

RESEARCH

Open Access



Video exhibition with adjustable augmented reality system based on temporal psycho-visual modulation

Pei-Yu Lin^{1*}, Bin You² and Xiaoyong Lu²

Abstract

Augmented reality (AR) has been widely applied in our lives. An AR tag is usually used and stacked on products for sake of being detected and recognized easily. With the recognized pattern, the AR system thereby can superimpose the corresponding virtual object on the AR tag. The AR tag, however, distort the perception of products. In this study, we proposed a mechanism that integrates the AR tag with digital screen by applying the temporal psycho-visual modulation (TPVM) technique. The new method utilizes the difference between the human-eye perception and the imaging of digital camera to produce an invisible AR tag on the digital screen. The AR tag can be detected by mobile devices but not yet by the human eyes. Based on the concepts of AR and TPVM, the new mechanism is practical in the related AR applications, such as the commercial, entertainment, and protection.

Keywords: Augmented reality, Temporal psycho-visual modulation, Perception, Monitor, Mobile device

1 Introduction

Augmented reality (AR) [1–3] is a new technology, which puts virtual object on the real-world environment via the help of auxiliary devices, such as mobile phone and camera. In general, according to the computer imaging technology, it makes real environment and virtual objects exist on the same screen or simultaneously combines them in the same space.

To exhibit the virtual object in the real-world environment, an AR tag (also called marker) is usually adopted within the marker-based AR system. The AR tag is a predefined pattern that is linked with the virtual object in the AR system [4–8]. Figure 1 shows an instance of an AR tag [9]. While the camera captures and recognizes the AR pattern, the AR system can estimate the rotation degree and translation between the camera and the AR tag. The virtual object thereby can be located on the AR tag appropriately via display monitors.

In the marker-based AR applications, the AR tag usually sticks on the surface of products. The irregular AR pattern inevitably distorts the quality and the

appearance of the products, such as the perceptions of the magazine, digital media, and digital signboard. Based on the observation, we aim to develop a new methodology that conveys the AR tag into the digital media with invisibility for preserving the integrity and appearance of the original media content. The human eyes are incapable of figuring out the AR tag from digital media directly. Moreover, the general user can reach the AR virtual object by simply scanning the invisible AR tag on the media via digital camera.

To achieve the invisible AR tag, we utilized the temporal psycho-visual modulation (TPVM) technique [10]. TPVM can be adopted in various refresh rates, such as 60, 120, and 240 Hz, via digital monitors or projectors. Especially, for the popular 3D media, the monitor/television is able to support up to 120-Hz refresh rate. However, in most cases, the human eyes are incapable of perceiving the optical signal that flashes frequency beyond 60 Hz [11]. Based on the observation, we aim to design an invisible AR tag approach with TPVM such that the human eyes can only observe the original media without the AR tag on the digital monitor or projector [12].

To exhibit the AR visual object, users can capture the video frames from the display monitor via digital camera

* Correspondence: pylin@saturn.yzu.edu.tw

¹Department of Information Communication, and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Chung-Li 32003, Taiwan
Full list of author information is available at the end of the article

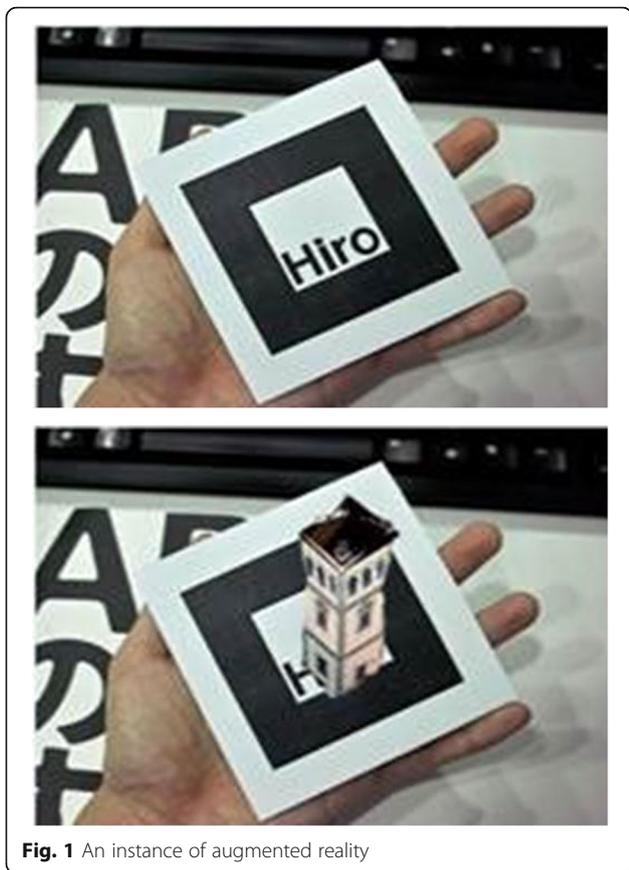


Fig. 1 An instance of augmented reality

or mobile phone. The camera normally can process and capture the display frames even the flicker frequency higher than 60 Hz. That is, a mobile camera can capture frames with an AR tag. However, human visual perception is incapable of distinguishing the embedded AR frames from the display monitor with higher than 60-Hz flicker frequency. The phenomenon is based on the differences between the human eyes and the digital camera image formation.

For the human visual system, it requires continuous light field to form a stable image. On the contrary, for

digital imaging camera, it has two types of sensors: CCD (charge-coupled device) and CMOS (complementary metal oxide semiconductor) [13]. For the camera of CCD, the first process is to close the mechanical shutter; meanwhile, the sensor will clean out the electric charge immediately. After cleanup, the mechanical shutter will turn on and activate the electronic shutter, at the same time, digital-imaging camera accepts the light of the image during the exposure time. After exposure, the mechanical shutter is automatically shutdown. During the period of processing time, the mechanical shutter will turn on until the operator pushes the button for the next data collecting. Due to the fact that the camera needs time to clean electric charge and set parameters, time delay is hard to avoid. That is the reason why differences exist between the human eyes and the camera image production. According to the phenomenon, the display media is not the same between the human perception and the digital camera.

The rest of the paper is organized as follows: Section 2 introduces the TPVM and the motivation of the new scheme. Section 3 describes the procedures of the proposed AR system. The simulation and experiment results are demonstrated in Section 4. The conclusions are made in Section 5.

2 Preliminary and motivation

According to the technology of digital light processing (DLP), the DLP refresh rate may be higher than 120-Hz [14]. The monitor/projector can emit two frames per second. In general, the flicker frequency of the human eye is around 60-Hz. That is, the human eyes cannot detect the complete two frames in 1 s through the 120-Hz monitor/projector.

Given a display video, let I be the original frame and T be the AR tag. To exhibit the video with the invisible AR tag via 120-Hz monitor, we firstly assume that A and B be the same frame I per second. To achieve the effect of AR, we aim to increase the distance between A and B . The concept of the display scheme with temporal psycho-

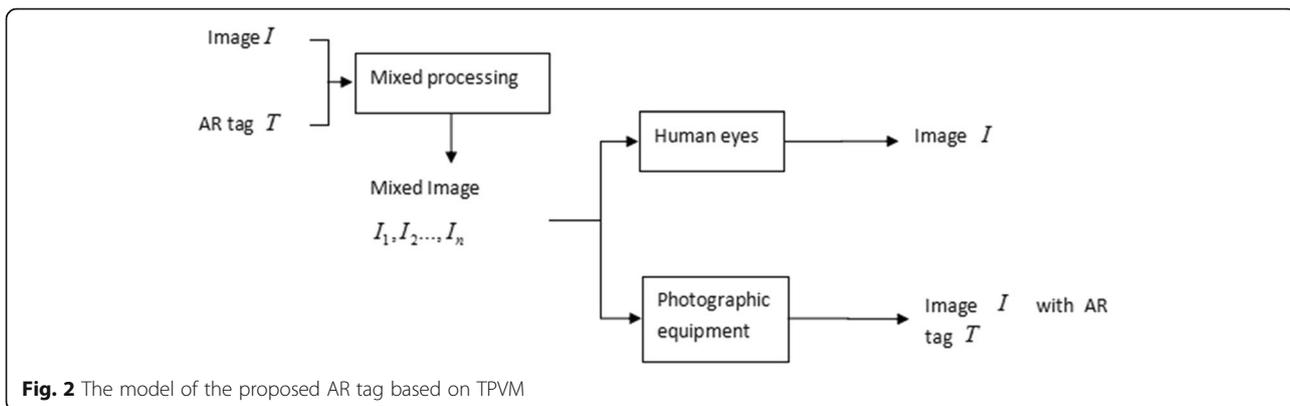


Fig. 2 The model of the proposed AR tag based on TPVM

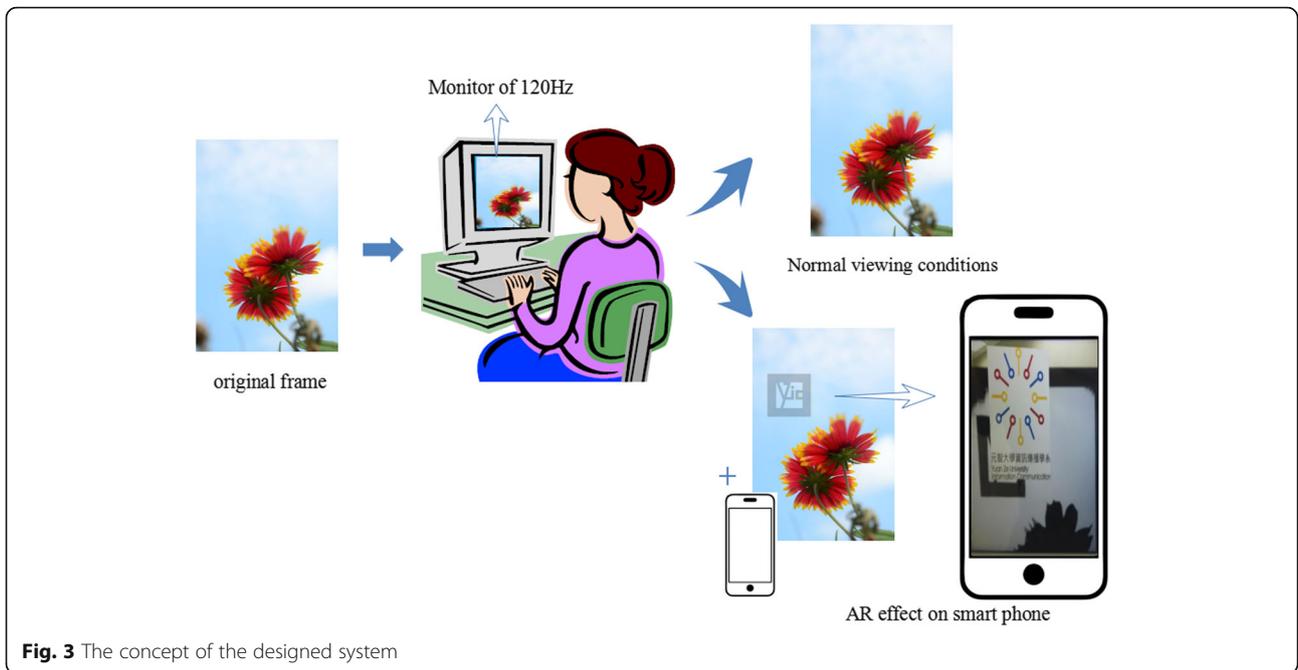


Fig. 3 The concept of the designed system

visual modulation (TPVM) is to satisfy the equations: $A = I - T$ and $B = I + T$, such that the human eyes would form complement images by mixing the images, $A + B = 2I$.

The pixel values of the RGB frames A , B , and I are ranged within $[0, 255]$. While stacking T on A and B , the pixel values of A and B need to be adjusted within the

range $[0, 255]$. However, for most monitors and projectors, the brightness and grayscale are nonlinear relationship [15]. That is, the display results by adjusting the grayscale values and brightness are unacceptable for the human eyes. Figure 2 shows the basic model of the proposed AR with TPVM.

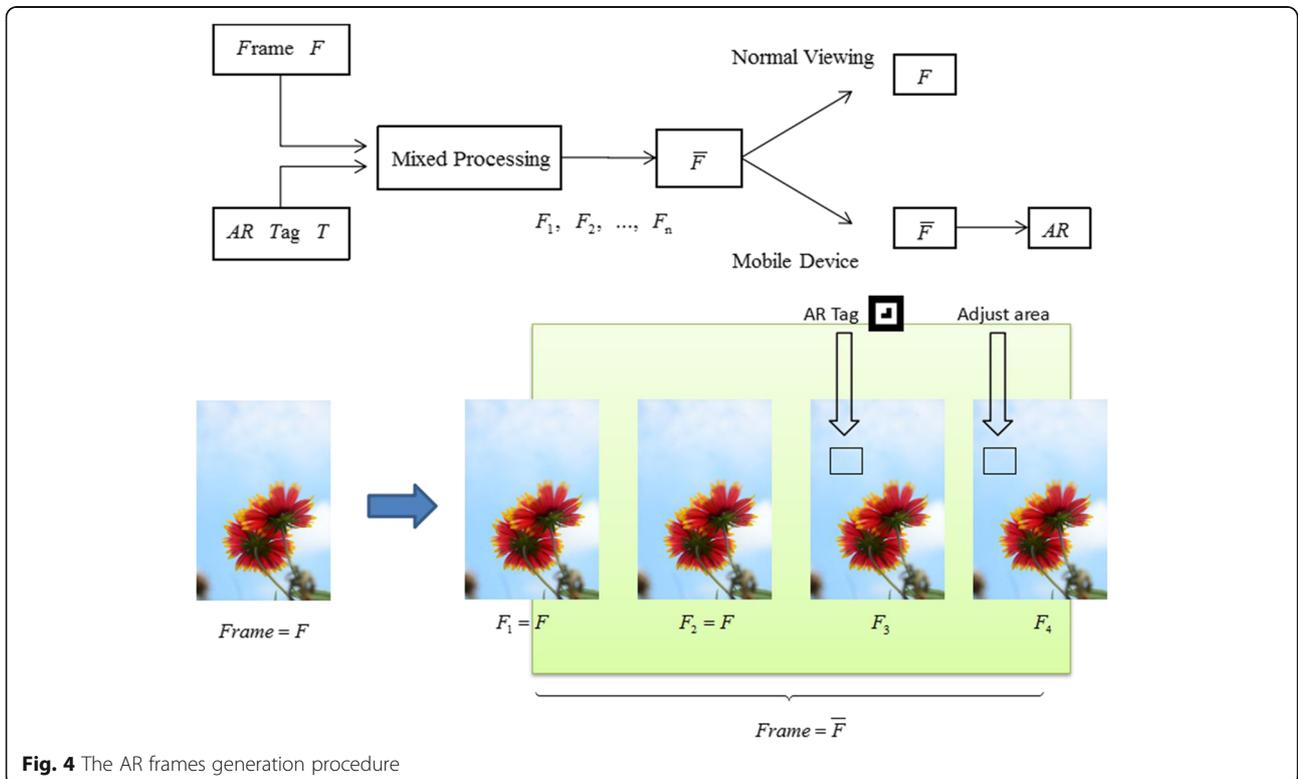


Fig. 4 The AR frames generation procedure

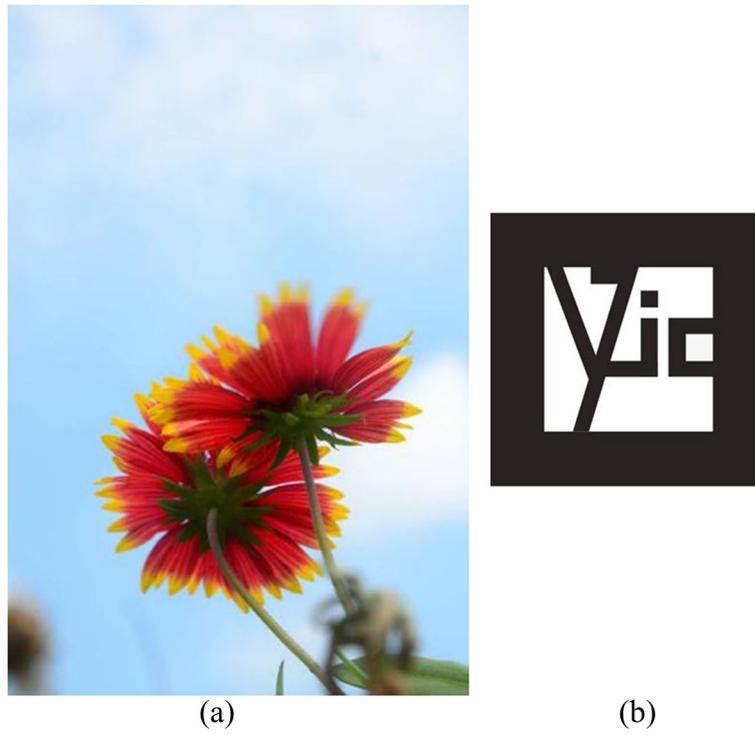


Fig. 5 The test frame F and AR tag T . **a** The test frame. **b** The AR tag

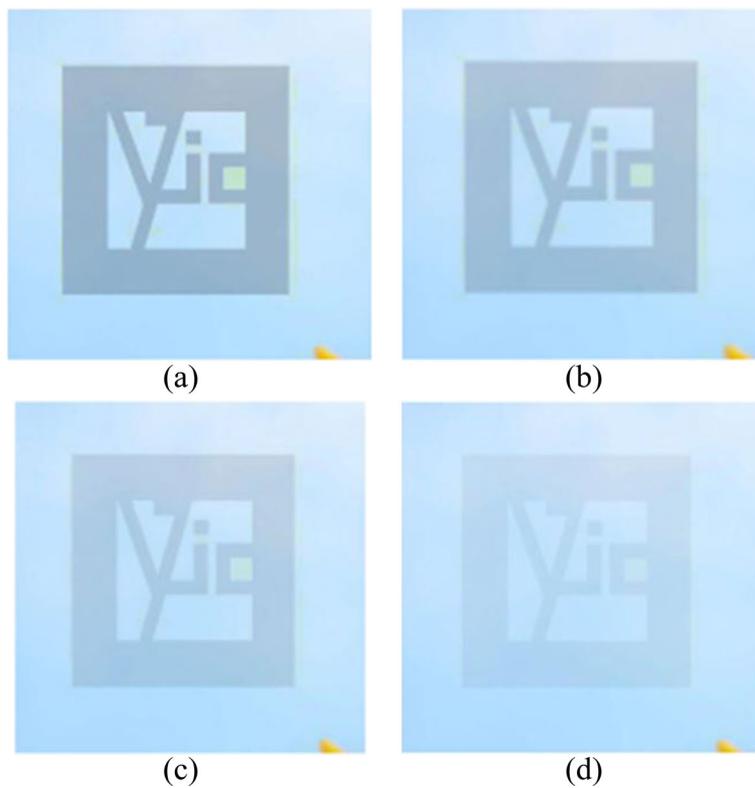


Fig. 6 The marked frame F_3 under different weights. **a** $\omega = 0.8$. **b** $\omega = 0.85$. **c** $\omega = 0.9$. **d** $\omega = 0.95$

We can note that the qualities of the modified frames A and B are unacceptable if the pixel values are manipulated directly. To mitigate the distortion of the modified video via displayer, the new scheme forms two pairs of frames in 120-Hz refresh rate by adding two original frames I with the modified frames A and B . The four frames are exhibited alternately on the displayer in order. This can weaken the sense of residual human vision.

To achieve the imperceptible AR pattern on the video frame, the new scheme aims to adopt the concept of a watermark technology by modifying the frame A corresponding to the tag T . On the contrary, for the frame B , the pixel values are supplemented corresponding to the tag T , for the sake of deriving two superimposed images to achieve the effect as $A + B = 2I$. Normally, the human naked eyes are incapable of realizing the presence of an AR tag with 120-Hz displayer.

3 The designed video with an augmented reality mechanism

In this article, we proposed a specific AR method via TPVM, which is based on different principles between the human eye and the camera sensor. Figure 3 shows the concept of the proposed mechanism. The original frames are displayed with a 120-Hz refresh-rate monitor. Normally, the user can observe the original frames without the AR tag on the monitor. With the mobile phone, the user can further capture the AR tag and thereby exhibit the corresponding AR object on the phone screen. Subsections 3.1 and 3.2 explain the designed procedures of the AR frame generation and the corresponding AR exhibition.

3.1 The AR frame generation procedure

Given a video with 30 frames per second and a 120-Hz monitor/projector, firstly, each frame F is extended to four sub-frames, F_1 , F_2 , F_3 , and F_4 . To stamp the binary AR tag, T , on the frame, the concept of watermarking technology is utilized to process the sub-frame F_3 with T to obtain the marked results \bar{F}_3 by the following steps. Figure 4 demonstrates the AR frame generation procedure.

Step 1: Let F_{3AR} be the selected smoothing region of the sub-frame F_3 .

Step 2: According to the block size of F_{3AR} , the AR tag T is resized to correspond with the same size of F_{3AR} .

Step 3: Let $F_{3AR}(x, y)$ and $T(x, y)$ be the coordinate position (x, y) of F_{3AR} and T , respectively. Here, the value of $T(x, y) = 0$ that represents the value of $T(x, y)$ is black, and $T(x, y) = 1$ that indicates the value of $T(x, y)$ is white. The binary AR tag $T(x, y)$ thereby can be concealed into

the corresponding $F_{3AR}(x, y)$ with a weight ω according to the formula:

$$\bar{F}_{3AR}(x, y) = \begin{cases} \omega \times F_{3AR}(x, y), & \text{if } T(x, y) = 0, \\ F_{3AR}(x, y), & \text{if } T(x, y) = 1. \end{cases} \quad (1)$$

The value of ω is selected by the user's demand and ranged within $[0, 1]$. The marked sub-frame \bar{F}_3 thereby can be derived by containing the result \bar{F}_{3AR} .

Step 4: In order to reduce the appearance of the distorted \bar{F}_{3AR} and to enhance the video effect, the neighboring frame F_4 is adopted. Let $F_{4AR}(x, y)$ be the corresponding position (x, y) of F_{4AR} , and let α be the adjustment value. We use the following formula to complement the region $F_{4AR}(x, y)$ and thereby learn the marked result:

$$\bar{F}_{4AR}(x, y) = \begin{cases} F_{4AR}(x, y) + \alpha, & \text{if } T(x, y) = 0, \\ F_{4AR}(x, y), & \text{if } T(x, y) = 1. \end{cases} \quad (2)$$

Here, the value of α is determined according to the value of ω .

Step 5: The adjustment process in Step 4 may cause overflow situations. To overcome the phenomenon, the



Fig. 7 The derived sub-frame F_4

value of $\bar{F}_{4AR}(x, y)$ is equal to 255, if $\bar{F}_{4AR}(x, y) > 255$. Finally, we can obtain the marked sub-frame \bar{F}_4 which is adjusted.

Step 6: With combining the marked sub-frames \bar{F}_3 and \bar{F}_4 with the original sub-frames F_1 and F_2 , we can derive the marked frame \bar{F} .

According to the above processing, we can conceal the AR tag into each video frame and eventually obtain the marked film with the 120 imperceptible AR frames.

This method combines TPVM with AR technologies to conceal the AR tag and reduce the discomfort of AR sense for the human eyes. When displaying the video with a 120-Hz flicker, the frequency for each sub-frame is set as 30 Hz, it can significantly reduce the residual sense of the AR tag on the human eyes and generate a satisfactory visual quality. Meanwhile, the mobile system can achieve the AR effect while capturing the video frames.

3.2 The AR frame exhibition procedure

Due to the human eyes that hardly recognize and trace each frame while the frequency of display device is higher than 60-Hz flicker, in the exhibition procedure, a display device with up to 120-Hz flicker is utilized in the system.

As the derived video in Subsection 3.1, the visual effect can be achieved similar to the original film without the embedded AR tag. According to the proposed AR frame generation procedure, the AR tag is concealed in a portion of film. To enhance the detection of a tag in the AR system, the mobile device captures four continuous images at a shot in the AR frame exhibition procedure. Let $F'_1, F'_2, F'_3,$ and F'_4 be the four captured images of a mobile phone via 120-Hz display device. The AR effect can be achieved according to the following steps:

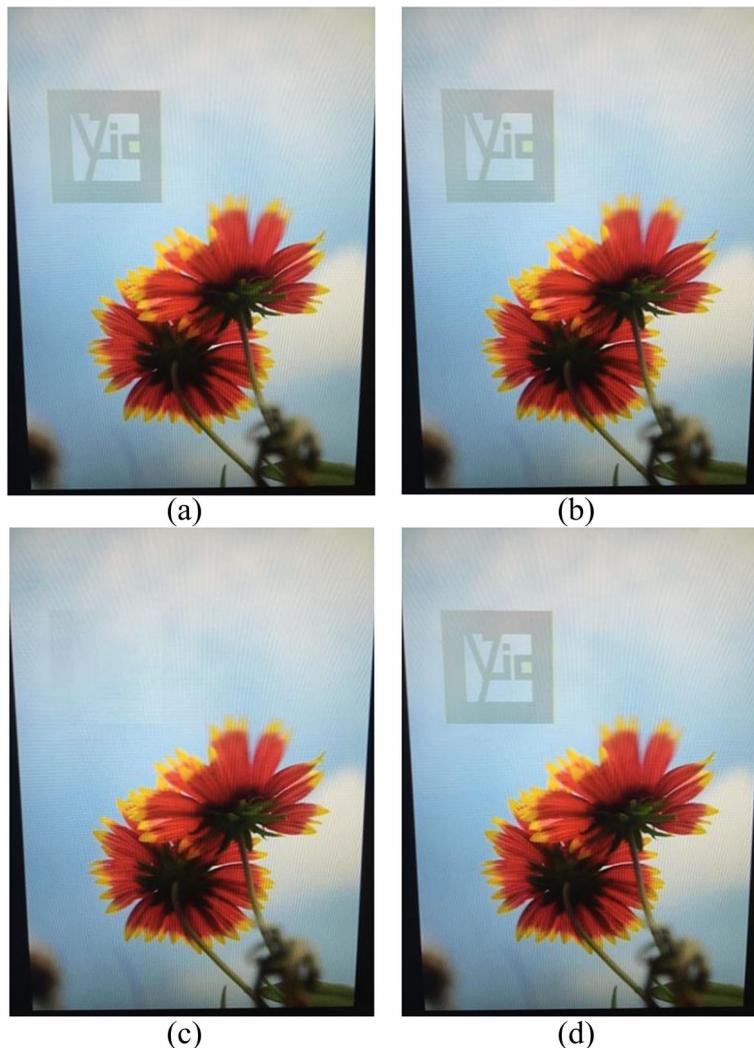


Fig. 8 The captured photos via iPhone6. **a** Photo 1. **b** Photo 2. **c** Photo 3. **d** Photo 4

Step 1: Transfer the four-color images to the corresponding grayscale images using color-to-grayscale image processing.

Step 2: Let $H_1, H_2, H_3,$ and H_4 be the calculated histograms corresponding to the four grayscale images, respectively.

Step 3: According to the histograms $H_i, i = 1, 2, 3,$ and 4, the interval $[a, b]$ of each H_i can be learned by selecting the two peak points in the histogram.

Step 4: To enhance the black and white contrast of the grayscale images and reduce the computational complexity in the AR system, the histogram operation of the gray stretch method [16] is used. Let $F'_i(x, y)$ be the coordinate position (x, y) of the four grayscale images $F'_i, i = 1, 2, 3,$ and 4. The following formula is adopted to stretch the histogram within the region $[0, 255]$:

$$F''_i(x, y) = \frac{255}{b-a} \times (F'_i(x, y) - a), i = 1, 2, 3, 4. \quad (3)$$

Hereafter, the enhanced four grayscale images $F''_i(x, y)$ and the corresponding histograms can be achieved.

Step 5: To binarize the grayscale images $F''_i(x, y), i = 1, 2, 3,$ and 4, let L be the lowest point in the interval $[a, b]$ of the corresponding histogram. The value of L is set to be the binary threshold and used to binarize the four grayscale images $F''_i(x, y)$ to the corresponding binary images \tilde{F}_i [16], $i = 1, 2, 3,$ and 4.

$$\tilde{F}_i(x, y) = \begin{cases} 0, & \text{if } F''_i(x, y) \leq L, \\ 1, & \text{if } F''_i(x, y) > L. \end{cases} \quad (4)$$

Here, $\tilde{F}_i(x, y)$ be the coordinate position (x, y) of the generated binary image \tilde{F}_i corresponding to the four grayscale images $F''_i(x, y), i = 1, 2, 3,$ and 4.

Step 6: With the generated binary image \tilde{F}_i , a mobile device thereby can scan and recognize the enhanced AR tag in the AR system. The required AR effect subsequently can be exhibited on mobiles.

4 Experimental results and discussions

To evaluate the proposed mechanism, the simulation frame and the AR tag in the AR system are shown in

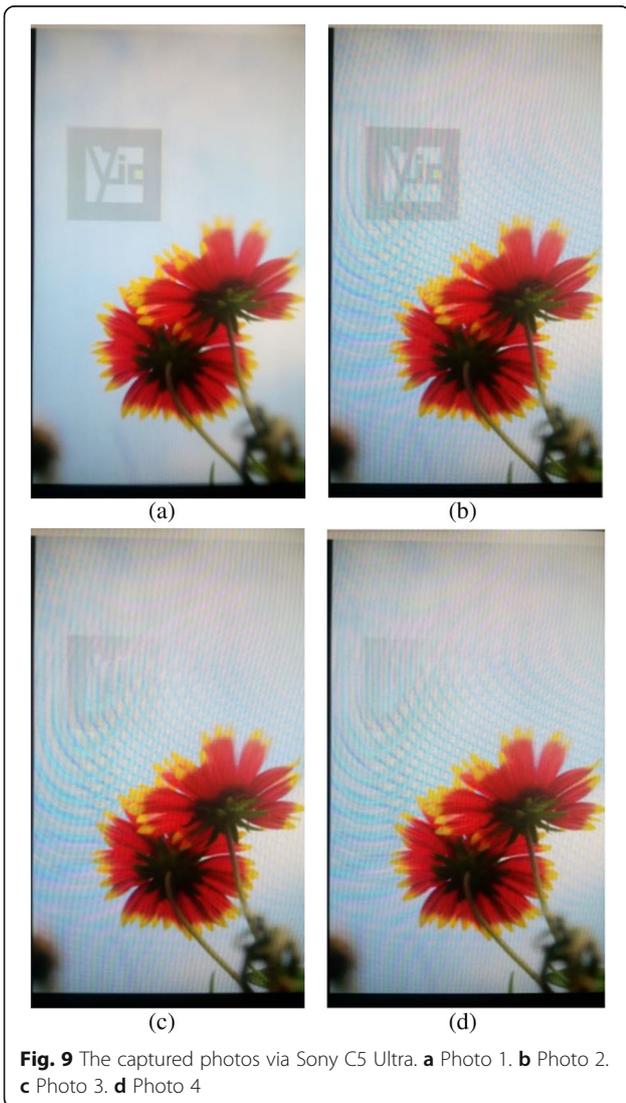




Fig. 11 The contrast-enhanced image of Fig. 10

Fig. 5a, b. According to the AR frame generation procedure, Fig. 5a is extended to four sub-frames, F_1 , F_2 , F_3 , and F_4 . The largest smooth area of the third sub-frame F_3 can be determined by image processing. Figure 5b thereby can be resized and embedded into the smooth

area of F_3 . Here, the weight ω is adjusted to derive acceptable visual quality of the marked frame by Eq. (1).

The weight ω in Eq. (1) can be adjusted for the sake of preserving the visual quality of the original frame. To achieve higher quality of the marked frame, the setting value of ω can be increased. On the contrary, the setting value of ω can be decreased to achieve clear visibility of the AR tag from the display monitor.

Figure 6 displays the generated marked results for the third sub-frame under the weights 0.8, 0.85, 0.9, and 0.95. In general, the quality of the marked F_3 is higher with larger value of ω . That is, the distortion between the original F_3 and the marked F_3 is small. However, the captured AR tag of the marked F_3 is susceptible and affected by the surrounding environment. The mobile device may be incapable of scanning and recognizing the AR tag from the marked F_3 .

On the contrary, the visual quality of the marked F_3 is lower with smaller weight value of ω . The mobile device can effectively capture the AR tag and demonstrate the AR model. According to the simulations, the value of ω is suggested to be ranged within [0.8, 0.9] for balancing the visual perception of the marked sub-frame and the recognition of the AR tag. Moreover, to balance the quality of the marked frame and the visibility of the AR tag, the weight $\omega = 0.85$ can normally achieve satisfactory result.

Figure 7 shows the complemented sub-frame F_4 with $\alpha = 10$. The modified sub-frame F_4 can improve the appearance of the distorted AR tag of the marked F_3 while displaying on a 120-Hz monitor. The human eyes can obviously sense the contrast difference between the modified F_3 and F_4 while the value of α is large than 20. According to the result, the value of α is suggested

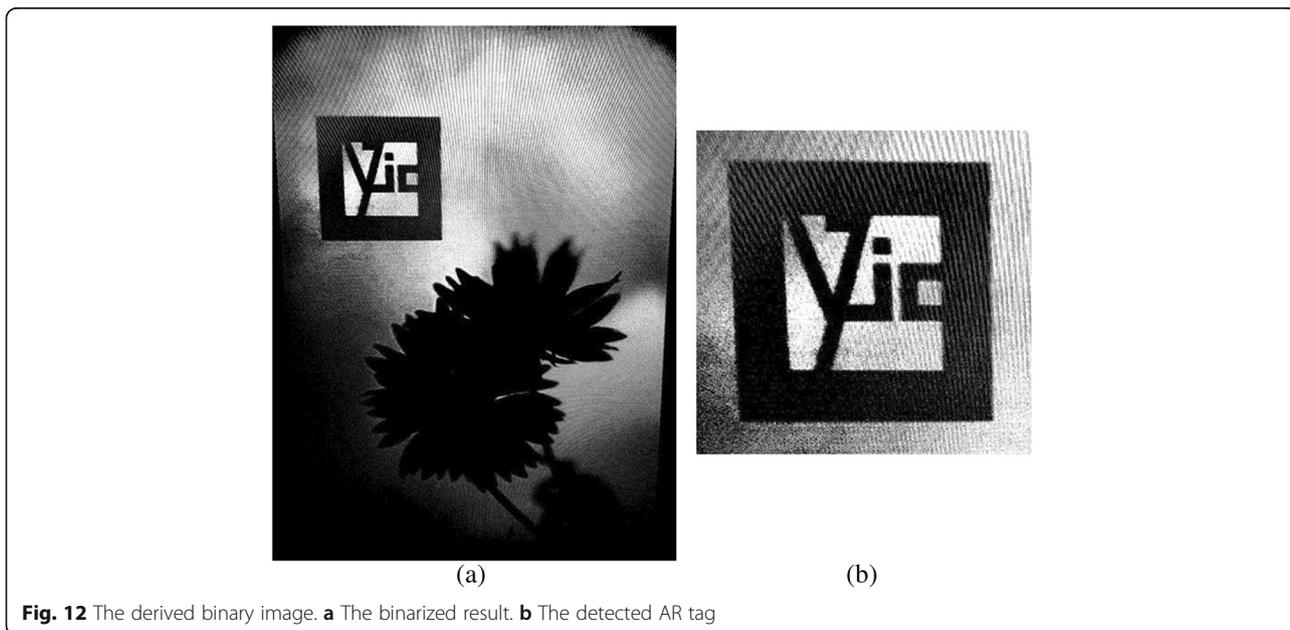


Fig. 12 The derived binary image. **a** The binarized result. **b** The detected AR tag

around ten for the sake of mitigating the perception and contrast between the modified F_3 and F_4 .

To demonstrate the feasibility of the proposed system, two different mobile devices, iPhone 6 and Sony C5 Ultra, are used in the simulation. The mobile devices take four images continuously while the modified film is displayed on a 120-Hz monitor. Figures 8 and 9 exhibit the four captured images by iPhone 6 and Sony C5 Ultra, respectively. We can observe that the modified frame with the AR tag can be captured when shooting four continuous images via mobile device.

The post-processing of the captured image via mobile devices can enhance the quality of the AR tag. To achieve recognizable AR tag, we first transform the captured color image to a binary image. Afterward, the de-noising and histogram operations are applied to help distinguish the white pixels and the black pixels from the binary image.

Figure 10 shows the transformed grayscale image corresponding to the captured color image by iPhone 6.

According to the AR frame exhibition procedure, the contrast of Fig. 10 can be enhanced as shown in Fig. 11. Figure 12a demonstrates the derived binary image by stretching and binarizing the histogram of Fig. 11. Figure 12b displays the detected AR tag in the AR system.

For the Sony mobile device, Figs. 13 and 14 exhibit the grayscale image and the corresponding enhanced result. Figure 15a shows the derived binary image. The detected AR tag of Fig. 15a is shown in Fig. 15b. Consequently, the mobile devices can link the AR tag and exhibit the AR object on the mobile devices. Figure 16 presents the AR effect captured by the mobile device.

5 Conclusion and future work

This paper designed a new AR mechanism that conceals the imperceptible AR tag into the video frame based on the concept of temporal psycho-visual modulation (TPVM). One can watch the marked video normally via a 120-Hz monitor without observing the embedded AR



Fig. 13 The processed grayscale image (Sony)



Fig. 14 The contrast-enhanced image of Fig. 13

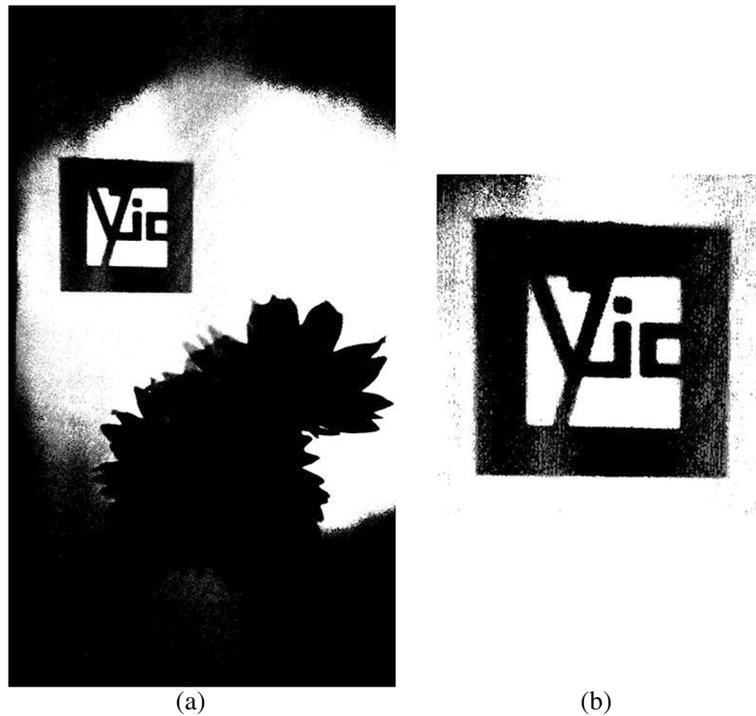


Fig. 15 The derived binary image. **a** The binarized result. **b** The detected AR tag

tag. With using a mobile device, the AR tag can be extracted and recognized due to the semiconductor camera sensors that can capture the high-frequency frames. The proposed system is feasible and can be applied in various AR application and entertainment. In

the future works, the post-image processing of the AR tag area, such as the noise elimination and the optimize binarization, can be considered to improve the quality of the marked frames and the identification of the AR tag in the AR system.

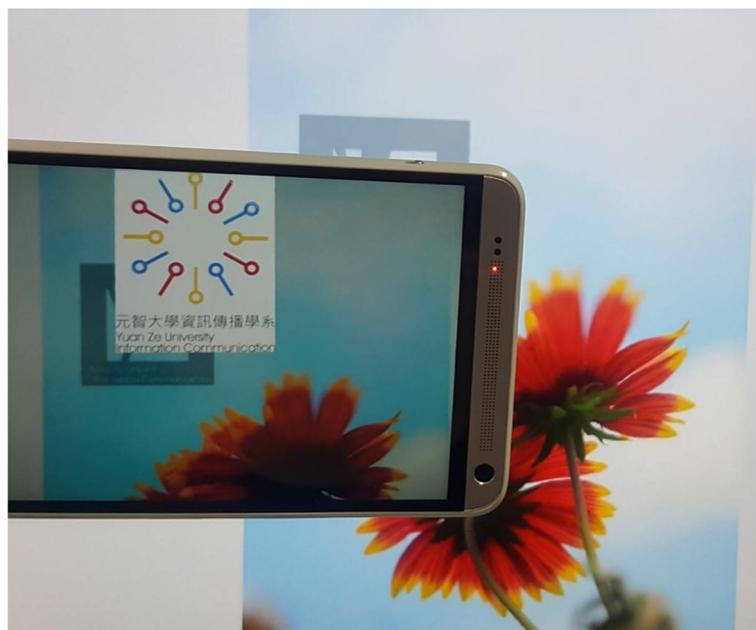


Fig. 16 The demonstration of the AR model

Acknowledgements

This research was supported by the Ministry of Science and Technology, Taiwan. The authors would like to thank the anonymous reviewers for their valuable comments to improve the quality of this work.

Funding

This research was supported by the Ministry of Science and Technology, Taiwan, and acknowledges funding from the contract No. MOST 105-2221-E-155-048 and MOST 105-2218-E-155-010.

Authors' contributions

BY carried out the main part of this manuscript. PL conceived of the study, participated in its design and coordination, and helped to draft the manuscript. JL participated in the discussion and corrected the English errors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Information Communication, and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Chung-Li 32003, Taiwan.

²School of Software Engineering, Nanchang University, Nanchang, Jiangxi, China.

Received: 2 June 2016 Accepted: 22 December 2016

Published online: 19 January 2017

References

1. R.T. Azuma, A survey of augmented reality. *Presence Teleop. Virt.* **6**(4), 355–385 (1997)
2. T. Nikolaos, T. Kiyoshi, QR code calibration for mobile augmented reality applications: Linking a unique physical location to the digital world, *ACM SIGGRAPH 2010*, no. 144, 2010
3. J. Carmigniani, B. Furht, Augmented reality: an overview. *Handbook of Augmented Reality Chapter 1*, 3–46 (2011)
4. T.H. Tsai, W.H. Cheng, C.W. You, M.C. Hu, A.W. Tsui, H.Y. Chi, *Learning and recognition of on-premise signs from weakly labeled street view images*, vol. 23, no. 3, pp. 1047–1059, 2014
5. M.C. Hu, C.W. Chen, W.H. Cheng, C.H. Chang, J.H. Lai, J.L. Wu, Real-time human movement retrieval and assessment with kinect sensor. *IEEE Transactions on Cybernetics* **45**(4), 742–753 (2015)
6. W.C. Jhou, W.H. Cheng, Animating still landscape photographs through cloud motion creation. *IEEE Transactions on Multimedia* **18**(1), 4–13 (2016)
7. C.H. Hsu, Y.L. Wu, W.H. Cheng, Y.J. Chen, K.L. Hua, HoloTube: a low-cost portable 360-degree interactive autostereoscopic display. *Multimedia Tools and Applications* (2016). doi:10.1007/s11042-016-3502-3
8. C.H. Hsu, W.H. Cheng, K.L. Hua, HoloTabletop: an anamorphic illusion interactive holographic-like tabletop system. *Multimedia Tools and Applications* (2016). doi:10.1007/s11042-016-3531-y
9. P.Y. Lin, C.H. Teng, Diverse augmented Reality exhibitions for differential users based upon private quick response code, *Asia-Pacific Signal and Information Processing Association (APSIPA ASC 2015)*, pp. 1121–1125, Hong Kong, 16–19 December, 2015
10. X. Wu, G. Zhai, Temporal psychovisual modulation: a new paradigm of information display. *IEEE Signal Process. Mag.* **30**(1), 136–141 (2013)
11. H. Qi, D. Zheng, J. Zhao, Human visual system based adaptive digital image watermarking. *Signal Process.* **88**(1), 174–188 (2008)
12. X. Lu, B. You, P.Y. Lin, Augmented reality via temporal psycho-visual modulation, *2016 IEEE International Conference on Multimedia & Expo (ICME 2016)*, Seattle, USA, 11–15 July, 2016
13. S.A. Taylor, CCD and CMOS imaging array technologies: Technology review, *Technical Report EPC-1998-106*, Xerox Research Centre Europe, pp. 1–14, 1998.
14. C. Hu, G. Zhai, Z. Gao, X. Min, Information security display system based on spatial psychovisual modulation, *2014 IEEE International Conference on Multimedia & Expo (ICME 2014)*, pp. 1–4, 2014
15. G. Zhai, X. Wu, Defeating camcorder piracy by temporal psychovisual modulation. *J. Disp. Technol.* **10**(9), 754–757 (2014)
16. R.C. Gonzalez, R.E. Woods, "Digital Image Processing," Prentice Hall, 2008

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com