The development of an augmented reality (AR) approach to mammographic training: overcoming some real world challenges

Qiang Tang, Yan Chen, Gerald Schaefer, Alastair G. Gale

SPIE.

Development of an Augmented Reality Approach to Mammographic Training: Overcoming some Real World Challenges

Qiang Tang[a], Yan Chen[*a], Gerald Schaefer[b], and Alastair G. Gale[a]

[a]Applied Vision Research Centre, Loughborough University, Loughborough, UK
[b]Department of Computer Science, Loughborough University, Loughborough, UK

ABSTRACT

A dedicated workstation and its corresponding viewing software are essential requirements in breast screener training. A major challenge of developing further generic screener training technology (in particular, for mammographic interpretation training) is that high-resolution radiological images are required to be displayed on dedicated workstations whilst actual reporting of the images is generally completed on individual standard workstations. Due to commercial reasons, dedicated clinical workstations manufactured by leading international vendors tend not to have critical technical aspects divulged which would facilitate further integration of third party generic screener training technology. With standard workstations, it is noticeable that the conventional screener training depends highly on manual transcription so that traditional training methods can potentially be deficient in terms of real-time feedback and interaction. Augmented reality (AR) provides the ability to co-operate with both real and virtual environments, and therefore can supplement conventional training with virtual registered objects and actions. As a result, realistic screener training can co-operate with rich feedback and interaction in real time. Previous work[1] has shown that it is feasible to employ an AR approach to deliver workstation-independent radiological screening training by superimposing appropriate feedback coupled with the use of interaction interfaces. The previous study addressed presence issues and provided an AR recognisable stylus which allowed for drawing interaction. As a follow-up, this study extends the AR method and investigates realistic effects and the impacts of environmental illumination, application performance and transcription. A robust stylus calibration method is introduced to address environmental changes over time. Moreover, this work introduces a completed AR workflow which allows real time recording, computer analysable training data and further recoverable transcription during post-training. A quantitative evaluation results show an accuracy of more than 80% of user-drawn points being located within a pixel distance of 5.

Keywords: AR, augmented reality, mammographic interpretation, training

1. INTRODUCTION

Breast cancer is one of the most common cancers that lead to death. As part of the NHS Breast Screening Programme in the UK, screener training plays a critical role in quality assurance. There are three primary breast cancer screening training formats in the UK: multi-disciplinary meetings (MDTs); interval cancer reviews and the PERFORMS® self-assessment scheme.[2] The first two are held in screening centres whereas the latter is a web-based training service which is remotely accessible.[3] As a biennial national self-assessment scheme which requires participants to interpret recent known difficult cases, PERFORMS®[4] has proven to be effective in maintaining radiological screening skills over more than 30 years of practice.

Traditional mammographic screening training depends highly on dedicated high-resolution workstations and their specific viewing software which, therefore, ensure sophisticated radiological image presence. The main barrier to develop further generic screener training technology (particularly for mammographic training) is the broad existence of commercial technological isolation. Leading international manufacturers of high-resolution radiological workstations keep their commercial edge by not divulging key technological aspects. Moreover, dedicated workstations and their standard DICOM viewing software are the only options preferred by radiologists

---

*Y.Chen@lboro.ac.uk

in the UK, even though both standard and alternative DICOM viewing software support interactions. As a result, radiologists use standard DICOM viewing software during image inspection but the actual screening reporting and feedback are completed on another device. To retain the confidence of participants during training, appropriate interaction and feedback is a prerequisite even though it is possible to complete training on various displays.[2] Therefore, developing a generic low-cost screener training technology has become an emerging challenge. At the same time, the value of traceable screener training data has also become important. Accurate transcription and recoverable training results is a particular requirement for mammographic training over time.

Augmented reality (AR) is a supplement of either the virtual or the real world. Therefore, AR can be a superior method of screener training which can: (a) supply image-associated information and feedback; (b) provide rich device-independent interactions; (c) enable automation of screener training workflows with fully automatic data acquisition, processing and analysis in real time. A previous study[1] has developed an AR prototyping approach which can indicate suspicious locations of feature types of an existing training database that contains known image-associated pathology descriptions as well as expert-annotated markers. As a follow-up, this study extends the AR approach with four aspects addressing realistic challenges during screener training: (a) dataflow and transcription; (b) interaction; (c) data storage; and (d) application performance.

2. METHOD

2.1 Dataflow and transcription

Conventional mammographic training relies on separate enclosed systems where only manual input (including transcription) and reporting results are allowed. Such an enclosed system has no integration with any dedicated workstation so that it requires individual interaction and feedback procedures. For the AR approach, a completely new workstation-independent dataflow is introduced in this study. At the beginning of reading each DICOM case, any existing historical training data (e.g. DICOM case associated description or annotations) will also be loaded. Every user manipulation will trigger a recording operation during the AR training and give options to select if the current training image is normal or abnormal. The final decisions that constitute historical training data for the next training session are saved with associated DICOM case ID and timestamps. As a result, each manipulation is recorded in real time and can be reloaded in subsequent training sessions if any interruption occurs. As indicated in Figure 1, the AR training approach has a cyclic dataflow that integrates interaction and feedback with the dedicated workstation, whereas conventional training methods use forward-directional dataflow that has no integration (therefore, traditional methods require an individual standard workstation for reporting and presenting feedback). Feedback (e.g., the yellow coloured circle and its associated annotation) are only visible via the AR headset.

In the process of the AR dataflow, only DICOM ID associated masks (training data except for the original DICOM cases) are required to be saved so that feedback can be re-superimposed on any type of dedicated workstation.

2.2 Interaction

It is very challenging to develop accurate recognition, compensation and registration for objects with irregular shapes under variable illumination conditions[5,6,7] Comparatively, there are more complicated issues for mammographic training scenes, in particular, for some aspects. Previous research[1] has established the feasibility of employing an AR approach during mammographic training and developed an AR recognisable handheld stylus whose input accuracy was within ± 5 pixels under appropriate visual cues guide. Nevertheless, there are complex challenging variances which can prevent the recognisable stylus performing as a real pencil in the real world. These variances can be, for instances, personal drawing skills, room illumination (e.g. moving persons, bulb flickers or light changes over time), reflections on the dedicated workstation and noise in images captured by the AR camera. Also, since mammographic training normally is carried in a dark environment, the confidence of detection and registration of the styles is reduced, while shape detection and registration are not always consistent with objects in the real world. Therefore, appropriate compensation is necessary for computer vision-based detection in practice.

The presented AR approach employs a series of calibration points distributed in the training scene to allow the stylus being used like a real pencil and to do so accurately. While calibration points can be randomly generated

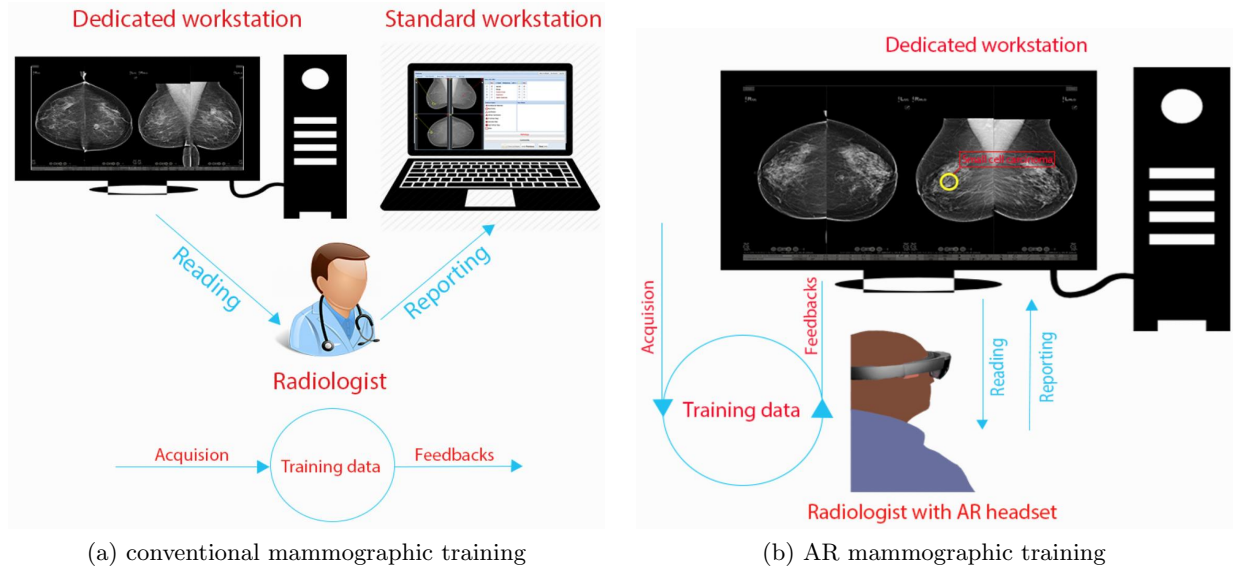(a) conventional mammographic training   (b) AR mammographic training

Figure 1. Conventional training vs. AR training.

and evenly distributed, for simplification the monitor screen corners are used. The size and orientation of the stylus can both change while illumination and the stylus depth can change over time in a 3-dimensional space. As a result, these calibration points can be treated as seeding factors so that appropriate size and orientation compensation changes for the stylus can be calculated based on the obtained displacements in terms of position and orientation between the captured stylus and the defined calibration points.

With appropriate calibration, the impact of lighting and depth changes are reduced and allow for an accurate match between the virtual and the real stylus. The calculated results are associated with gradual changes of illumination and depth changes, although sudden sharp illumination changes remain a challenge (on the other hand, these are practically impossible to find in a screening scene and are therefore excluded in the consideration of realistic mammographic training). Furthermore, the depth information of the stylus can be retrieved in the viewpoint of the camera with this method.

Additionally, an overlaid virtual menu, show in Figure 2, at the bottom of a dedicated workstation monitor has been introduced to allow various operations. This virtual menu can accept position-triggered input and currently supports to quickly indicate if a DICOM case is normal or abnormal for experimental purposes.

Figure 2. The virtual menu allows quick decisions: Y indicates the case is normal, N indicates it is abnormal.
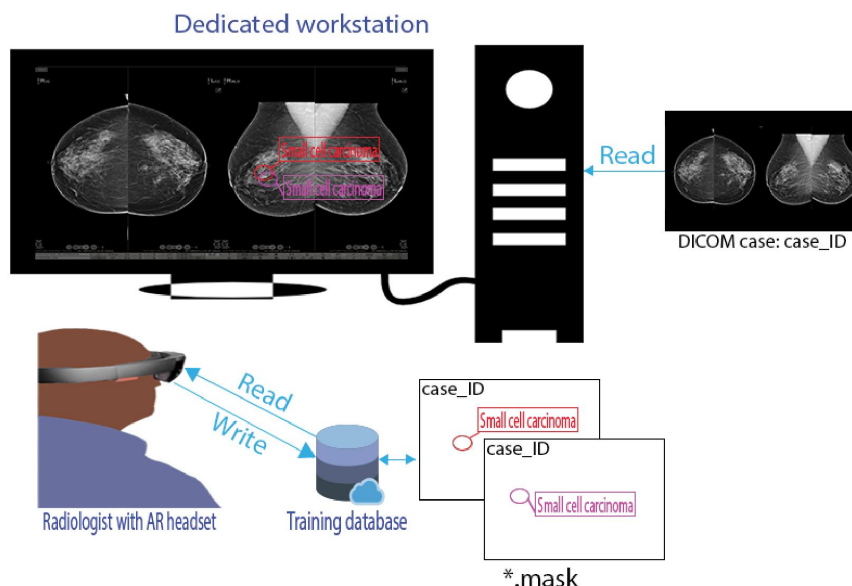
Figure 3. The AR storage method keeps DICOM case ID associated marks and annotations separately (as *.mask files in the training database) from the dedicated workstation that keeps their corresponding DICOM cases.

2.3 Data storage

The improved AR approach allows different roles (expert, trainee, etc.) to register and log in. Decisions are saved separately according to the different roles and are completely independent of the original radiological images. Therefore, position-associated annotations and markers made on a workstation can be remapped on another different type of workstation. Moreover, training decisions are immediately comparable among different roles which are considered to be important in improving radiologist skills. Consequently, AR training is only case ID associated and historical training is available on any kind of workstation (including tablets and laptops). As indicated in Figure 3, the expert data (red coloured) and trainee data (purple coloured) are saved in an individual training database which is fully independent of the dedicated workstation.

2.4 Application performance

It is well known that computer vision for AR heavily relies on hardware performance and the corresponding algorithms. In particular, the AR mammographic training approach employs a visual-based interaction method (namely a handheld recognisable stylus) and requires complex object recognition and tracking. For improved efficiency, the AR approach has thus been redesigned based on C++ and CUDA (a parallel computing platform and programming model) to take advantage of hardware accelerated parallel computing and to provide acceptable performance. Consequently, the implemented vision algorithms can be executed on a conventional laptop with a dedicated graphics processor and allow processing at a speed of more than 20 frames per second depending on the complexity of surroundings. Even though there are many further possibilities to improve efficiency, it has been shown that AR-based mammographic training is practically feasible and has the ability to enable real time interaction and feedback.

3. RESULTS

Previous work[1] has established an accuracy of ± 5 pixels for visual cues guided input (based on identifying individual random points). The follow-up experiment evaluated the motion performance of the visual-based interaction method to annotate different DICOM cases. The experiment consisted of a set of 10 real DICOM cases. Three participants were required to log in under appropriate roles and to examine every case. The decisions for each case were given by annotating suspicious areas and rating the corresponding case as normal or abnormal (using the virtual menu). For each participant, there were 2 extra DICOM cases to acquaint them with the AR training method.
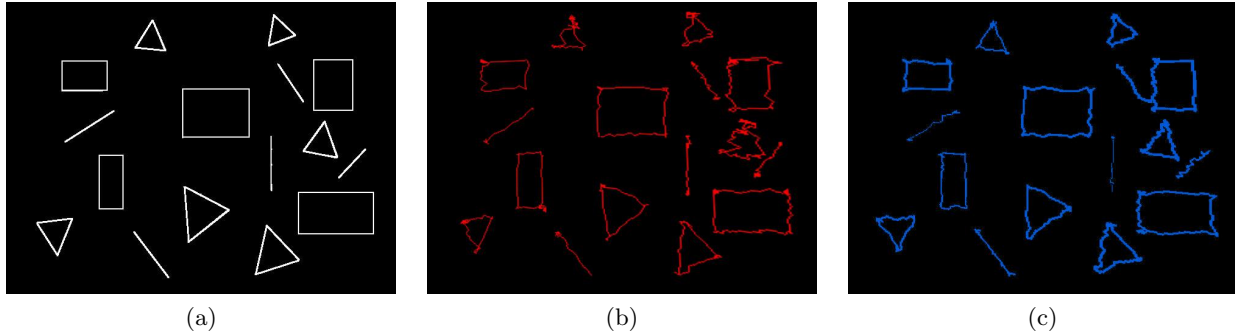
Figure 4. Accuracy evaluation experiment: (a) template DICOM case, (b) results by first role, (c) results by second role.

Moreover, a template DICOM case, shown in Figure 4(a), containing different shapes was displayed on the dedicated workstation to examine the accuracy of workstation displayed markings. A participant was required to re-draw each shape on the template case displayed on the workstation. The template and the user-drawn annotations were then compared based on an appropriate metric.

Clearly, a user-drawn shape highly depends on the user's drawing skills and experience, and a hand-drawn shape and its corresponding template shape are not necessarily the same, either in terms of its vertices, lines or size. As a Hausdorff metric[8] or other simple edge detection evaluation method are less suitable, Pratt's Figure of Merit[9] was used. The Figure of Merit between the template and hand-drawn annotation is calculated as

$$R = \frac{1}{\max(I_I, I_A)} \sum_{i=1}^{I_A} \frac{1}{ad^2}, \tag{1}$$

where I_I is the number of template DICOM case edge points, I_A is number of hand-drawn edge points, d indicates the displacement (based on Euclidean distance) between template and user-drawn edge points, and a is scaling constant ($a = 1/9$ in the experiments).

Taking Figure 4(a) as the template, Pratt's Figure of Merit scores between 0.454 and 0.461 were obtained for comparing hand-drawn annotations to the original template, while scores between 0.766 and 0.882 were obtained when swapping template and user points in the calculation.

threshold	% of points within threshold
1	66% − 92%
5	81% − 96%

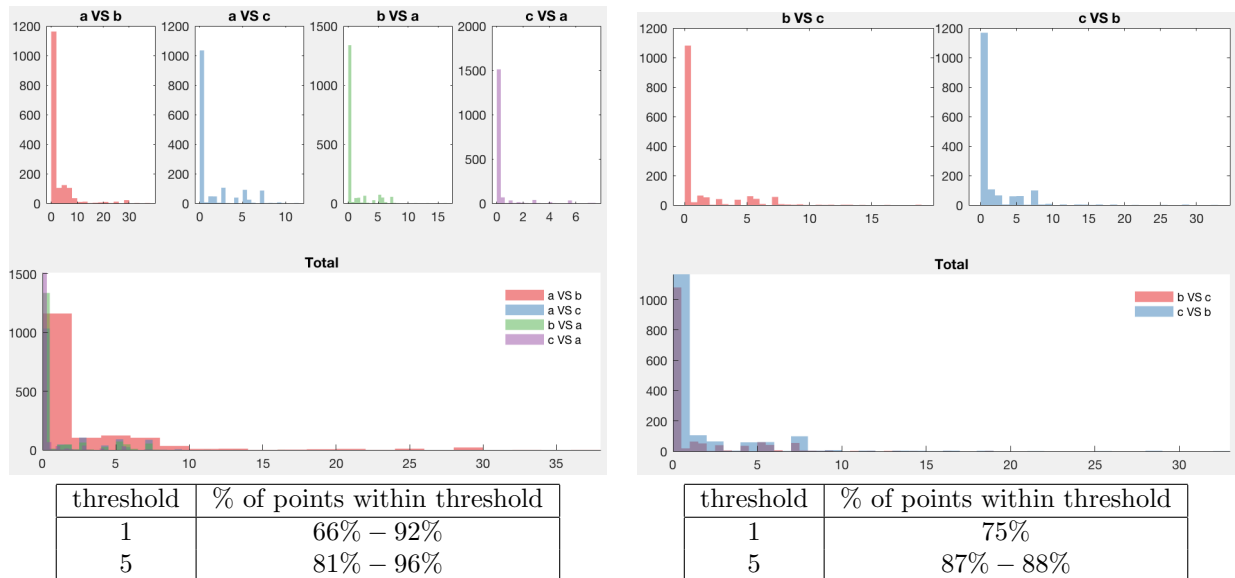| threshold	% of points within threshold
1	75%
5	87% − 88%

Figure 5. Displacement threshold results. Left: user vs. template; right: user vs. user.
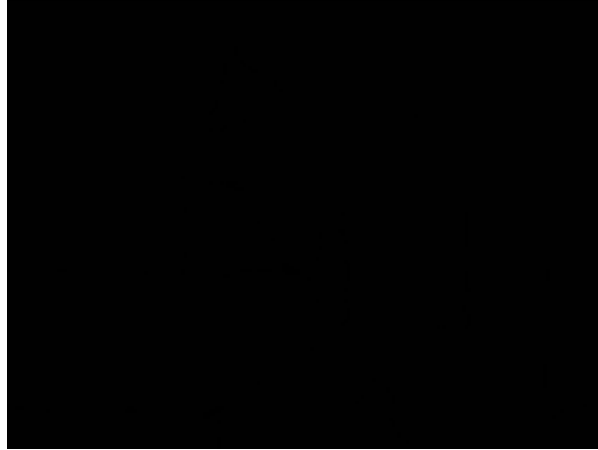
Figure 6. Overlay of DICOM template and two user annotations.

In addition, displacement thresholds were considered to evaluate the accuracy of user-drawn annotations with points within a threshold distance considered as accurate. The results are given in Figure 5 both in terms of comparing user annotations to the template and in terms of comparing the consistency between users. When taking Figure 4(a) as the template, 66% to 92% of the points are within a displacement distance of 1 and 81% to 96% of the points within a displacement of 5. Comparing the annotations from two users (roles), 75% of the points have a displacement of at most 1, while 87% to 88% exhibit a displacement of at most 5.

In addition, Figure 6 shows an overlay of the template and the two user-drawn markings.

4. DISCUSSION

Clearly, hardware performance and cost are still critical factors when developing a practical usable AR mammographic training method. Even though the current AR training system runs at about 20 frames/second after being migrated to C++ and CUDA, the computationally intensive tasks including data acquisition, image/video processing, and post-processing depend on appropriate hardware specifications and thus cost. However, there is also still a lot of room to improve the existing algorithms and data processing strategies so that even a conventional laptop can be employed for the AR training method. The fully workstation-independent dataflow has the ability to reproduce immersive training experience on any kind of workstation. The AR method has the ability to digitally record and distinguish every single operation so that each breast cancer feature in a case can be analysed individually and automatically. The AR training dataflow integrates training data acquisition, processing and analysis in an accumulative way and furthermore has shown that an AR training method[1] can also be automatic.

However, the complexity of realistic training scenarios can constrain the correctness and accuracy of the AR training method without appropriate calibration or compensation. It is a challenge to eliminate or reduce the impact of real world challenges, in particular for occasional situations. In the case of non-sudden environment changes, the AR training method ensures that more than 80% of user-drawn points are located within a pixel distance of 5. Since the employed metrics do not distinguish between interior and exterior displacements, if the radius of a suspicious area (assuming the area is circular) is within a distance of 5, then the centre point of the area and a concentric circle with a radius of 10 can lead to the same scores. While this is an extreme example, better metrics might be necessary as the size of a suspicious area is an important attribute.

Moreover, an AR method is naturally affiliative with location-associated interaction methods. By integrating an overlaid virtual menu, it provides a new quick decision method by which its location is able to be signalled.

5. CONCLUSIONS

Even though the AR method relies on appropriate hardware, it has been shown that a normal dedicated laptop can perform the necessary computations. However, there are still significant real world challenges due to the

complexity of mammographic training scenarios. Considering the uncertainty of drawing, for the current AR method more than 80% of user-drawn points are located within a pixel distance of 5 while the improved interaction method proves to be effective. At the same time, an overlaid virtual menu is practically effective to support quick decisions by the user. By providing a cyclic dataflow (which can be furthermore expanded as an AR training framework and integrate automation), the AR training can be accumulated and fully independent of any kind of workstation so that it can facilitate situated learning in screening education in the long term.

REFERENCES

[1] Tang, Q., Chen, Y., and Gale, A. G., "The implementation of an AR (augmented reality) approach to support mammographic interpretation training: an initial feasibility study," in [*SPIE Medical Imaging*], Kupinski, M. A. and Nishikawa, R. M., eds., 1013604–1013604–8 (2017).

[2] Chen, Y., Gale, A. G., and Scott, H., "Mammographic interpretation training in the UK: current difficulties and future outlook," in [*SPIE Medical Imaging*], Sahiner, B. and Manning, D. J., eds., 72631C–12 (2009).

[3] Gale, A. and Chen, Y., "Radiological informatics in breast screening: Towards improved radiologists' performance," in [*5th International Conference on Biomedical Engineering and Informatics*], 1104–1107 (2012).

[4] Gale, A. G., "PERFORMS: a self-assessment scheme for radiologists in breast screening," *Seminars in Breast Disease* **6**(3), 148–152 (2003).

[5] Liu, S.-C., Fu, C.-W., and Chang, S., "Statistical change detection with moments under time-varying illumination," *IEEE Transactions on Image Processing* **7**(9), 1258–1268 (1998).

[6] Martinkauppi, B., Soriano, M., and Pietikainen, M., "Detection of skin color under changing illumination: a comparative study," in [*12th International Conference on Image Analysis and Processing*], 652–657 (2003).

[7] Belhumeur, P. N. and Kriegman, D. J., "What is the set of images of an object under all possible illumination conditions?," *International Journal of Computer Vision* (1998).

[8] Bossomaier, T. and Loeff, A., "Parallel computation of the hausdorff distance between images," *Parallel Computing* **19**(10), 1129–1140 (1993).

[9] Abdou, I. E. and Pratt, W. K., "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proceedings of the IEEE* **67**(5), 753–763 (1979).