Computer-Assisted Culture Learning in an Online Augmented Reality Environment Based on Free-Hand Gesture Interaction

Mau-Tsuen Yang and Wan-Che Liao

Abstract—The physical-virtual immersion and real-time interaction play an essential role in cultural and language learning. Augmented reality (AR) technology can be used to seamlessly merge virtual objects with real-world images to realize immersions. Additionally, computer vision (CV) technology can recognize free-hand gestures from live images to enable intuitive interactions. Therefore, we incorporate the latest AR and CV algorithms into a Virtual English Classroom, called VECAR, to promote immersive and interactive language learning. By wearing a pair of mobile computing glasses, users can interact with virtual contents in a three-dimensional space by using intuitive free-hand gestures. We design three cultural learning activities that introduce students to authentic cultural products and new cultural practices, and allow them to examine various cultural perspectives. The objectives of the VECAR are to make cultural and language learning appealing, improve cultural learning effectiveness, and enhance interpersonal communication between teachers and students.

Index Terms—Augmented reality, gesture recognition, computer-assisted language learning, VECAR

1 Introduction

Today, people live and work in a globalized society; hence, understanding other cultures is a critical skill. Paige et al. [21] defined cultural learning as a process of acquiring the knowledge, skills, and attitudes required for effective communication and interaction with people from other countries. Cultural learning, immersion, and simulation attract researchers from a variety of fields, including anthropology, cognitive psychology, business, and education. In particular, cultural learning is tightly coupled with language learning because language is the principal component of cultural contexts, and learners cannot truly master the target language until they have also mastered the cultural contexts. Paige et al. [21] argued that it is essential for language learners to also be effective cultural learners. Language instruction must provide opportunities for students to be exposed to or immersed in the target culture to gain knowledge of the cultural meanings of time, place, person, and circumstances.

Immersion is a critical factor in cultural and language learning. Baker and MacIntyre [7] found that immersion students indicated higher willingness to communicate, lower communication anxiety, higher perceived communicative competence, and more frequent communication. Physically being or studying in another country offers an immersive experience in its language and culture. By maximizing that immersive experience, the awareness, acquisition, and understanding of a country's language and culture follows naturally. Interaction is another essential factor in cultural and language learning. Freed [6] found that the benefits derived from an overseas experience depended on the type of contacts that students had during their overseas stay: those who engaged in interactive encounters gained more than those who engaged in non-interactive behaviors.

Whereas immersion and interaction in cultural and language learning can be experienced by studying abroad, most learners cannot afford the time and cost to travel to foreign countries for an extended stay. Augmented reality (AR) technology can overlay virtual objects in the real world to create a feeling of immersion. Additionally, computer vision (CV) technology can recognize free-hand gestures from captured live images to enable intuitive interactions. The combination of AR-type immersion and CV-based interaction shows promise for educational applications. Introducing the latest technologies into a classroom provides both novel opportunities and new challenges. Because most students are familiar with or interested in the latest technologies, language and culture teachers have the opportunity to use the learners' interests in technology to promote learning experiences.

To explore the potential of how these state-of-the-art technologies could suit educational purposes, we extended our virtual English classroom [25] by using AR and CV algorithms, abbreviated as VECAR, for foreign students to learn English and its corresponding cultures online. As shown in Fig. 1, participants can be either in a physical classroom or at distinct remote locations with Internet access. Each participant wears a head-mounted camera to capture live images that are processed using the proposed AR and CV algorithms. Subsequently, the output images are shown on a projection screen (usually for teachers) or a head-mounted display (usually for students). All participants' devices are connected in a network based on a client-server architecture. The VECAR enables students to

• *The authors are with the Department of Computer Science and Information Engineering, National Dong-Hwa University, Hualien, Taiwan. E-mail: {mtyang, m9921019}@mail.ndhu.edu.tw.*

Manuscript received 25 July 2013; revised 10 Jan. 2014; accepted 4 Feb. 2014. Date of publication 26 Feb. 2014; date of current version 2 July 2014.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TLT.2014.2307297

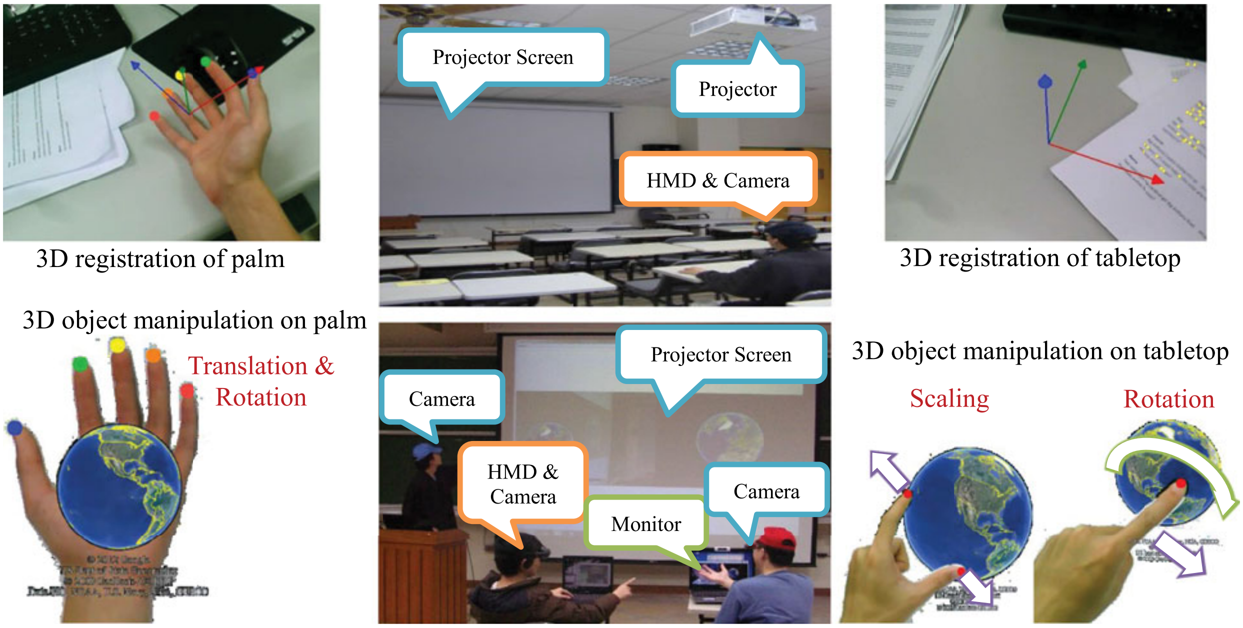


Fig. 1. System overview of the VECAR.

interact with relevant and authentic cultural content by using natural hand gestures. Fig. 2 shows the technical flowchart of the VECAR. Currently, individual hardware components are employed in our prototype; however, they can all be embedded in a compact wearable computing device (such as a pair of intelligent glasses) in the near future. The goal of the VECAR is to enhance the English language cultural learning experiences, improve cultural learning effectiveness, and promote interpersonal communication between teachers and students.

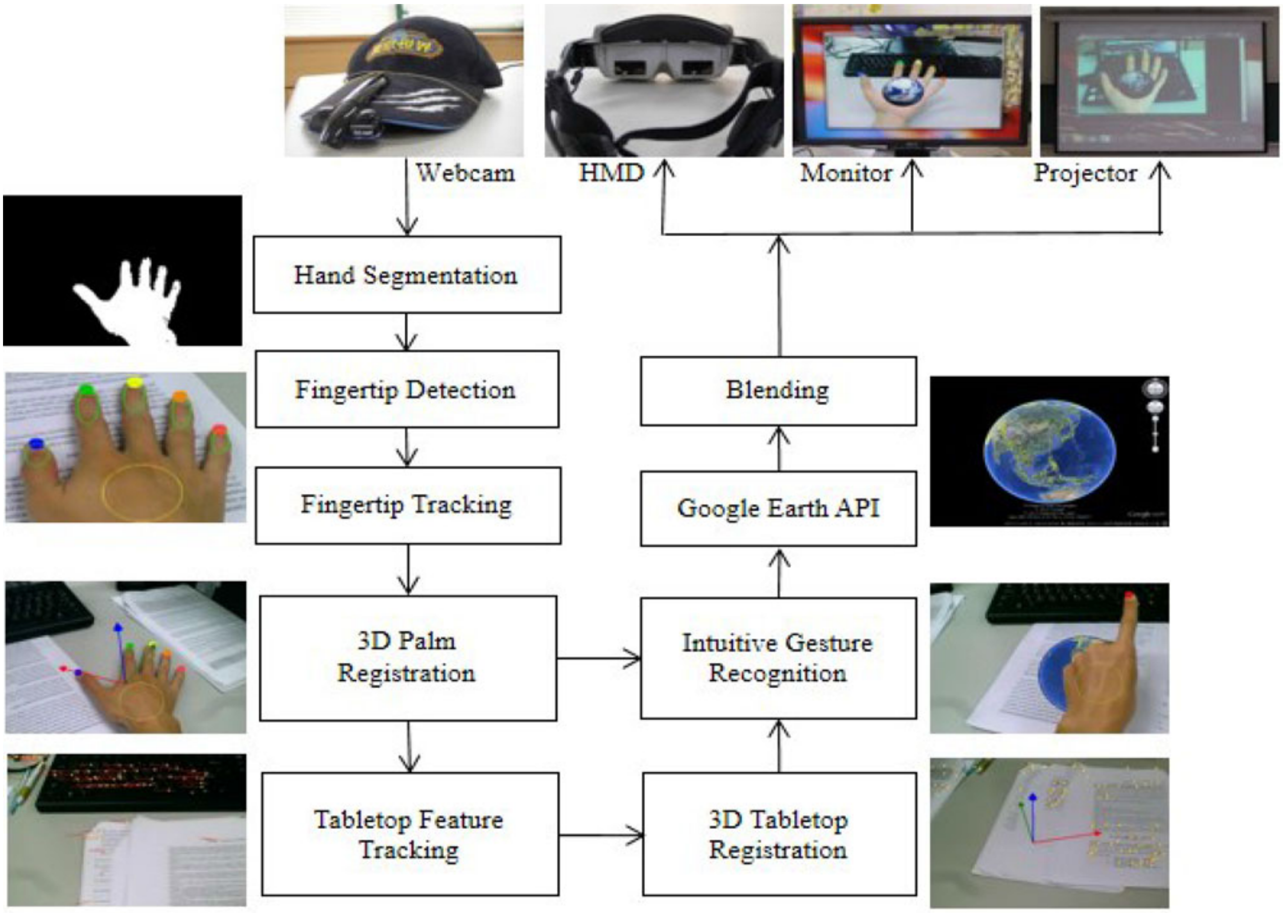


Fig. 2. Technical flowchart of the VECAR.

This paper discusses the system development, curriculum design, and user evaluation of the VECAR. Particularly, we evaluate the students' learning effectiveness of cultural contents, assess the interpersonal communication behaviors between teachers and students, and verify the usability and learnability of the VECAR. The remaining parts of this paper are organized as follows. Section 2 is a review of the relevant literature. Section 3 designs three cultural and language learning activities. Section 4 provides an explanation of the markerless three-dimensional registration of palm and tabletop. An intuitive free-hand gesture interaction is also developed. Section 5 is an explanation of the setup of the learners' evaluations. Section 6 provides a discussion of the evaluation results, and Section 7 offers conclusions.

2 Background

Virtual reality (VR) is a computer-simulated environment in which users can navigate in a virtual world and manipulate virtual objects. VR can simulate immersion and interaction in a target culture and language. It can also offer learners an audio-visual dimension that provides more contextual and linguistic information than a standard textbook can provide. Several virtual learning environments were constructed for improving cultural understanding, communication skills, and interaction with people of other countries. Croquelandia [28], which was developed at the University of Minnesota in 2008, is a 3D virtual environment designed specifically for learning how to make requests in Spanish. Second China [9], which was developed at the University of Florida in 2008, is a virtual island in Second Life that has been designed to mimic cultural and visual aspects of China. The tactical language and culture training system (TLCTS) [11], developed by Alelo in 2008, was designed to teach functional skills in four languages and cultures: Iraqi, Dari, Pashto, and French. Bi-lateral negotiation (BiLAT) [13], which was developed at the University of Southern California Institute for Creative Technologies in 2009, is a game-based simulation and tutoring system for practicing negotiation by using one's awareness of an Arab cultural context. These virtual environments offer users a visual representation of visible aspects of culture, such as streets, buildings, and artifacts. They provide users with a sense of cultural immersion; however, rich cultural context and natural cultural interaction are relatively limited in these virtual learning environments.

Mixed reality (MR) is an extension of VR in which the virtual world is combined with the real environment. According to Milgram and Colquhoun [19], MR can be classified based on the relative balance between virtual and real components in the system. Augmented reality is a subgroup of MR in which a few virtual objects are integrated into an abundant real world. AR supplements reality rather than completely replacing it, as in the case with VR. Azuma [1] argued that a typical AR system has three essential characteristics: First, the system combines virtual objects and the real world; second, users can interact in real time; third, the mix of virtual and real components should be registered in a 3D space.

The use of AR has become prevalent in numerous applications because of its ability to seamlessly merge virtual objects with a real environment. Combined with computer vision technology, AR enables users to interact with virtual objects intuitively and does not require the use of a keyboard or mouse. Three types of AR interaction exist. First, users interact with AR by using a two-dimensional (2D) fiduciary marker containing a black square with a white interior [12], [18]. The pattern inside the white region can be used to identify different markers through template matching. Second, users interact with AR by using special equipment, such as data gloves, a tracker [4], or Kinect [22]. Third, users interact with AR by using bare-hand gestures [10]. In this case, a camera is used to capture live images, and computer vision algorithms are required for gesture recognition. Because of their remarkable immersion and interactivity, AR and CV are ideal technologies for e-learning and future classrooms.

AR has been applied to education in a variety of subjects. Juan et al. [12] used AR markers to represent alphabets for children to learn spelling. Huang et al. [10] taught piano lessons by providing guidance to a learner through the registration of the learner's bare hands and a physical keyboard. Mannus et al. [18] used AR to enable learners to visualize the changes in magnetic fields when interacting with multiple physical magnets. Sugimoto [27] proposed a system called GENTORO, in which a physical robot and a set of projected virtual objects were combined to support children's storytelling activities. Despite ongoing efforts to incorporate AR and CV technologies to support cultural and language learning [30], the effects of their use have not been evaluated from the viewpoint of learners. Table 1 compares various virtual learning systems using distinct technologies with different applications.

TABLE 1
Comparison of Various Virtual Learning Systems

Method	Technology	Application
[28]	VR	Spanish
[9]	VR	Chinese
[11]	VR	Iraqi, Dari, Pashto, & French
[13]	VR	Arabic culture
[12]	AR (fiduciary marker)	Spelling
[18]	AR (fiduciary marker)	Magnetic field
[27]	AR (magic marker) & robot & projector	Storytelling
[10]	AR (markerless)	Piano lesson
Proposed	AR (markerless) & gesture interaction	English language & culture

3 Cultural Learning Activities

There are five interconnected standards designed for foreign language learning [20]: communication, cultures, connections, comparisons, and communities. Among them, culture contains three interrelated components. First, philosophical *perspectives* represent the viewpoints and the underlying beliefs, values, attitudes, ideas, and meanings of a society. Second, behavioral *practices* refer to the patterns of behavior that are considered socially acceptable. Third, cultural *products* are tangible or intangible creations that pervade and reflect the cultural contents of a society. We designed three cultural learning activities by introducing students to authentic cultural products and to new

TABLE 2
Detailed Information Regarding the Three Cultural and Language Learning Activities

Activities	Walk in London	American West Exploration	American Football
Theme	Follow a girl and her mother as they board the classic red bus and begin a whirlwind tour of some of London's most iconic landmarks.	Seaman, the Newfoundland dog that took the long journey with Lewis and Clark, serves a crucial role in the expedition's success, and reaches the Pacific at the end of the journey.	American football is the most popular sport in the United States. It is a sport played by two teams of eleven players on a rectangular field.
Goal	Students participate in a virtual field trip to London and improve their communication skills.	Students discover spirit of exploring the American West and improve their communication skills.	Students understand American football culture and improve their communication skills.
Material	*A Walk in London*, by S. Rubbino.	*Seaman: The Dog Who Explored the West with Lewis and Clark*, by G. Karwoski	*Football in the USA: American Culture and the World's Game*, by P. Morris
Expectation: Communication	The students' ability to read, interpret, discuss, and present a landmark with their partner in a target language facilitates their personal progress in *interpretive*, *interpersonal*, and *presentational* skills.	The students' ability to read, interpret, discuss, and summarize a book chapter with their partner in a target language helps their personal progress in *interpretive*, *interpersonal*, and *presentational* skills.	The students' ability to read, interpret, discuss, and summarize a topic with their partner in a target language helps their personal progress in *interpretive*, *interpersonal*, and *presentational* skills.
Expectation: Culture	The students can analyze and evaluate the relationship between the cultural *perspectives* and *products* of London.	The students can understand the relationships among the cultural *products*, *practices*, *perspectives* involved in the exploration of the American West.	The students can explore the relationships between the cultural *perspectives* and *practices* of American football.
Assessment: Communication	Assess the student's communication skills by asking them to discuss and present a landmark in pairs.	Assess the student's communication skill by asking them to discuss and summarize a book chapter in pairs.	Assess the student's communication skill by asking them to discuss and summarize a topic in pairs.
Assessment: Culture	Assess the student's cultural knowledge by asking them about the history, functions, characteristics, and architecture of a landmark.	Assess the student's cultural knowledge by asking them about the characters, plot, setting, theme, and description of a book chapter.	Assess the student's cultural knowledge by asking them about the history, rules, gameplay, and tradition of American football.
Phase 1: Preparation	Provide readings in advance to ensure that students arrive with a deep understanding of the topics.		
	Introduce background information and provide an overview of the theme to the students.		
Phase 2: Comprehension	The students, working independently, read the sections describing an assigned landmark. They identify the principal features of the landmark.	The students, working independently, read their assigned chapter. They list the main characters in the book chapter and the role that they played.	The students, working independently, read their assigned topic. They identify the main ideas and write an outline on the assigned topic on football.
Phase 3: Interpretation	The students work in pairs to express impressions of the landmark. The paired students identify the words that indicate the attitude of the author toward the landmark.	The students work in pairs to describe the overall theme, setting, and plot of the chapter. The paired students create a summary of the assigned chapter and a route map on Google Earth.	The students work in pairs to express why Americans like football. The paired students formulate questions about the topic for others to discuss.
Phase 4: Application	The paired students role-play as tour guides to introduce the landmark to the class. Ask students to introduce a landmark in their home town.	The paired students summarize a chapter to the class by using a route map. Ask students to share, compare, and combine their route map with other groups.	The paired students ask the class relevant questions. Solicit students to respond to questions with appropriate information.
Phase 5: Extension	The students look for Internet sites that present other London attractions and landmarks.	The students keep a journal during their vacation. They write a biography of their pet or their favorite animal.	The students discuss the various methods used to encourage a team and celebrate its victory.

cultural practices, and allowing them to examine different cultural perspectives.

Backward design [29], which involves the incorporation of cultural elements in language learning activities, was used as a basis for our curriculum planning. First, we determined the information that we wanted students to learn and the goals that we expected them to accomplish as a result of each learning activity. Subsequently, we sought methods for assessing the students' understanding of the products, practices, and perspectives of the target culture, and their interpretive, interpersonal, and presentational competence in the target language. Finally, we determined the detailed procedures of the learning activities that students would use to acquire cultural knowledge and improve their communication skills. Our instructional procedures followed an interactive model [26] containing five steps. The first step was the *preparation phase*, which orients the students to the topic and nature of the theme. The second step was the *comprehension phase*, which familiarizes students with the main ideas of the content. The third step was the *interpretation phase*, which allows students to extract meaningful details. The fourth step was the *application phase*, in which students can demonstrate their comprehension by applying their knowledge or by revising the content in another form. The fifth step was the *extension phase*, in which students can extend the concept to similar themes or circumstances.

The details of the three cultural and language learning activities are described in Table 2. In all activities, reading materials were provided to students in advance to ensure

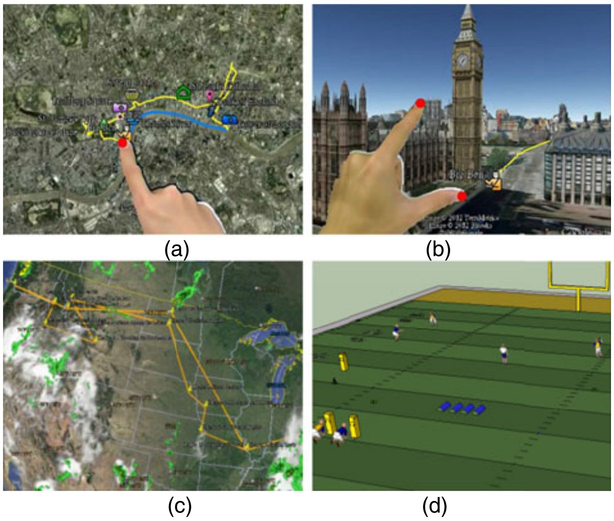


Fig. 3. Screenshots of three learning activities. (a) 2D map with placemarks in LA1 (b) 3D architecture of Big Ben in LA1 (c) 2D route map of west exploration in LA2 (d) 3D football field in LA3.

that they attained a deep understanding of the topics. The content knowledge and background information were introduced by an instructor at the beginning of the activities. The students were subsequently grouped into pairs to accomplish learning tasks. In the first learning activity (LA1), the students took a virtual field trip of London based on the picture book *A walk in London*, written by S. Rubbino. The second learning activity (LA2) was designed to provide an experience of the early exploration of the American West based on the book *Seaman: The Dog Who Explored the West with Lewis and Clark*, written by G. Karwoski. The third learning activity (LA3) introduced American football, the most popular sport in the United States, based on the article "Football in the USA: American Culture and the World's Game" written by P. Morris and Wikipedia football materials. These activities guided the learners to walk inside the story, travel with the characters, and experience the cultures of the target language. In addition, Google Lit Trips (GLTs) [3] were used to organize the 3D contents in these learning activities. GLTs contain a collection of Google Earth placemark files (*.kmz) that mark the journeys described in famous novels or momentous events. Google Earth API [8] was employed to combine online 3D contents in Google Earth with the VECAR. Fig. 3a shows a 2D map containing placemarks in London in LA1. Fig. 3b shows the 3D architecture of Big Ben in LA1. Fig. 3c shows a 2D route map of the exploration of the American West in LA2. Fig. 3d shows a 3D American football field in LA3. Students could mark route maps in 2D and translate, rotate, and scale architecture models in 3D by using hand gestures. AR-based physical-virtual immersions and CV-based gesture interactions in the VECAR were utilized to provide students with authentic learning experiences.

4 Technical Implementation

To fulfill the requirements of immersion and interaction in the aforementioned learning activities for foreign students to learn English and its corresponding cultures online, we developed a virtual English classroom, called VECAR, by incorporating AR and CV algorithms. As shown in Fig. 2, the technical implementation of VECAR can be divided to three parts: three-dimensional registration of palm and tabletop, and an intuitive free-hand gesture interaction.

4.1 3D Free-Hand Registration

Instead of using 2D AR markers, we achieved markerless AR by using the detection, tracking, and 3D pose estimation of the free-hand of users. First, the captured images are transformed from RGB to HSV color space to perform skin color segmentation. A distance transform algorithm is applied to extract the largest component in the skin regions. The pixel exhibiting the maximal distance in the largest component is selected as the center of the hand. Second, the curvature is computed for each boundary point along the hand contour. Ellipse-fitting algorithms are performed along the boundary points that exhibit high curvatures to locate the finger tips. Third, the five finger tips are labeled and tracked online to register the 3D position and orientation of the palm [14]. Supposing that (x_i, y_i) is the 2D image coordinate and (x_w, y_w, z_w) is the 3D world coordinate, a 3D to 2D projection can be formulized using a 3×4 projection matrix P that can be further decomposed into two matrices in which $[r]$ represents 3D rotation, $[t]$ represents 3D translation, (x_0, y_0) indicates the center of the image plane, and (f_x, f_y) are the camera's focal lengths in the horizontal and vertical directions respectively. Assuming that the 3D world coordinate is aligned with the palm so that $z_w = 0$, the 3×4 projection matrix P can be further reduced to a 3×3 homography matrix H. Using this simplification, even if one fingertip is occluded or miss-detected, data obtained from the other four finger tips are sufficient to solve H for the 3D registration.

As shown in the upper left corner of Fig. 1, after a 3D coordinate system is registered using the palm, a virtual object can be projected and rendered by using the estimated projection matrix, thus appearing to be on the top of the palm, as shown in the lower-left corner of Fig. 1. The users can translate and rotate a virtual object in a 3D space by simply translating and rotating their palms in 3D.

4.2 3D Tabletop Registration

In addition to the capability of translating and rotating a virtual object by using their palms, users can place the virtual object at any designated location on their real tabletop. To achieve this goal, we designed algorithms to detect, track, and register the physical tabletop in a 3D space. First, a shift invariant feature transform (SIFT) [16] is performed on each captured image to extract feature points. Second, a Lucas–Kanade optical flow algorithm [17] is applied on temporal consecutive frames to track these feature points continuously. Third, a RANSAC algorithm [5] is adopted to distinguish the feature points on the tabletop plane from those outside the plane. In each RANSAC iteration, four pairs of corresponding feature points are randomly selected and used to estimate a 3×3 homography matrix. For each corresponding pair of the remaining feature points, an error score is defined as the distance between the detected coordinate and the estimated coordinate by homography. The

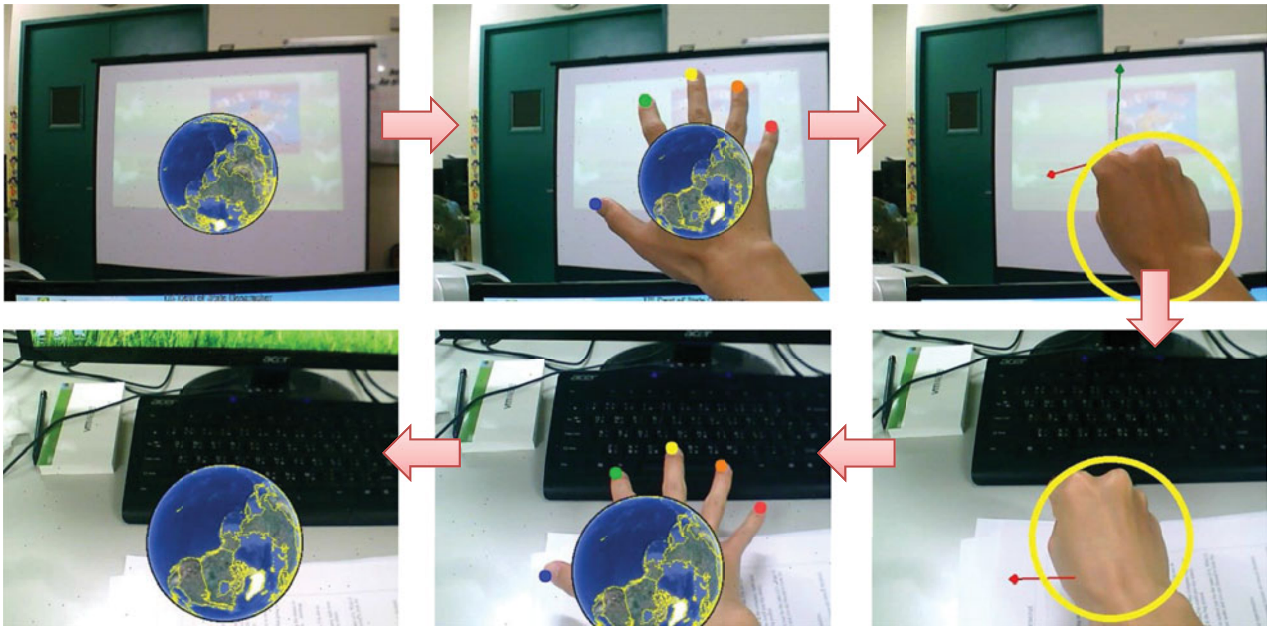


Fig. 4. Getting a virtual object from the projector screen and dropping it on the tabletop.

corresponding pairs with error scores smaller than a predefined threshold are accepted as candidate feature pairs on the tabletop plane. After a sufficient number of iterations, the homography matrix containing the maximal number of accepted candidates is selected. Finally, a 3×4 projection matrix representing the 3D position and orientation of the tabletop is estimated based on the selected homography matrix and its accepted corresponding pairs of feature points. The upper-right corner of Fig. 1 shows a 3D coordinate system that is registered with the tabletop. After the 3D registration, users can obtain a virtual object from the projection screen and drop it at any location on the tabletop, as shown in Fig. 4. Because the virtual object was rendered using the estimated projection matrix, it appears to be on the top of the table. As shown in the lower-right corner of Fig. 1, users can scale and rotate a virtual tabletop object in 3D by using the intuitive hand gestures discussed in next section.

4.3 Intuitive Gesture Interaction

To enable natural and intuitive interaction, a set of hand gestures were designed to manipulate virtual objects on either the tabletop or on the user's palm. Fig. 5 depicts the state diagram of the proposed free-hand gestures for virtual object manipulation. A set of states is defined as follows: $S = \{S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$ where S_0 is the initial state. To model the transfer between states, a set of events is defined as follows: $E_{f0} = \{\textit{no hand detected}\}, E_{f1} = \{\textit{one finger detected}\}$, $E_{f2} = \{\textit{two fingers detected}\}$, $E_{f5} = \{\textit{five fingers detected}\}, E_{f6} = \{\textit{fist detected}\}$, $E_{cc} = \{\textit{center changed}\}, E_{dc} = \{\textit{distance changed}\}, E_{tf} = \{\textit{tracking failed}\}$. Furthermore, a set of context-free grammar rules are defined as follows: $\{S_0 \rightarrow E_{f1}S_1|E_{f2}S_3|E_{f5}S_6\}, \{S_1 \rightarrow E_{cc}S_2\}$, $\{S_2 \rightarrow E_{tf}S_0\}, \{S_3 \rightarrow E_{dc}S_4\}, \{S_4 \rightarrow E_{cc}S_5|E_{tf}S_0\}, \{S_5 \rightarrow E_{dc}S_4|E_{tf}S_0\}, \{S_6 \rightarrow E_{f6}S_7|E_{tf}S_0\}, \{S_7 \rightarrow E_{f0}S_8|E_{tf}S_0\}, \{S_8 \rightarrow E_{tf}S_0\}$.

The proposed state diagram is simple and intuitive. First, if one finger is detected, the rotation mode (S_2) is active and users can rotate a virtual object in 3D by moving the finger (usually the index finger) in the desired direction. Second, if two fingers are detected, the scaling mode (S_4) is active and users can scale a virtual object by stretching two fingers (usually the thumb and the index fingers) apart or closer to each other. Moving two finger tips apart increases the scale, and moving two finger tips closer reduces the scale. In addition, if the center point between two finger tips moves, the virtual object is rotated in the desired direction (S_5). Third, if five fingers are detected, the palm-top manipulation (S_6) is active and users can rotate and translate the virtual object by rotating and translating their palm in 3D, as shown in the lower-left corner of Fig. 1. Fourth, users can grasp a virtual object by clenching a fist (transferring from five fingers to a fist). Subsequently, users can drop the virtual object at another location on the tabletop by spreading their hand (transferring from a fist to five fingers). In this case, the tabletop manipulation (S_8) is active and users can rotate or scale the virtual object by using one or two fingers, as shown in the lower-right corner of Fig. 1. Finally, if a hand is not detected or the tracking fails at any time, the system returns to its initial state (S_0).

Because all of the participants' devices were connected through a network and were organized using a client-server architecture, three types of interactions were supported in the VECAR. The first type is interaction between students and teachers. As shown in Fig. 4, students could grasp a virtual object from the teacher-controlled projection screen and drop it at any location on their tabletop. Conversely, students could grasp a virtual object from their tabletop and drop it on the teacher-controlled projection screen. The second type is interaction between students and curricula. Students could translate, rotate, and scale a virtual object in 3D by using the aforementioned hand gestures. In addition, students could modify the color, shape, and status of the virtual object, or even add messages to it. The third type is interaction between classmates. Students could share a

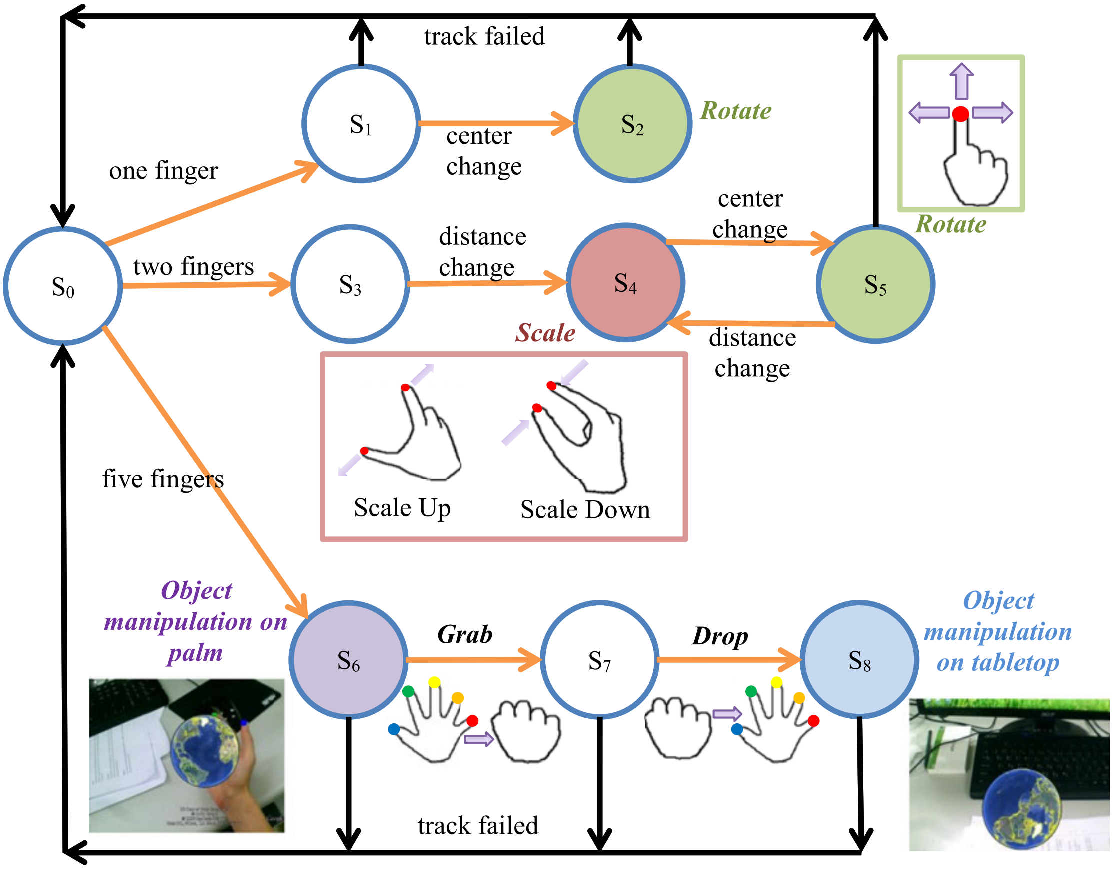


Fig. 5. State diagram of hand gestures for virtual object manipulation.

virtual object with neighboring classmates by taking the virtual object from their tabletop and dropping it on their classmates' tabletop.

5 Evaluation

The goal of our research was threefold. The first research objective was to evaluate the students' learning effectiveness of cultural contents depending on whether the VECAR was used. The second research objective was to assess the interpersonal communication behaviors between teachers and students depending on whether the VECAR was used. The third research objective was to verify the usability and learnability of the VECAR. The participants included forty-four engineering graduate students who learn English as a second language. Their ages ranged from 22 to 26 years. These students were randomly divided to two groups (the experimental group and the control group) with equal size (twenty-two). The instructor was a native English speaker with more than 10 years of experience in teaching English as a second language.

The evaluation was accomplished in stages. In the first stage, participants in both groups took the pre-test in 5 min, the instructor spent 5 min to provide the background information on the themes of the learning activities, and another 5 min to demonstrate the basic functionality of the interaction with the VECAR. In the second stage, a learning activity was conducted and guided by the instructor for 25 min. All participants were equipped with notebook computers with an Internet connection. Participants in the experimental group accessed the online virtual world through the VECAR with the AR-based physical-virtual immersion and CV-based hand gesture interaction. Participants in the control group used a mouse and keyboard to access the online Google Earth and other web materials containing static images and interactive maps through an ordinary web browser. In the final stage, participants in both groups were required to complete the post-test and two questionnaires in 10 min. A pre-test and a post-test were designed to evaluate the learning effectiveness of cultural contents of the students. Both the pre-test and post-test contain the same set of 20 multiple-choice questions giving a full score of 100. The tests were evaluated by three language learning educators and researchers for expert validity.

The Teacher Communication Behavior Questionnaire (TCBQ) [24] was employed to assess the interpersonal communication behaviors between teacher and students depending on whether the VECAR was used. The TCBQ contains 40 questions with five-point Likert-scale items. The questions in TCBQ can be further divided into five subscales: challenging, encouragement and praise, non-verbal support, understanding and friendly, and controlling. The TCBQ questionnaire was slightly modified to adapt its terminology to the VECAR. With eight items in each subscale, all five subscales of the TCBQ were found to display satisfactory internal consistency reliability, discriminant validity, and factor validity [24].

The system usability scale (SUS) [2] was used to verify the usability of the online learning system with the VECAR. The SUS is a simple, efficient, and reliable tool

TABLE 3
Mean Scores, Standard Deviations and Reliability Estimates for Pre-Test and Post-Test of Cultural Learning Effectiveness

Test scores	Control group			Experimental group			Difference
	M_a	SD_a	Cronbach's α	M_b	SD_b	Cronbach's α	$M_{a\text{-}b}$
Pre-test	38.2	18.9	0.76	39.1	14.8	0.80	-0.9
Post-test	78.6	13.6	0.79	85.9	13.7	0.82	-7.3
Difference	40.5			46.8			

TABLE 4
The Results of the t-Tests of Cultural Learning Effectiveness for within-Group and between-Group

Types of *t*-tests	DOF	*t* stat.	*p*=P(*T*<=*t*) one tail
Control group paired *t*-test	21	-8.570	1.343 E-08
Experimental group paired *t*-test	21	-22.080	2.572 E-16
Pre-test independent *t*-test	42	-0.178	0.430
Post-test independent *t*-test	42	-1.771	0.042

TABLE 5
Mean Scores, Standard Deviations and Reliability Estimates for all Subscales in TCBQ

TCBQ Subscale	Control group			Experimental group		
	M_a	SD_a	Cronbach's α	M_b	SD_b	Cronbach's α
Challenging	3.40	0.86	0.95	3.84	0.86	0.97
Encouragement and Praise	3.68	0.81	0.97	3.93	0.78	0.96
Non-Verbal Support	3.38	0.63	0.93	3.82	0.69	0.94
Understanding and Friendly	3.85	0.62	0.84	4.14	0.54	0.86
Controlling	3.09	0.79	0.93	3.29	1.00	0.96
Overall	3.48	0.79	0.96	3.80	0.84	0.98

to measure user's subjective experience of satisfaction. The SUS contains 10 questions with five-point Likert-scale items. Each positive question has a score ranging from 0 to 4 and each negative question has a score ranging from 4 to 0. The overall SUS score ranges from 0 to 100, and is equal to 2.5 times the sum of scores over 10 questions. A higher SUS score indicates superior user experiences when interacting with the system. According to the analysis of the factor structure of the SUS, Lewis and Sauro [15] indicated that the SUS can be further divided into two factors: learnability and usability.

6 Result and Discussion

The first objective was to determine differences in the learning effectiveness of cultural contents depending on whether the VECAR was used. Cronbach's alpha statistics were calculated to estimate the reliability of the collected data. The Cronbach's alpha values for the test results of the three learning activities were 0.83, 0.79, and 0.76, respectively, implying that the questionnaire was reliable. As shown in Table 3, when comparing pre-test and post-test point scores for the experimental group ($N = 22$), the average pre-test score was 38.2 and the average post-test score was 78.6, yielding a percentage difference of +40.5. For those students in control group, the average pre-test score ($N = 22$) was 53.57 and the average post-test score was 70.11, yielding a percentage difference of +16.54. As shown in Table 4, the results of the two within-group paired t-tests yield p values far less than 0.05, meaning that the difference between the pre-test average score and the post-test average score was statistically significant in both groups, i.e., the learning activities improve the cultural knowledge of students in both groups.

Moreover, two independent sample t-tests were conducted to compare differences in the test scores between two groups. For the pre-tests between two groups, the p value for the t-test was higher than 0.05, indicating that the difference of the pre-test between two groups was not statistically significant ($t(42) = -0.178$, $p = 0.430$), i.e., the two groups of students had approximately equivalent basic cultural knowledge. For the post-tests between two groups, the p value for the t-test was lower than 0.05, indicating that the difference of the post-test between two groups was statistically significant ($t(42) = -1.771$, $p = 0.042$). These t-test results indicate that the learning effectiveness of cultural contents was better in the experimental group using the VECAR, compared to those in the control group without the VECAR. VECAR enables students to translate, rotate, scale, and modify virtual objects in 3D by using the intuitive hand gestures, thus it helps students to visualize and play around some cultural contents, such as football rules, that are difficult to understand solely based on verbal explanation.

The second objective was to assess the interpersonal communication behaviors between teachers and students depending on whether the VECAR was used. Table 5 shows descriptive statistics for the five subscales in the TCBQ used to measure the interpersonal communication behaviors. For both the control and experimental groups, the overall Cronbach's alpha values for all subscales were higher than 0.8; hence, the internal consistency of the collected data was good. For all subscales, the mean scores corresponding to the experimental group were higher than

TABLE 6
The Results of Independent Sample t-Tests for all Subscales in TCBQ

TCBQ Subscale	DOF	t stat.	p=P(T<=t) one tail
Challenging	42	-1.915	0.031
Encouragement and Praise	42	-1.174	0.124
Non-Verbal Support	42	-2.682	0.005
Understanding and Friendly	42	-2.376	0.011
Controlling	42	-0.857	0.198
Overall	42	-2.056	0.023

those corresponding to the control group. The largest differences among mean scores were yielded by the *Non-Verbal Support* ($M_{a-b} = -0.44$) and *Challenging* ($M_{a-b} = -0.44$). This observation implies that the VECAR promotes the interpersonal communication between teacher and students, particularly in the *Non-Verbal Support* and *Challenging* subscales.

For each of the five subscales, a Shapiro–Wilk test [23] of normality distribution was applied to examine the distribution of the difference between groups. The p value of the Shapiro–Wilk test for each subscale, was higher than .05, hence the difference for each subscale may be attributed to a normal distribution. Six f-tests were performed to test the equality of population variances of the two groups over five factors and the overall. The f-tests results show that all the p values were higher than .05 (overall: $f(1, 42) = 0.679$, $p = 0.191$), thus we assumed equal variances for the subsequent t-tests. Six independent sample t-tests were conducted to compare the TCBQ differences between the two groups over five factors and the overall. As shown in Table 6, the p values for three subscales were lower than .05, indicating that the difference was statistically significant in *Challenging* ($t(42) = -1.915$, $p = 0.031$), *Non-Verbal Support* ($t(42) = -2.682$, $p = 0.005$), and *Understanding and Friendly* ($t(42) = -2.376$, $p = 0.011$). Overall, a significant difference was observed in the TCBQ scores for the control group without the VECAR (M = 3.48, SD = 0.79) and the experimental group with the VECAR ($M = 3.80$, $SD = 0.84$); ($t(42) = -2.056, p = 0.023$). These t-test results indicate that the interpersonal communication between teachers and students were better in the experimental group using the VECAR, compared to those in the control group without the VECAR. VECAR enables students to move virtual objects back and forth between their tabletops and the projection screen. This non-verbal interaction encourages the interpersonal communication between teachers and students.

The third research objective was to verify the usability of the VECAR. Table 7 shows descriptive statistics for the two subscales used to estimate the system usability by using the SUS. The Cronbach's alpha values for the usability and learnability subscales were higher than 0.7; hence, the internal consistency of the collected data was acceptable. Generally, the learnability subscale value (M = 73.30) was lower than the usability subscale value (M = 80.68), indicating that more tutorials or training can be provided for learners who are unfamiliar with the hand gesture interface in the VECAR. Nevertheless, the overall SUS value (M = 79.20) indicates that the subjective experience of using the VECAR is good. After some practices to familiarize the students with the VECAR, they are attracted by the novel interactive and immersive interface, and enjoy the learning experience using VECAR.

The instructor and learners in the experimental group were interviewed regarding their experiences in the learning activities using the VECAR. Their comments indicated some problems of using the VECAR in cultural and language learning. Instead of using a keyboard and mouse, students in the experimental group need to learn a new interface using hand gestures. Besides, the vision-based algorithms in the VECAR are sensitive to visual occlusions, background noises, and illumination changes, so occasional miss-detections and false alarms can disengage learners' concentrations. Examples of comments regarding these problems are as follows:

"How can I translate/rotate/scale the virtual object in 3D?"
"How can I reset the virtual object to its original state?"
"The system does not respond to my gestures sometimes."
"I feel dizzy after long time use of the mobile computing glasses."

Despite these problems, most participants soon got used to the new interface and successfully completed the assigned tasks in the VECAR. The technical problems were not sufficiently substantial to diminish the enthusiasm of students to try new technologies and accomplish the learning activities in the VECAR. Moreover, learners are not required to attend school and can freely access the VECAR from remote locations. The comments from learners indicated that they enjoy learning through VECAR

"The virtual object seems to be right on my palm."
"After getting familiar with the system, I find it is easy to use."

TABLE 7
Mean Scores, Standard Deviations and Reliability Estimates for Two Subscales in SUS

SUS Subscale (Observations: 22)	M	SD	Cronbach's α
Usability	80.68	10.14	0.88
Learnability	73.30	14.07	0.74
Overall	79.20	9.27	0.85

"I am happy I can interact with the 3D model as I wish."

"It is fun throughout the activity; I don't feel bored at all."

We found an ideal instructor with the enthusiasm to promote cultural and language learning using new technologies. He prepared the online curriculum and familiarized himself with the system in advance. He was also well-prepared to face the confronting troublesome technical problems in class. His comments revealed the potential to enhance online cultural and language learning using the VECAR:

"VECAR can be a tool to visualize 3D cultural contents and get learners to produce language."

"Using lightweight glasses can make the learning experience more natural and less intrusive."

"Learners' familiarity with technology makes them receptive to using it in the classroom."

7 Conclusion

We explored the potential to enhance online language and cultural learning by using the latest AR and CV technologies. Each learner wears a pair of mobile computing glasses with a head-mounted display and camera. The captured live images are subsequently processed by the VECAR using the proposed AR and CV algorithms. Equipped with the capability of maneuvering virtual objects by using free-hand gestures and combining virtual objects with a physical tabletop, the VECAR can create immersive and interactive experiences for language and cultural learners. Three cultural learning activities were designed by introducing students to authentic cultural products and to new cultural practices, and allowing them to examine different cultural perspectives. The evaluation results showed that the VECAR improved the cultural learning effectiveness and promoted interpersonal communication between teachers and students. The goal of the proposed work was to elucidate the interface between linguistic and cultural learning, and the latest CV and AR technologies.

Acknowledgments

This research was supported by the National Science Council, Taiwan.

References

[1] R. Azuma, "A survey of augmented reality," *Presence: Teleoperators Virtual Environ.*, vol. 6, no. 4, pp. 355–385, 1997.

[2] J. Brooke, "SUS - a quick and dirty usability scale," in *Usability Evaluation in Industry*, P. Jordan, B. Thomas, B. Weerdmeester, and A. McClelland, eds., New York, NY, USA: Taylor and Francis, 1996.

[3] J. Burg. (2013 Jul.) Google Lit Trips. [Online]. Available http://www.googlelittrips.com/GoogleLit/Home.html

[4] K. Dorfmuller-Ulhaas and D. Schmalstieg, "Finger tracking for interaction in augmented environments," in *Proc. IEEE ACM Symp. Augmented Reality*, 2001, pp. 55–64.

[5] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, 1981.

[6] B. Freed, *Second language Acquisition in a Study Abroad Context*. Philadelphia, PA, USA: John Benjamins, 1995.

[7] S. Baker and P. Maclntyre, "The role of gender and immersion in communication and second language orientations," *Lang. Learn.*, vol. 50, no. 2, pp. 311–341, 2000.

[8] Google. (2013 Jul) Google Earth API. [Online] Available: https://developers.google.com/earth.

[9] J. Henderson, P. Fishwick, E. Fresh, F. Futterknecht, and B. Hamilton, "Immersive learning simulation environment for Chinese culture," in *Proc. Interservice/Ind. Training, Simul. Educ. Conf.*, 2008.

[10] F. Huang, Y. Zhou, Y. Yu, Z. Wang, and S. Du, "Piano AR: A markerless augmented reality based piano teaching system," *Proc. Int. Conf. Intell. Human-Mach. Syst. Cybern.*, pp. 47–52, 2011.

[11] W. Johnson and A. Valente, "Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures," in *Proc. 20th Nat. Conf. Innovative Appl. Artif. Intell.*, 2008, pp. 1632–1639.

[12] C. Juan, E. Llop, F. Abad, and J. Lluch, "Learning words using augmented reality," in *Proc. Int. Conf. Adv. Learn. Technol.*, 2010, pp. 422–426.

[13] J. Kim, R. Hill, P. Durlach, H. C. Lane, E. Forbell, M. Core, S. C. Marsella, D. V. Pynadath, and J. Hart, "BiLAT: A game-based environment for practicing negotiation in a cultural context," *Int. J. Artif. Intell. Educ.*, vol. 19, pp. 289–308, 2009.

[14] T. Lee and T. Höllerer, "Handy AR: Markerless inspection of augmented reality objects using fingertip tracking," in *Proc. Int. Symp. Wearable Comput.*, 2007, pp. 83–90.

[15] J. Lewis and J. Sauro, "The factor structure of the system usability scale," in *Proc. Int. Conf. Human Comput. Interaction*, pp. 94–103, 2009.

[16] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[17] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. DARPA Imaging Understanding Workshop*, 1981, pp. 121–130.

[18] F. Mannus, J. Rubel, C. Wagner, F. Bingel, and A. Hinkenjann, "Augmenting magnetic field lines for school experiments," in *Proc. Int. Symp. Mixed Augmented Reality*, 2011, pp. 263–264.

[19] P. Milgram and H. Colquhoun, "A taxonomy of real and virtual world display integration," in *Mixed Reality—Merging Real and Virtual Worlds*, Chiyoda-ku, Tokyo, Japan: Ohmsa, 1999 pp. 1–16.

[20] *National Standards for Foreign Language Education Project*. Standards for foreign language learning in the 21st century,1999.

[21] R. Paige, J. Jorstad, L Siaya, F. Klein, and J. Colby, "Culture learning in language education: A review of the literature," in *Culture as the core: Integrating culture into the Lang. education*, D. Lange and R. Paige, eds., Greenwich, CT, USA: Information Age, 2003 pp. 173–236.

[22] E. Santos, E. Lamounier, and A. Cardoso, "Interaction in augmented reality environments using kinect," in *Proc. Symp. Virtual Reality*, 2011, pp. 112–121.

[23] S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.

[24] H. She and D. Fisher, "The development of a questionnaire to describe science teacher communication behavior in Taiwan and Australia," *Sci. Educ.*, vol. 84, pp. 706–726, 2000.

[25] Y. Shih and M. Yang, "A collaborative virtual environment for situated language learning using VEC3D," *Educ. Technol. Soc.*, vol. 11, no. 1, pp. 56–68, 2008.

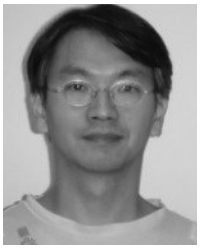
[26] J. Shrum and E. Glisan, *Teacher's Handbook: Contextualized Language Instruction*. second ed. Boston, MA, USA: Heinle & Heinle, 1999.

[27] M. Sugimoto, "A mobile mixed-reality environment for children's storytelling using a handheld projector and a robot," *IEEE Trans. Learn. Technol.*, vol. 4, no. 3, pp. 249–260, Jul.–Sep. 2011.

[28] J. Sykes, A. Oskoz, and S. Thorne, "Web 2.0, synthetic immersive environments, and mobile resources for language education," *CALICO J.*, vol. 25, pp. 528–546, 2008.

[29] G. Wiggins and J. McTighe, *Understanding by Design*, Alexandria, VA, USA, Assoc. Supervision Curriculum Develop., 1998.

[30] M. Yang, F. Yang, and Y. Shih, "VECAR: Virtual English classroom with markerless augmented reality and intuitive gesture interaction," *Proc. Int. Conf. Advanced Learn. Technol (2013)*, pp. 439–440.



Mau-Tsuen Yang received the PhD degree in computer science and engineering from the Pennsylvania State University, in 2000. He is currently an associate professor in the Department of Computer Science and Information Engineering, National Dong-Hwa University, Taiwan. He is the director of computer vision and virtual reality laboratory in NDHU and the author of more than 20 journal papers and 30 conference papers. His research interests include computer vision, augmented reality, and e-learning.

Wan-Che Liao received the MS degree in computer science and information engineering from National Dong-Hwa University, Taiwan, in 2012. His research interests include e-learning and augmented reality.