TRAINING A VISION-BASED AUTONOMOUS ROBOT FROM MATERIAL BENDING ANALYSIS TO DEFORMATION VARIABLES PREDICTIONS WITH AN XR APPROACH

CHI-FU HSIAO1, CHING-HAN LEE2, CHUN-YEN CHEN3 and TENG-WEN CHANG4
1Tamkang University;
2,3,4National Yunlin University of Science and Technology.
1chifu.research@email.com, 0000-0002-9018-635X
2chinghanglee@gmail.com, 0000-0002-9001-5972
3chunyan056@gmail.com, 0000-0001-9503-2766
4tengwen@gmail.yuntech.edu.tw, 0000-0003-4798-4580

Abstract. This paper proposes a "Human Aided Hand-Eye System (HAHES)" to aid the autonomous robot for "Digital Twin Model (DTM)" sampling and correction. HAHES combining the eye-to hand and eye-in hand relationship to build an online DTM datasets. Users can download data and inspect DTM by "Human Wearable XR Device (HWD)", then continuous updating DTM by back testing the probing depth, and the overlap between physics and virtual. This paper focus on flexible linear material as experiment subject, then compares several data augmentation approaches: from 2D OpenCV homogeneous transformation, autonomous robot arm nodes depth probes, to overlap judgement by HWD. Then we train an additive regression model with back-testing DTM datasets and use the gradient boosting algorithm to inference an approximate 3D coordinate datasets with 2D OpenCV datasets to shorten the elapsed time. After all, this paper proposes a flexible mechanism to train a vision-based autonomous robot by combing different hand-eye relationship, HWD posture, and DTM in a recursive workflow for further researchers.

Keywords. Digital Twin Model; Hand-Eye Relationship; Human Wearable XR Device; Homogeneous Transformation; Gradient Boosting; SDG 4; SDG 9.

1. Introduction

The integration of artificial intelligence, autonomous robot aided fabrication, and "Human-Computer Interaction (HCI)" workflow present a rising tendency in nowadays "Architecture, Engineering and Construction (AEC)" workflow.

Traditional HCI robot arm operation needs operators to take a lot of time in configuring the tedious, point-to-point processing, and path teaching planning before the work starts. Moreover, when the AEC workflow getting much more complex, there
had to divide work stream into several stages, not only each period of robot operation must be recalibrated for different configured environment, but also the entire production will be interrupted for a lot of time-losing. Besides, real construction environments are usually much more chaotic and unpredictable, how to measure the deformation features of operating materials need to be processed before they are going to be operated within CAD-CAM system. In other words, autonomous robot construction is chasing for a kind of more flexible, sensible inference mechanism, which can apply to more complicated, unpredictable in-site environment.

Thus, this paper proposes a "Human Aided Hand-Eye System (HAHES)" to aid autonomous robot before AEC environment by establishing a digital model experiment to sample and update material features. HAHES mixes eye-to hand, and eye-in hand relationship of robot arm, then use "Human Wearable XR Device (HWD)" to observe the linear material for examining the difference between optical results and reality.

In addition, HAHES keeps updating the "Digital Twin Model (DTM)" from different perspective views from camera, HWD, and back-testing values by executing the autonomous robot to probe the deformation variables of material, and then make sure DTM is as precisely as the real scene configurations. With the recorded sampling datasets from above recursive examining tasks, HAHES then train an additive regression model to inference the approximate 3D coordinates of nodes distribution on bending materials for reducing the time losing of the latter examining works.

2. Precedents

2.1. HAND-EYE RELATIONSHIPS

There are two kinds of observing and executing system in autonomous robot construction relationships, known as "eye-in hand" and "eye-to hand" (Flandin et al., 2000). Eye-to hand means that the camera and the arm are fixed in two positions separately, camera pictures overall environments in a pre-settled configuration, but inevitable occlusions by moving arms or involving team members.

In comparison, eye-in hand hangs the camera on the end-effector of robot arm, camera takes picture when the arm is driven to achieve the executing command, so that users can control the arm dynamically by locating the relative relationship with homogeneous coordinates calculations in real time. Batliner had showed several projects to combine eye-in hand and projection with real-time robotic motion operations (Batliner et al., 2015).

Different from eye-in hand, eye-to hand system is fixed in a static transformation matrix, when camera capturing and recognizing, the arm is moving and executing commands at the same time. Therefore, there is better cycle time, but the end-effector and the camera view field must be ensured in a certain and clear moving path.

Otherwise, eye-in hand maintains a relative relationship with homogeneous coordinates though, the flexibility of free moving end-effector also causes several optical deviation issues of picture taken, such as focal length, chromatic aberration, contrast sensibility or overexposed… etc. which will raise the difficulty of the artificial intelligence sampling pre-processing model built.
2.2. OBJECT-RECOGNITION

Cheng presents an autonomous robot mechanism practice with eye-in-hand object-recognition. They use pre-trained 2D depth computer vision to recognize the targets on a plan surface, then use Computer in the Human Interaction Loop (CHIL) to choose desired object. CHIL insert an initialize eye-to-hand stage in to the eye-in-hand setting. They configure the robot arm and camera posture into a specific perpendicular state to plan plate each time human selection occurs. Thus, they can build a stable digital twin model, and segmented monitor the task process (Cheng et al.).

2.3. ACTIVE BENDING SIMULATION

Chang propose a XR relative digital twin model: Cofab, to simulate a mixed-reality weaving structure, record physical nodes states disturbs, and adjust the morphing shapes via gesture at the same time. They use a variable parameter (Pa) to modulate the degree of active-bending states globally, then adjust this Pa value to simulate the morphing between the virtual and physical changes and observe the displacement of the connective nodes cloud with XR.

Because the structure was relative connected and shifted overall structure displacement. Afterall, they also build a practical weaving structure then using robot arm to measurement the precision in virtual and the physical expression (Chang et al., 2020).

2.4. AI PREDICTION IN 2D LINEAR GEOMETRY

Yang proposed a tool to choose predictable linear natural material. In amount of 4,000 mm length, 50 mm width, 10mm thickness unprocessed bamboo strips. This tool predicts both the number and locations of nodes on nature material in 2D distribute pattern and apply to desired geometry deformation simulation. He setups a bending experiment environment workflow with 5 stages:

- Pre-marked nodes on material, holding clamps,
- Setup a static eye to hand position, and initialize the computer vision background with black fabric and black PVC boards,
- Corresponding pressure the active bending material with 2 stabilize robot end effectors,
- Processing the computer vision output into curve and nodes into automatic "Neural Architecture Search (NAS)" datasets to approach the reality curvature,
- Training and predict material node distribution patterns, then post-test for promising accuracy for further deformation design task (Yang & Xu, 2021).

Yang utilising the NAS machine learning techniques to predict natural material behaviours by 2D camera setting. However, a bending curvature material may actually twist or offset by 3D direction forces, which is hard to apply in NAS vision algorithm.

2.5. 3D LINEAR GEOMETRY KEY FACTOR SEARCHING
We propose another approach to describe the linear material nodes in 3D distortion point datasets. By combining eye-to-hand and eye-in-hand camera projection datasets, we can complement the depth losing of 2D camera by homogenous equation transformation, theoretically. But we then figure out that DTM still need to adjust several key factors to approximate the robot back-testing results by probe depth after several time retests.

From 2D coordinates retrieving, homogeneous equation calculation, to recorrecting DTM, the sequences of the processes have closely related each other from previous to after task results.

We introduce additive regression models by sequentially fitting 2D eye-to-hand and eye-in-hand sample coordinates into the back-testing 3D coordinates at each iteration. Friedman proposed that randomly selected subsample from coordinates data type is then effective in place of the full sample to fit the base learner and compute the model update for iteration. Randomized approach also increases the robustness against overcapacity of the base learner (Friedman, 2002).

Figure 1. framework of HAHES.
3. Autonomous Robotic Sampling and Training Workflow

The process of HAHES workflow can be divided into 4 stages (shown as Figure 1):

- Eye-to-hand sampling: Homogenous matrix transform from 2D to 3D calculation.
- Eye-in-hand complement: Upload robot arm probe plumbs depth and camera inputs, then complement the Z-axis correction of previous 3D calculated oriental plane in DTM.
- HCI correction: Relocated the overlap of marked nodes with HoloLens multi-view scopes.
- AI prediction: Train additive regression model with gradient boosting algorithm from the datasets of previous 4 stages.

3.1. SAMPLING

The consideration of each stage provides different datasets in a sequence of matrix transformation. Eye-to-hand camera (as shown in Figure 2) fix the camera on the tripod in front of the workspace and use OpenCV to find out the relative coordinates of bending material nodes in camera 2D projection.
3.2. COMPLEMENT OF HOMOGENEOUS MATRIX

After uploading these 2D projection coordinates into cloud datasets, we then use the eye-in-hand camera projection to process the 3D coordinates from projective coordinates (shown as Figure 3). The equation of homogeneous coordinates transformation matrix will be inferred from below equations:

\[
D = H \cdot M \cdot H \cdot C
\]  
\[
\begin{bmatrix}
1 & 0 & -w/2 \\
0 & 1 & -h/2 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
m_{00} & m_{01} & m_{02} & m_{03} \\
m_{10} & m_{11} & m_{12} & m_{13} \\
m_{20} & m_{21} & m_{22} & m_{23} \\
m_{30} & m_{31} & m_{32} & m_{33}
\end{bmatrix}
\begin{bmatrix}
f & 0 & w/2 & 0 \\
f & h/2 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x' \\
y' \\
z'
\end{bmatrix}
\]

\[
f = \sqrt{x^2 + y^2} \cdot \sin(\theta)^{-1}
\]

Among equation (1) and (2), M is the homogeneous coordinates transformation matrix we are looking forward, D is the recognized 3D Cartesian coordinate in digital twin space, C is the camera input coordinates nodes values, H is the appearance of the variable f which corresponds to the focal length of our camera. Equation (3) presents the forward or backward states from the operation scene with a distance ΔW.
Equation (4) abbreviated our approaching transforming matrix M with translation, rotation, scale, and shear transformation. For the situation in eye-to-hand relationship, each of above matrixes will be static, because the camera is fixed in front of the workspace. But for the camera on the robot arm end effector, the matrixes are constantly changing through the process of robot arm moving.

\[ M = T \cdot R \cdot S_c \cdot S_h \]  

(4)

Translation matrix expresses as equation (5), and the rotation matrix would be calculated by equation (6) and (7). The \( \theta \) is the changing value from the direction of 6-axes robot TCP and will relate to the camera plane orientations.

\[
T = \begin{bmatrix}
1 & 0 & 0 & t_x \\
0 & 1 & 0 & t_y \\
0 & 0 & 1 & t_z \\
0 & 0 & 0 & 1
\end{bmatrix} = (t_x \cdot t_y \cdot t_z)
\]  

(5)

\[
R = R_x \cdot R_y \cdot R_z
\]  

(6)

\[
R = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_z) & -\sin(\theta_z) & 0 \\
0 & \sin(\theta_z) & \cos(\theta_z) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} \cdot \begin{bmatrix}
\cos(\theta_y) & 0 & -\sin(\theta_y) & 0 \\
0 & 1 & 0 & 0 \\
\sin(\theta_y) & 0 & \cos(\theta_y) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} \cdot \begin{bmatrix}
\cos(\theta_x) & -\sin(\theta_x) & 0 & 0 \\
0 & 1 & 0 & 0 \\
\sin(\theta_x) & 0 & \cos(\theta_x) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(7)

The \( S_c \) in equation (8) represents the scale transformation from operation scene to DTM, in this case we equally scaling up the object recognition pixel values into same parameter after measuring the physical scene in millimetres unit.

\[
S_c = \begin{bmatrix}
S_{c_x} & 0 & 0 & 0 \\
0 & S_{c_y} & 0 & 0 \\
0 & 0 & S_{c_z} & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(8)

\[
S_h = S_{h_x} \cdot S_{h_y} \cdot S_{h_z}
\]  

(9)

\[
S_h = \begin{bmatrix}
1 & 0 & 0 & 0 \\
\tan(\phi_z) & 1 & 0 & 0 \\
\tan(\phi_y) & 0 & 1 & 0 \\
\tan(\phi_x) & 0 & 0 & 1
\end{bmatrix} \cdot \begin{bmatrix}
1 & 0 & \tan(\phi_z) & 0 \\
0 & 1 & 0 & \tan(\phi_y) \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(10)

For the perspective configuration of eye-in-hand camera, we also need to add a shearing modification, the abbreviated matrix is shown as equation (9) and expended 3 direction axes as equation (10). We introduce \( S_h \) matrix to modify the losing depth of eye-to-hand 2D camera projectivity sample when the robot arm is moving, then complement 3D coordinates with this reverse inference process.
3.3. COORECTION

After above period of workflow, we map both 2 camera results on an external spreadsheet table synchronously online to monitor the material bending status, then build a primitive DTM to describe the experiment scene.

![Image](image.png)

*Figure 4. Use HWD to observe the difference between reality scene and DTM, then control the probe on the robot arm to update confirm the depth of delta vectors.*

After settling the DTM environment, we use HWD to examine the difference between nodes marked on physical material, and generated model (shown as Figure 4). Observers need to fit the bending model upon the reality structure in every perspective by gesture, then HAHES will upload modified values of each node into the spreadsheet sample library online as Table 1.

|----|---|---|------|------|--------|--------|------|--------|-----|-----|------|------|--------|--------|------|--------|

3.4. PREDICTIONS

From camera capturing pictures, uploading 2D datasets, generates 3D coordinates, HCI confirming to robot arm probes back-testing, the elapsed time will be 3-5 minutes for a complete detection task.

We then use "GradientBoostingRegressor" package from scikit-learn library to
build an additive model in a forward stage-wise, which allows the optimization of arbitrary differentiable loss functions. In our case, we need to rearrange previous sampling, complement, and correction stages of all nodes in each repeating operation results into a list of row data (shown as Table 2).

Table 2. Executing previous Table 1 multiple times to train an additive model, then predict the relationship between datasets X (2D calculation) and datasets Y (3D correction results).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>X (2D Calculation)</th>
<th>Y (3D Correction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.23</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>1.34</td>
<td>0.56</td>
</tr>
<tr>
<td>3</td>
<td>1.45</td>
<td>0.67</td>
</tr>
<tr>
<td>4</td>
<td>1.56</td>
<td>0.78</td>
</tr>
<tr>
<td>5</td>
<td>1.67</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>1.78</td>
<td>0.99</td>
</tr>
<tr>
<td>7</td>
<td>1.89</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>1.90</td>
<td>1.01</td>
</tr>
<tr>
<td>9</td>
<td>1.91</td>
<td>1.02</td>
</tr>
<tr>
<td>10</td>
<td>1.92</td>
<td>1.03</td>
</tr>
<tr>
<td>11</td>
<td>1.93</td>
<td>1.04</td>
</tr>
<tr>
<td>12</td>
<td>1.94</td>
<td>1.05</td>
</tr>
<tr>
<td>13</td>
<td>1.95</td>
<td>1.06</td>
</tr>
<tr>
<td>14</td>
<td>1.96</td>
<td>1.07</td>
</tr>
<tr>
<td>15</td>
<td>1.97</td>
<td>1.08</td>
</tr>
<tr>
<td>16</td>
<td>1.98</td>
<td>1.09</td>
</tr>
<tr>
<td>17</td>
<td>1.99</td>
<td>1.10</td>
</tr>
<tr>
<td>18</td>
<td>2.00</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Furthermore, because of the predict datasets Y are sequence of coordinates, we also import the "MultiOutputRegressor" package for the multivariate output target of parallelize regressors. Afterall, HAHES is able to inference approximate coordinates for a shorter time duration. In our tasks, we repeated HCI approaches for 37 times, and the time elapsed from 3-5 minutes into 3-5 seconds; the accuracy of AI prediction coordinates in X is 93.64%, in Y is 90.84%, and is 98.65% in Z (shown as Figure 5).
4. Conclusion

This paper proposes another co-existing approach of "Multi-View Scopes (MVS)". The sequence proceeds from marking the nodes distribution on material, use 2D vision based DTM with different hand-eye relation configured camera, to physical back-testing probes. Then we confirm the nodes position in 3D immersive perspective by HWD. To reduce the time elapsed of bending material inputs, we also use previous back-testing datasets to train an additive regression model and predict parameterized nodes distribution by gradient boosting algorithms. However, further adjustments must be examined in detail. The integration of HWD adjustment is of focus as the authors intend the observer to be the intervene aid to recheck the effectiveness of DTM during the detecting process. An advantage of using interactive stage instead of fully automated is that fewer iterations are needed to conclude, but more importantly there would be a dialogue between the algorithm and the user is introduced. Currently the determination of the coordinates is hypothesis the pregiven nodes is able to describe the bending materials, we can use add more possible nodes on HWD, and recheck the nodes by pasting another physical mark on the bending material.

With the integration of the additive regression model, HAHES proposes a more flexible occasion to decide when to teach system with HWD reflecting behaviours via human gesture sampling, or general use the hand-eye relationship AI prediction results. We expect the consequences of this paper can provide future researchers another kind of human aided approaching in training a vision-based autonomous robot from material bending analysis to deformation variables predictions.

References


