Model based 3D vision and analysis for Production Audit purposes

(Invited Paper)

K.Vaiapury, A.Aksay (*Senior Member, IEEE),* X.Lin and E.Izquierdo (*Senior Member, IEEE)*

***Abstract*— This paper describes a new methodology for 3D measurement and model matching for installation audit in industrial environment. The problem is addressed by using both 3D information from 2D images and semantic meta-data of the installation engine parts by comparing with that of corresponding base CATIA installations. In this research, we deploy independent (vision- based) method to accelerate the convergence toward optimal system architecture that integrates safety constraints.**

***Index Terms*—model knowledge, installation and production audit, aircraft safety audit and analysis, three-dimensional measurement, model matching,**

I. INTRODUCTION: PRODUCTION AUDIT

WITH the availability of cheap sensors and high end 3D vision technology from the market, industries show lot of interest in automating the industrial audit process. Upon a careful analysis, one can note that there is indeed a gap that exists between the industrial need/requirement for safety analysis and existing commercial solutions (for example scanners).

The prior knowledge about industrial installations is the model information available for each of the subparts of the installation components so called DMU (Digital Mock-Ups). In this research, we took initiatives to address the auditing requirement using safety engineer input and CATIA model information as a key input for discrepancy checking.

The increase of aerospace systems complexity has meant that by using existing methods for systems development, industry has reached a barrier to innovation and a risk to the competitiveness of products. This is characterized by an increasing time to market for new technologies, increasing costs to demonstrate proof of safety, a greater demand for skilled resources and a limitation on design iterations, which means there is less time to optimize designs that are compliant with safety targets. In this work, it is aimed at filling this gap and providing methods and infrastructure that accelerate the convergence toward optimal system architecture that integrate safety constrains.

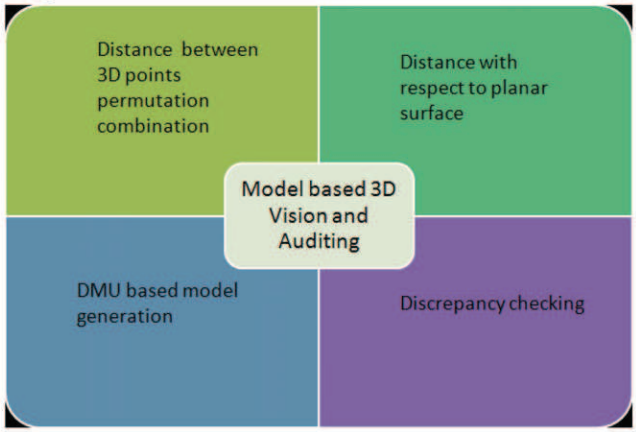

Fig 1.Functional matrix in proposed 3D Vision Auditing framework

Reliable 3D information can be extracted just using the input from 2D images of the industrial installations. As shown in Figure 1, our framework contains four functional components a) measuring distance of any permutation pair of points selected in 2D images to give measure in 3D space b) measure distance with respect to planar surface detecting using non collinear point selection c) DMU based model generation d) discrepancy checking between the visual and model information. It should be noted that functionality c is used for discrepancy checking process.

II. DISTANCE BETWEEN 3D POINTS

A. *User point selection from safety engineer perspective*

The first functionality is where a user provides points of interest with which distance constraints need to be checked.

B. *Point based disparity:*

Point based disparity is based on each of the points identified from in two images that show a different perspective of the same scene. While selecting points, first the user is shown the left image where they select the point coarsely (refer Figure 2 (a)). Then, the system shows the point location on left image that is zoomed so that the user can select the point

Manuscript received August 17, 2011. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° ACP7-GA-2008-212088-MISSA.

Authors are with the MMV Research Laboratory, School of Electronic Engineering and Computer science, Queen Mary University of London, United Kingdom. Phone: 02078827132; Fax: 02078827997; email:{karthike.vaiapury,anil.aksay,xinyu.lin,ebroul.izquierdo}@eecs.qmul.ac.uk).

location very accurately. Similarly, a user will be shown the right image where they will select the point coarsely (refer Figure 2 (b)). Then the point location on the right image will be zoomed so that the user can select the point location very accurately. Since disparity is obtained manually, there is no need for outlier removal and mm level accuracy can be achieved.

Fig.2(a). Selected Points in Pipe (Left View)

Fig.2(b). Selected Points in Pipe (Right View)

C. 3D Stereo Triangulation

As stated by Izquierdo et al, two corresponding points represent the projection onto the image planes of the same object point [3]. 3D position is the intersection of both viewing lines and can be estimated using the coordinates of its projection in both images and the camera parameters. Triangulation is the process of determining the location of a point by measuring angles to it from known points at either end of a fixed baseline, rather than measuring distances to the point directly. The point can then be fixed as the third point of a triangle with one known side and two known angles.

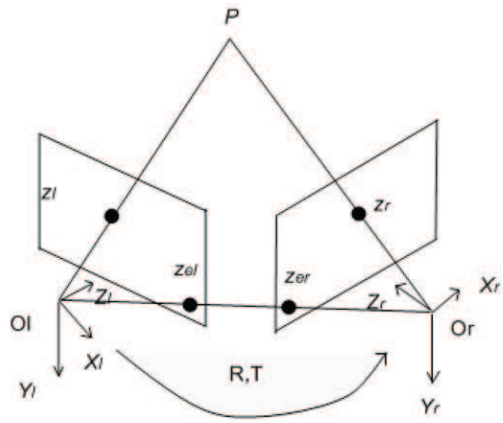

Figure 3 Epipolar stereo geometry

As shown in Figure 3, we retrieve P in space from observed projection $z_l = (u_l, v_l)$ and $z_r = (u_r, v_r)$ onto image planes. u, v represent the coordinate system used in computer or digitized image. It can also be noted that triangulation is not possible if P lies on O_l, O_r or $z_l = z_{el}$ and $z_r = z_{er}$ where z_{el} and z_{er} are epipoles.

Given 3×4 camera projection matrices M_l, M_r and z_l, z_r which represents the corresponding points in stereo images, then mathematically triangulation can be written as a function

$$P = \tau(M_l, M_r, z_l, z_r)$$
$$M_i = K_i[R_i \mid t_i]$$

Where i is the index representing number of cameras n_c

K is the calibration matrix. R_l and R_r represent the rotation matrixes of an object relative to the first camera and to the second camera. The rotation between them (R_{lr}) can be calculated as

$$R_{lr} = R_r \times R_l^{-1}$$

R_{lr} can be written as

$$R_{lr} = \begin{pmatrix} \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & \cos\beta\sin\gamma & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ -\cos\alpha\cos\gamma + \sin\alpha\sin\beta\cos\gamma & \cos\beta\cos\gamma & \sin\alpha\sin\gamma + \cos\alpha\sin\beta\cos\gamma \\ \sin\alpha\cos\beta & -\sin\beta & \cos\alpha\cos\beta \end{pmatrix}$$

Where α, β and γ are rotation angles around X, Y and Z axis. Similarly given two translation vectors t_l and t_r, translation between two cameras can be obtained as

$$T_{lr} = t_r - R_{lr} \times t_l$$

The 3D position of a point P can be reconstructed from the perspective projection of M on the image planes of the cameras, once the relative position and orientation of the two cameras are known.

Let $\hat{X}_l = (X_l, Y_l, Z_l)$ and $\hat{X}_r = (X_r, Y_r, Z_r)$ represent the 3D world coordinate points of point P in left and right camera coordinate systems.

$U_l = \frac{\hat{X}_l}{Z_l} = \begin{pmatrix} u_l \\ v_l \\ 1 \end{pmatrix}$, $U_l = \frac{\hat{X}_r}{Z_r} = \begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix}$ are the coordinate vectors of perspective projection of P on the image.

$\hat{X}_l$ and $\hat{X}_r$ are related by rigid motion equation as

$$\hat{X}_l = R_{lr}\hat{X}_r + T_{lr}$$

$$U_l Z_l = R_{lr} Z_r U_r + T_{lr}$$

$$\begin{bmatrix} U_l & -R_{lr}U_r \end{bmatrix} \begin{bmatrix} Z_l \\ Z_r \end{bmatrix} = T_{lr}$$

With each of two cameras, we get linear equations in unknown coordinates of P, which can be written as $AP = T_{lr}$ where $A = \begin{bmatrix} U_l & -R_{lr}U_r \end{bmatrix}$

where A is 3×4 matrix involving projection matrix M_l, M_r of the camera.

In order to find the best reconstructed 3D point, linear method minimizes the criterion

$\|AP - T_{lr}\|^2$ with respect to P.

This can be written as sum of squared error criterion which needs to be minimized.

$$J(P) = \|AP - T_{lr}\|^2$$

where a and k represent the samples and number of samples in A respectively.

The above equation can be solved by closed form solution based on

$$\Delta J = A^T(AP - T_{lr})$$

Setting the above to zero, we get

$$A^T AP = A^T T_{lr}$$

This provides a closed form pseudo inverse solution

X can be determined as

$$P = (A^T A)^{-1} A^T T_{lr}$$

where $A^T A$ is non-singular and $(A^T A)^{-1} A^T$ is pseudo inverse of A

Given the 3D points, measurement error can be calculated as follows.

$$\%Measurementerror = \frac{abs(estdist - actdist)}{actdist} \times 100$$

The 3D measurement results for the points selected in Figure 2 are summarized in Table.1 The mean error is 4.2895%.

Points	Actual Dist (mm)	Estimated Dist(mm)	3D Measurement Error %
(P1,P2)	64± **0.5**	65.7198	2.6872
(P1,P3)	28± **0.5**	30.0682	3.2316
(P4,P5)	56± **0.5**	56.8278	1.2934
(P5,P6)	28± **0.5**	26.8777	1.7536
(P6,P7)	54± **0.5**	63.6137	15.0214
(P8,P9)	22± **0.5**	20.8802	1.7497
Average Error			4.2895

TABLE.1. DISTANCE OF SELECTED POINTS IN PIPE

III. DISTANCE WITH RESPECT TO PLANAR SURFACE

A plane can be defined just using 3 non-collinear points. The user is required to select 3 non-collinear points in the stereo images in order to obtain the plane (refer Figure 4).

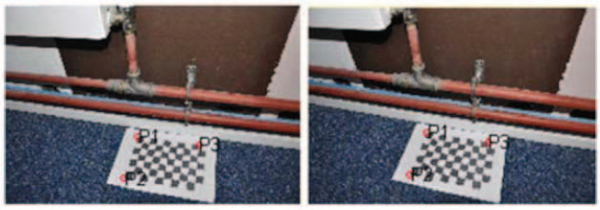

Figure 4 Noncollinear points selected in Pipe

The system of equations $(ax + by + cz + d = 0)$ with the selected points can be solved using Cramer's rule.

Let three non collinear points of 3D data be $x_1, y_1, z_1, x_2, y_2, z_2$ and x_3, y_3, z_3. Then, the parameters of plane can be obtained as

$$\theta = \tan^{-1}(b/a)$$

$$\phi = \tan^{-1}(-\cos(\theta)./a)$$

$$\rho = c \times \sin(\varphi)$$

where a, b, c are calculated as follows

$$a = \left(\frac{-d}{D}\right) \times \begin{vmatrix} 1 & y_1 & z_1 \\ 1 & y_2 & z_2 \\ 1 & y_3 & z_3 \end{vmatrix}$$

$$b = \left(\frac{-d}{D}\right) \times \begin{vmatrix} x_1 & 1 & z_1 \\ x_2 & 1 & z_2 \\ x_3 & 1 & z_3 \end{vmatrix}$$

$$c = \left(\frac{-d}{D}\right) \times \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

Where D is determinant of the 3D data and $d = 2$

Figure 5(a). Selected Points in Pipe (Left View)

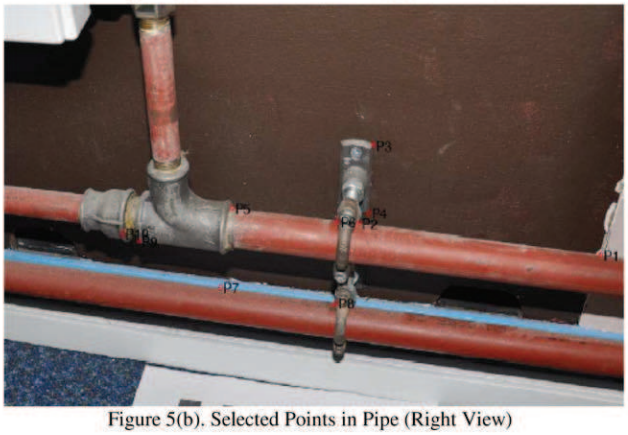

Figure 5(b). Selected Points in Pipe (Right View)

As provided in Table 2, the average error of distance between points selected in Figure 5 and the plane obtained using three non collinear points is 6.5988%. At the cost of around .5% more error than that of using automatic plane detection algorithm [1,4,5], this method is preferable and faster since it needs only 3 points selection rather estimating automatic disparity (SIFT[7], SURF[8]) which relatively takes longer time.

		Manual Selection	
3D Points	**Actual (mm)**	**Manual Selection**	**Error (%)**
P1	225	189.0557	15.1664
P2	216	223.5899	3.2025
P3	237	256.3304	8.1563
P4	176	192.6959	7.0447
P5	216	221.1379	2.1679
P6	230	234.2832	1.8073
P7	136	184.2114	20.3424
P8	160	159.3372	0.2797
P9	210	194.1538	6.6862
P10	213	210.3114	1.1344
Average Error			6.5988 %

TABLE.2. DISTANCE OF SELECTED POINTS IN PIPE WITH REFERENCE TO PLANE

IV. DMU BASED MODEL GENERATION

A. *DMU based knowledge*

The semantic information is used for building the part primitives with exact object component. 3D cloud is segmented using metadata knowledge where information such as color, location and shape class labels etc., are available. The 3D cloud of points need to be compared and fitted with DMU (Digital Mock–Up) model shape primitives thereby enabling further discrepancy analysis.

Model based matching is based on fact that whole object is a transformation (projection) of a preconceived model, or else it can be broken into parts that are. There exists numerous literature for 3D model matching, however the work which is close to our research is Georgeli et al. 2007., The authors [2] propose augmented reality solution for discrepancy check for identifying the differences between the planned 3D model and built items. Initially, anchor–Plates are used as reference information to obtain the pose in the coordinate system of the 3D model. Once pose is obtained, an augmented CAD is created by positioning the image into the 3D view. Upon positioning, a transparency level is used to view the deviation which is done using 2D information. However, we use 3D information from multiview images and we use model knowledge to perform discrepancy analysis in 3D space.

B. *Geometry model generation*

Geometry primitive are generated for each of the class labels available from the prior knowledge model representation. The sample generated primitive for a rectangular prism, pyramid and cube is shown in Figure 6.

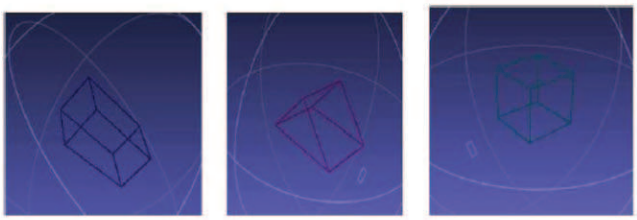

Figure 6. Generated Geometry Primitives

V. DISCREPANCY CHECKING AND ANALYSIS PROCESS

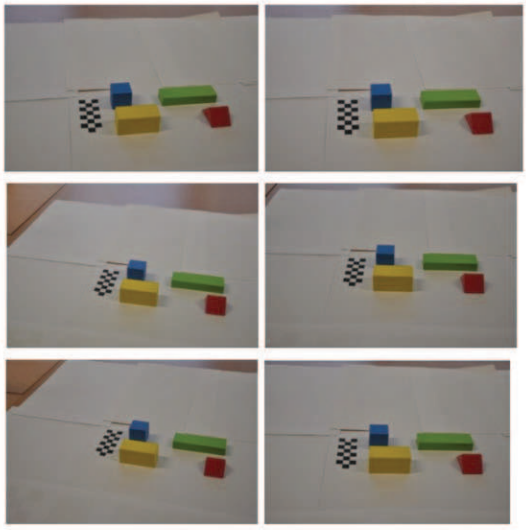

Figure 7.Sample Multiview Images

In this framework, we use calibration information to a) estimate camera pose information and b) perform model alignment. A 3D model of the environment is obtained using a Patch-based Multi-view Stereo algorithm [6] by Furukawa et al 2009. In order to get more accurate parameters, we use automatic calibration pattern recognition instead of bundle adjustment using automatic correspondences. In order to align the whole model with the cloud, we can use either a calibration pattern or one of the objects in the environment as a datum.

A. *Pose Estimation*

In order to perform the projection, we need to know the pose information of the objects present in the environment. Given the reconstructed computer vision Model of the Datum shape (MD) and Model of the Geometry shape (MG), the problem we solve is to minimize the difference between clouds of points and find the best alignment of MD with MG to obtain the pose information. This problem can be formulated based on least square criterion. The points are associated by nearest neighbor criteria and transformation parameters are estimated using a mean square cost function. We use the ICP (Iterative Closest Point) algorithm [9] for estimating this transformation matrix and it is fed to the discrepancy checking module for deviation analysis of the installation environment.

Let M be the reconstructed CV model of the datum shape and $M^{'}$ represent the geometry model shape. $\{m1_i\}$ and $\{m2_i\}$ represent the point sets of models.

$$M = \{m1_i\}_{i=1}^{N1} \text{ and } M^{'} = \{m2_i\}_{i=1}^{N2}$$

This problem can be formulated based on least square criterion as follows.

$$\min_{R,Tj\in\{1,2..N_2\}} \sum_{i=1}^{N1} \left\| Rm1_i + T - m2_j \right\|$$

where $R^T R = I_m$ and $|R| = 1$

where R and T are rotation and translation paramete
The two main steps of ICP algorithm are as follows.

The correspondence between two point sets M and based on $(p-1)_{th}$ rigid transformation are achieved as

$$cp(i) = \arg\min_{j\in\{1,2..N_2\}} \left\| R_{p-1} m1_i + T_{p-1} - m2_{j(i} \right.$$

The rotation and translation parameters are obtain minimizing squared distance

$$(R_k, T_k) = \arg\min_{R^T R = I_m, \det(R)=1} \sum_{i=1}^{N1} \left\| Rm1_i + T - m2 \right.$$

the set of points of the given argument for which the va the given expression attains its minimum value. The ob R and T transform the CV model to the geometry mo order to transform from geometry to CV model, the para such as $R^{'}$ and $-T^{'}$ can be used.

B. *Point Cloud Segmentation*

Since we know the semantic metadata description reg each of the objects such as color, position, and class lab use it as key information to aid segmentation. For exam this setup, each object has distinct color information, wh highly useful in discriminating between those object hence we use segmentation based on color palette inform The location information can be used to localize the s space of the model within the vicinity. The supervise based (R,G,B) color segmentation is used [1].

Pseudo code:

1. $Input: \quad 3D \quad point s(d)$
2. $for \quad i = 1 : size(d,1)$
3. $if \quad d(i,indR) <= rh \,\&\, d(i,indR) >= rl \,\&\, ...d$
   $<= gh \,\&\, d(i,indG) >= gl \,\&\, ...d(i,indB)$
   $d(i,indB) >= bl$
4. $extract \quad d(i,:)$
5. end
6. $Output: \quad Segmented \quad cloud\ based$

The cloud segmented based on color palette informat

shown in Figure 8.

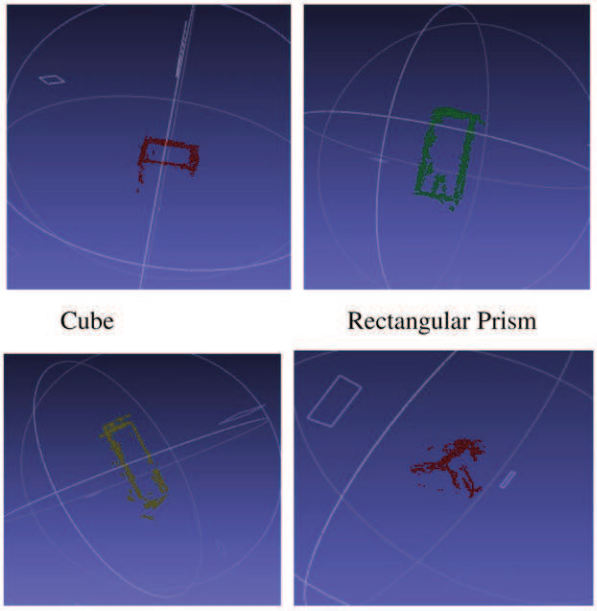

Figure 8. Segmented color clouds

If the segmented cloud still contains outliers, post processing is done using component analysis to get reliable segmented cloud. The algorithm for connected component analysis is as below.

1. *number of points* $(N = 100)$
2. *minimum distance threshold* $(m = 0.5cm)$
3. *find the distance between eachpoint*
4. *Create list for labels of the points a vector of size N. Label all to* -1.
set label_counter $= 0$
5. *loop* :
6. *find minimum distance in matrix indexes i and j*
if min dist $>$ *m stop*
7. *check the labels of both*
8. *if they have no labels set their labels to label_counter and increment labelcounter*
9. *if one of them have a label, set the other one with the same label.*
10. *if they both have labels, choose label with smaller label set the other to the smaller label*
11. *replace the other label with smaller one in label vector*
12. *go to* 5

C. Pose estimation of the Individual objects

For each of the segmented object and corresponding geometry primitive, pose $P_i\{i = 1 : S_n\}$ is recovered using method discussed earlier. S_n is the number of segmented objects. This transformation matrix is used to project geometry primitives onto the 3D digital reconstruction. We use a single pose recovered from the whole system and a pose for each of the object in a system so that the entire system is available.

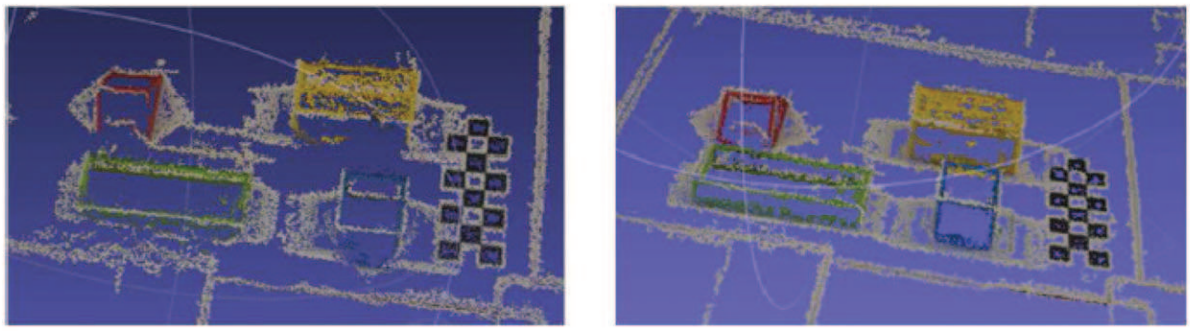

Figure 9-1 Original and Blue shifted cloud

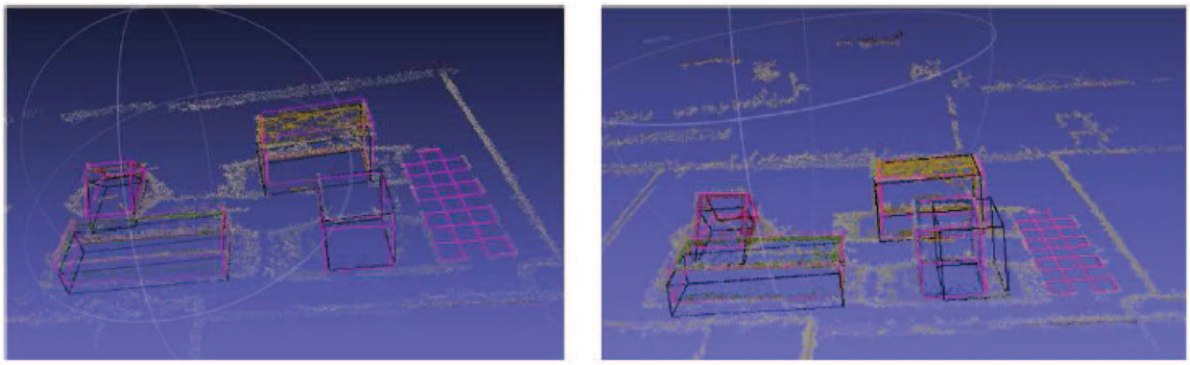

Figure 9-2 Original Fitted and Blue shift Fitted cloud

Class	X	Y	Z
Cube	**0.0597**	**0.0165**	**0.0997**
Rectangular Prism1	0.0061	0.0315	0.5877
Rectangular Prism 2	0.6549	0.2610	0.0746
Triangular Prism	0.2232	0.1043	0.8892

Set (A)

Class	X	Y	Z
Cube	**0.0663**	**0.9556**	**0.1275**
Rectangular Prism1	0.3401	0.0435	0.4443
Rectangular Prism 2	0.1190	0.2941	0.0225
Triangular Prism	0.4288	0.0945	0.7117

Set (B)

TABLE.3 DIFFERENCE BETWEEN CENTER OF OBJECTS ALIGNED BY ICP AND CALIBRATION LOCATION

The original (SetA) and blue shifted cloud (SetB) used for experimentation is shown in Figure 9-1. We use single pose recovered from the whole system and pose for each of the object in a system so that entire system is available. The model and CV cloud aligned using the pose information is shown in Figure 9-2. It can be clearly seen that the shift of 1 cm in blue cube is visible. This deviation from the model would be useful for automatic verification analysis. The difference between the center of the objects aligned by ICP and aligned by calibration location for both normal(Set A) and blue shifted(Set B) cloud is provided in Table 3.

V. CONCLUSIONS

In this paper, we presented a new 3D vision based auditing framework which uses safety engineer input and CATIA model knowledge information. Firstly, we showed that distance between any points in a given installation environment with or without planar surface reference can be measured. Secondly, we provided a novel model matching strategy that uses input from digital camera and semantic metadata knowledge available from geometry models which can be used for verification tasks. The discrepancy analysis of CV model with DMU is demonstrated. A novel framework constituting the various algorithms for achieving the model matching and discrepancy analysis is presented with results. Ideally, the framework acted as proof of concept for safety analysis and verification and has been tested in a controlled environment dataset for model matching. 3D object structures with respect to other objects position in the scene can be extracted. In future, experiments would be conducted in real industry setup.

ACKNOWLEDGMENT

Our thanks to Dr. Chris Papadopoulos for his help throughout the project.

REFERENCES

[1] K.Vaiapury, A.Aksay, X.Lin, E.Izquierdo, C.Papadopoulos, A Vision Based Audit Method and Tool that Compares a Systems Installation on a Production Aircraft to the Original Digital Mock-Up, with a Focus on Safety Driven Installation Constraints, such as Segregation, Proximity and Orientation, *SAE 2011 AeroTech Congress & Exhibition*, 2011.

[2] P.Georgeli, P.Schroederi, S.Benhimanei, S.Hinterstoisseri, M.Appel, N.Navab, An Industrial Augmented Reality Solution For Discrepancy Check, *Proceedings of the 6th International Symposium on Mixed and Augmented Reality*, 2007.

[3] E.Izquierdo and J.R.Ohm, 2000, Image based rendering and 3D Modeling, A complete framework, *Signal processing: Image Communication*, Vol.15, Issue 10, pp. 817-858, 2000.

[4] F.T.Kurdi, T.Landes, and P.Grussenmeyer, Hough Transform and extended RANSAC algorithms for automatic detection of 3D Building Roof Planes from LIDAR data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, ISPRS, Volume 3, Issue Part 3/W52, p.407-412,2007.

[5] R.Schnabel, R.Wahl, R.Klein,Efficient RANSAC for Point-Cloud Shape Detection, *In Computer Graphics Forum*, pp.214-226, 2007.

[6] Y.Furukawa and J.Ponce, Accurate, Dense, and Robust Multi-View Stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[7] Lowe D. G, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004.

[8] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346--359, 2008.

[9] P.J. Besl and N.D. McKay. 1992. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis.and Machine. Intelligence*, Vol.14, Issue 2, ,pp. 239-256, February 1992.

Karthikeyan Vaiapury has obtained BTech (Information Technology) from School of Engineering and Technology, Bharathidasan University (currently Anna University,Trichy), India and M.S (Computer Science) from National University of Singapore, Singapore in 2003 and 2007 respectively. He has worked as software engineer (R&D) at Mahindra Satyam, Singapore for 2 years.He has also worked as Lecturer at the School of Engineering and Technology, Bharathidasan University from Oct 2003 to Dec 2004. Currently, he is a PhD student at Queen Mary, University of London, working on Model based 3D Reconstruction and Recognition using multiview stereo images.

Dr.Anil Aksay received the B.S., M.S and Ph.D degrees in Electrical and Electronics Engineering, from Middle East Technical University, Ankara, Turkey in 1999, 2001 and 2010, respectively. His research interests include image, video and 3D coding, streaming, transmission.

Dr. Xinyu Lin graduated from University of Central Lancashire with BSc (Hons) in Electronics in 1997. He continued his graduate study in the same University and earned PhD degree in image compression in 2004. Currently he is a post-doc researcher in Multimedia and Vision Group in Queen Mary University of London. His research interests include computer vision, medical image processing and object oriented software development.

Prof.Ebroul Izquierdo (M'95–SM'03) received the B.Sc and M.Sc degrees in applied maths from Humboldt University, Berlin, Germany, in 1987 and 1989, respectively. He received the Dr. Rerum Naturalium (Dr.rer.nat) degree for his thesis on the numerical approximation of algebraic-differential equations in 1993 from the same institution. From 1990 to 1992, he was a Teaching Assistant with the Department of Applied Mathematics,Technical University Berlin, Berlin, Germany. From 1993 to 1997, he was an Associate Researcher with Heinrich-Hertz Institute for Communication Technology, Berlin, Germany. From 1998 to 1999, he was a Senior Research Officer with the Department of Electronic Systems Engineering, University of Essex, Essex, U.K. Since 2000, he has been with the Department of Electronic Engineering, Queen Mary, University of London, London, U.K., where he is currently the Chair of Multimedia and Computer Vision and the Head of the Multimedia and Vision Research Group with the Department of Electronic Engineering and Computer Science. He has been involved and coordinating several large European research projects over the last 12 years. He has published over 300 technical papers, including different chapters in books. His current research interests include video coding, video analysis, and visual information retrieval. Dr. Izquierdo is an Associate Editor of several journals in the field of multimedia signal processing, including the IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) and the EURASIP Journal on Image and Video Processing. He has served as a Guest Editor of three special issues of the IEEE TCSVT, three special issues of the Journal Signal Processing: Image Communication, two special issues of the EURASIP Journal on Applied Signal Processing and Image and Video Processing, and several special issues in other journals in the same field. He is a Chartered Engineer, a Fellow Member of the Institution of Engineering and Technology, and a Member of the British Machine Vision Association. He has chaired and co-organized several conferences and workshops in multimedia communications, visual information retrieval, and coding technology.