construction engineering and facility
214 management.

215 **Problem statement and research objectives**

216 As mentioned in the previous section, there are three major issues associated with the current
217 approaches for automatically calculating the absolute scale factor for monocular settings.
218 First, adding extra sensors to the setup defeats the value of monocular setups and is not
219 always cost effective (precise accelerometer sensors usually cost more than \$300), thus is
220 not a feasible alternative to stereo setups for routine tasks in the A/E/C domain. Second,
221 acquiring extra information from specific classes of objects in the scene is not a reliable
222 approach since objects vary from one built infrastructure scene to another (Rashidi et al.
223 2013). Finally, there is no guarantee that certain classes of objects can be successfully
224 reconstructed during the processing stages. As the result, there is significant demand for a
225 simple, accurate, yet practical solution applicable for regular built infrastructure scenes
226 (Nutzi et al. 2011).

227 The research objective of this paper is to test whether the method proposed by the authors
228 is able to successfully and accurately compute the absolute scale of various built
229 infrastructure scenes in both indoor and outdoor environments. The presented solution relies
230 on using predefined objects, with known dimensions, for each indoor and outdoor scenario in
231 order to extract the necessary prior knowledge about the scene. Theoretically, our approach is
232 similar to other existing methods using pre-defined objects for extracting absolute
233 measurements. However, the following advantages differentiate our work compared to the
234 existing methods within the literature:

235 - We have tried to simplify the calibration objects as much as we can. The chosen objects
236 could be easily found, or built, in almost all jobsites with lowest efforts and costs.

237 - By implementing robust techniques for detecting and reconstructing calibration objects,
238 accurately computing the absolute scale is guaranteed in almost all cases.

239 **Proposed solution for automated absolute scale computation for outdoor settings**

240 Many A/E/C practices take place in outdoor settings, so it is necessary to choose a simple,
241 consistent object which is easily detectable and easy to use at most job sites. Among
242 geometrical objects, a cube is the simplest. The dimensions of a cube are equal and it is
243 typically possible to view three of its surfaces from various perspectives simultaneously. We
244 chose a cube made of plywood, which is solid and light weight, noting that it can be built at
245 nearly any job sites. The size of the cube should be big enough to use in large scale
246 infrastructure scenes, yet small enough to be carried out and handled by only one person.
247 Considering those factors we choose 0.8 meter as the standard dimension for the cube.

248 In order to better detect the object in the scene we chose three different colors for the
249 cube's surfaces. Two criteria should be considered while choosing the right colors for the
250 cube surfaces: 1) the colors should be distinct from the colors of existing features in the
251 scene, and 2) there should be a maximum difference between RGB (HSV) values of the
252 selected colors so they can easily be identified using color detection algorithms. Considering
253 the above constraints, we remove colors close to blue and green since those colors frequently
254 appear in outdoor settings. Examining what remains, and distributing the color values as
255 evenly as possible across the remaining spectrum, leads to the three distinct colors whose
256 HSV values are depicted in Figure 2.

257 Given the selected colors, the overall method for calculating absolute scale mainly relies
258 on detecting the cube in video key frames; identifying, matching and reconstructing the cube

259 vertices along with other feature points of the scene; and scaling the obtained PCD given the
260 known dimensions of the cube (distances between the vertices). Figure 3 depicts the proposed
261 framework for absolute scale estimation.

262 Insert Figure 2 here

263 Insert Figure 3 here

264 The proposed algorithm consists of the following three steps:

265 **Step 1: Detection of the cube's vertices**

266 Figure 4 describes the necessary steps for detecting the vertices of the cube in 2D video
267 frames captured from the scene.

268 Insert Figure 4 here

269 The procedure starts with detecting the surfaces of the cube by filtering the HSV values.
270 For each detected surface, the connected components are analyzed and an opening
271 morphology operator (size of structuring element = 3×3 pixels; two iterations) is applied to
272 remove small areas with the same color values which do not belong to the cube's surface (Chi
273 and Caldas 2011). To ensure that detected areas belong to the cube surfaces, the following
274 constraints should be met:

- 275 - The area of the surface should be bigger than 0.005 times the area of the entire image. This
- 276 criterion removes false detections of small areas that might match, and also ignores detected
- 277 boxes that are too far from the camera which often introduce estimation error. As explained
- 278 later, the threshold value, 0.005, was experimentally obtained.

279 - It is assumed that each surface of the cube should look neither too long nor too circular in
280 the image. Accordingly, the roundness of the surface, calculated by the following equation,
281 should be located between an upper and a lower threshold:

$$282 \quad \text{Roundness} = \frac{4\pi \times \text{Area}}{(\text{Perimeter})^2} \quad (1)$$

283 - Due to the perspective projection equations describing image formation, the imaged
284 surfaces of a cube are trapezoidal in shape, which is convex. To isolate potential cubes by
285 removing non-convex objects, the real area of the surface should be approximately equal to
286 the convex hull of the surface (Figure 5).

287 After identifying the surfaces of the cube, the edges of the cube are detected using a
288 modified version of the Hough transform. Due to nonlinear lens distortions, the cube edges
289 may not appear straight in the 2D images, but will be slightly curved. In order to address the
290 issue, a modified Hough transform algorithm was implemented. The details of the modified
291 algorithm are below:

292 A dilation procedure, which is a common function in image processing applications, is
293 applied to remove some of the noises. In the modified Hough transform algorithm, all edges
294 in different directions with a radial resolution equal to 2 degrees are recognized in the polar
295 coordination system (range: $[-\frac{\pi}{2}, \frac{\pi}{2}]$)

296 The other approach for dealing with this type of distortion is using undistorted images by
297 applying the lens radial distortion factors computed through the SfM.

298 Finally, the cube vertices are identified by determining neighboring edges through their
299 intersection points. To this end, edges on all different surfaces are extended into both

300 directions until they intersect the first other edge (neighboring edge) . It is possible that 3
301 edges do not exactly intersect at the same point so we consider the point with the minimum
302 distance to all corresponding edges as the intersection point.

303 Insert Figure 5 here

304 **Step 2: Matching the cube’s vertices across key frame views**

305 In parallel with extracting cube’s vertices, other feature points of the scene are also
306 recognized using SURF feature detection algorithm (Rashidi et al. 2013). As the next step,
307 camera intrinsic and extrinsic parameters are computed using two standard approaches:
308 camera calibration and structure from motion (SfM). In our study, we calibrated the camera
309 offline (using a calibration pattern); however, In the case of processing images, instead of
310 manually calibrating the camera, it is possible to automatically extract the initial values of the
311 intrinsic parameters using the Exchangeable image file format (Exif) (Golparvar-Fard, et al.
312 2012). Values obtained from the Exif tags are then used as the initial estimates for the bundle
313 adjustment procedure. In this case, the camera calibration step, which might be a slightly
314 challenging task for job site personal, is eliminated.

315 After detection of the cube’s vertices and calculating the camera parameters, the next step
316 is to match these vertices within two key frame views. For this purpose, we followed a
317 specific matching strategy explained below. Our matching strategy consists of two
318 components:

319 1) The corresponding point for each vertex in one key frame view should be located on the
320 epipolar line for the other view (Dias 2006). If P and P' are the camera matrices for the first
321 and second view, the ray which is projected onto the point x in the first view is defined as:

322
$$X(\lambda) = P^+x + \lambda C \quad (2)$$

323
 324 Where C is the common camera center for both P and P' , λ is a scalar, P^+ is the pseudo
 325 inverse to P , i.e, $PP^+=I$ and $PC=0$. The line $X(\lambda)$ intersects the points P^+x and C . These
 326 points are mapped into the other camera P' at $P'P^+x$ and $P'C$. The epipolar line l' intersects
 327 these projected points and can be written as:

328
$$l' = (P'C) \times (P'P^+x) \quad (3)$$

329 The point $P'C$ is the epipole e' or the projection of the first camera center into the second
 330 camera. Thus the epipolar line can be formulated as:

331
$$l' = [e']_{\times}(P'P^+)x = Fx \quad (4)$$

332 Where, $[e']_{\times}$ is the corresponding skew-symmetric of e' and F is a 3×3 non-zero matrix
 333 known as the fundamental matrix. Applying this criterion always limit the search area into a
 334 few candidates (usually 1 or 2) located on the corresponding epipolar line on the second view
 335 (Figure 6).

336
 337 Insert Figure 6 here

338
 339 2) Applying the color differences is the second criterion. We consider a rectangular window
 340 around each vertex. Since the motion of the camera between two consecutive key video
 341 frames is small, we expect that the corresponding window in the other frame also contains
 342 similar color values. In other words, the best corresponding window is selected by following
 343 a differentiation and cross correlation approach between the color values of the two windows
 344 in two consecutive frames and calculating the similarity score as following (Rashidi et al.
 345 2011):

346
$$\text{Col - Diff } (W, W') = \sum_1^n \sum_1^m (|I_{xy} - I'_{xy}| + |R_{xy} - R'_{ixy}| + |G_{xy} - G'_{xy}| + |B_{xy} - B'_{xy}|) \quad (5)$$

347
348
$$\text{Corr}(W, W') = \sum_1^n \sum_1^m (|I_{xy} I'_{xy}| + |R_{xy} R'_{xy}| + |G_{xy} G'_{xy}| + |B_{xy} B'_{xy}|) \quad (6)$$

349
350
$$\text{Similarity Score}(W, W') = \frac{\text{Corr}(W, W')}{1 + (\text{Col-Diff}(W, W'))} \quad (7)$$

351
352 Where R_{xy} , G_{xy} , B_{xy} and I_{xy} are the individual color channel and intensity values of the
353 neighborhood pixels of the windows constructed around each vertex and n is the size of the
354 window in pixels. W and W' refer to the first and second windows respectively.

355 It is necessary to emphasize that using fiduciary markers or more distinguishable
356 patterns on the sides of cube would improve the performance of the detection algorithms;
357 however, for two reasons we did not choose this solution. First, it is more practical to keep
358 the calibration object as simple as possible. Second, our experiments indicate that the
359 performance of the proposed algorithm for detecting the cube in current shape is very
360 promising.

361
362 **Step 3: 3D reconstruction of the cube's vertices along with other features of the scene**

363 We use a standard 3D reconstruction pipeline, as introduced in (Rashidi et al. 2013), to
364 reconstruct the vertices of the cube as well as other features of the scene. We used the Patch-
365 Based Multi-view Stereo (PMVS) approach to reconstruct the entire scene and compute the
366 PCD. Assuming that the dimensions of the cube are known, we can scale up the entire PCD.
367 As explained in the previous sections, the matches for the vertices come from using epipolar
368 geometry + window search, while the others come from standard SURF matching algorithm.
369 Since the number of reconstructed edges is usually more than one, a least square error (LSE)
370 approach is applied to obtain a unique scaling factor for the entire scene as described below:

371 Assuming n is the number of reconstructed edges, X_i is the i^{th} computed dimension with the
 372 actual length of Y_i ; the scale factor (S.F.) relates X_i and Y_i as:

$$373 \quad Y_i = (\text{S.F.}) \times X_i + B \quad (8)$$

374 Where B is the computed error (in ideal situation: $B=0$) and we assume that the distribution
 375 of errors in the 3D space is uniform. Considering the linearity assumption, the scale factor
 376 (S.F.) is calculated using the following regression-based equations (Montgomery et al, 2012):

$$377 \quad S.F. = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \quad (9)$$

$$378 \quad B = \frac{\sum_{i=1}^n Y_i - S.F. \sum_{i=1}^n X_i}{n} \quad (10)$$

380
 381 One important issue that needs to be taken into account is the drift problem. It is well known
 382 that scaling a large infrastructure scene using a relatively small object is error prone (Botterill
 383 et al., 2012). To address the issue, a weighting function has been added to the cost function of
 384 the Bundle Adjustment. The cost function of the Bundle Adjustment is the sum of the
 385 distance between detected points and projected points. We set the weight of the cost function
 386 as 2 for vertices of the cube and kept the cost function weight of other points of the scene as
 387 1; this way we give priority to the important points of the scene, corner points and vertices,
 388 and reconstruct them more accurately. Another feasible solution to handle the drift problem
 389 is using multiple objects located in different parts of the scene. Using multiple objects would
 390 result in more uniform distribution of errors instead of cumulative. That being said, numbers,
 391 locations and sizes of calibration objects play important roles in drift problem. The authors
 392 plan to focus more on this issue in future research.

393 **Proposed solution for automated absolute scale computation for indoor settings**

394 Our suggestion for a proper object for use in indoor settings is a simple letter-size sheet of
395 paper. Letter-size paper can be found in almost every indoor environment, including homes
396 and offices. The paper should be placed on a darker uniform surface to maximize detection
397 (Figure 7).

399 Insert Figure 7 here

400 The algorithm for detecting, matching and reconstructing the corners of the sheet of the
401 paper is the same as those of the cube with the exception of the matching stage. All four
402 corner points of the paper have almost the same color values; thus, it is not possible to
403 effectively use the color differentiation criterion. The solution is straight forward: since we
404 are only dealing with four points as the corners of the paper, it suffices to implement the
405 epipolar geometry constraint, and taking note that the four corners in the first view and their
406 correspondences in the second view are located based on a same clockwise order (Figure 8).

407 Insert Figure 8 here

408 It is important to mention that using more distinctive objects such as printed sheets with
409 elaborated patterns and codes might also lead to very accurate results, but the advantage of
410 our method lays on the simplicity of the chosen object, as well the sufficient accuracy of the
411 results.

412 **Implementation and experimental setup**

413 A C# based prototype was implemented to test the validity of the proposed algorithm. It was
414 written in Visual Studio 2010 using Windows Presentation Foundation (WPF) and publicly
415 available libraries such as OpenCV 2.0 (wrapped by EmguCV) for access to computer vision
416 tools and DirectX 10 for the graphic display of results. The Open CV's image structure was

417 the primary data structure. It removed the conversion needs of the image processing tools
418 from that library, which significantly reduced the processing speed. The aim of the
419 experimental setups is two folds: 1) identifying the thresholds for applying in the proposed
420 algorithms and 2) evaluating the performance of the implemented algorithms as well as the
421 overall performance of the proposed method. Each step is explained in the following
422 sections:

423 **Identifying thresholds for the minimum acceptable area of the cube in images**

424 As previously explained, if the areas of the cube surfaces in images were too small, i.e. the
425 cube is located too far from the camera, the estimated errors in detecting and reconstructing
426 the cube corner points would increase significantly. To tackle this issue, we implement a
427 specific threshold as the minimum acceptable area of a surface of the cube, compared to the
428 total area of the image. Frames including the cube surfaces smaller than the calculated
429 threshold are removed from further processing. It is important to mention that discarding
430 some frames from further processing might have effects on different part of the algorithms;
431 however, smooth, sequential videotaping the scene would minimize those effects (e.g.
432 instead of arbitrary moving the camera, we either move forward or backward toward the
433 cube). On the other hand, different faces of the cube are sufficiently differentiable so
434 disregarding some of the frames or changes in cube surfaces' views does not affect the
435 performance of the matching algorithm.

436 In order to identify a proper threshold, we conducted a number of experiments.
437 Considering the variety in built infrastructure scenes, we placed the cube and the sheet of
438 paper in 10 outdoor and 10 indoor built infrastructure scenes. The scenes were videotaped

439 from different views with varying distance of the camera from the calibration object. As the
440 first step, the video clips were processed and the surfaces of cubes were detected. The success
441 rates of detecting the surfaces were measured using the precision and recall values as defined
442 in the following equations:

$$443 \quad Precision = \frac{TP}{TP+FP} \quad (10)$$

$$444 \quad Recall = \frac{TP}{TP+FN} \quad (11)$$

445 In these equations, TP is the number of correctly detected cube surfaces' (paper) pixels;
446 (TP+FP) is the number of detected cube surfaces' (paper) pixels; and (TP+FN) is the number
447 of actual cube surfaces' (paper) pixels. Precision basically means the area of correctly
448 recognized cube region divided by the total area of recognized cube regions and measures the
449 "exactness" of the detection algorithm. Recall is known as the area of correctly recognized
450 cube regions divided by the area of actual cube regions and shows the "completeness" of the
451 detection algorithm.

452 The results of calculating precision and recall ratios for different sizes of the calibration
453 objects compared to the entire size of the frames are illustrated in Figure 9.

454 As the next step, the corner points of the calibration objects were detected and
455 reconstructed. The average errors in computing the 2D locations of the extracted corner
456 points compared to the actual locations, as well as the re-projection errors for calculating the
457 3D locations of the corner points in the space were computed and demonstrated in Figure 10.
458 In this study, the 2D location error (%) was calculated by dividing the distance between the
459 computed and actual locations of the vertex on the image to the length of the longer edge of

460 the cube (paper) to where the vertex is located. The same approach, but in 3D, was
461 implemented for computing the re-projection errors.

462 To determine the threshold, the minimum precision and recall rates set to 95% and 90%
463 respectively (based on the collective evaluations of Figures 9 and 10). In addition, maximum
464 allowable error in 2D location of corner points and re-projection errors are considered as 2%
465 and 1%. As shown in figures 9 and 10, the smallest ratio for achieving the above mentioned
466 levels of accuracy is between 0.5-1 percentages. As the result, the minimum ratio of each
467 component surface to the entire image surface was set to 0.005 (0.5%).

468 Insert Figure 9 here

469
470 Insert Figure 10 here

471 **Identifying thresholds for the maximum and minimum roundness factors**

472 Using the same video data as the previous section, the roundness factors for the cube surfaces
473 (paper) in 437 frames were computed. Upper and lower thresholds for the roundness factor
474 can be identified by calculating the confidence intervals for this set of the measured
475 roundness factors:

$$476 \quad \text{upper and lower thresholds} = \left(\mu - 1.96 \frac{\sigma}{\sqrt{n}}, \mu + 1.96 \frac{\sigma}{\sqrt{n}} \right) \quad (13)$$

477 Where the confidence level is 95%, μ is the mean and σ is the standard deviation of the
478 measured roundness factors. After plugging the observed values, the upper and lower
479 thresholds were set to 0.85 and 0.1 respectively.

480 481 **Validation of the proposed methodology**

482 The validation procedure took place in two steps:

483

484 **Step 1: Validating the performance of the corner points' detection and matching**
485 **algorithm**

486 To evaluate the performance of the corner points' detection and matching algorithms, we
487 selected ten indoor and ten outdoor cases as our case studies (these case studies are different
488 from the initial scenes which were used for computing different thresholds). The indoor cases
489 include offices and different locations of homes, e.g. bathroom, living room and kitchen,
490 while the outdoor cases cover a variety of civil infrastructure scenes including campus
491 buildings, highway bridges, a train station building, a sport facility and an under-construction
492 wall in a construction jobsite. Each scene was videotaped as completely as possible, with
493 sensing from multiple viewpoints to minimize occlusions. An off-the-shelf Canon Vixia-HF
494 S100 was utilized for data collection purposes. The corners point detection and matching
495 algorithms were implemented for each captured video clip separately (Figure 11) and the
496 associated errors were measured in terms of precision and recall values for the surface
497 detection algorithm, deviation between computed and actual 2D location of corner points for
498 corner point detection algorithm and percentage of successfully corresponded corner points
499 for the matching algorithm. The summary of the results are presented in Table 2.

500 As shown in Table 2, the performance of the detection algorithm was the best for yellow
501 surfaces. It is necessary to highlight that we do not need to detect and reconstruct all the cube
502 vertices in all frames. It is only sufficient to successfully detect and reconstruct three vertices
503 of the cube for the entire video clip.

504 Insert Figure 11 here

505 Insert Table 2 here

506 **Step 2: Validating the overall performance of the proposed algorithm for computing the**
507 **absolute scale PCD of the scenes**

508 To validate the overall performance of the proposed methods, the captured video clips were
509 processed and the absolute scale PCD for each built infrastructure scene was generated
510 following the procedures explained in the methodology section. For each case study, we
511 consider the deviation between a number of real dimensions and computed dimensions of the
512 scene as the metric for measuring the accuracy of the presented methods. For each scene,
513 several dimensions and distances were identified and measured by a TC805 total station for
514 outdoor cases and a Leica DISTO D5 Laser measurer for indoor environment (Figure 12).
515 The average measuring time for measuring each dimension of the outdoor setting is around
516 15 minutes. This time includes possible traversing between different locations within the
517 jobsite (for large scale jobsites or the cases that data should be collected from different sides
518 of a building), setting up and adjusting the total station, conducting measurements, converting
519 the files into the computer, manually finding the corresponding dimensions on the PCD and
520 calculating the scale factor.

521 Samples of generated PCD for both indoor and outdoor case studies are presented in
522 Figures 13 and 14.

523 The results of computing the accuracy of the proposed methods in measuring different
524 dimensions within built infrastructure case studies are summarized in Table 3.

525 Insert Figure 12 here

526 Insert Figure 13 here

527 Insert Figure 14 here

528

Insert Table 3 here

529 Illustrated results in table 3 indicate that the performance of the algorithm is promising (<4
530 mm per meter error for outdoor settings and <2 mm per meter error for indoor case studies).
531 Compared to other common measurement devices, e.g. measurement tape and total station,
532 this approach is not the most accurate method. However, based on experts' opinions, the
533 obtained level of accuracy is sufficient for a number of applications in the area of A/E/C. For
534 example, the obtained level of accuracy would suffice for rough quantity take offs, e.g.
535 calculating surfaces of wall for painting or surface of the floor for carpeting; or interior layout
536 design, e.g. comparing the dimensions of different elements in a room or office and making
537 decisions about new furniture which fits properly. Automating the procedure is the biggest
538 advantage of the proposed approach over the traditional measurement devices.

539

540 **Summary and conclusions**

541 Calculating the absolute scale of PCD generated by monocular photo/videogrammetry is a
542 challenging task for practitioners in the field of A/E/C. The potential solution should entail
543 the following characteristics:

- 544 - It should not rely on any specific hardware settings or extra sensors for measurements so
545 it can be easily applied in almost all built infrastructure job sites.
- 546 - It should be simple, yet general enough to cover a variety of applications in both indoor
547 and outdoor environments.
- 548 - The solution should be cost effective with the minimal amount of human involvement in
549 the pre/post processing stages.

550 - In the case of using predefined objects as the registration targets, the applied objects
551 should be easily used in almost every job site. In addition, considering the dynamic and
552 cluttered environments of built infrastructure job sites, high success rates for detecting
553 and reconstructing the registration targets, as well as minimized amounts of error in
554 computing absolute scale, is crucial.

555 In this paper, an effective method for automatically computing the absolute scale of
556 PCD's obtained from indoor/outdoor built infrastructure scenes was presented and validated.
557 Computing the absolute scale of PCD is a major issue faced by civil engineers and facility
558 managers since they need to extract the real measurements from video-generated PCD with
559 scale uncertainty. The proposed algorithm is based on detecting, matching and reconstructing
560 the corner points of two simple categories of objects: a letter size piece of paper for indoor
561 applications and a plywood cube for outdoor, large scale cases. The average length
562 measurement errors resulted by implementing the proposed method for indoor and outdoor
563 scenarios were 0.14cm and 0.37 cm per meter respectively. The experiment results revealed
564 that the proposed method enables A/E/C practitioners to accurately scale up PCD with least
565 amount of manual work and without the need for extra sensor/prior knowledge about the
566 scene. As the extension of the current research, the authors will conduct more experiments in
567 both indoor and outdoor settings to better evaluate the performance of the method and reduce
568 the errors. In particular, the authors will focus on the drift problem and the effects of the
569 number, size and location of calibration objects on the accuracy of computed measurements.

570 **References**

571 Anati, R., Scaramuzza, D., Derpanis, K., and Daniilidis, K. (2012). "Robot localization using

572 soft object detection.” *Proc. IEEE Int. Conf. Robot. Autom.*, Minnesota, USA, 4992-4999.

573 Becerik-Gerber, B., Siddiqui, M., Brilakis, I., El-Anwar, O., El-Gohary, N., Mahfouz, T.,
574 Jog, G., Li, S., and Kandil, A. (2013). "Civil Engineering Grand Challenges:
575 Opportunities for Data Sensing, Information Analysis, and Knowledge Discovery."
576 *Journal of Computing in Civil Engineering*, 10.1061/(ASCE)CP.1943-5487.0000290

577 Botterill, T., Mills, S., and Green, R. (2012). "Correcting scale drift by object recognition in
578 single-camera SLAM." *IEEE Transactions on Systems, Man, and Cybernetics*, PP(99), 1-
579 14.

580 Brilakis, I., Fathi, H., and Rashidi, A. (2011). "Progressive 3D reconstruction of
581 infrastructure with videogrammetry." *Automation in Construction*, 20(7), 884-895.

582 Carozza, L., Tingdahl, D., Bosché, F., and Van Gool, L. (2012). "Markerless vision-based
583 augmented reality for urban planning." *Computer-Aided Civil and Infrastructure*
584 *Engineering*, Early View (Online Version of Record published before inclusion in an
585 issue).

586 Castellanos, J. A., Montiel, J. M. M., Neira, J., and Tardós, J. D. (2000). "Sensor influence in
587 the performance of simultaneous mobile robot localization and map building." *Proc., In*
588 *6th International Symposium on Experimental Robotics*, Sydney, Australia, 203-212.

589 Chi, S., and Caldas, C. H. (2011). "Automated object identification using optical video
590 cameras on construction sites." *Computer-Aided Civil and Infrastructure Engineering*,

591 26(5), 368-380.

592 Coaker, L.H. (2009). "Reflector-less total station measurements and their accuracy, precision
593 and reliability." Dissertation, USQ Project, University of Southern Queensland, Australia.

594 Dai, F., Rashidi, A., Brilakis, I., and Vela, P. (2013), "Comparison of image-based and time-
595 of-flight-based technologies for three-dimensional reconstruction of infrastructure." *ASCE*
596 *Journal of Construction Engineering and Management*, 139(1), 69-79.

597 Dias, J. M. S., Bastos, R., Correia, J., and Vicente, R. (2006). "Semi-automatic 3D
598 reconstruction of urban areas using epipolar geometry and template matching." *Computer-*
599 *Aided Civil and Infrastructure Engineering*, 21(7), 466-485.

600 Eudes, A., Lhuillier, M., Naudet, S., and Dhome, M. (2010). "Fast odometry integration in
601 local bundle adjustment-based visual SLAM." *Proc., In 20th International Conference of*
602 *Pattern Recognition (ICPR)*, Istanbul, Turkey, 290-293.

603 Fathi, H., and Brilakis, I. (2013) "A videogrammetric as-built data collection method for
604 digital fabrication of sheet metal roof panels", *Journal of Advanced Engineering*
605 *Informatics*, 27(4), 466-476

606 Ganci, G. and Brown, J. (2008). "Developments in Non-Contact Measurement Using
607 Videogrammetry." *Boeing large scale metrology seminar*, Melbourne, FL.

608 Golparvar-Fard, M., Tang, P., Cho, Y., and Siddiqui, M. (2013) Grand Challenges in Data
609 and Information Visualization for the Architecture, Engineering, Construction, and

610 Facility Management Industries. Proceedings of the ASCE International Workshop on
611 Computing in Civil Engineering, Clearwater Beach, CA.

612 Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2012), “Automated progress
613 monitoring using unordered daily construction photographs and IFC-based building
614 information models.” *Journal of Computing in Civil Engineering*, in press.

615 Gutierrez-Gomez, D., Puig, L., and Guerrero, J. J. (2012). “Full scaled 3D visual odometry
616 from a single wearable omnidirectional camera.” *International Conference on Intelligent
617 Robots and Systems (IROS)*, Vilamoura, Portugal, 4276-4281.

618 Jahanshahi, M. R., Masri, S. F., Padgett, C. W., and Sukhatme, G. S. (2013). “An innovative
619 methodology for detection and quantification of cracks through incorporation of depth
620 perception.” *Machine Vision and Applications*, 24(2), 227-241.

621 Jung, M. J., Myung, H., Hong, S. G., Park, D. R., Lee, H. K., and Bang, S. W. (2004).
622 “Structured light 2D range finder for simultaneous localization and map-building (SLAM)
623 in home environments.” *Proc., In IEEE 4th International Symposium on Micro-
624 NanoMechatronics and Human Science*, 371-376.

625 Kneip, L., Martinelli, A., Weiss, S. Scaramuzza, D., and Siegwart, R. (2011). “A closed-
626 form solution for absolute scale velocity determination combining inertial measurements
627 and a single feature correspondence.” *Proc., In IEEE International Conference on
628 Robotics and Automation (ISRA)*, Shanghai, China, 4546-4553.

629 Kuhl, A., Wöhler, C., Krüger, L., d'Angelo, P., and Groß, H. M. (2006). "Monocular 3D
630 scene reconstruction at absolute scales by combination of geometric and real-aperture
631 methods." *Lecture Notes in Computer Science*, 4174, 607-616.

632 Li, H., Huang, Y., Ou, J., and Bao, Y. (2011). "Fractal dimension-based damage detection
633 method for beams with a uniform cross-section." *Computer-Aided Civil and*
634 *Infrastructure Engineering*, 26(3), 190-206.

635 Montgomery, D. C., Peck, E. A. and Vining, G. (2012) "Introduction to Linear Regression
636 Analysis, 5th Edition" Wiley: Wiley Series in Probability and Statistics

637 Nützi, G., Weiss, S., Scaramuzza, D., and Siegwart, R. (2011). "Fusion of IMU and vision for
638 absolute scale estimation in monocular SLAM." *Journal of Intelligent and Robotic*
639 *Systems (JIRS)*, 61(1-4), 287-299.

640 Pavlidisa, G., Koutsoudisa, A., Arnaoutogloua, F., Tsioukasb, V., and Christodoulos, C.,
641 (2007) "Methods for 3D digitization of Cultural Heritage", *Journal of Cultural Heritage*,
642 8(1), 93-98.

643 Photo modeler, (2013). "Overview of applications for Photo-Modeler."
644 <http://www.photomodeler.com/applications/default.html>.

645 Rashidi, A., Fathi, H., and Brilakis, I. (2011). "Innovative stereo vision-based approach to
646 generate dense depth map of transportation infrastructure." *Transportation Research*
647 *Record, Journal of the Transportation Research Board*, 2215, 93-99.

648 Rashidi, A., Dai, F., Brilakis, I., and Vela, P. (2013). “Optimized selection of key frames for
649 monocular videogrammetric surveying of civil infrastructure.” *Journal of Advanced*
650 *Engineering Informatics*, 27(2), 270-282.

651 Saidi, K., Cheok, G., Franaszek, M., Brown, C., Swerdlow J., Lipman, R., Katz, I.,;
652 Golparvar-Fard, M., Goodrum, P., Akula, M., Dadi, G, and Ghadimi, B (2011),
653 Development and Use of the NIST Intelligent and Automated Construction Job Site
654 Testbed, NIST TN – 1726 report, 2011

655 Scaramuzza, D. Fraundorfer, F. Pollefeys, M., and Siegwart, R. (2009). “Absolute scale in
656 structure from motion from a single vehicle mounted camera by exploiting nonholonomic
657 constraints.” *Proc., In IEEE International conference on computer vision*, Kyoto, Japan,
658 1413-1419.

659 Olson, E. (2011). “AprilTag: A robust and flexible visual fiducial system”, *Proceedings of*
660 *the IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China

661 Tribou, M. (2009). “Recovering scale in relative pose and target model estimation using
662 monocular vision.” M.S. thesis, University of Waterloo, Waterloo, Ontario, Canada.

663 Zhang, G., and Wang, Y. (2013). “Optimizing Coordinated Ramp Metering: A Preemptive
664 Hierarchical Control Approach.” *Computer-Aided Civil and Infrastructure Engineering*,
665 28(1), 22–37.

666

667 **List of Tables**

668 **Table 1.** Different types of 3D reconstruction approaches

669 **Table 2.** Summary of the results obtained from implementing the corner detection and
670 matching algorithms

671 **Table 3.** Summary of the results obtained from evaluating the overall performance of the
672 proposed method

673

674

675

676

677

678

679

680

681

682

683

684

685

686 Table 1: Different types of 3D reconstruction approaches
687

Known parameters	Reconstruction level
Intrinsic and extrinsic	Absolute scale reconstruction
Only intrinsic	Metric reconstruction (up to an unknown scale)
No information	Projective reconstruction

688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707

708 Table 2: Summary of the results obtained from implementing the corner detection and
 709 matching algorithms.

Experimental setting		Average error in 2D corner points detection algorithm*		Average error in 2D corner points detection algorithm *	Average accuracy of 2D matching algorithm (%)
	Surface	Precision (%)	Recall (%)		
Outdoor setting (cube)	Red	92.1	90.8	0.03	98.7
	Yellow	96.5	94.1		
	Purple	91.8	89.9		
Indoor setting (sheet of paper)		98.3	92.1	0.01	100

710 *error is calculated as Δ/l where Δ is the deviation between actual and computed 2D locations of the corner
 711 points (in pixel) and l is the longest associated vertex.
 712

713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727

728 Table 3: Summary of the results obtained from evaluating the overall performance of the
729 proposed method

Experimental setting	Indoor	Outdoor
Average number of measurements for each case study	107	281
Average error* (mm per meter)	1.4	3.7
Maximum error (mm per meter)	4.2	8.5
Standard Deviation	0.7	1.8

730 *error is measured based on the ratio of computed dimensions to actual dimensions per unit of length (meter)

731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750

751

752

753

754

755

756

757

758 **List of Figures**

759 **Fig. 1.** Projector and camera setup for extracting absolute measurements in manufacturing
760 industry (Ganci and Brown, 2008)

761 **Fig. 2.** Selected colors for surfaces of the cube (top) and snapshots of the cube (bottom)

762 **Fig. 3.** Overall workflow of the proposed algorithm for computing absolute scale of PCD

763 **Fig. 4.** Necessary steps for detection of the cube vertices

764 **Fig. 5.** Convex hull algorithm: a) non-convex surface, c) equal convex hull surface

765 **Fig. 6.** Corresponding corner points for the first view (left), are located on epipolar line in the
766 next view (right)

767 **Fig. 7.** Possible locations for the letter-size sheet of paper in indoor settings

768 **Fig. 8.** Locations of corner points of the sheet of paper follow the same clockwise order in
769 different views

770 **Fig. 9.** Precision and recall ratios for detection of the cube surfaces (top) and sheet of paper
771 (bottom)

772 **Fig. 10.** Errors in 2D locations (top) and re-projection errors in computing 3D locations
773 (bottom) of the cube vertices and corners of the paper

774 **Fig. 11.** Sample of the implementation results for the cube corners detection algorithm: from
775 top left: the original image of the cube-the result of filtering the image based on HSV
776 thresholds- Detected red, yellow and purple surfaces- detected lines based on the improved
777 Hough transform - and the intersections of the cube edges as the final result.

778 **Fig. 12.** Actual distance measurements and preparation of ground truth: Leica TC805 total
779 station (left) and Leica DISTO D5Laser measurer (middle and right)

780 **Fig. 13.** A samples of the generated PCD for indoor settings: bathroom

781 **Fig. 14.** Samples of the generated PCD for outdoor settings: Campus building (top) and
782 construction wall (bottom)

783

784

785

786

787

788

789

790

791

792

793

794

795



796

797 Figure 1: Projector and camera setup for extracting absolute measurements in manufacturing
798 industry (Ganci and Brown, 2008)

799

800

801

802

803

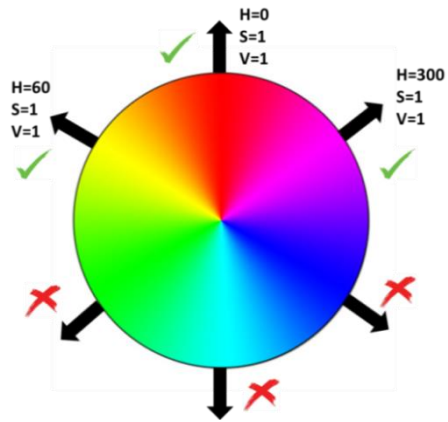
804

805

806

807

808



809



810

811 Figure 2: Selected colors for surfaces of the cube (top) and snapshots of the cube (bottom)

812

813

814

815

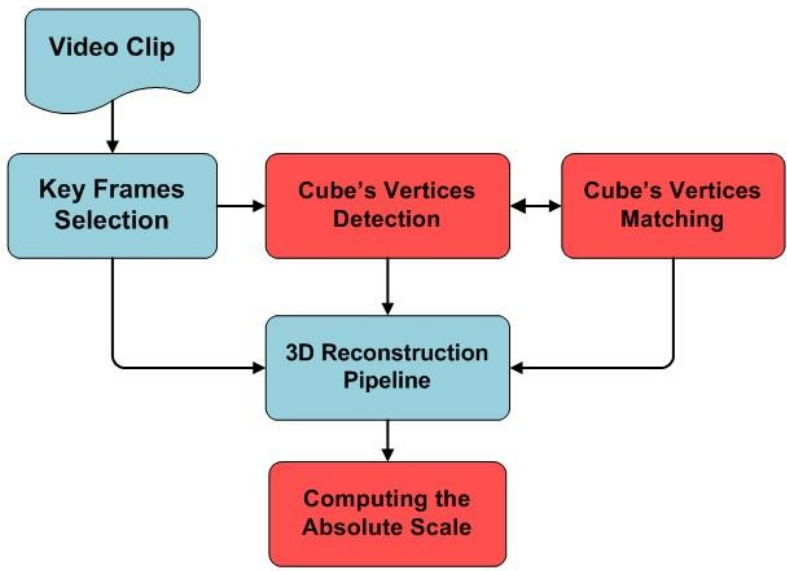
816

817

818

819

820



821

822 Figure 3: Overall workflow of the proposed algorithm for computing absolute scale of PCD

823

824

825

826

827

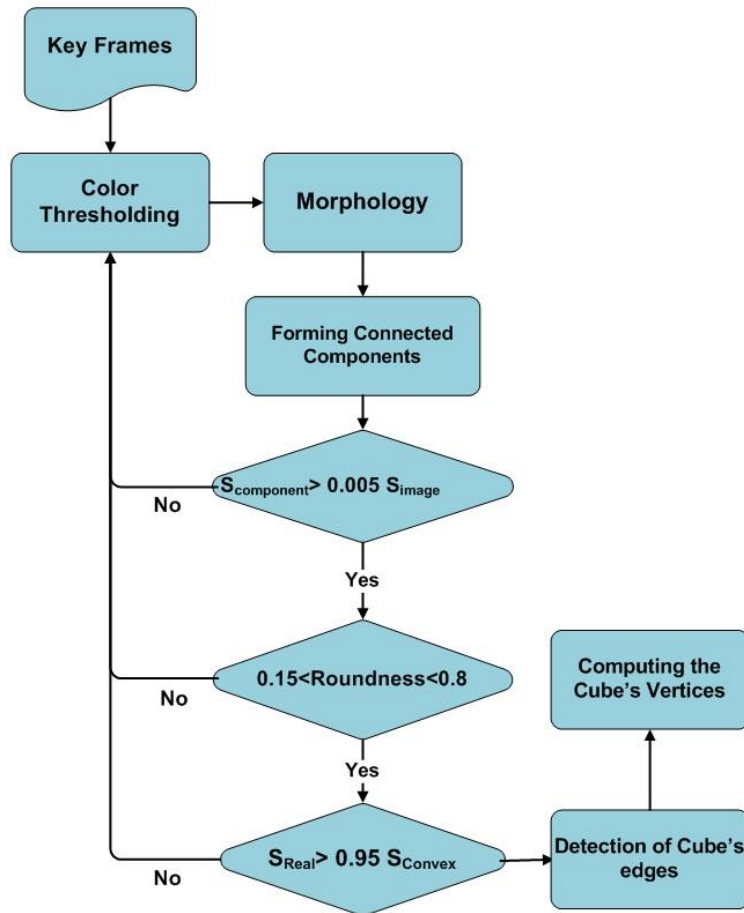
828

829

830

831

832



833

834

Figure 4: Necessary steps for detection of the cube vertices

835

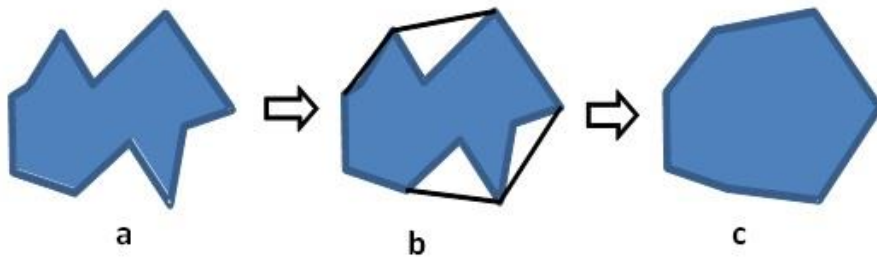
836

837

838

839

840



841

842 Figure 5: Convex hull algorithm: a) non-convex shape, b) constructing an equal convex hull

843 for the initial shape and c) reconstructed convex hull shape

844

845

846

847

848

849

850

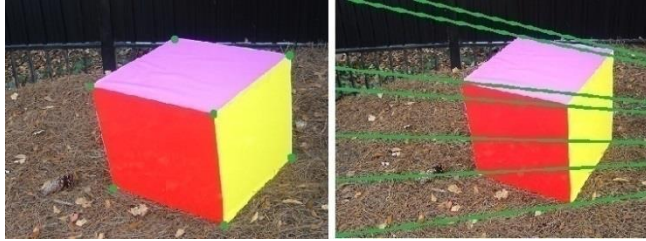
851

852

853

854

855



856

857 Figure 6: Corresponding corner points for the first view (left), are located on epipolar line in

858

the next view (right)

859

860

861

862

863

864

865

866

867

868

869

870



871

872

Figure 7: Possible locations for the letter-size sheet of paper in indoor settings

873

874

875

876

877

878

879

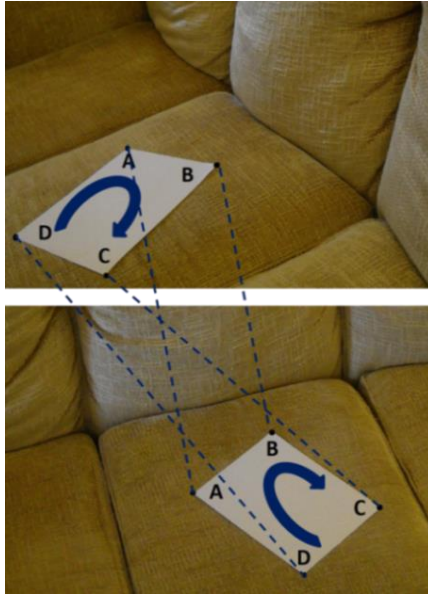
880

881

882

883

884



885

886 Figure 8: Locations of corner points of the sheet of paper follow the same clockwise order in

887

different views

888

889

890

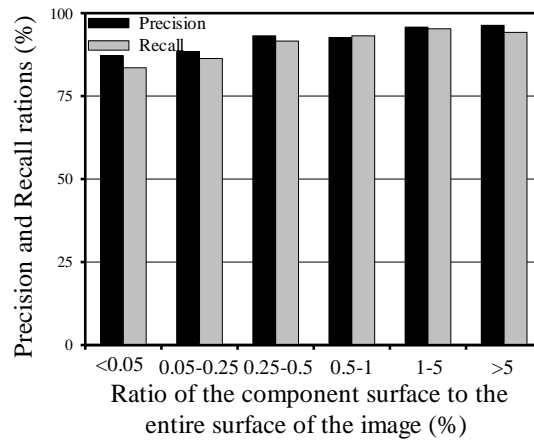
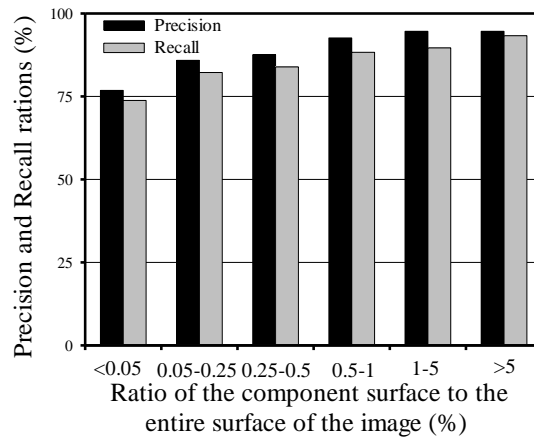
891

892

893

894

895



896

897 Figure 9: Precision and recall ratios for detection of the cube surfaces (top) and sheet of paper

898

(bottom)

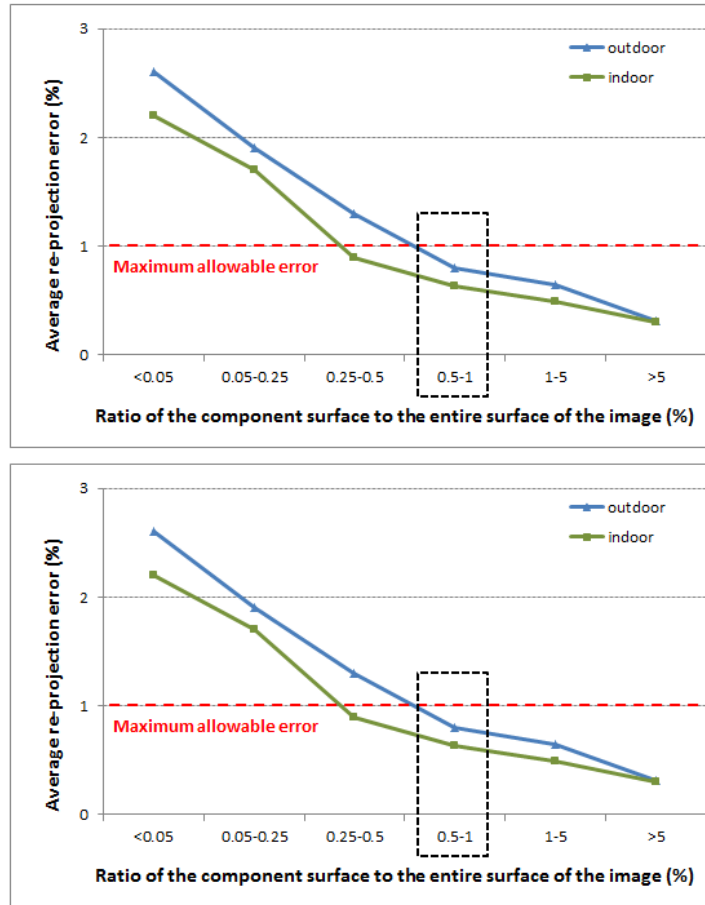
899

900

901

902

903



904

905

Figure 10: 2D location errors (top) and re-projection errors (bottom) for both indoor and

906

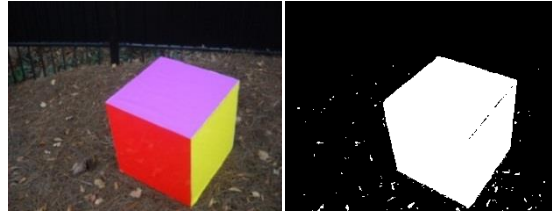
outdoor settings

907

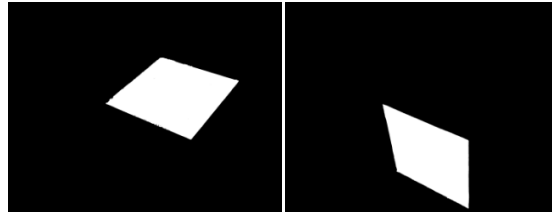
908

909

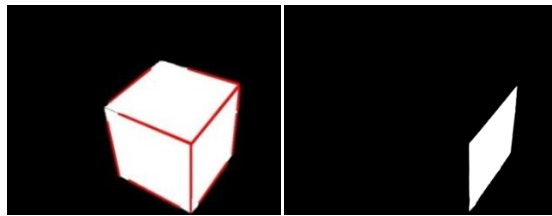
910



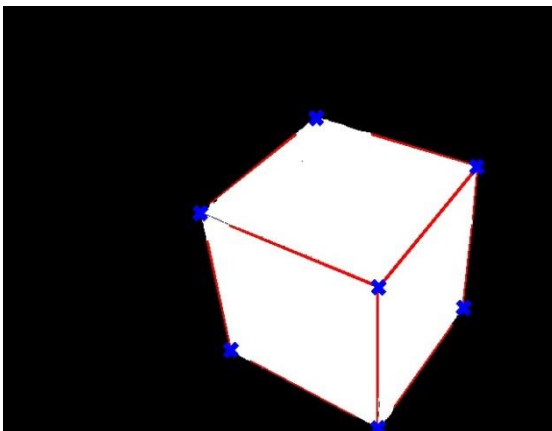
911



912



913



914 Figure 11: Sample of the implementation results for the cube corners detection algorithm:
915 from top left: the original image of the cube-the result of filtering the image based on HSV
916 thresholds- Detected red, yellow and purple surfaces- detected lines based on the improved
917 Hough transform - and the intersections of the cube edges as the final result.

918

919



920

921

Figure 12: Actual distance measurements and preparation of ground truth: Leica TC805 total

922

station (left) and Leica DISTO D5Laser measurer (middle and right)

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

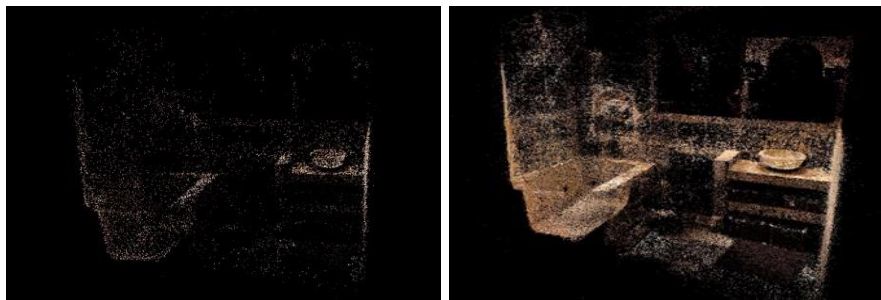
943



944

945

946



947

Figure 13: A sample of the generated PCD for indoor settings: bathroom- Sparse PCD

948

generated by SfM (left) and PCD generated by PMVS (right)

949

950

951

952

953

954

955

956

957

958

959

960



969 Figure 14: Samples of the generated PCD for outdoor settings: Campus building (top) and
970 construction wall (bottom)

971