



# Projector-based augmented stacking framework for irregularly shaped objects

Andrea Settimi<sup>1</sup> · Qianqing Wang<sup>2</sup> · Edward Andò<sup>3</sup> · Julien Gamarro<sup>1</sup> · Katrin Beyer<sup>2</sup> · Yves Weinand<sup>1</sup>

Received: 21 December 2022 / Accepted: 10 April 2023  
© The Author(s) 2023

## Abstract

Augmented reality in additive fabrication is predominantly applied to the manufacturing of structures from regularly shaped materials. Our work however embraces natural heterogeneity, by focusing on the assembly of irregularly shaped elements such as mineral scraps. We introduce a computer-aided framework with a 3D stacking engine and an augmented reality interface capable of guiding users through the balanced positioning of discrete and highly nonuniform objects according to an on-the-fly computed model and without prior trimming of the building unit. The pipeline is tested by the construction of two dry-stone structures (i.e., lacking any adhesives), and its accuracy is validated with a comparative study between the point cloud of the as-built digitized artifacts and the generated execution model, which shows an average  $2.9 \pm 1.8$  cm error between the stones of the two models. We finally show how the proposed framework could be improved both in terms of software and hardware. In the interests of reproducibility, all the methods are shared as open source with the community.

## Graphical abstract



**Keywords** Augmented reality · Stacking algorithm · Computer-assisted assembly · Human–machine interaction

✉ Andrea Settimi  
andrea.settimi@epfl.ch

<sup>1</sup> IBOIS, EPFL: Laboratory for Timber Construction, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

<sup>2</sup> EESD, EPFL: Earthquake Engineering and Structural Dynamics, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

<sup>3</sup> IMAGING CENTER, EPFL: Imaging Center, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

## Abbreviations

AR	Augmented reality
LiDAR	Light detection and ranging
HMD	Head-mounted display
CLI	Command-line interface
FoV	Field of view
IR	Infra-red
UI	User interface
OBB	Oriented bounding box
CoM	Center of mass

AABB	Axis-aligned bounding box
FPS	Frame-per-second
ICP	Iterative closest point
HUD	Heads-up display
FPS	Frames-per-second
ROI	Region of interest
CD	Chamfer distance
EMD	Earth mover's distance
RMSE	Root mean square error
SLAM	Simultaneous localization and mapping

## 1 Introduction

Augmented reality (AR) applied to manual additive fabrication has grown in popularity in recent years (Wang et al. 2022) and revealed itself to be a profitable human–machine collaboration format (Hoc 2000) to enhance the efficiency of already present personnel (Bottani and Vignali 2019). Nonetheless, the majority of AR assembly applications are centered on numerically produced elements or regular and standardized components, e.g. bricks (Mitterberger et al. 2020). In opposition, minimally processed or reclaimed materials are characterized by unpredictable shapes which ultimately complexify their adoption in modern augmented digital processes.

Besides, the use of irregular materials in additive tasks has been mainly targeted by the robotic construction domain. In particular, robotic alternatives to stacking processes have been theorized (Thangavelu et al. 2018), experimented in controlled environments (Aejmelaeus-Lindström et al. 2016; Furrer et al. 2017; Wermelinger et al. 2018; Liu et al. 2021) and applied to real scenario demonstrators built by a mobile robot (Aejmelaeus-Lindström et al. 2020) and a robotic excavator in a fully fledged unstructured environments (Johns et al. 2020). Such applications have been proven to be successful but for either one of the extremes of the irregular material's size spectrum like gravel or boulders.

A manual system guided by computational intelligence could help overcome limitations faced by the robotic assembly of irregular geometries, notably, dealing simultaneously with a very large number of differently sized elements, in site characterized by severe spatial constraints in e.g. concentrated urban areas, and for densely packed target volumes such as dwelling's walls. Human cognitivity and dexterity have also the potential in adding flexibility and adaptability to autonomous processes otherwise too brittle to unexpected events or adjustments.

We present a projection-based AR guidance system for a digitized version of the manual stacking of unaltered rigid irregular materials. The worker will be assisted by instructions generated from a numerically computed packing model. The paper describes its application in a real-life scenario for

an alternative version of traditional dry stone assembly, a complex manufacturing technique usually requiring highly skilled personnel and intense unit processing. Overviews are provided on the general setup, the stone dataset digitization, and how the stacking algorithm computes each unit's best pose within the scanned as-built landscape. We describe the processing of sensing data and the design of the augmented projected interface. The AR framework is also evaluated by constructing two single-layer walls of 1.7 m length, 0.6 m height, and 0.7 m width, fabricated out of as-found mineral scraps issued by quarry extraction processes. The structures are first digitized with light detection and ranging (LiDAR) scanning, and the obtained models are compared with their ground-truth data. We finally outline potential improvements to the current system's limitations and future developments.

## 2 Relevant works

AR-assisted manufacturing of unpredictable geometries is a topic explored in limited research works (Arena et al. 2022; Syed et al. 2023). Among those we can report Larsson et al. Larsson and Yoshida (209) who propose an audio-visual tracking system based on fiducial markers able to guide users through the machining of moderately sized irregular tree branches with a 6-axis router. Jahn et al. Jahn et al. (2022) showcase how introducing a feedback loop logic based on depth sensing streams can successfully assist a user during a sculpting operation.

Nevertheless, AR-based additive applications tend to focus rather on the assembly of geometries with homogeneous shapes. Their pose can be easily detected and monitored during the process with either CAD-referenced fiducial markers (Hughes et al. 2021) or object-based visual-inertial tracking techniques (Sandy and Buchli 2018). Concerning the latter approach, the demonstrator proposed by Gard et al. Gard et al. (2022) illustrates how, given a known geometry, it is possible to provide AR-assisted assembly of multi-state objects and multiple object-based tracking of less predictable, untextured elements. Several other kindred object-based algorithms can be applied to guide the correct positioning of an object in generic additive manufacturing (Huang et al. 2021; Stoiber et al. 2022). Nevertheless, such methods present technical limitations in monitoring numerous and similarly textured, irregular, densely packed, elements.

Before the recent appearance on the market of advanced head-mounted displays (HMDs), researchers have often opted for projection-based AR interfaces to guide users through precise subtractive (Rivers et al. 2013) or large-scale additive operations (Yoshida et al. 2015). Although lacking in ergonomics and ease of deployment, spatial projected interfaces remain to this day a well-suited alternative for multi-user AR applications by avoiding most of the typical

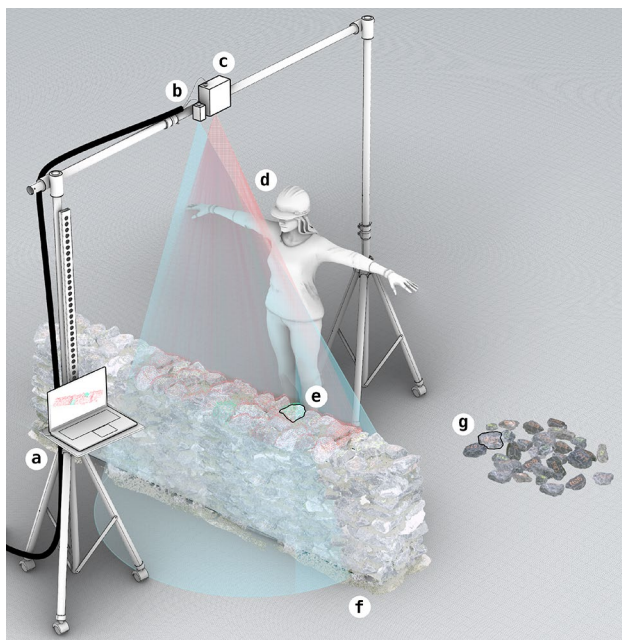
AR limitations in such a scenario: per-device spatial anchor sharing, need for stable network connectivity and displays' lag between users caused by rendering bottlenecks. This is proven by the recent project of Mitterberger et al. Mitterberger et al. (2022), which showcases how projector-based AR can be an effective interface in designing and guiding contemporary large-scale manufacturing operations.

Robotics projects have also offered numerous approaches to the geometric packing and planning of nonstandard rigid materials in recent years. Johns et al. Johns et al. (2020) propose a model generator that can build multiple layers following an arbitrarily specified bounding box and an updated model of the as-built landscape. Also in Thangavelu et al. Thangavelu et al. (2020) a sensing system is used to inform the planner of the actual state of the overhaul target structure's morphology to compute the next best pose. In Furrer et al. Furrer et al. (2017) and Liu et al. Liu et al. (2021), the entire dataset of objects is pre-scanned only to be thereafter fed on-the-fly as single elements to the planner.

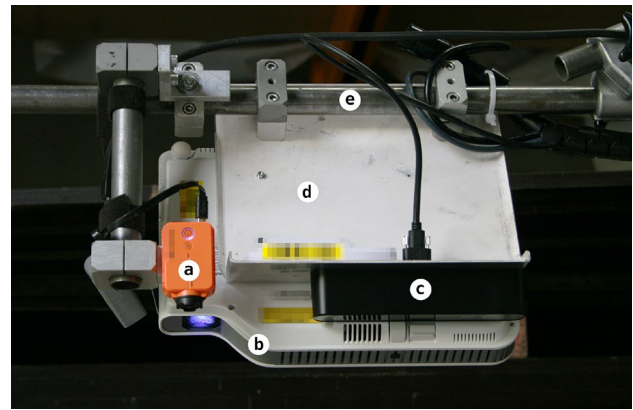
### 3 Methodology

#### 3.1 System setup

The assembly station is designed as a demountable tubular structure built from two lateral, vertical supports and a single horizontal beam placed 2.3 m from the ground.



**Fig. 1** Overview of the proposed digital manual fabrication setup: (a) Computational unit, (b) stereo camera, (c) LED/laser beamer, (d) operator, (e) currently placed stone, (f) assembly area, (g) pool of available stones

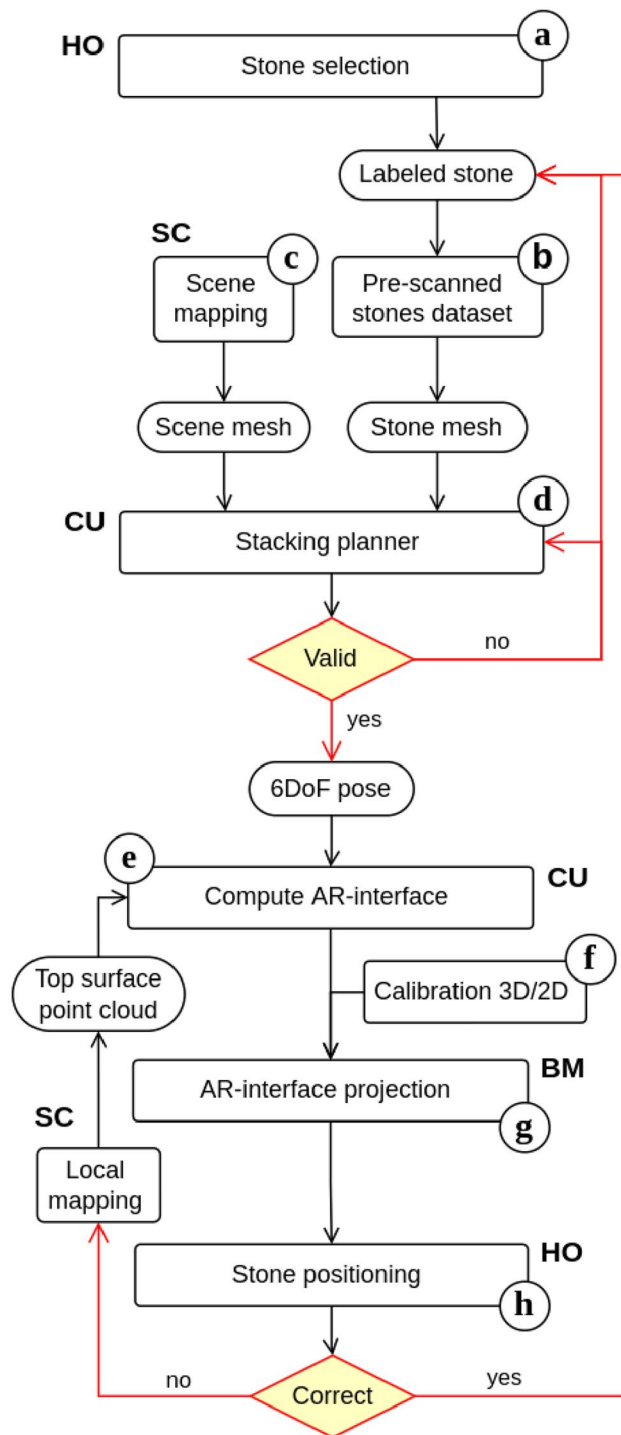


**Fig. 2** Close-up detail of the hardware mounted on the overhanging portal beam: (a) RGB camera, (b) LED/laser beamer, (c) stereo camera, (d) metal support plate, (e) portal beam

The middle section of the main spanning element hosts the on-board computer vision devices (Fig. 1b, c) featuring an RGB camera (Fig. 2a), a 3D stereoscopic camera (Stereo-labs Inc. 2022), and a XJ-A252 hybrid LED/laser beamer with a maximal brightness power of 3000 ANSI-Lumens (CASIO COMPUTER CO. 2014) (Fig. 2b). All the hardware components are connected to one computational unit (Fig. 2a) responsible for receiving 3D data feeds from the stereo camera, processing and outputting projected-visual stimuli for the operators to receive. The sensing back ends and the major base code body is written in Python, whereas the stacking planner has been implemented in C++. Direct access to the application is made possible via a simple command-line interface (CLI) developed on the Linux-based operative system Ubuntu 22.04 LTS. The source code has been released and is openly accessible under a permissive MIT license (Settimi et al. Oct. 2022). The working area is defined by the portion of ground where the field of view (FoV) of both the sensing device and projector overlap (see Fig. 2f). The surroundings are occupied by stones stockpiled within the operator's reach to be constantly selected along the entire building process (Fig. 2g).

#### 3.2 Digital-physical workflow

The diagram of Fig. 3 illustrates the broad sequencing and the dataflow of all human or machine agents involved in the proposed fabrication hybrid system. Before beginning the assembly, a calibration is required to later convert computed data from 3- to 2-dimensional raster frames to be projected (Fig. 3g). Additionally, the entire stone dataset at the operator's disposal has been labeled and digitized beforehand



**Fig. 3** Diagram of the digital-physical workflow: (HO) Human Operator, (SC) stereoscopic camera, (CU) Computational Unit, (BM) LED/Laser Beamer

(Fig. 3b). As the operator selects one stone from the pile without specific computational guidance (Fig. 3a), its associated 3D model is queried and fetched from the cloud dataset by feeding the label as a user input to the software. The

retrieved mesh is next fed to the geometric planner (Fig. 3d), together with a 3D mapping of the current scene (Fig. 3c). The planner computes the input stone's best position within a hard-coded bounding box corresponding to the targeted structure's dimensions. Once the output pose is compiled and visualized, the operator is asked to whether accept the pose, restart the geometric node or select a different stone. If the pose is categorized as validated, an interactive augmented interface is generated (Fig. 3e) and projected as an overlay to the assembly area (Fig. 3g). During the assembly, the interface node instructs the operator with different guidance visuals on the correct positioning and orientation of the designated stone (Fig. 3h). Once the placement criteria are met, the operator validates the stone's location and the current as-build model is updated. Ultimately, the described workflow is repeated until the structure is completed.

### 3.3 Stones entry-dataset

Prior to any assembly operations, all the building units need to be labeled and digitized via hand-scanning (Fig. 4). Stones are grouped in batches of approximately 10 and digitized via a manual infra-red (IR)-based scanning procedure with a handheld scanner (FARO Technologies Inc. 2022). Through a previously tested post-processing pipeline, separate point cloud fragments of the same stone are registered, merged, cleaned from outliers, down-sampled, and finally, water-tight meshes are obtained via Poisson surfacing (Settimi et al. 2022).

In the current version, the dataset consists of 444 entities with a total number of polygon face not exceeding the threshold of 500 when consumed by the software (Settimi et al. 2022). Reducing the definition of digital models appears to be a necessary trade-off between computational processing time and accuracy both for the geometric planner and the AR-interface generator. Selected stone lengths span from a minimum of 10 cm to a maximum of 60 cm (Fig. 5). The upper limit of the selected range is mostly set by the payload capacity of an average human worker, which is approximately 25 kg. Stones with a diameter inferior to 10 cm are classified as rubble. Thus, they are not individually scanned and are used indiscriminately as filling material for the cavities of the assembled structure as in traditional assembly techniques. The inventory is characterized by an inconsistent geometrical variation. Although the majority of the stones' shapes could be categorized as 2.5D since presenting at least two parallel faces, the overhaul irregularity contributes greatly to the categorization of such material stock as waste since regularization operations result in extra time and costs for the fabrication pipeline. By informing the geometric planner of the geometric model of each assembly unit, it is possible to avoid any trimming normally required for dry-stone stacking.



**Fig. 4** Picture of mineral scraps from quarry's water-sawing processes. This typology is the most geometrical by-product of quarries' extraction and transformation operations. As all the more irregular waste typologies present in this study, each element has been labeled, scanned, and stored in a dataset



### 3.4 Stacking planning

At the beginning of each construction step, the stone's index of the selected stone from the pile is manually inserted through the command line interface of the software (Fig. 7e). This operation is required to retrieve the corresponding 3D geometrical model of the stone. Then with the point cloud of the current scene captured by the stereo camera (as shown in Fig. 6), the geometric planning aims to solve the translation and rotation of the stone such that the transformation can be executed by the operator successfully. To this end, a stacking algorithm is developed, solving the problem with an optimization formulation, written as the following:

$$\min_M \quad \text{CoM}_{S'_t}, \quad (1a)$$

$$\min_M \quad \text{AABB}_{S'_t}, \quad (1b)$$

$$\text{s.t.} \quad S'_t = MS_t, \quad (1c)$$

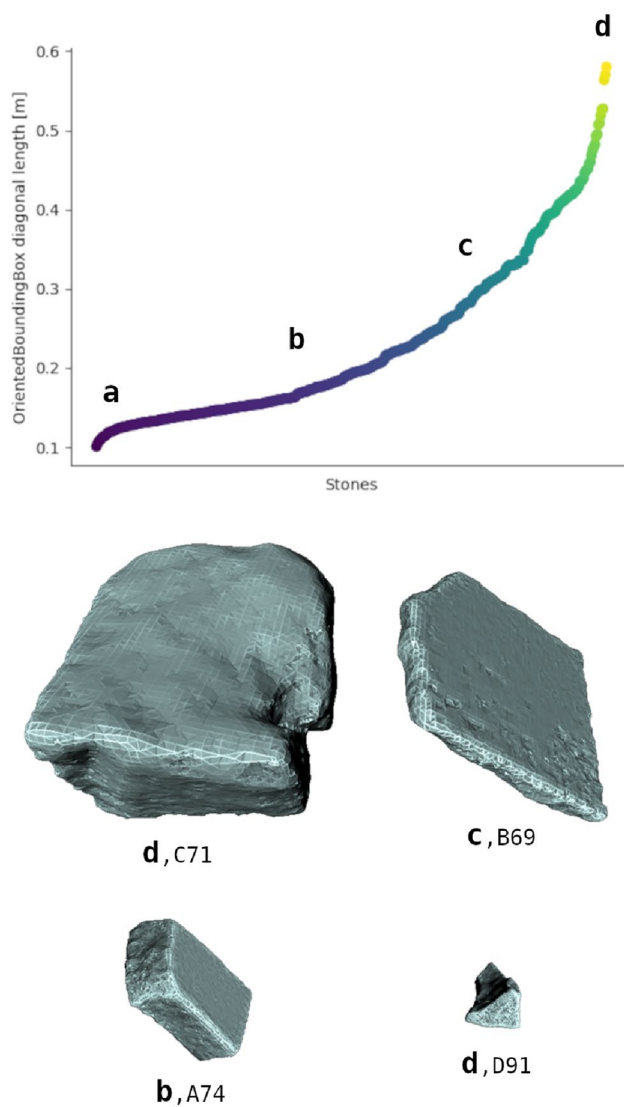
$$S'_t \cap S_c = \emptyset, \quad (1d)$$

$$S'_t \in \text{BB}_{\text{wall}}. \quad (1e)$$

Here  $S'_t$  refers to the transformed stone after applying the transformation solution  $M$  to the original stone  $S_t$ , as shown in Eq. 1c. There are two objectives in the proposed formulation: Eq. 1a aims to place the stone as low as possible by minimizing the height of the center of mass (CoM) of the current stone. Equation 1b minimizes the volume of the axis-aligned bounding box (AABB, refers to the bounding box that is aligned with the axes of the

global coordinate system) of the placed stone, as a smaller value leads to a better geometric index in the evaluation of masonry wall panels (Almeida et al. 2016). Apart from the objective, two constraints are considered. First, the stone should not overlap with any objects in the current scene  $S_c$  (Eq. 1d), which is a mesh generated from the point cloud obtained from the stereo camera (see Fig. 7b). Also, the stone should be transformed into the desired wall space, which is defined by a bounding box (Eq. 1e).

The formulated optimization problem is solved by a two-fold process. Since the volume of the axis-aligned bounding box of a stone is independent of the translation, the first step consists of rotating the stone such that Eq. 1b is satisfied. Full enumeration is applied to solve the problem and the obtained rotated stone is pose-optimal. The second step is solving Eq. 1a by translating the stone to the as-built half wall to minimize the height of the center of mass. We discretized the solution space by 0.01 m (for translation) and 0.01 rad (for rotation) and used a heuristics optimization solver proposed by Shaqfa et Beyer Shaqfa and Beyer (2021) to find the best position. The solver samples a random population of possible solutions at every iteration and narrows down the solution space gradually such that the global optimum can be found in an efficient way. In the current experiment, it took 20 s on average, and on a processor Intel Core i5-11300 H to solve the best placement of one stone. Once the optimal solution is found, the stone model from the dataset is translated into the scene, as shown in Fig. 7c. The translated stone in the form of a surface mesh is recorded (as shown in Fig. 7d) to facilitate future comparison of the geometry of the planned wall and the as-built wall. The scene geometry (Fig. 7c) and the stone assembly (Fig. 7d) are not perfectly aligned in this case because it was obtained at the step when we



**Fig. 5** Top: graph illustrating the size distribution of stones within the dataset. The stone size value is approximated as the diagonal of the oriented bounding box (OBB). Below: (a–d) specimens of meshes from various shapes and dimensions of mineral scraps. The dataset identifier is reported next to the figure label

placed the timber frame on the stones to start the next layer. The role of the timber frame for each compartment of dry-stone rocks is allowing the construction of the final wall to be perfectly plumb and avoiding the accumulation of a talus during the building process.

### 3.5 Spatial calibration

The heart of the augmented reality system is the ability to precisely highlight areas of physical space, based on measured (e.g., the measured height map) or computed (e.g., a given stone's computed destination position) quantities.

The 3D coordinate system provided by the stereo-vision camera is used as the master coordinate system  $\vec{X} = [XYZ]$ , meaning that the identification of the corresponding 3D lines to the corresponding 2D pixel coordinate  $\vec{x} = [xy]$  can solve the calibration between the projector and the 3D sensor.

The projector is modeled as the mathematical equivalent of a pinhole camera, with a punctual light source with perfect divergence and no distortion (although taking these imperfections into account would be easy if they are required for better performance). The parameters requiring calibration are therefore the projector's "extrinsics" (i.e., its 3D position and orientation in the coordinate system defined by the stereo-vision camera) and the projector's "intrinsics" (i.e., in this case, the focal length and scaling of the projector). Equation 3.5 shows the equation that maps  $\vec{x}$  into  $\vec{X}$  (projector pixels coordinates into 3D coordinates in meters) (Intrinsic camera parameters calibration 2022) where  $R$  is the projector's rotation matrix,  $t$  is its position.

$$s \cdot \vec{x} = K[Rt]\vec{X} \quad (2)$$

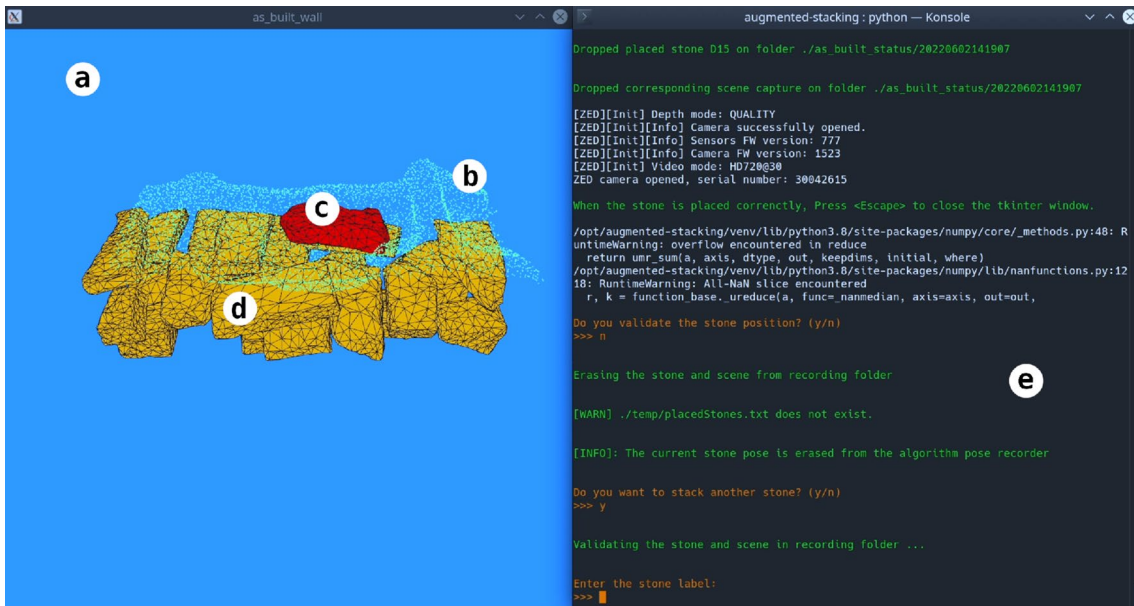
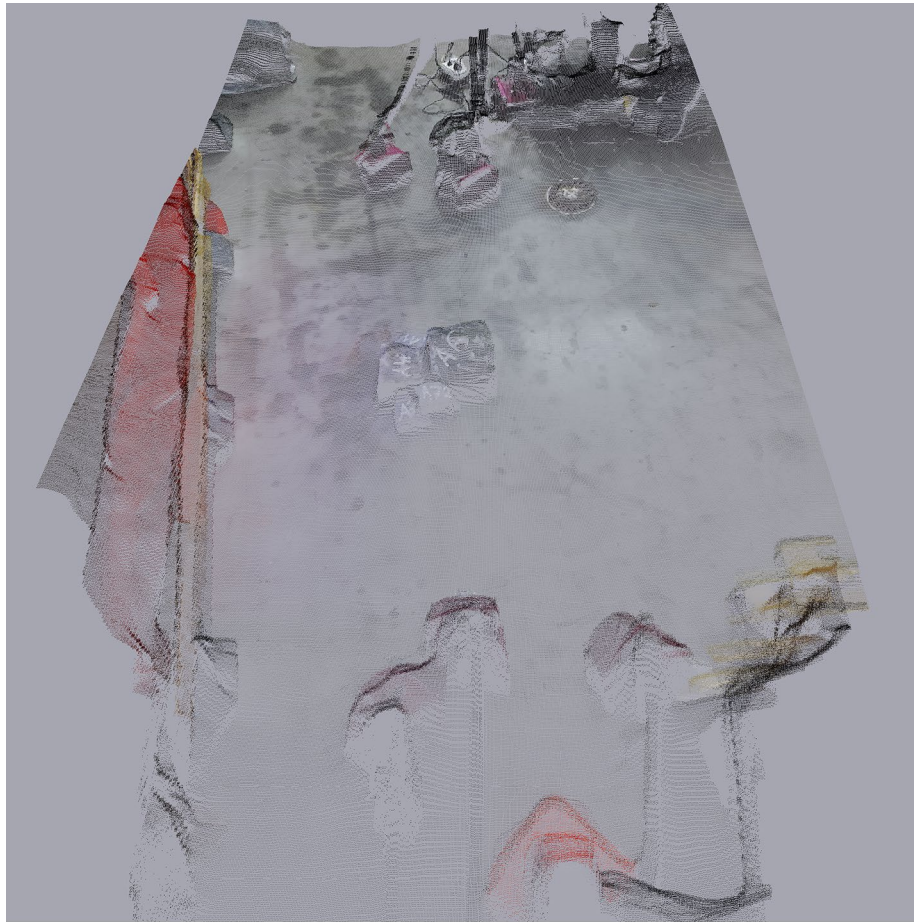
A calibration procedure was developed to collect pairs of  $\vec{x}$  and  $\vec{X}$  in order to start a minimization procedure to find all the unknowns. A calibration target illustrated in Fig. 8c was built such that it was easy to identify directly in the  $\vec{X}$  field coming from the stereo-camera—the disc-on-a-stick can easily be detected with a threshold in the  $Z$  direction since it is higher than the background. The  $X$ ,  $Y$ ,  $Z$  pixels selected by the threshold on  $Z$  are then separately averaged to give a precise location for the target. The projector is set to project a green grid at known  $\vec{x}$  (visible in Fig. 8), and the mean  $\vec{X}$  is recorded for each one of these.

Thereafter all unknowns in the above equation are searched for with a Powell minimization (Scipy documentation for minimization function 2022). If the procedure converges, a check allows the user to place the calibration target anywhere and its center is computed and highlighted on the projector. Users are guided through the procedure as visible in Fig. 9.

### 3.6 Projected AR interface

Once the geometric planner has completed computing the stone pose in the execution model, the operator is asked to match the physical rigid transformation to the generated virtual one. Following the calculation of the stone pose in the execution model, the operator is required to match the physical pose with the virtual one. In general, the set of visual stimuli required to guide all operator's manipulations is what defines an AR interface. In the proposed pipeline, the UI is composed of a series of projected raster images generated from the computed model and the sensing data.

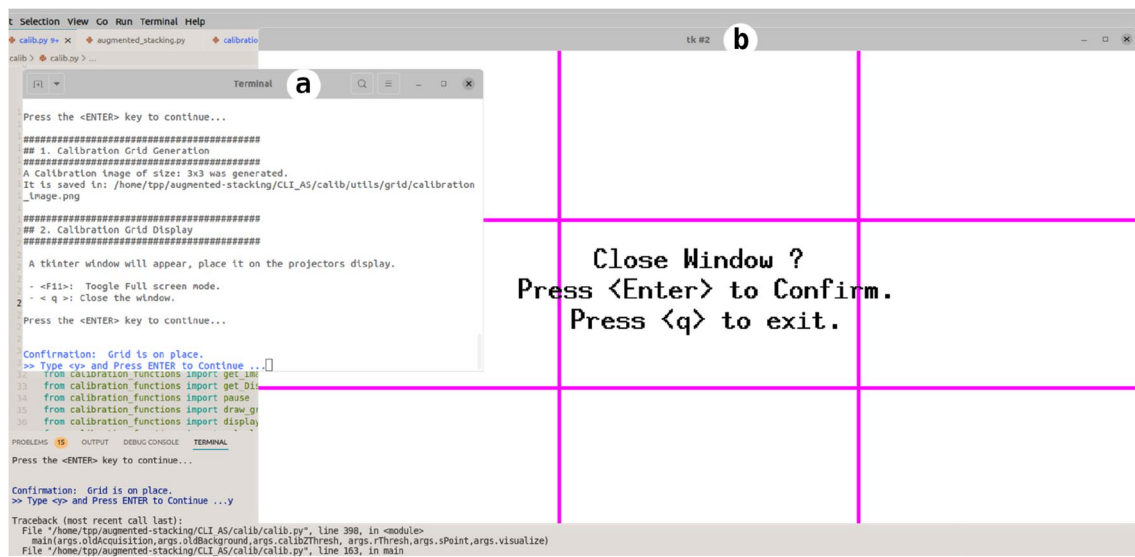
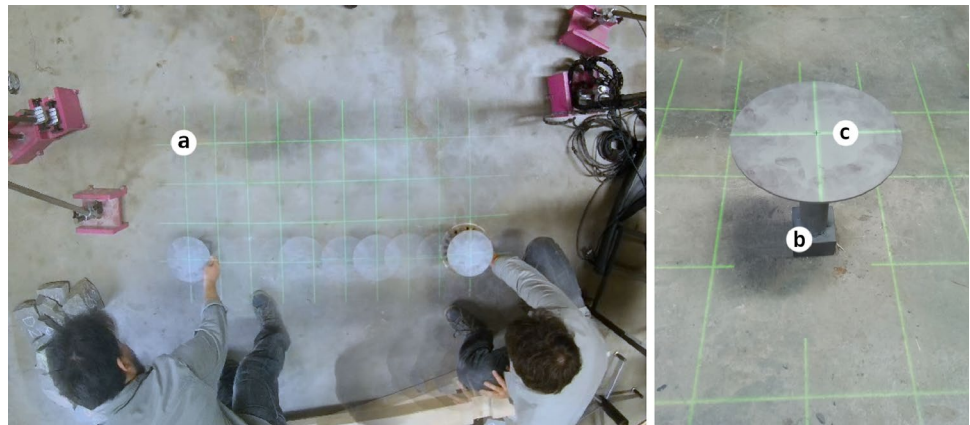
**Fig. 6** Raw point cloud data obtained from the stereo camera. The depth matrix is later converted to a mesh to be input to the geometric planner. To avoid camera flickering and reduce the overhaul noise, multiple frames' depth matrices are averaged together, resulting in a mean refresh rate of ~21 frames-per-second (FPS)



**Fig. 7** Image of the application user interface (UI) during the assembly: (a) As-built 3D model live-visualizer, (b) point cloud from sensing device, (c) current stone, (Lausanne) stones previously placed and validated, (e) terminal interface



**Fig. 8** Projected UI for calibration phase. **a** Ground projection of the generated grid with 10 vertical and 4 horizontal lines for a total of 40 calibration sequences, **b** portable metal disk, **c** grid node overlaid with the pre-marked center of the elevated disk



**Fig. 9** Calibration UI from the terminal. **a** CLI program guiding the operator during the calibration, **b** raster image of the generated grid to be projected

Projection-based AR interfacing presents major challenges compared to HDMs or other screen-based applications. Building elements can obfuscate portions of the UI, objects with hard edges can deform and make widgets or heads-up displays (HUDs) unreadable, and external sources of light e.g., an overload or on the contrary an insufficient amount of projected widgets, could result in a diminution of the system's overhaul visibility and guiding capability (Fig. 10). These factors oblige the interface design to limit the amount of information displayed to diminish the operator's cognitive load while being effective in conveying precise assembly instructions.

Furthermore, specifically in our scenario, the UI must instruct the user on a spatial manipulation characterized by 6 axes of freedom with limited sensing coverage. The stereo camera can only provide 3D data of the built artifact's upper

portion (Fig. 11a). Thus, all computed feedback relies solely on the point cloud stream of only a fragment of the actual physical geometry.

To address these challenges the proposed guidance module is designed as a two-component interface. First, a projected green contour provides a visual landmark suggesting the correct planar detection of the stone's localization (translation in the  $x$  and  $y$  axes) and the  $yaw$  axis (rotation in the  $z$  axis) without any live correcting feedback (Fig. 11c). On the other hand, the local refinements for the rotation of  $x$  and  $y$  axes are obtained by following 3 punctual indicators actively guiding the user in the achievement of the correct heights (direction and intensity of the translation in the  $z$  axis, Fig. 11d). The 3 key points define a plane defining the missing orientation's *roll* and *pitch* rotations (Fig. 11e). The previously computed calibration matrix will convert 3D data



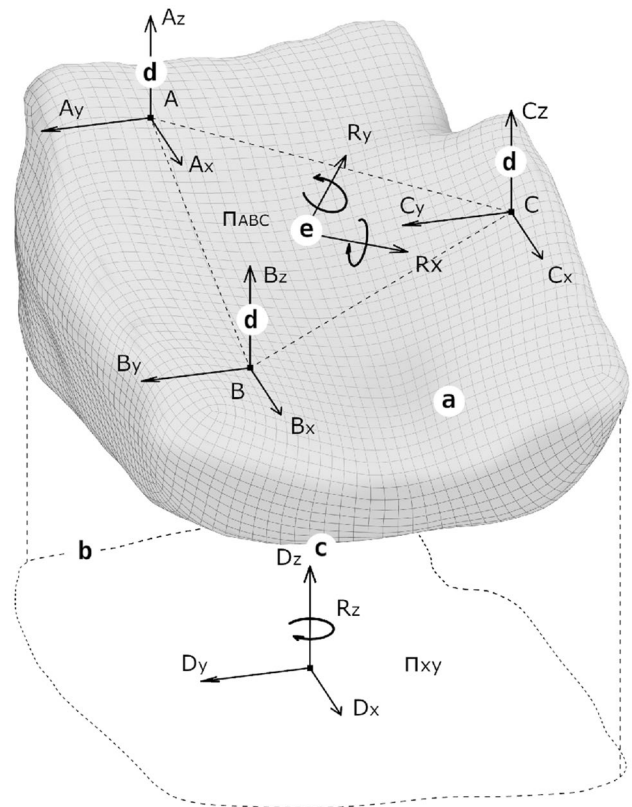


**Fig. 10** Preliminary tests for the developing of an efficient projected mapping interface: (a) overload of projected information, (b) insufficient highlighted pixels to effectively display instructions

to bi-dimensional pixels at every stage of the sensing data processing.

The complete description of the sensing data transformation into a viable UI takes place in the main rendering thread every  $\sim 21$  frames-per-second (FPS), and it is as follows:

1. Upon selecting a stone, the stacking planner calculates its corresponding digital twin's transformation. The rigid transformation is applied to the mesh and it is eventually flattened only to extract its 2D contour. The obtained polyline is later on projected to localize the region of interest (ROI) designated for the current stone (Fig. 12a, 13a). This allows the user to manually match the physical stone's contours with the projected boundary.
2. The mesh is next sub-sampled into a point cloud  $pcd_v$  with a definition inferior to 1000 points. Although only its visible top portion from the camera origin perspective is retained. As the physical stone enters the ROI, the RGB-D video stream from the stereo camera is averaged out every 10 frames, converted into a point cloud  $pcd_c$ , and then cropped to fit the ROI (Fig. 12b).
3. After applying a  $K$  - Mean clustering algorithm to  $pcd_v$ , we obtain a set of 3 sub-point clouds  $pcd_v^{i,ii,iii}$ . For each of these, we average the center and we recover 3 key points  $K_v^{i,ii,iii}$  defining the target plane  $\Pi_v$ . From every  $K_v^{i,ii,iii}$  cropping areas are generated with a width of  $\sim 10$  cm. The points of  $pcd_c$  falling into the projection of the cropping areas are segmented out from the parent geometry creating 3 distinct point clouds  $pcd_c^{i,ii,iii}$ . Simi-



**Fig. 11** Illustration of the 6 degrees of freedom to match for correctly positioning the object with the correct transform indicated by the geometric planner. Note that only the upper area of the stone can be monitored due to the vertical position of the 3D camera. Legend: (a) the stone, (b) the contour of the stone's 2D projection following the  $z$  axis, (c) yaw axes defined by the projected object's outline, (d)  $z$  axes of the identified key points, (e) the plane constructed from the 3 key points define implicitly the roll and pitch axes

larly to the processing applied to the geometric planner's  $pcd_v$ , 3 key points  $K_c^{i,ii,iii}$  are also identified (Fig. 12c).

4. While the operator manipulates the stone within the ROI, the vectors  $\overline{K_v^i K_c^i}$ ,  $\overline{K_v^{ii} K_c^{ii}}$ , and  $\overline{K_v^{iii} K_c^{iii}}$  vary mostly in intensity and direction. Since we estimate the angle differences as negligible, we simplify these vectors to their  $z$  components. 3 graphical probes with the shape of a square are added to the UI and anchored to the key points' 2D projection coordinates (Fig. 12d). The function of the widget is to convert their own vector's length and signs respectively to a proportioned surface and colors:

$$W_{Area} \neq 0, \text{ if : } |K_c^n z - K_v^n z| \neq 0, \tag{3a}$$

$$W_{Color} = (255, 0, 0), \text{ if : } (K_c^n z - K_v^n z) < 0, \tag{3b}$$

**Fig. 12** Sequence of the sensing data processing into the augmented UI: (a) projection of the stone contours, (b) RGB-D capture of the stone's upper geometry, (c) point cloud clustering to select 3 key points which defines the target plane in the 3D model and the captured planes in the captured point cloud, (d) widgets for refining the final stone's pose

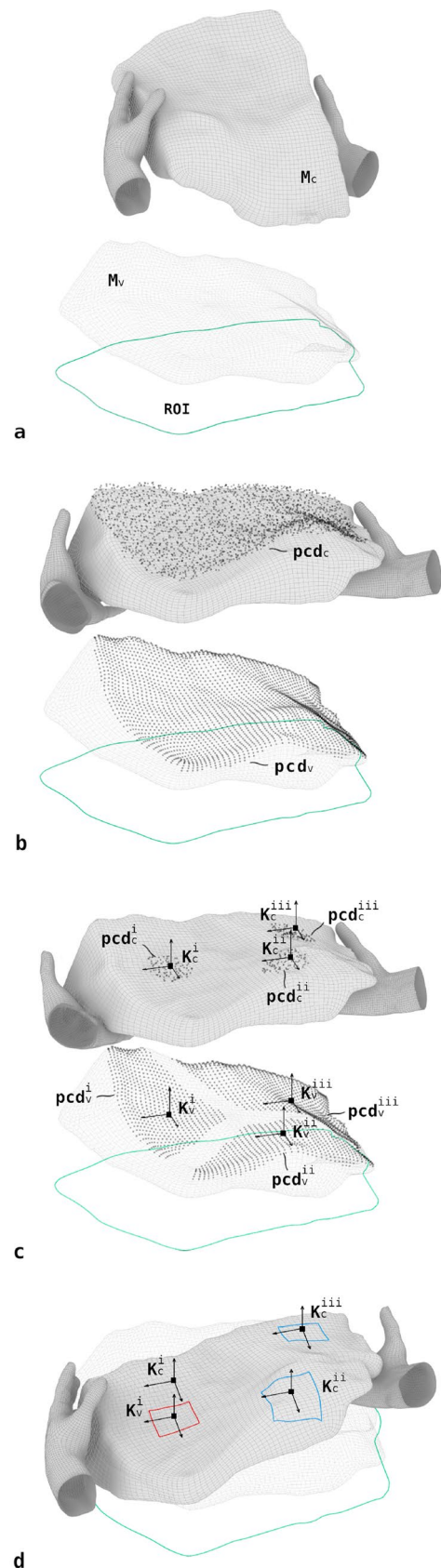
$$= (0, 0, 255), \text{ if } : (K_c^n z - K_v^n z) > 0 \quad (3c)$$

With  $W$  as the graphic widget and  $K_{c,v}^n z$  as the  $z$  components of key points belonging to the captured geometry of the physical stones or their virtual, transformed meshes. The widget's surface will result as larger if its correlated vector's module grows (Eq. 3a). If the vector's direction is negative the widget's graphics will turn red (Eq. 3b), hence the user needs to elevate the local stone's portion. Likewise in the opposite direction, the UI element will turn blue (Eq. 3c) and the user needs to lower the corresponding stone's corner (Fig. 13a,b,c). The pose is considered correct when none of the colored widgets is visible (Fig. 13e)

As a result, only those measurements and information that are crucial and key are being converted into 2D indicators and used to steer a process otherwise characterized by a great deal of complexity (Fig. 14). This seems sufficient to simplify a 3D process that is otherwise characterized by many factors that are at play (Fig. 11).

## 4 Experimental results

Two prototypes of one-layered dry-stone structures with a 1.7 m length, 0.6 m height, and 0.7 m width each with a different collection of building units, were realized to test the developed AR-based methodology. The first wall presents 40 monitored mineral by-products uniquely from by-products of quarry processes involving sawing and water-cutting (Fig. 15). In contrast, the second wall features 30 tracked mineral scraps issued from all mixed quarry's transforming operations (Figs. 15, and 16a). The as-built artifacts are compared to their respective as-expected geometric planner's recorded models. As stones are positioned and validated by the operator during the AR-assembly sequence, the system records their pose in a multi-dimensional file. The recording containing all the tracked stones is stored as the fabrication is concluded (Fig. 16b). Physical structures undergo a destructive scanning procedure to obtain their corresponding digital models. For each stone to be removed, geometric data of





**Fig. 13** Example of a stone placed with the proposed AR guiding framework: height probe (a) consistently, (b) moderately, and (c) close to the computed target, (d) UI widget to approximate the initial stone position, (e) interface state indicating a satisfactory pose. (f) To prepare a placement, the operator can also sense the targeted height indicated by each probe by inspecting the region with their hand. This haptic procedure allows the user to foresee the final height before start manipulating heavy building units

the current landscape is acquired. The procedure continues until the entire wall is dismantled. The collected raw scans are subsequently realigned and post-processed in a supervised manner to finally register each stone's pose within the 3D model of the walls (Settimi et al. 2022) (Fig. 17). By comparing the misalignment between each stone of both models, it is possible to produce a quantitative indication of the proposed digital pipeline's reliability. All the captured data employed in the presented evaluation are open-source and publicly accessible (Settimi et al. 2022).

With the aim of evaluating the deviation of the as-built, ground truth artifact from its execution model, a spatial alignment of both is required. The translation defined by the vector approximates this transformation originated by the two point clouds' centers and refined by a RANSAC and Iterative Closest Point (ICP)-based registration.

The tolerance induced by the proposed method could be evaluated by state-of-the-art metrics for mesh

comparison via Hausdorff distance (Aspert et al. 2002), the chamfer distance (CD) or the Earth mover's distance (EMD) between two point clouds. To avoid meshing the scanned point cloud of the artifact and in order to obtain additional metrics about the overhaul overlap portions between the two sets we used the following metrics described below.

First, the fitness measure indicates the overlapping area between both models. The higher this parameter appears to be, the more closely the two models overlap.

$$fitness = \frac{\#inlier\ correspondences}{\#points\ in\ ground\ truth}$$

Second, the inlier\_RMSE measures instead the root mean square Error (RMSE): the distance between the two sets of inlier points. In this scenario, a low value is recommended.

$$inlier\_rmse = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n}}, i\ in\ inliers$$

with

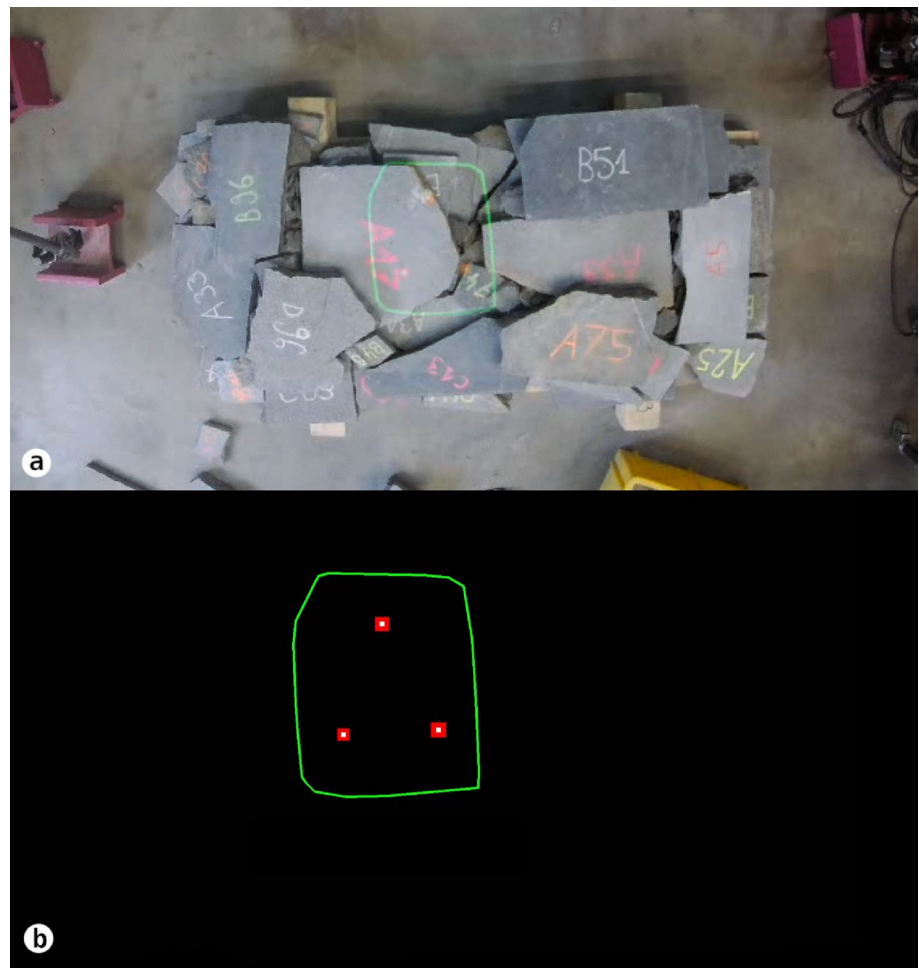
- $y_i$  is a point defined as an inlier, hence it exists a correspondence point in a range of 0.05 m.
- $\hat{y}_i$  the corresponding ground truth point.
- $n$  the number of inlier points.

The first evaluation was done on the entirety of the two sets of point clouds per wall without distinction of stone as if we had only two elements to obtain a global accuracy of the realignment. As described above, the fitness, the inlier\_rmse, and the RMSE on the entire wall were calculated (Table 1). The average error of the realignment was estimated to be around 3 cm for both walls. If we compare this value of 3 cm in terms of volume (a cubic volume of 3 cm on a side) to the overall volume of the wall (0.714 m<sup>3</sup>), this represents an error of only 0.004 %.

The next step was to measure the deviation for each stone, which are significantly smaller objects compared to the whole wall, separately using the same method. All the results are listed in Table 2 and visible in Fig. 17c,f. The average pose error of the stones was  $2.9 \pm 1.8$  cm. This error is therefore more important if we compare in terms of volume on one stone only. However, we understand that this error is not cumulative because it is very close to the global error made on the first analysis. We can therefore easily understand that this error comes partially from the realignment method but that the stones were laid in a repeatable and precise manner within a few centimeters of accuracy. This result also shows that



**Fig. 14** The projected AR interface is composed of procedurally generated raster images synthesized from the computed 3D model and the captured 3D data from the stereo camera: **(a)** projected UI, **(b)** the corresponding 2D template



our method of calculating the optimal placement for each new stone to be placed prevents the accumulation of too much error compared to a CAD model already established beforehand. If we improve our precision via the different tracks elaborated in the next section, the final precision error will therefore be improved by the same value due to this non-accumulation of error.

## 5 Limitations and improvements

In spite of the fact that our approach shows some promise, there are a number of challenges that need to be properly addressed to scale the system to a production-ready state. The system presents a height limitation of 1 m, after which the calibration fails to represent the generated 2D interface without distortions. This influences the precision at which visual stimuli are displayed, ultimately jeopardizing the reliability of the execution. To this extent, a different calibration methodology, together with multiple 3D cameras in different spots of the frame

and characterized by larger FoV may represent a substantial improvement to the current version. In the current state, the structure hosting the sensing hub is stationary. This implies that the building area must fit in a volume of interest limited by the sensing coverage of the onboard sensors. To extend the manufacturing station coverage to a linear system, a movable structure seems to be a possible solution. Nevertheless, this modification will engender a complex need for simultaneous localization and mapping (SLAM) to self-localizing and re-calibrating the sensors' positions and orientations at any given time during the fabrication. Seemingly, the use of HMDs, although providing better sensing covering and a more immersive guidance system for the stone's placement, will also generate complexities in terms of anchoring and syncing multiple devices and 3D information to the commonly shared scene.

In future developments, we will focus also on a dedicated on-the-fly, potentially in-hand scanning additional feature to the current pipeline which could replace the tedious and unrealistic operations of scanning and labeling a dataset prior to the assembly as it is in the proposed

**Fig. 15** Assembly time-lapse for the two prototypes: on the left the structure constituted by scraps originated from cutting operations in quarries. On the right: the one realized from not categorized mineral by-products



study. For outdoor improvements, due to signal interference in the stereo camera and low visibility of the projected interface, a safety one-classed laser projector, as well as high-quality infra-red 3D cameras, could represent a better setup than the proposed one. Regarding the sensing coverage, multi-sided and higher-quality sensors will most surely make the system more robust to cluttered scenes.

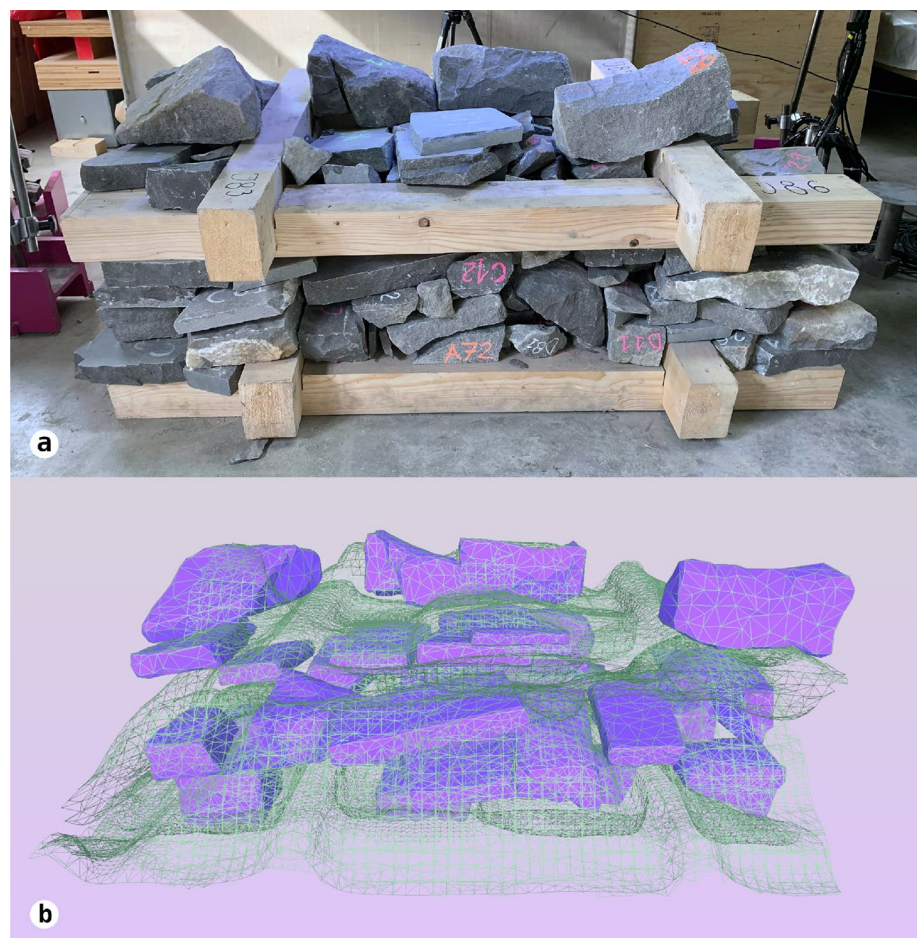
To what concerns the digitization of the stone library, more advanced compression methods via, i.e. encoding in latent vector space would open new directions to the characterization of irregularly shaped elements and their packaging via AI-based state-of-the-art methods.

Finally, an effective qualitative and quantitative analysis of the human factor involved in the manufacturing process could also unveil new insights into the actual effectiveness of the proposed interface.

## 6 Conclusions

We presented a projection-based AR guidance system for the manual stacking of rigid irregular materials following a numerically computed packing model. We detail the processing of sensing data, the adopted geometric planner, and the design of the augmented projected interface. The evaluation of the developed fabrication pipeline was designed for testing its efficiency with the construction of full-scale artifacts. Its overall precision has been measured to be close to the range of  $2.9 \pm 1.8$  cm. Although fully functional, we outlined all the limitations and possible improvements to the current system, which might scale it to a future version that can perform in real-life production scenarios both indoors and outdoors. With the presented experimental setup, we demonstrated how it is possible to accomplish additive manufacturing operations

**Fig. 16** **a** Frontal view of the physical model realized from a mixed set of mineral scraps. The timber frame helps to erect the wall with a 90 degrees angle by defining separate compartments with traditional heights of 40 cm. **b** The recorded model is composed by the major stones validated by the geometric planner and the as-built captured landscapes



**Table 1** Evaluation of the realignment between two sets of point clouds without distinction of stone

Wall ID (-)	Fitness (%)	inlier_rmse (m)	RMSE (m)
1	93.6	0.021	0.030
2	91.6	0.018	0.029

such as stacking with irregular objects by following projected instructions from a limited sensing coverage. Considering the state of the described system we can conclude that the proposed pipeline can be generalized

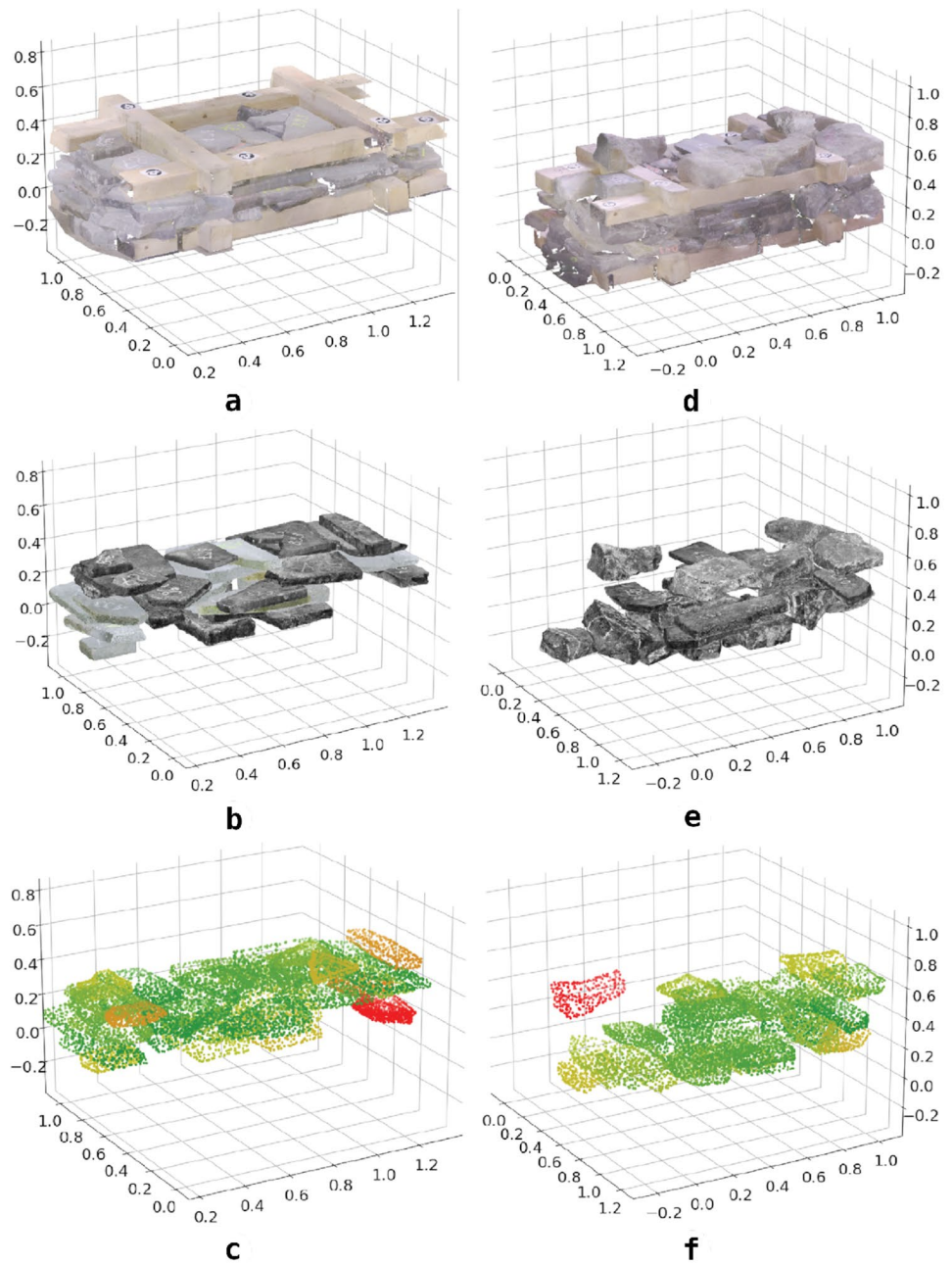
to any additive stacking operations involving the irregular object. AR applied to manual operations can offer an economically attractive, accessible development, and socially acceptable entry-level digitization to numerous entities across multiple sectors counting on complex manual activities in their processes. Projection-based AR in particular can be considered a valuable medium in the contemporary technological panorama to instruct multiple operators at once and foster collaboration and communicability through a commonly shared sensory interface, all by improving the precision of execution and the reliability of human-machine craftsmanship.

**Table 2** Evaluation of the pose deviation of each stone (despite the relative realignment precision between two sets of point clouds)

Wall ID (-)	Nbr Stones (-)	Mean RMSE (m)	Median RMSE (m)	SD RMSE (m)	Min RMSE (m)	Max RMSE (m)
1	30	0.028	0.023	0.016	0.010	0.100
2	40	0.030	0.024	0.021	0.011	0.127
Total	70	0.029	0.023	0.018	0.010	0.113



**Fig. 17** 3D-plot of the distribution of the two walls. **a, d**: 3D point cloud of the wall constructed during the pipeline. **b, e**: 3D meshes of the stones posed using the pipeline. **c, f**: Deviation measure of the placed stones, comparison between ground truth model and expected model



**Funding** Open access funding provided by EPFL Lausanne.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aejmelaeus-Lindström P, Rusenova G, Mirjan A, Ibáñez JM, Gramazio F, Kohler M (2020) Rock print pavilion: robotically fabricating architecture from rock and string. *Construct Robot* 4(1–2):97–113. <https://doi.org/10.1007/s41693-020-00027-8>
- Aejmelaeus-Lindström P, Willmann J, Tibbits S, Gramazio F, Kohler M (2016) Jammed architectural structures: towards large-scale

- reversible construction, *Granular Matter* 18 (2) (Apr. 2016). <https://doi.org/10.1007/s10035-016-0628-y>
- Almeida C, Guedes J. P, Arêde A, Costa A (2016) Geometric indices to quantify textures irregularity of stone masonry walls, *Construction and Building Materials* 111 199–208. <https://doi.org/10.1016/j.conbuildmat.2016.02.038><https://www.sciencedirect.com/science/article/pii/S0950061816300988>
- Arena F, Collotta M, Pau G, Termine F (2022) An overview of augmented reality, *Computers* 11 (2). <https://doi.org/10.3390/computers11020028>. <https://www.mdpi.com/2073-431X/11/2/28>
- Aspert N, Santa-Cruz D, Ebrahimi T (2002) Mesh: measuring errors between surfaces using the hausdorff distance, in: *Proceedings. IEEE International Conference on Multimedia and Expo*, Vol. 1, pp. 705–708. <https://doi.org/10.1109/ICME.2002.1035879>
- Bottani E, Vignali G (2019) Augmented reality technology in the manufacturing industry: a review of the last decade. *IJSE Trans* 51(3):284–310. <https://doi.org/10.1080/24725854.2018.1493244>
- CASIO COMPUTER CO., CASIO XJ-A252 hybrid LED/laser beamer, mA2008-F (2014). [https://support.casio.com/storage/en/manual/pdf/EN/007/A257\\_Setup\\_EN.pdf](https://support.casio.com/storage/en/manual/pdf/EN/007/A257_Setup_EN.pdf)
- FARO Technologies Inc., FARO Freestyle 2 and Mobile PC, user Manual (2022). <https://downloads.faro.com/index.php/s/sqCRBipgSy9GaEq?dir=undefined&openfile=138985>
- Furrer F, Wermelinger M, Yoshida H, Gramazio F, Kohler M, Siegart R, Hutter M (2017) Autonomous robotic stone stacking with online next best object target pose planning, in: *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Institute of Electrical and Electronics Engineers (IEEE), pp. 2350–2356. <https://doi.org/10.1109/icra.2017.7989272>
- Gard N, Hilsman A, Eisert P (2022) Combining local and global pose estimation for precise tracking of similar objects, in: *2022 17th International Conference on Computer Vision Theory and Applications (VISAPP)*, Vol. 4, VISAPP, pp. 745–756. <https://doi.org/10.5220/0010882700003124>
- Hoc J-M (2000) From human-machine interaction to human-machine cooperation. *Ergonomics* 43(7):833–843. <https://doi.org/10.1080/001401300409044>
- Huang H, Zhong F, Qin X (2021) Pixel-wise weighted region-based 3d object tracking using contour constraints. *IEEE Trans Vis Comput Graphics* Early Access. <https://doi.org/10.1109/tvcg.2021.3085197>
- Hughes R, Osterlund T, Larsen NM (2021) Integrated design-for-manufacturing and AR-aided-assembly workflows for light-weight reciprocal frame timber structures. *Construction Robot* 5(2):147–157. <https://doi.org/10.1007/s41693-020-00048-3>
- Intrinsic camera parameters calibration, [https://mphy0026.readthedocs.io/en/latest/calibration/camera\\_calibration.html](https://mphy0026.readthedocs.io/en/latest/calibration/camera_calibration.html), accessed: 2022-06-23
- Jahn G, Newnham C, Berg N (2022) Depth camera feedback for guided fabrication in augmented reality, in: A. Masoud, A. Dorit, J. Hina, S.-S. Robert (Eds.), *Proceedings of ACADIA Conference Hybrids & Haecceities*, [http://acadia.org/content/conferences#past\\_conferences](http://acadia.org/content/conferences#past_conferences)
- Johns RL, Wermelinger M, Mascaro R, Jud D, Gramazio F, Kohler M, Chli M, Hutter M (2020) Autonomous dry stone. *Construction Robot* 4(3–4):127–140. <https://doi.org/10.1007/s41693-020-00037-6>
- Larsson M, Yoshida H (2019) Igarashi T Human-in-the-loop fabrication of 3D surfaces with natural tree branches, in: *Proceedings of the ACM Symposium on Computational Fabrication, SCF '19*, Association for Computing Machinery, New York, NY, USA, <https://doi.org/10.1145/3328939.3329000>
- Liu Y, Choi J, Napp N (2021) Planning for robotic dry stacking with irregular stones, in: G. Ishigami, K. Yoshida (Eds.), *Field and Service Robotics*, Springer Singapore, Singapore, pp. 321–335. [https://doi.org/10.1007/978-981-15-9460-1\\_23](https://doi.org/10.1007/978-981-15-9460-1_23)
- Mitterberger D, Dörfler K, Sandy T, Salveridou F, Hutter M, Gramazio F, Kohler M (2020) Augmented bricklaying, construction. *Robotics* 4:151–161. <https://doi.org/10.1007/s41693-020-00035-8>
- Mitterberger D, Jenny S. E, Vasey L, Lloret-Fritsch E, Aejmelaeus-Lindström P, Gramazio F, Kohler M (2022) Interactive robotic plastering: Augmented interactive design and fabrication for on-site robotic plastering, in: *CHI Conference on Human Factors in Computing Systems*, no. 174, ACM, pp. 1–18. <https://doi.org/10.1145/3491102.3501842>
- Rivers A, Rothmyer Alec, Durand F (2013) Augmented manual fabrication methods for 2d tool positioning and 3d sculpting, Ph.D. thesis, Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science. (Jun. 2013). <http://hdl.handle.net/1721.1/79215>
- Sandy T, Buchli J (2018) Object-based visual-inertial tracking for additive fabrication. *IEEE Robot Auto Lett* 3:1370–1377. <https://doi.org/10.1109/lra.2018.2798700>
- Scipy documentation for minimization function, <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>. Accessed: 2022-06-25
- Settimi A, Gamero J, Weinand Y (2022) Cockroach official documentation. <https://ibois-epfl.github.io/Cockroach-documentation/docs/tutorials/digital-twin-of-masonry-stone-wall-b/>
- Settimi A, Vestartas P, Gamero J, Weinand Y (2022) Cockroach: an open-source tool for point cloud processing in CAD, in: N. G. J. van Ameijde, K. H. Hyun, D. Luo, U. Sheth (Eds.), *POST-CARBON – Proceedings of the 27th CAADRIA Conference*, pp. 325–334. <https://doi.org/10.52842/conf.caadria.2022.2.325>
- Settimi A, Wang Q, Andò E, Andò E, Gamero J, Bayer K, Weinand Y (Oct.2022). Scans dataset for evaluation of AR-based assembly pipeline of half-timber dry-stone structures. <https://doi.org/10.5281/zenodo.7200022>
- Settimi A, Wang Q, Gamero J, Beyer K, Weinand Y (2022). Dataset of mineral scraps for augmented stacking. <https://doi.org/10.5281/zenodo.7189478> [github.com/ibois-epfl/augmented-stacking](https://github.com/ibois-epfl/augmented-stacking)
- Settimi A, Wang Q, Rebiffé D, Andò E, Gamero J, Beyer K, Weinand Y (Oct.2022). Augmented-stacking <https://doi.org/10.5281/zenodo.5817181087> [github.com/ibois-epfl/augmented-stacking](https://github.com/ibois-epfl/augmented-stacking)
- Shaqfa M, Beyer K (2021) Pareto-like sequential sampling heuristic for global optimisation, *Soft Computing* 25 (07 2021). <https://doi.org/10.1007/s00500-021-05853-8>
- Stereolabs Inc., Zed 2i stereo camera, rev. 1 (2022). <https://cdn.stereolabs.com/assets/datasheets/zed-2i-datasheet-feb2022.pdf>
- Stoiber M, Sundermeyer M, Boerdijk W, Triebel R (2022) A multi-body tracking framework — from rigid objects to kinematic structures, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Aug. 2022). <https://doi.org/10.48550/arXiv.2208.01502>
- Syed TA, Siddiqui MS, Abdullah HB, Jan S, Namoun A, Alzahrani A, Nadeem A, Alkhodre AB (2023) In-depth review of augmented reality: tracking technologies, development tools, ar displays, collaborative ar, and security concerns. *Sensors* 23 (1) . <https://doi.org/10.3390/s23010146>. <https://www.mdpi.com/1424-8220/23/1/146>
- Thangavelu V, da Silva M. S, Choi J, Napp N (2020) Autonomous modification of unstructured environments with found material, in: *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Institute of Electrical and Electronics Engineers (IEEE), pp. 7798–7804. <https://doi.org/10.1109/icra40945.2020.9197372>
- Thangavelu V, Liu Y, Saboia M, Napp N (2018) Dry stacking for automated construction with irregular objects, in: *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018. <https://doi.org/10.1109/icra.2018.8460562>
- Wang Z, Bai X, Zhang S, Billingham M, He W, Wang P, Lan W, Min H, Chen Y (2022) A comprehensive review of augmented reality-based instruction in manual assembly, training and repair. *Robot*

Comput-Integrated Manuf 78:102407. <https://doi.org/10.1016/j.rcim.2022.102407>

Wermelinger M, Furrer F, Yoshida H, Gramazio F, Kohler M, Siegart R, Hutter M (2018) Greedy stone tower creations with a robotic arm, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Int Joint Conf Artif Intell Organization, <https://doi.org/10.24963/ijcai.2018/760>

Yoshida H, Igarashi T, Obuchi Y, Takami Y, Sato J, Araki M, Miki M, Nagata K, Sakai K, Igarashi S (2015) Architecture-scale human-assisted additive manufacturing. *ACM Trans Graph* 34(4):1–8. <https://doi.org/10.1145/2766951>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.