Journal of Universal Computer Science, vol. 18, no. 20 (2012), 2832-2850
submitted: 22/8/11, accepted: 20/7/12, appeared: 1/12/12 © J.UCS

Non-Marker based Mobile Augmented Reality and its Applications using Object Recognition

Daewon Kim
(Dankook University, Cheonan, South Korea
drdwkim@dku.edu)

Doosung Hwang
(Dankook University, Cheonan, South Korea
dsung.hwang@gmail.com)

Abstract: As the augmented reality technology has become more pervasive and applicable, it is easily seen in our daily lives regardless of fields and scopes. Existing camera vision based augmented reality techniques depend on marker based approaches rather than real world information. The augmented reality technology using marker recognition has limitations in its applicability and provision of proper environment to guarantee user's immersiveness to relevant service application programs. This study aims to implement a smart mobile terminal based augmented reality technology by using a camera built in a terminal device and image and video processing technology without any markers so that users can recognize multimedia objects from real world images and build an augmented reality service, where 3D content connected to objects and relevant information are added to the real world image. Object recognition from a real world image is involved in a process of comparison against preregistered reference information, where operation to measure similarity is reduced for faster running of the application, considering the characteristics of smart mobile devices. Furthermore, the design allows users to interact through touch events on the smart device after 3D content is output onto the terminal screen. Afterward, users can browse object related information on the web. The augmented reality technology appropriate for the smart mobile environment is proposed and tested through several experiments and showed reliable performances in the results.

Key Words: Mobile, Augmented Reality, Image, Object, Recognition

Category: I.0, I.4.9, I.5.4, J.0

1 Introduction

Smart terminal devices have evolved into ever more intelligent gadgets fitted with computer and internet features in addition to their original mobile phone functions. Additional features have enabled smart phones to perform a wider range of functions such as email, web browsing, fax, mobile banking, and games, let alone providing simple communication. Higher performance and diversified functions on multimedia terminal devices have drawn more attention to smart phones, leading to an increase in user base. At the same time, the size of smart phone market has grown, and in particular, the extensive market for smart phone apps has allowed not only business entities but also general users to generate income. As a result, interests in this field are growing further, and the app store environment is rapidly developing. The developing market for smart phone content has naturally complicated and diversified the content operating on smart phones. Over time, apps for smart phones have broken away from the simple and one way programs and required two way high quality services based on new technology. Augmented reality is an alternative to meet such user demands in general. Augmented reality is part of virtual reality. Existing virtual reality replaces the whole real world, whereas augmented reality provides users with an environment where the real world converges into the virtual one with a view to better reality, immersiveness, and interactivity. Interactivity here refers to using a device and augmented reality information to communicate and interact with the real world [Yuan *et al.* 2006]. As the performance of multimedia terminal devices fitted with mobility and augmented reality technology improves, the augmented reality technology is diversely applied to smart phones. To implement an augmented reality, optical flow, marker based, NFT (Natural Feature Tracking) based and sensor based technologies are used [Neumann and You 1999]. These technologies have been widely investigated to solve problems arising in augmented reality when moving while using smart phones and to provide users with efficient UI/UX (User Interface/User eXperience) and appropriate content. In particular, the marker based augmented reality technology that is mostly applied to smart phones at the present is a duplicate of the findings reported in studies on existing desktop environments, thereby leaving lots of differences in I/O, camera usability, system performance, etc [Diverdi and Hollerer 2008]. Also, the widely used image processing treats images given as 2D signals and applies the corresponding standard signal processing techniques [Samavi *et al.* 2006]. This study uses the image processing technique to implement a non marker based augmented reality technology instead of the simple marker based approach, ultimately to improve the quality of augmented reality service technology on smart terminals and to provide general users with further immersiveness and better interactivity. Built with the non-marker based augmented reality based on the image recognition and detection technology, the mobile application software is tested here on Google Android platform based smart terminals in light of efficient content disposition and faster operating speed [Kline and Murray 2010]. Further, to implement a core algorithm as well as to overcome the limited environment of smart phones, the application program is designed with an independent screen structure and UI/UX. The core technology described here relevant to the smart terminals is an app built in the Android based smart phones.

2 Augmented Reality with Multimedia Object Detection and Recognition

2.1 Augmented Reality

Unlike virtual reality, where users are immersed in a virtual world without seeing the real world, the augmented reality provides more than a perfect virtual space. Augmented reality refers to a field of virtual reality, where real objects or places are given additional information or meaning to generate objects that look just like real ones [Cheok *et al.* 2007][Kalkofen *et al.* 2009]. Figure 1 shows a continuum of relations between the real world, virtual reality, and augmented reality.

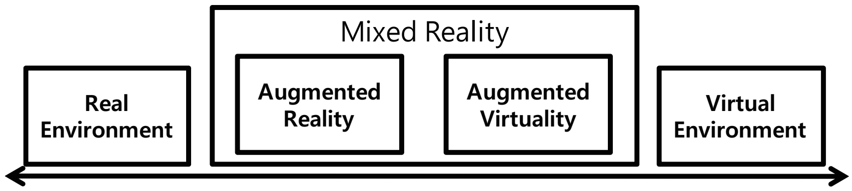


Figure 1: Relationship among reality, virtuality, and augmented reality

As a rule, augmented reality is defined as a technology that, first, combines real images with virtual ones; second, engages in real time interactions; and third, registers virtual images in the real world. Augmented reality is diversely and widely applicable to such fields as medical service [Azuma *et al.* 2001][Birkfellner *et al.* 2002][Harders *et al.* 2009], military, education [Lee *et al.* 2009], games, clothing, real estate, distribution, auto, finance, air service, films, design, architecture, SNS (Social Network Service), etc. Augmented reality involves marker detection, marker matching, tracking, 3D object rendering techniques, etc. The common denominator required of augmented reality is the congruency of objects between real and 3D worlds [Farbiz *et al.* 2005][Kutulakos and Vallino 1998][Seo and Hong 2000]. For this, it is necessary to recognize and track down the target in an image. Object tracking, in general, is realized as sensor based, computer vision based [Hariharakrishnan and Schonfeld 2005], and sensor plus computer vision based hybrid technology. Of the object tracking technologies, the most widely used technology is to track artificial markers, which are fit to ease the object tracking with minimal error, and to further implement an augmented reality. Yet, such markers have drawbacks in terms of limitations hindering their application to ever more diverse and refined fields. This study adopts a non-marker based method to make up for such drawbacks; extracts features like points, lines, edges, and texture of images input from a camera; and applies them to an augmented reality application program. As it extracts features, user defined patterns can be used along with the markers. The non-marker based method is regarded as the most difficult and important technology in the field of augmented reality on account of its outstanding performance in recognizing and detecting rotation angles, orientation, changes in lighting, and partial over laps based on features of objects. Combined with the outdoor tracking functions using information from GPS (Global Positioning System), inertial sensors, and computer vision sensors, image information collected as such will lead to more satisfactory object recognition and tracking results [Bichlmeier 2009][Pressigout and Chaumette 2006].

2.2 Image and Video Objects Detection and Recognition

Object detection refers to a series of processes to match input images and model object data to search a model most similar to the object included in the input image or to find out the location of target objects in case any image is input when information on multiple model objects are kept [Lee and Hollerer 2009][Raghavan *et al.* 1999][Zhong *et al.* 2000]. This process is sensitive to the state of input images because the size, orientation, and location of the object within the input image may differ; because several objects can be included in an image; and because the target objects may be seen partially or distorted if they are blocked by other objects. Two methods are normally used to establish an object recognition system. One is the appearance based approach, where an image in a certain area of the input image is used. This approach uses the brightness value of the area as it is, thereby expressing the shape or texture information of the object implicitly. Also, it uses the vector defined as the brightness values of all pixels within a certain area to express the object or a pattern, where the vector can grow in proportion to the size of the area chosen, thus a dimension reduction technique is applied. This approach is widely used in the field of facial recognition. The other approach is a feature based determination, where the input images are preprocessed to extract features to be used [Mohan *et al.* 2001][Wagner and Reitmayr 2010]. In this approach, the variance of image brightness values is calculated to detect the edges that are prone to change, based on which the contours of the object or the line segments or curves approximate to the contours are found. Also, the corner points present within the input image, the invariant points against certain transformation, and certain points invariant against rotation or size change are extracted and used to detect objects [Lee and Choi 2008][Gavrielides *et al.* 2006]. In addition to object detection, the image recognition technology takes input images to gain information, through which feature points are extracted and compared with the data registered in the system so that the given object can be identified. This recognition technology is applicable to medical services, finger marks, face, counterfeit notes, vehicles committing traffic violations, and specific targets, and it has lately been converged into a range of other technologies [Monro and Rakshit 2007][Zhang 2010]. This study uses the image recognition and detection technology through feature points extraction to determine an object, which is in turn applied with the augmented reality technology, to deal with and implement an applicable service technology. Also, upon object recognition and detection being completed, preregistered 3D content corresponding to the image is called and displayed on the screen followed by some designated operations to search and acquire information related to the content on the web. Figure 2 shows the sequence of the process including object recognition and detection, matching feature points, convergence into augmented reality technology, final information display, and access to online server for information search.

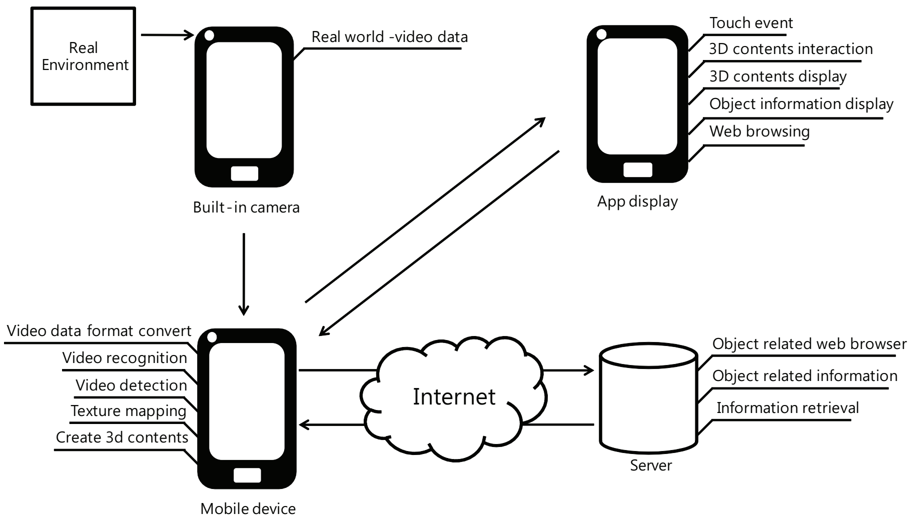


Figure 2: Structure of object recognition, detection and integration of augmented reality technology process

3 Augmented Reality with a Video Object Recognition for Mobile Environments

This study aims to learn and implement a non-marker based augmented reality technology through recognition of images input by a camera built in a smart phone. The mobile application software displays the input video on the screen from the camera and disposes a control box in the middle to mark the extraction of certain parts from the input video. The core technology of image recognition and detection is used to set certain areas on 20 flags of G20 (Group of 20) members and to form their images, which are in turn linked to the web sites to search and browse information on corresponding countries. In addition, several objects round our daily life are also applied and tested. Then, augmented reality technology is used to display the 3D content relevant to a certain country and each objects, and with a few interactions, the users can move to the web site representing the country and objects. Figure 3 shows the process of running the application algorithm where objects are recognized from the camera video and applied with the augmented reality service.

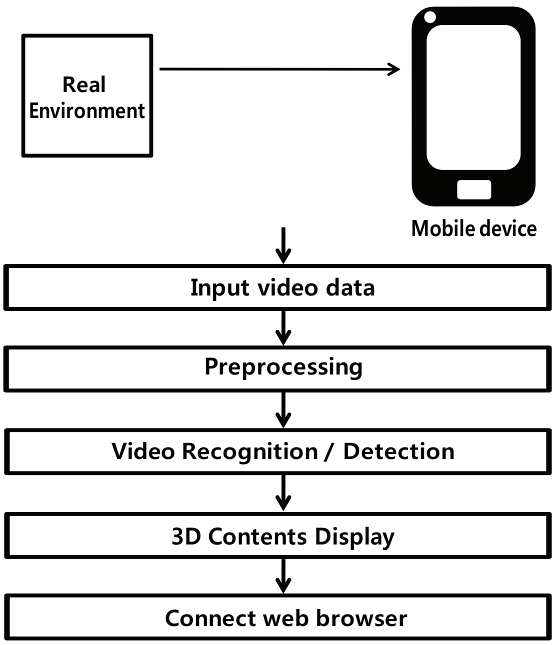


Figure 3: Application process of the video recognition and augmented reality

3.1 Object Detection and Recognition

Image processing for recognition and detection is a technology accounting for a very significant part in the field of vision based augmented reality and has been diversely studied. Desktop based image processing technology has developed a lot thanks to the advancement of device specs and working environment.

Still, more research and development are required in relation to the image contour recognition and detection on some mobile terminal devices with limited resources. On the mobile terminal device used in this study, the target objects are fixed in the pre set area, and a touch event is generated on the screen to start the preprocessing and then receive the input images of YUV420 format from the camera built in the smart terminal. Then, the format is converted into RGB format, and the signals for certain parts that are recognized are extracted. The input image, due to the characteristics of the smart terminal camera, undergoes rotation and reduction for the sake of saving resources for operation and improving the processing speed. Once the information on input image is accurately transmitted and preprocessing is completed without failure, the next step of image recognition and detection starts. In this step, the preprocessed image information and predetermined reference image are compared and analyzed to measure similarity, through which objects are extracted.

$$R(x,y) = \sum_{x',y'} (T^{'}(x^{'},y^{'}) \cdot I^{'}(x+x^{'},y+y^{'})) \tag{1}$$

$$\text{where, } x^{'} = 0, ..., w-1, y^{'} = 0, ..., h-1$$

$$T^{'}(x^{'},y^{'}) = T(x^{'},y^{'}) - \frac{\sum_{x'',y''} T(x^{''},y^{''})}{(w \cdot h)} \tag{2}$$

$$I^{'}(x+x^{'},y+y^{'}) = I(x+x^{'},y+y^{'}) - \frac{\sum_{x'',y''} I(x+x^{''},y+y^{''})}{(w \cdot h)} \tag{3}$$

In the expression (1), T is a preprocessed input image. I is a reference image prepared in advance, and R is an output image after image recognition and detection. w is the horizontal length of image T, and h the vertical one, both of which are used to find the covariance that tells the orientation and aspect of changes in the two values. If the covariance is bigger than 0, then the pixel values of image T and I increase concurrently. Covariance smaller than 0 leads to a concurrent decrease in the pixel values of images T and I. Here, the size of image T is designed to be smaller than that of I and to have a covariance value somewhere between $-\infty$ and ∞. Depending on the measurement units of covariance, the value can vary to a significant extent. The larger the size of the absolute value, the higher the relevance between the two images. For a diverse comparison between input and reference images, the covariance value is divided by image T's standard deviation and that of I's in the corresponding area to yield and compare correlation coefficients. With this method, image T scrutinizes the whole range of I and finds a covariance in each location. To extract the object input in the camera on the mobile terminal device, feature extraction with correlation coefficients is used and the similarity of images is estimated. In the process of feature extraction, the average values between two images are calculated and compared and matches or mismatches are determined. 1 is the output when two images meet the matching reference value designated, and 0 is the output when the two images lack in relevance and are far from meeting the reference value.

$$Z(x,y) = \sqrt{\sum_{x',y'} T'(x',y')^2 \cdot \sum_{x',y'} I'(x+x',y+y')^2} \quad (4)$$

$$R(x,y) = \frac{\sum_{x',y'}(T'(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T'(x',y')^2 \cdot \sum_{x',y'} I'(x+x',y+y')^2}} \quad (5)$$

The expression (4) is the process to yield the standard deviations of the images, T and I, in a given area to turn the covariance into correlation coefficients. Here, the standard deviation is equal to the arithmetic mean distance of the image data from the average values of images, T and I, in the corresponding area. As the covariance values can significantly differ depending on the measurement units despite the same values are used to measure the same data, standardization is used to convert the values for a relatively easier comparison. The expression (5) is to divide the expressions (1) by (4) to find the correlation coefficients. The operation in expression (5) makes it easier to compare values and to reduce the influence of the interference of light while the image is processed. The correlation coefficients are the covariance standardized. Upon completing this process, information on the location can be found, and the most similar value is generated from the correlation coefficient map of the image, R. Using this location information, the most similar object is extracted and the 3D content of the object is called later on. In this process, the input object's feature value and its most similar reference object information are compared before going to the next step. In this step, in case the information on objects is transmitted right away, an error may take place and unwanted items may be delivered to the next process. To prevent such an error, the similarity value information is designed to be compared with the preset reference value. If the measured value indicating similarity is not above the reference value, then the object is determined not to have been registered as a reference object. In this case, the system is designed to send a warning message to the user and run again starting from the image input prior to preprocessing. Once every process goes well without an error, the reference object's 3D content is called and extracted to correspond to the object of information input from the camera. Figure 4 is a schematized block diagram of the aforementioned process.

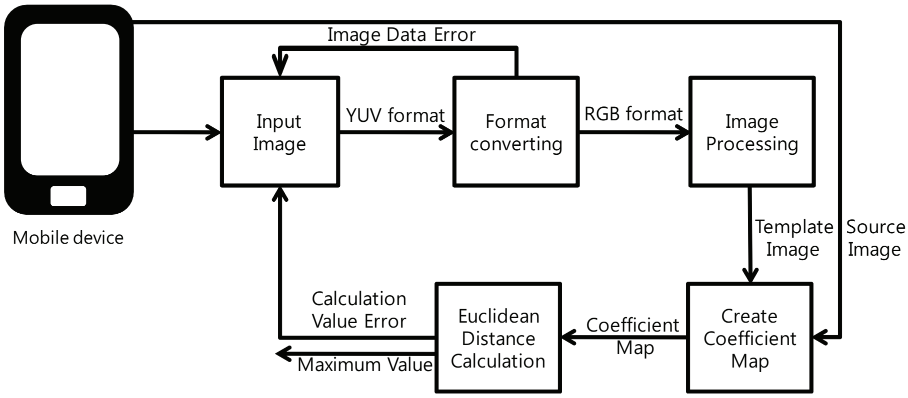


Figure 4: Block diagram of the image object detection and recognition process

3.2 3D Contents Interaction using the Mobile Augmented Reality

To run the application described here, the mobile device must have a built in camera to obtain real world images and be fitted with a touch screen for interactions with the 3D content. In addition, a 3G or 4G wireless network or Wi-Fi environment needs to be set to link to the web site to search or provide the relevant information. When implementing the augmented reality on a mobile environment, restraints must be considered for users' better immersiveness. Such restrictions include the small screen of the mobile equipment, the hardware specification lower than that of the desktop environment, and the relatively slower network environment than wired one due to the characteristics of wireless networks. Moreover, owing to the mobile terminal's small screen, some restraints follow in the process of designing the user interface and the size of 3D content, so the users are given descriptions on how to run it for interaction, here. The application program implemented following the proposed method uses a 3D cube and each corresponding shapes as contents, which are the texture mapping of the objects recognized, as the hardware specification of the mobile terminal that is relatively lower than that of desktop environment could get in the way of user's immersiveness. The application program built as such has a function to access the web carrying the information corresponding to the object recognized from the 3D contents display and is designed to be connected if 3G or 4G or Wi-Fi environment is fitted. Figure 5 shows a series of the process mentioned up to now. The non-marker based augmented reality technology using vision recognition uses the information gained through features of the real world to provide users with augmented information and to improve their immersiveness. The previous section 3.1 in this paper describes image recognition and extraction, through which information on objects is transmitted followed by the corresponding 3D contents displayed. As aforesaid, the 3D contents are comprised of 3D cubes and each corresponding shapes, and for better immersiveness of users, the object related texture is mapped for display, while equivalent related sound source plays.

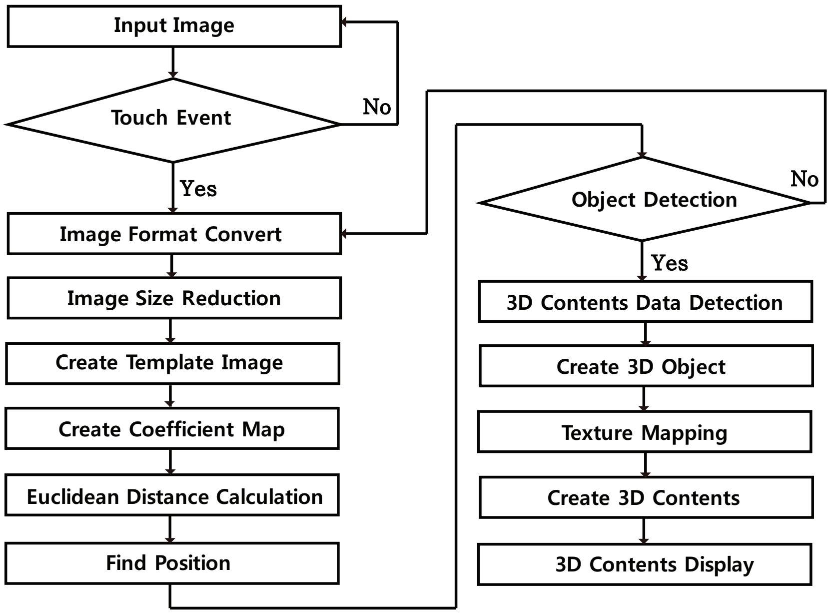


Figure 5: Flowchart of the 3D contents involved augmented reality

Texture mapping is a method which two-dimensional patterns such as an image is applied to a three-dimensional object's surface that is composed of a point, line and a surface to make a solid image with maximum reality. There are mainly two ways of texture mapping. One is the method of projecting an image to an object like a slide projector. The other is the method of coating an image on the surface of a three-dimensional object. Many of other kinds of texturing methods exist in terms of material's shape and quality. In this research, we have used the method of plating texture maps on a three-dimensional object's surface. 3D graphics are processes of transformation to display an object, which has location coordinates in three-dimensional space, on two-dimensional screen and those are implemented with the OpenGL ES (Embedded Systems) methods in Android mobile environments. The 3D graphics technology calculates the position of an object in three-dimensional space and displays on the screen through matrix computations. The OpenGL ES provides various useful methods and functions so we can implement somewhat complicated 3D graphics relatively easily. Figure

6 shows the pipeline structure for processing 3D graphics with the OpenGL ES. In this research, 3D models are produced and saved in *.obj file format by employing the 3D MAX tool. The essential vertex and texture informations are extracted from the *.obj formatted file and saved in arrays for utilizing as object's location coordinates.

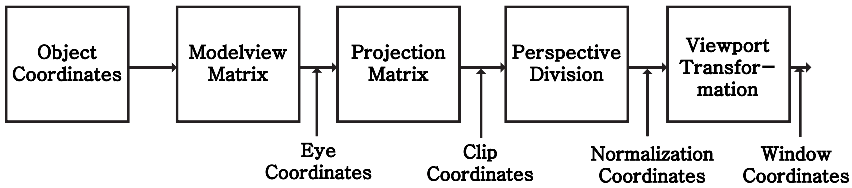
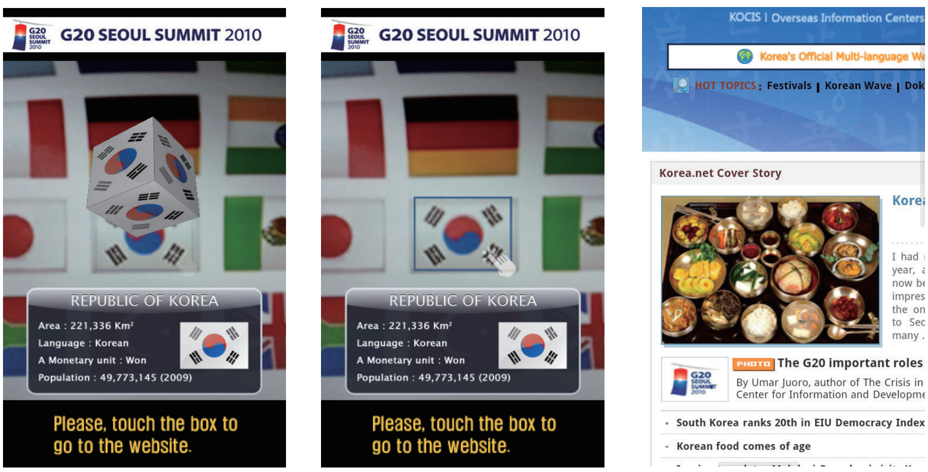


Figure 6: Pipeline structure for processing 3D graphics

We have implemented and tested several samples such as a fire extinguisher, a Bluetooth USB, an iphone and its electric charging adapter. Figure 7 shows the operation of the app implemented on the smart terminal. The screens in Figure 7 are the ones captured by the application program run on the smart terminal.



(a) (b) (c)

Figure 7: 3D contents display and control in real time for the augmented reality: (a) 3D cube contents display, (b) 3D contents touch interaction, (c) Web site browsing

The background of Figure 7(a) and Figure 7(b), respectively, is the flags of G20 member states printed in the information leaflet of the summit and shown through the camera built in the smart phone in real time. It is designed not only to view the 3D content on the terminal display but also to control it while engaging in interactions through the augmented information and simple touch events. In Figure 7, (a) shows the 3D cube content corresponding to flags recognized as an image and relevant information output. When the 3D cube is displayed on the smart phone screen, it can be rotated as necessary by fling events. Next, in (b) of Figure 7, the 3D content, applied with a double tab event, disappears. Lastly, a touch event leads to a web site representing the country as in Figure 7(c).

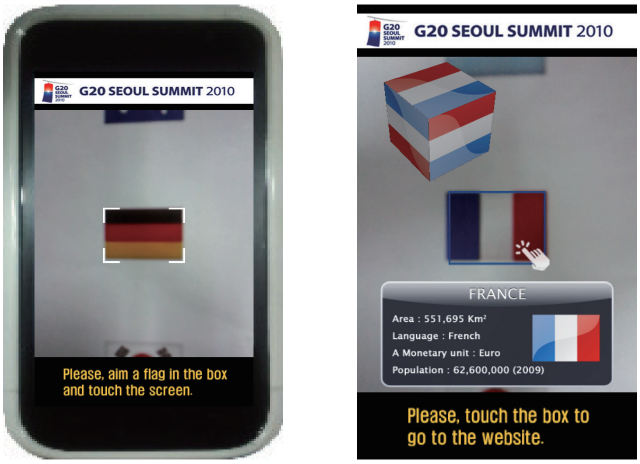
4 Experimental Results and Performance Evaluation

The proposed system is developed under Google Android mobile SDK and Eclipse 3.5 environments, and the experiment is done using real target terminals. Terminal equipments used to implement the system are Samsung SHW-M110S (Galaxy S), SHW-M250S (Galaxy S2) and LG-SU660 (Optimus 2X). For image recognition and extraction, each terminal equipment uses a built in camera. Owing to different product specifications, the speed differs between terminal equipments but in regard of the performance of the camera, there is no significant difference.

Table 1. Comparison of the hardware device specifications used in experiments

	SHW-M110S	SHW-M250S	LG-SU660
CPU	A8 Cortex 1GHz	SS 1.2GHz Dual Core	nV Tegra2 AP20H
OS	Android 2.2	Android 2.3	Android 2.2
RAM	512 MB	1GB	512MB
Memory	16GB		
Screen	4" S-AMOLED	4.3" S-AMOLED Plus	4" IPS 2.5D
Touch	Capacitive Multi-Touch		
Resolution	800×480		
Video	Divx XviD(720p)	MPEG-4, H.263, H.264 Divx, XviD	MPEG-4 H.263, H.264
Wi-Fi	802.11 b/g/n type		
Battery	1500 mAh	1650 mAh	1500 mAh

Table 1 briefs on the specifications of the wireless terminal equipments used in the experiment. The proposed system, as in Table 1, is installed on an Android OS 2.2 or 2.3 based smart phones. Figure 8 shows a screen of running the application program during the simulation. In Figure 8, (a) and (b) show how the proposed system operates. In Figure 8(a), the system starts and performs recognition of the national flag of Germany among G20 countries. In Figure 8(b), along with the information on the country whose flag is recognized upon the screen being touched in the previous step, a 3D cube of flag texturing is displayed on the screen, which is the national flag of France in this case.



(a) (b)

Figure 8: Experimental results using a SHW-M110S: (a) Object recognition of the national flag of Germany, (b) Corresponding 3D contents display of the national flag of France

In the experiment, the static image input through the built in camera in the smart phone underwent a format conversion for image processing. Generally the proposed method is faster than the existing marker based method. The difference in application running speed here is basically ascribable to the different amount of operation for image processing. The marker based existing technology finds out if a marker exists throughout the input image and outputs 3D content on a location of the mobile equipment, whereas the proposed technology extracts a certain part from the image input from the camera in favor of an optimal smart mobile environment and analyzes its features for object recognition, decreasing the running time for operation. It should be noted that the marker based existing technology uses the whole image for marker recognition and binary image, while the proposed technology uses the color image input in YUV420 format and processes the image partially for feature extraction. Compared with black and white images, the YUV420 format color images include much more information in each pixel, leading to better image processing. Nevertheless, the proposed technology demonstrated an outstanding operating performance, in which sense, it is considered to be very effective for the operating environment of smart terminals. For the experiment of measuring accuracy, the proposed technology was to randomly recognize flags of countries. Here, the initial recognition of flags did not go well in some cases such as those of the U.S. and Saudi Arabia, which was attributable to the information loss due to lack of clarity of flag image output in color used as input data for recognition. Another cause of the initial recognition error was the shadow included in the process of image input such as the flags of Argentina, Japan, and Korea. Next, the proposed technology was measured to see if it recognized flags of those countries that were not preregistered and objects completely irrelevant to 20 flags of G20 registered beforehand for recognition. The experiment found that the proposed technology did not recognize any wrong objects.

Figure 9: Experimental results using a SHW-M250S in front of a clean background: (a) Object recognition of a fire extinguisher through the phone camera, (b) Corresponding 3D contents display of (a), (c) Object recognition of an iphone through the phone camera, (d) Corresponding 3D contents display of (c)

In order to extend the feasibility test, two of other kinds of target terminals were also used to verify the method proposed in this paper. Each of Fig. 9, Fig. 10 and Fig. 11, Fig. 12 shows the results of applying the augmented reality application to a SHW-M250S (Samsung Galaxy S2) and a LG-SU660 (LG Optimus 2X), respectively. In Figure 9 and Figure 10, each of (a) and (c) shows the phone LCD focusing on a fire extinguisher and an Apple iphone to recognize the ob-

(a) (b) (c) (d)

Figure 10: Experimental results using a SHW-M250S in front of a noisy background: (a) Object recognition of a fire extinguisher through the phone camera, (b) Corresponding 3D contents display of (a) augmented to the reality, (c) Object recognition of an iphone through the phone camera, (d) Corresponding 3D contents display of (c) augmented to the reality

(a) (b) (c) (d)

Figure 11: Experimental results using a LG-SU660 in front of a clean background: (a) Object recognition of a Bluetooth USB through the phone camera, (b) Corresponding 3D contents display of (a), (c) Object recognition of an electric charger through the phone camera, (d) Corresponding 3D contents display of (c)

(a) (b) (c) (d)

Figure 12: Experimental results using a LG-SU660 in front of a noisy background: (a) Object recognition of a Bluetooth USB through the phone camera, (b) Corresponding 3D contents display of (a) augmented to the reality, (c) Object recognition of an electric charger through the phone camera, (d) Corresponding 3D contents display of (c) augmented to the reality

jects through the camera embedded on the smart phone in front of a clean and a noisy background respectively for the variety of testing environments. In Figure 9 and Figure 10, each of (b) and (d) shows the 3D contents of the fire extinguisher and the iphone correspondingly invoked from the 3D contents library and augmented to the reality by touch events on the LCD of the target smart phones in front of two different backgrounds (clean and noisy). In Figure 11 and Figure 12, each of (a) and (c) also shows the phone LCD focusing on a Bluetooth USB and an electric charging adapter to recognize the objects through the phone camera in front of a clean and a noisy background respectively. In Figure 11 and Figure 12, each of (b) and (d) also shows the 3D contents of the Bluetooth USB and the electric charging adapter correspondingly invoked from the 3D contents repository and augmented to the reality in front of two different backgrounds. Consequently, the proposed technology is highly reliable in terms of running speed and object recognition rate and highly applicable to smart terminal environment. Also, the proposed technology is designed to use non-marker based image recognition and detection features to output the 3D content recognized on the display as part of the augmented reality and to move the content to a web site related through a designated operation. As a range of smart terminals including smart phones and tablet PCs are widely in use now, application program services are on the rise using the images input through the camera embedded in the terminals. Likewise, the proposed technology here is expected to be conducive to relevant applications to a certain extent.

5 Conclusion

This study describes the augmented reality service where the image input through a camera built in a multimedia based smart terminal is determined using non-marker based object recognition and detection algorithm, and the resulting 3D content is output on the terminal display. In addition to the image acquired from the real world image content, to provide users with augmented information and implement image overlay, a technology to recognize the input image and determine accurate matching is required. Existing marker based augmented reality technology is not fully optimized for wireless multimedia terminal equipment specifications in terms of the small display screen of smart terminals, basic resource and equipment performance, battery consumption, etc. Also, using a marker limits the applicability and decreases the immersiveness of users. Therefore, this study calculates the covariance and correlation coefficients between reference images saved as pre data and the image inputs from the camera to determine the similarity between two images. Also, image recognition and feature extraction are used to improve the overall operation speed and users' immersiveness, and to make the information exchange easier through a touch operation on the 3D content considering the small screen of the smart terminal. The experiment on the Android based smart terminals show a significant competitive advantage of the proposed system over the old marker based technology. In the near future, as vision based augmented reality technology operating on smart terminals increases, a meaningful relative performance test will become possible. This study proposes a non-marker based augmented reality technology, where the input video plays while the 3D content is displayed, and in the course of extracting the object from the input video, the object size control or rotation is performed. Future studies will deal with a real time processing of the video input using the current technology while users can engage in real time interactions with the 3D video. Furthermore, an algorithm needs development that is least affected in terms of size or degree of rotation in image recognition when extracting the objects after capturing the feature points.

Acknowledgement

This work was supported by the NGTMS Lab. in Dankook University, South Korea. Dr. Doosung Hwang is the corresponding author of this paper.

References

[Azuma et al. 2001] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B.: "Recent advances in augmented reality"; IEEE Computer Graphics and Applications, 21, 6 (2001), 34-47.

[Bichlmeier 2009] Bichlmeier, C.: "The Virtual Mirror: A New Interaction Paradigm for Augmented Reality Environments"; IEEE Transactions on Medical Imaging, 21, 9 (2009), 1498-1510.

[Birkfellner et al. 2002] Birkfellner, W., Figl, M. and Huber, K.: "A Head-Mounted Operating Binocular for Augmented Reality Visualization in Medicine-Design and Initial Evaluation"; IEEE Transactions on Medical Imaging, 21, 8 (2002), 991-997.

[Cheok et al. 2007] Cheok, A.D., Qiu, Y., Xu, K. and Kumar, K.G.: "Combined Wireless Hardware and Real-Time Computer Vision Interface for Tangible Mixed Reality"; IEEE Transactions on Idustrial Electronics, 54, 4 (2007), 2174-2189.

[Diverdi and Hollerer 2008] DiVerdi, S. and Hollerer, T.: "Heads Up and Camera Down: A Vision-Based Tracking Modality for Mobile Mixed Reality"; IEEE Transactions on Visualization and Computer Graphics, 14, 3 (2008), 500-512.

[Farbiz et al. 2005] Farbiz, F., Cheok, A.D. and Wei,L.: "Live Three-Dimensional Content for Augmented Reality"; IEEE Transactions on Multimedia, 7, 3 (2005), 514-523.

[Gavrielides et al. 2006] Gavrielides, M.A., Sikudova, E. and Pitas, I.: "Color-Based Descriptors for Image Fingerprinting"; IEEE Transactions on Multimedia, 8, 4 (2006), 740-748.

[Harders et al. 2009] Harders, M., Bianchi, G., Knoerlein, B. and Szekely, G.: "Calibration, Registration, and Synchronization for High Precision Augmented Reality Haptics"; IEEE Transactions on Visualization and Computer Graphics, 15, 1 (2009), 138-149.

[Hariharakrishnan and Schonfeld 2005] Hariharakrishnan, K. and Schonfeld, D.: "Fast Object Tracking Using Adaptive Block Matching"; IEEE Transactions on Multimedia, 7, 5 (2005), 853-859.

[Kalkofen et al. 2009] Kalkofen, D., Mendez, E. and Schmalstieg, D.: "Comprehensible Visualization for Augmented Reality"; IEEE Transactions on Visualization and Computer Graphics, 15, 2 (2005), 193-204.

[Klein and Murray 2010] Klein, G. and Murray, D.W.: "Simulating Low-Cost Cameras for Augmented Reality Compositing"; IEEE Transactions on Visualization and Computer Graphics, 16, 3 (2010), 369-380.

[Kutulakos and Vallino 1998] Kutulakos, K.N. and Vallino, J.R.: "Calibration-Free Augmented Reality"; IEEE Transactions on Visualization and Computer Graphics, 4, 1 (1998), 1-20.

[Lee and Choi 2008] Lee, D.J. and Choi, K.T.: "Recognizable-Image Selection for Fingerprint Recognition With a Mobile-Device Camera"; IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 38, 1 (2008), 233-243.

[Lee et al. 2009] Lee, S.H., Choi, J.Y. and Park, J.I.: "Interactive E-Learning System Using Pattern Recognition and Augmented Reality"; IEEE Transactions on Consumer Electronics, 55, 1 (2009), 883-890.

[Lee and Hollerer 2009] Lee, T.H. and Hollerer, T.: "Multithreaded Hybrid Feature Tracking for Markerless Augmented Reality"; IEEE Transactions on Visualization and Computer Graphics, 15, 3 (2009), 355-368.

[Mohan et al. 2001] Mohan, A., Papageorgiou, C. and Poggio, T.: "Example-Based Object Detection in Images by Components"; IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 4 (2001), 349-361.

[Monro and Rakshit 2007] Monro, D.M. and Rakshit, S.: "DCT-Based Iris Recognition"; IEEE Transactions on Pattern Analysis and Machine Intelligence, 29, 4 (2007), 586-595.

[Neumann and You 1999] Neumann, D. and You, S.: "Natural Feature Tracking for Augmented Reality"; IEEE Transactions on Multimedia, 1, 1 (1999), 53-64.

[Pressigout and Chaumette 2006] Pressigout, M. and Chaumette, F.: "Real-Time Markerless Tracking for Augmented Reality: The Virtual Visual Servoing Framework"; IEEE Transactions on Visualization and Computer Graphics, 12, 4 (2006), 615-628.

[Raghavan et al. 1999] Raghavan, V., Molineros, J. and Sharma, R.: "Interactive Evaluation of Assembly Sequence Using Augmented Reality"; IEEE Transactions on Robotics and Automation, 15, 3 (1999), 435-449.

[Samavi et al. 2006] Samavi, S., Shirani, S. and Karimi, N.: "Real-Time Processing and Compression of DNA Microarray Images"; IEEE Transactions on Image Processing, 15, 3 (2006), 754-766.

[Seo and Hong 2000] Seo, Y.D. and Hong, K.S.: "Calibration-Free Augmented Reality in Perspective"; IEEE Transactions on Visualization and Computer Graphics, 6, 4 (2000), 346-359.

[Wagner and Reitmayr 2010] Wagner, D. and Reitmayr, G.: "Real-Time Detection and Tracking for Augmented Reality on Mobile Phones"; IEEE Transactions on Visualization and Computer Graphics, 16, 3 (2010), 355-368.

[Yuan et al. 2006] Yuan, M.L., Ong, S.K. and Nee, A.Y.C.: "Registration Using Natural Features for Augmented Reality Systems"; IEEE Transactions on Visualization and Computer Graphics, 12, 4 (2006), 569-580.

[Zhang 2010] Zhang, J.: "Low-Resolution Gait Recognition"; IEEE Transactions on Systems, Man, and Cybernetics-PART B: Cybernetics, 40, 4 (2010), 986-996.

[Zhong et al. 2000] Zhong, Y., Jain, A.K. and Dubuisson-Jolly, M.P.: "Object Tracking Using Deformable Templates"; IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 5 (2000), 544-549.