# An Architecture for Ubiquitous and Collaborative 3D Position Sensing for Ubiquitous 3D Drawing

Shun'ichi Tano, Byongchol Pei, Junko Ichino, and Tomonori Hashiyama University of Electro-Communications Tokyo, Japan tano@is.uec.ac.jp

Abstract— Many methods for 3D position sensing using magnetic sensors, the ultrasonic sensors, special cameras, and so on have been proposed. Most of them, however, require specialized equipment that must be installed in the environment in question. This drawback prevents prototype 3D systems from reaching practical use in real fields. "Ubiquitous 3D drawing" (namely, drawing a sketch in any 3D space in front of a designer), which appears as a common feature in a conventional 3D system, is investigated, and a new method for 3D position detection, called "UbiC3D (<u>ubiquitous and collaborative 3D</u> position sensing) architecture, is proposed. UbiC3D uses only a single camera with sensors and generates a global 3D map of visual feature points through users' collaboration. A preliminary experiment shows that the UbiC3D architecture is feasible.

Keywords: 3D position sensing; ubiquitous 3D drawing; collaboration

#### I. INTRODUCTION

The progress of IT (information technology) is making it possible to handle various forms of media. We have been studying a "creativity-centered media system" that ensures systems truly support creative and intelligent human activities. The systems we have studied range from those used by designers, knowledge workers, medical doctors, office workers, children, and handicapped persons. One of the most attractive media is 3D space, so many support systems for utilizing 3D space have been built, and we have developed prototypes of several such systems that use 3D space.

In this study, we focus on "ubiquitous 3D drawing (namely, drawing a sketch in any 3D space in front of the designer)," which appears as the common feature in the 3D systems, and propose a new method for sensing positions in 3D drawings (hereafter, "3D position sensing"). First, the current status and problems regarding 3D position sensing are briefly explained. Secondly, the requirements that a 3D position-sensing method must meet are summarized in terms of three previously developed prototypes (called systems 1, 2, and 3 hereafter). Thirdly, as the solution to meet these requirements, a "UbiC3D" (ubiquitous and collaborative 3D position sensing) architecture is proposed. In the case of the UbiC3D architecture, short movies, collaboratively taken by many users, are transformed and merged into a global 3D map of visual feature points. Lastly, the feasibility of UbiC3D architecture is demonstrated by conducting a preliminary evaluation.

Mitsuru Iwata Tokyo Metropolitan College of Industrial Technology Tokyo, Japan

#### II. RELATED WORKS ON 3D POSITION SENSING

#### A. Overview

Design support systems using 3D space must be able to detect 3D positions. Since they work within a small area, i.e., within the accessible range of the user's hands and body, magnetic sensors, ultrasonic sensors, and camera-based sensors are used for 3D position sensing [3-9].

Similarly, medical support systems that project medical information directly onto the patient's body need 3D position detection. Since the detection area is limited, magnetic sensors, ultrasonic sensors, and camera-based sensors are used too [16, 17].

As for long-range detection, a star sensor can detect a 3D position in space, and GPS can output a position on the earth. As for micro-level detection, a laser range finder and a special device that uses the characteristic of light waves have been developed.

### B. Our approach

As summarized above, many methods for 3D position detection have been developed and are being widely used. However, current methods for 3D position sensing are not sufficient in regard to a support system for creative and intelligent human activities due to the low accuracy, the requirement for the space to be heavily instrumented and so on. This drawback prevents such support systems from being practically used.

Since most of the current methods require that magnetic sensors, ultrasonic sensors, and cameras are installed in the target environment, the applicable field is quite limited. Moreover, each method has its own strengths and weaknesses; accordingly, to ensure the methods compensate each other, a combination of several sensing algorithms must be used. It, however, results in a complex sensing mechanism.

In most cases, because a prototype system works well in a limited field, its users want to use it in real fields. Insufficient 3D position detection, however, prevents practical use of the prototype. To solve this problem, we focus on "ubiquitous 3D drawing (drawing sketch in any 3D space in front of the user)" as a common function of systems for supporting creative and intelligent human activity by using 3D space. Then the requirements of 3D position sensing are deduced by analyzing three previous prototypes (called systems 1, 2, and 3 hereafter).

## III. RESEARCH MOTIVATION AND REQUIREMENTS OF 3D POSITION SENSING

#### A. Our Development History behind Motivation

The three prototype systems (systems 1, 2, and 3) are explained as follows as typical examples of systems for supporting creative and intelligent human activity by using 3D space. They are quite different from many viewpoints. For example, the target users are ordinary people in the case of system 1, designers in the case of system 2, and medical doctors in the case of system 3. The target field of system1 is nationwide, and that of system 2 and system 3 is room sized. A common feature of each system is "3D drawing".

## B. System 1: Ubiquitous "informal" information-sharing system [9]

The importance of "informal" communication has lately drawn considerable attention because "informal" information can be intuitive, emotional, personal, honest, and subjective-traits that may be very difficult to get from "formal" media. Accordingly, the features of "informal" communication were analyzed, and the requirements of an "informal" communication support system were clarified. On the basis of the results of the analysis, a new "informal" communication support system (system 1, hereafter) was designed [9]. The basic architecture of this system is a combination of a wearable computer, a see-through HMD, a bone-conduction speaker, and a 3D-pen, as shown in Fig. 1. The user of the system can write graffiti such as a handwritten or voice memo and leave it in a real-world context, such as location (GPS), environmental sound, and real objects (RFID). The informal information is shared with others on a peer-to-peer network that is gradually self-organized.



Figure 1. Overview of system 1

A snapshot of a user of system 1 is shown in Fig. 2 on the left, and an example of the user's view through his see-through HMD is shown in Fig. 2 on the right. In this example, the user wrote the memo "The wait for the level crossing here is too long, so take a detour along this road." and attached this memo to the context of the location (GPS) and the warning sound of the railroad crossing.

In the case of system 1, the user must be able to write and leave a hand-written memo at any place. Accordingly, his global location, precise trajectory, and orientation of the memo are sensed by GPS, a pair of ultrasound position devices (Mimio), and magnetic field sensor respectively.



Figure 2. Experiment on informal memo

#### C. System 2: 3D sketch system [10]

There are many researches on 3D sketch systems that support the design of 3D objects by drawing them in a 3D space directly, that is, sketching them in midair in front of a designer. However, they are not used by designers in real fields; they are treated as a mere attraction in an amusement park.

We previously found that 3D sketches must have the characteristics of "life-size and operability" if they are to overcome this miserable situation. Accordingly, we designed a simple architecture for a system for creating 3D sketches. This system ("system 2" hereafter) allows the user to sketch in a 3D space directly [10].

Operating system 2 involves the following steps. First, the designer imagines a shape of a copy machine, asking him or herself, "What's a smart design for a copy machine?" and drawing the idea in life-size. Second, the designer imagines the operation of the sketched machine, thinking to him or herself, "The tray will move in this direction. If I push this button, the paper is ejected." while formulating the operation rules by grasping and moving the sketch shown in 3D. Third, the designer checks the machine's usability by operating the sketch while sitting down, stooping down, extending an arm, and so on. At that time, the designer may find that a button is hard to push because it is inconveniently located, the tray is hard to pull out because you have to get into an uncomfortable position, and so on. In such a case, the designer simply erases the 3D sketch and redraws it.

The current prototype consists of a see-through HMD, a head tracker, a 3D pen, and a palette (command board), as shown in Fig. 3. A sense of 3D position is created by combining an ultrasonic and magnetic sensor to handle a large 3D sketch such as a control room. To promote cooperative design, three and more HMDs (maximum of six) are connected.



Figure 3. Hardware of system 2

Examples of what a user sees through the HMD are shown in Fig. 4. As shown in Fig. 4(a), a virtual pen is displayed on the user's real-world pen. As shown in Fig. 4(b), the designer draws a 3D sketch by referring to the size of a real object.



Figure 4. Views seen by the designer

The target field of system 2 is not limited to the design room described in the above examples (Figs. 3 and 4). It may be used in a large chemical plant, the operation room of a nuclear power plant, or even outside in case of city planning.

### D. System 3: 3D medical system [11]

System 3 is a medical AR (<u>Augmented Reality</u>) system that uses 3D position sensing. The fundamental function of the prototype AR system is to display an ultrasonic image on exactly the corresponding plane in the body of the patient [11]. As a result, a doctor can view the ultrasonic image of the inside of the patient's body as if the ultrasonic image were inserted into the organ on the scanned plane.

The simplest realization of the prototype is shown schematically in Fig. 5. The doctor wears a see-through 3D HMD fitted with a camera. An AR-marker is attached to the ultrasonic probe. One camera-and-marker combination detects the relative 3D position between the doctor's head and the probe. The ultrasonic image can show exactly the same plane in the body of the patient only at the relative 3D position because the current ultrasonic image at the top of the probe can be displayed.



Figure 5. Simple architecture using see-through HMD and single camera and marker combination

To confirm the effectiveness of the system, a prototype was developed (as shown Fig. 6) and experimentally tested. The prototype consists of a 3D-HMD (a Z800 3D visor) and a standard web camera. In the experiment, first, the evaluator (i.e., a doctor) familiarized himself with all the functions of the prototype system and used it to examine the liver of the subjects. The doctor gave positive comments about the usability of the system (i.e., interaction) as well as its overall effectiveness were received. In the next phase of this evaluation testing, a new function, which enables doctors to write a memo on the 3D object, will be added to system 3.



Figure 6. Current prototype system

#### E. Summary of requirements concerning 3D position sensing

Systems 1, 2, and 3 have been successfully evaluated by professional users. However, they are still prototype systems, although their functionalities are quite satisfactory as long as the 3D position sensing works. The success of systems 1, 2, and 3 depends heavily on the 3D position sensing.

The 3D position sensing of system 1 requires more precise detection than that provided by GPS. For example, the required resolution of the location (i.e., where the user wants to attach a memo to a specific location) is about a few decimeters and that of the handwritten memo is less than a few millimeters. Moreover it is required to work anywhere the user wants to leave a memo.

System 2 is currently working in a design room. It is, however, required to work in a much larger area and outside the building. The 3D position sensing must thus also work in a larger area and outside. And the sensing accuracy should be less than a few millimeters.

System 3 has been implemented by combining a marker and a camera. Its 3D position sensing can output only the relative 3D distance between the doctor's eye and the probe. It should therefore be extended to locate the absolute 3D position at the origin of the soft organ of the patient.

The requirements concerning 3D position sensing are thus fourfold:

- (i) global (nationwide) coverage (system 1),
- (ii) operability in ordinal places without prior preparation (system 1),
- (iii) easy adaptability to new places (systems 2 and 3), and
- (iv) variable accuracy (sub-millimeters, a few millimeters, and a few decimeters) according to the situation (systems 1, 2, and 3)
- IV. OVERVIEW OF UBIQUITOUS AND COLLABORATIVE 3D POSITION SENSING ARCHITECTURE: UBIC3D
- *A.* Basic architecture: 3D map of visual features captured by single camera

The first design decision of the UbiC3D is not to depend on a special signal (i.e. ultrasound, laser and magnetism) but a visual signal. Our goal is to support intellectual and creative human activities. So our architecture might be able to depend on the same signal that the human's 3D perception relies on, because our architecture works in the same environment where humans behave.

An architecture that relies on the visual signal can be realized only by a passive camera, needs no special sensors installed in world-wide, and may meet requirements (i), (ii), and (iii).

Recently, a method for calculating the 3D position of the visual feature points has been proposed and almost reached practical application. For example, SLAM [1] and PTAM [2] can generate a 3D map of visual feature points from a camera image in real-time. Although it is not a real-time method, but it can be regarded as 3D map generation, a method of reproducing pursuits from many photos taken in Rome by general users was proposed [12, 13].

The basic architecture of UbiC3D is shown in Fig. 7. The target (a house) is shot by a single camera while moving. The movies are transformed into a 3D map of the visual features. If the corresponded images are given, the 3D map of the visual features is easily calculated by the well-known image processing method. In the case of our architecture, since a movie provides with a close series of images, the matched

visual points are easily tracked. The trajectory of the camera is also estimated.

PTAM [2] utilizes an algorithm based on sophisticated parallelisms, so it can handle 3D maps with more than a few hundred visual features in real-time on a laptop PC.



Figure 7. Basic architecture: 3D map generation by single camera

#### *B. Extension 1: Absolute distance calculation by acceleration and gyro sensors equipped with camera*

As shown in Fig. 7, a 3D map is constructed from information taken by a single camera. However, the value of the position is relative; in other words, in the right image in Fig. 7, there is no unit of length, which is a non-dimensional value.

As shown in Fig. 8, Ubic3D utilizes an accelerometer sensor and a gyro sensor equipped with a camera. These devices give absolute physical values, so a 3D map and camera trajectory can have real units (such as millimeters and meters)

However, in general, the reconstruction of the trajectory obtained from the acceleration and gyro sensor results in very low accuracy. To compensate this low accuracy, two shapes of the trajectory are compared. The trajectory given by the camera image has the good trajectory shape but relative. On the other hand, the trajectory given by the acceleration and gyro sensor is inaccurate but absolute. These characteristics increase the accuracy.



Figure 8. Calculation of absolute distance by sensors

## C. Extension 2: Collaborative construction of a 3D map of the whole world

It takes one user a long time to create 3D maps of the whole world. Therefore, as shown in Fig. 9, many users collaboratively take pictures at their places and the proposed Ubic3D system merges them into a global 3D map. That is, it is an attempt to create a 3D map of the earth in a collaborative manner by the user. It can be seen as a similar attempt to reconstruct buildings and cities from users' pictures [12, 13] mentioned above in a sense of "collaboration". But Ubic3D gathers 3D maps rather than photos, builds up a global 3D map of visual feature points, and uses it for 3D position sensing. So Ubic3D is completely different.



Figure 9. Merge of partial 3D maps

#### D. Extension 3: Fast 3D map matching by hash

To merge 3D maps as shown in Fig. 9, a big hurdle must be overcome. A partial map given by one user must be matched to the global 3D map. This task is a matching problem between a small 3D map and a huge 3D map, and it may be time consuming.

An example of a document search is shown in Fig. 10(a). Recently, millions of pages of printed books have been scanned and stored as a huge 2D map of visual feature points. If a certain page of a particular book (at an arbitrary angle and portion) is photographed with a camera, the original book can be found by real-time matching of a partial 2D map against a huge 2D map on the basis of a hashing algorithm [14].

The UbiC3D system has to match 3D maps. Since the handled data volume expands from 2D to 3D, the matching process may be regarded as more difficult and time consuming. That, however, is not the case. 3D data contain so informative cues that the computational complexity of matching does not grow so fast.



Figure 10. Fast matching of maps by hash

#### E. Extension 4: Accuracy assurance and instruction of camera direction

In UbiC3D system, 3D position is sensed by detecting the camera's position.

The partial 3D map obtained by the user's camera is constantly matched to the global 3D map. The matched location is the user's position.

For examples, in case of a 3D sketch system in a room, the camera is attached to the pen, and the global 3D map stores the 3D positions of the visual points in the room as shown in Fig. 11. The rectangle shows the target zone in which the designer draws lines. The accuracy of the 3D position sensing depends on the number of visual features that the camera, attached to the pen, captures. When the camera captures dense visual features (shown as case A), the accuracy of the position is high. On the other hand, in the case B (sparse visual features), accuracy is low, and in case C, (lack of visual features) no position is given.

There must be sufficient visual features in all directions from the target zone. Accordingly, to assure the accuracy of the position, UbiC3D always estimates the error rate and instructs the user where to turn the camera.



Figure 11. Assuring accuracy of position detection and indicating camera direction

#### F. Extension 5: Treatment of movable features

The images captured by the camera contain two types of visual features: immovable features and movable features. An example of the former is a house; that of the latter is a person. In general, UbiC3D tries to ignore movable features, concentrates on immovable features, and builds a 3D map only by using immovable features. Some applications, however, need to track movable features. For example, the tracking of solid and movable visual features is useful in the case of the ultrasonic probe tracing of system 3. If soft (topological) and movable visual features are properly handled, a soft environment (such as inside the stomach) can be transformed to a soft 3D map.



Figure 12. Immovable, movable, solid and soft (topological) visual feature

#### G. Summary of ubiquitous and collaborative 3D position sensing architecture: UbiC3D

Camera+Sensor

The architecture of UbiC3D is shown in Fig. 13. Ubic3D consists of many ubiquitous clients and a single server (data

center). The ubiquitous clients exist anywhere the 3D position sensing is required. The server manages the global 3D map.

The server (data center) receives partial 3D maps sent by the clients and merges them into the global 3D map (Sever A in Fig. 13) and sends back the appropriate 3D map to the clients upon receiving their requests (Sever B). The client usually initializes the 3D map (Client 1), calculates the current position by matching the current partial 3D map given by their own camera image and the 3D map sent by the server (Client 2), checks the accuracy of the matching, if insufficient, instructs the user in which direction to turn the camera (Client 3), and, finally, requests the server to register the client's partial 3D map and update the global 3D map (Client 4). Usually, the client can calculate the current 3D position on the basis of only the 3D map sent by the server. However, in the case that the server does not have the appropriate 3D map (new place), the client builds the 3D map of the new place by itself and requests to register it. When the client visits that place again, the appropriate 3D map is sent to the client by the server.



#### V. PRELIMINARY FEASIBILITY STUDY OF UBIC3D

### A. Basic process flow of feasibility study

The feasibility of the UbiC3D architecture was experimentally verified according to the process flow shown in Fig. 14. First, a real-time 3D map was generated. The accuracy of the map was then checked, and additional snapshots were taken to improve the 3D map. Finally, 3D drawing application was demonstrated.



Figure 14. Basic process flow of feasibility study

#### B. PTAM

Example results given by PTAM are shown in Fig. 15(a) [2]. The edges of colors and shapes are the visual feature points. As shown in Fig. 15(b), a partial 3D map was gradually generated according to the movement of the camera in realtime. PTAM is used in the first and fourth steps shown in Fig. 14.



(a) Visual Features of PTAM

Figure 15. Data examples generated by PTAM

#### C. Prototype system

The equipment composing the prototype system is listed in Table 1, and the experiment environment is shown in Fig. 16. In this experiment, a small camera was attached to a pen. According to the first step in Fig. 14, to make a good 3D map, the user turned the pen in all directions, from the side wall to side wall, toward the ceiling, toward the furniture in the room, and so on. In the second and third steps in Fig. 14, UbiC3D checks the accuracy of the position and instructs the user to capture further images if necessary. The accuracy of the position, namely, whether a 3D position can be obtained above the desk (the target drawing area), was evaluated. Finally, the user drew a cubic shape by moving the pen in the 3D space over the desk.

| TABLE I. | EXPERIMENT ENVIRONMENT                    |
|----------|---|
| OS       | Windows vista                             |
| CPU      | Intel® Core <sup>TM</sup> 2 6700 2.67 GHz |
| Video    | NVIDIA Quadro FX 540                      |
| Camera   | Point Grey FL2-03S2C                      |



Figure 16. Experiment environment

#### D. Result 1: Accuracy assurance

The third step in Fig. 14 was implemented by the processing flow shown in Fig. 17. First, Ubic3D analyzed the distribution of visual features. For example, as for the distribution of visual features in Fig. 18(a), the left and bottom portions do not have sufficient visual features. When the camera points in this direction, the 3D position is lost. In that case, UbiC3D instructs the user to take additional snapshots.

The sphere is shown in Fig. 18(b), that indicates which direction the visual features exist, whose center is the drawing point. According to the figure, the visual features are concentrated in two directions. When the camera points in these two directions, the accuracy of the position is high; otherwise, the accuracy is low.



Figure 17. Assurring accuracy by instructing the user to take more snapshots



Figure 18. Example of accuracy assurance

#### E. Result 2: 3D drawing

The results of the fourth step (in Fig. 14) are shown in Fig. 19. The user was asked to draw a cube (dice). The experimental environment is shown in Fig. 19(a). The user wore the seethrough HMD while observing his drawing lines and the environment (desk and room). The cube drawn by the user wearing this HMD set up is shown in Fig. 19(b). Since this example result came from the first trial, the shapes are not clear. Even so, it undoubtedly shows that Ubic3D is feasible.



(a) Experimental environment

Figure 19. Drawing of a cube

#### VI. CONCLUDING REMARKS

"Ubiquitous 3D drawing" (namely, drawing a sketch in any 3D space in front of a designer)," which is a common feature of a 3D system, was investigated and experimentally verified. A new method for detecting 3D position, called "UbiC3D" (ubiquitous and collaborative 3D position sensing), which uses only a single camera with sensors and generates a global 3D map of the visual feature points by users' collaboration, was proposed and successfully demonstrated.

Although the feasibility of Ubic3D was demonstrated by a preliminary experiment, various problems are still unsolved. We are now evaluating some alternatives of fast 3D map matching, how to guarantee accuracy of the position, how to handle soft (topological) and movable visual features, and global 3D map DB design.

#### REFERENCES

- Andrew J. Davison, "Real-Time Simultaneous Localisation and [1] Mapping with a Single Camera", ICCV'03, pp. 1-8, 2003.
- G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR [2] Workspaces", In Proc. International Symposium on Mixed and Augmented Reality ISMAR, 2007.
- G. Wesche and H.-P. Seidel, "Free Drawer: A Free-Form Sketching [3] System on the Responsive Workbench", VRST'01, pp. 167-174, 2001.
- S. Schkolne, M. Pruett, and C. P.Schroder, "Surface Drawing: Creating [4] Organic 3D Shapes with the Hand and Tangible Tools", CHI 2001, pp. 261-268, 2001.
- Wayne et al., "Interactive Augmented Reality Techniques for Construction at a Distance of 3D Geometry", Eurographics 2003, pp. 19-28 2003
- Daniel et al., "A Fully Immersive 3D Artistic Medium and Interactive [6] Experience", Proceedings 2001 ACM Symposium on Interactive 3D Graphics, pp. 85–93, 2001.
- Tovi et al., "Creating Principal 3D Curves with Digital Tape Drawing", [7] CHI2002, pp. 121-128, 2002.
- H. Perkunder, J. H. Israel, and M. Alexa, "Shape modeling with [8] sketched feature lines in immersive 3D environments", In Proc. of SBIM 2010, pp. 127-134, 2010.
- S. Tano, T. Takayama, M. Iwata, and T. Hashiyama, "Wearable [9] Computer for Ubiquitous Informal Communication", Sixth International Workshop on Smart Appliances and Wearable Computing -IWSAWC 2006- (at 26th IEEE International Conference on Distributed Computing Systems ICDCS), pp. 1-8, 2006.
- [10] S. Tano, et al., "Three Design Principles Learned through Developing a Series of 3D Sketch Systems: "Memory Capacity", "Cognitive Mode", and "Life-size and Operability", IEEE-SMC2012, 2012.
- [11] S. Tano, K. Suzuki, K. Miki, N. Watanabe, M. Iwata, T. Hashiyama, J. Ichino, and K. Nakayama, "Simple Augmented Reality System for 3D Ultrasonic Image by See-through HMD and Single Camera and Marker Combination", Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2012), pp. 464-467, 2012.
- [12] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski, "Reconstructing Building Interiors from Images", ICCV 2009, pp. 80-87, 2009.
- [13] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski, "Building Rome in a Day", ICCV 2009, pp. 72-79, 2009
- [14] Kazutaka Takeda, Koichi Kise, and Masakazu Iwamura, "Real-Time Document Image Retrieval for a 10 Million Pages Database with a Memory Efficient and Stability Improved LLAH", 2011 International Conference on Document Analysis and Recognition, pp. 1054-1058, 2011.
- [15] B. Pei, S. Tano, T. Hashiyama, J. Ichino, and M. Iwata, "Proposal of architecture for ubiquitous direct drawing in 3D real space using monocular camera", 12th International Symposium on Advanced Intelligent Systems (ISIS 2011), pp. 425-428, 2011.
- [16] M. Das, F. Sauer, U. J. Schoepf, A. Khamene, S. K. Vogt, S. Schaller, R. Kikinis, E. vanSonnenberg, and S. G. Silverman, "Augmented Reality Visualization for CT-guided Interventions: System Description, Feasibility, and Initial Evaluation in an Abdominal Phantom", Radiology, pp. 230-235, 2006.
- [17] D. G. Trevisan, L. P. Nedel, B. Macq, and J. Vanerdonckt, "Detecting Interaction Variables in a Mixed Reality System for Maxillofacialguided Surgery", SVR2006, pp. 39-50, 2006.