Applying Saliency-Based Region of Interest Detection in Developing a Collaborative Active Learning System with Augmented Reality

Trung-Nghia Le[1,2], Yen-Thanh Le[1], and Minh-Triet Tran[1]

[1] University of Science, VNU-HCM, Ho Chi Minh city, Vietnam
[2] John von Neumann Institute, VNU-HCM, Ho Chi Minh city, Vietnam
nghia.le.ict@jvn.edu.vn, lythanh@apcs.vn,
tmtriet@fit.hcmus.edu.vn

Abstract. Learning activities are not necessary to be only in traditional physical classrooms but can also be set up in virtual environment. Therefore the authors propose a novel augmented reality system to organize a class supporting real-time collaboration and active interaction between educators and learners. A pre-processing phase is integrated into a visual search engine, the heart of our system, to recognize printed materials with low computational cost and high accuracy. The authors also propose a simple yet efficient visual saliency estimation technique based on regional contrast is developed to quickly filter out low informative regions in printed materials. This technique not only reduces unnecessary computational cost of keypoint descriptors but also increases robustness and accuracy of visual object recognition. Our experimental results show that the whole visual object recognition process can be speed up 19 times and the accuracy can increase up to 22%. Furthermore, this pre-processing stage is independent of the choice of features and matching model in a general process. Therefore it can be used to boost the performance of existing systems into real-time manner.

Keywords: Smart Education, Active Learning, Visual Search, Saliency Image, Human-Computer Interaction.

1 Introduction

Skills for the 21st century require active learning which focuses on the responsibility of learning on learners [1] by stimulating the enthusiasm and involvement of learners in various activities. As learning activities are no longer limited in traditional physical classrooms but can be realized in virtual environment [2], we propose a new system with interaction via Augmented Reality (AR) to enhance the attractiveness and collaboration for learners and educators in virtual environment. To develop a novel AR system for education, we focus on the following two criteria as the main guidelines to design our proposed system, including real-time collaboration and interaction, and naturalness of user experience.

R. Shumaker and S. Lackey (Eds.): VAMR 2014, Part II, LNCS 8526, pp. 51–62, 2014.

The first property emphasizes real-time collaboration and active interaction between educators and learners via augmented multimedia and social media. Just looking through a mobile device or AR glasses, an educator can monitor the progress of learners or groups via their interactions with augmented content in lectures. The educator also gets feedbacks from learners on the content and activities designed and linked to a specific page in a lecture note or a textbook to improve the quality of lecture design. Learners can create comments, feedback, or other types of social media targeting a section of a lecture note or a page of a textbook for other learners or the educator. A learner can also be notified and know social content created by other team members during the progress of teamwork.

The second property of the system is the naturalness of user experience as the system can aware of the context, i.e. which section of a page in a lecture note or a textbook is being read, by natural images, not artificial markers. Users can also interact with related augmented content with their bare hands. This helps users enhance their experience on both analog aesthetic emotions and immersive digital multisensory feedback by additional multimedia information.

The core component to develop an AR education environment is to recognize certain areas of printed materials, such as books or lecture handouts. As a learner is easily attracted by figures or charts in books and lecture notes, we encourage educators to exploit learners' visual sensitivity to graphical areas and embed augmented content to such areas, not text regions, in printed materials to attract learners. Therefore in our proposed system, we do not use optical character recognition but visual content recognition to determine the context of readers in reading printed materials.

In practice, graphical regions of interest that mostly attract readers in a page do not fully cover a whole page. There are other regions that do not provide much useful information for visual recognition, such as small decorations or small texts. Therefore, we propose the novel method based on saliency metric to quickly eliminate unimportant or noisy regions in printed lecture notes or textbooks and speed up the visual context recognition process on mobile devices or AR glasses. Our experimental results show that the whole visual object recognition process can be speed up 19 times and the accuracy can increase up to 22%.

This paper is structured as follows. In Section 2, the authors briefly present and analyze the related work. The proposed system is presented in Section 3. In Section 4, we present the core component of our system – the visual search engine. The experiments and evaluations are showed in Section 5. Then we discuss potential use of the system in Section 6. Finally, Section 7 presents conclusion and ideas for future work.

2 Related Work

2.1 Smart Educational Environment

Active learning methods focus on the responsibility of learning on learners [1]. To create an environment for leaners to study efficiently with active learning methods, educators should prepare and design various activities to attract learners. The educators also keep track of the progress for each member in team-work, and simulate the enthusiasm and collaboration of all learners in projects.

Learning activities are not necessarily to be in traditional physical classrooms but virtual environment as well [3]. An educator is required to use various techniques to attract learners' interest and attention to deliver knowledge impressively to them. Augmented Reality (AR) is an emerging technology that enables learners to explore the world of knowledge through the manipulation of virtual objects in real world.

AR has been applied in education to attract learners to study new concepts easily. With handheld displays, users can see virtual objects appearing on the pages of MagicBook [4] from their own viewpoint. After the work was published, several implementations of AR books were created for education, storytelling, simulation, game, and artwork purposes such as AR Vulcano Kiosk and S.O.L.A.R system [5].

AR has also shown great potential in developing and creating an interactive and a more interesting learning environment for the learners. Therefore, useful methods such as interactive study, collaboration study are proposed to enhance this. The classroom environment can be implemented in many ways: collaborative augmented multi user interaction [2] and mixed reality learning spaces [3].

However, these systems still have some limitations. First of all, they do not explicitly describe mechanism and processes for educators and learners to interact and collaborate efficiently in virtual environment with AR. Second, the educators may not have the feedbacks from learners to redesign or organize augmented data and content that are linked to sections in a printed material to improve the quality of education activities. Third, although AR system permits different users to get augmented information corresponding to different external contexts, all users receive the same content when looking at the same page of the book. And the last limitation is that these systems usually give unnatural feeling due to using artificial markers.

The mentioned problems motivate us to propose our smart education environment with AR and personalized interaction to enhance the attractiveness and immersive experience for educators and learners in virtual environment to improve efficiency in teaching and learning. In our proposed system, educators can receive explicit and implicit feedbacks from learners on the content and activities that are designed and linked to a specific lecture in a printed material.

2.2 Visual Sensitivity of Human Perception

A conventional approach to evaluate the attraction of objects in an image is based on textural information. In this direction, regional structural analysis algorithms based on gradient are used to detect features. However, saliency is considered better to reflect sensitivity of human vision to certain areas on an image thus benefits context awareness systems [6]. Visual saliency [7], human perceptual quality indicating the prominence of an object, person, or pixel to its neighbors thus capture our attention, is investigated by multiple disciplines including cognitive psychology, neurobiology, and computer vision. Salient maps are topographical maps of the visually salient parts of scenes without prior knowledge of their contents and thus remains an important step in many computer vision tasks.

Saliency measures are factors attracting eye movements and attention such as color, brightness, and sharpness, etc. [8]. Self-saliency is a feature that expresses the inner region complexity, which includes color saturation, brightness, texture, edginess, etc. Whereas, relative saliency indicates differences bet1een a region and its surrounding regions such as color contrast, sharpness, location, etc. Saliency measures can be combined with different weights to determine important regions more efficiently.

Most of saliency object detection techniques can be characterized as bottom-up saliency analysis, which is data-driven [9], or top-down approach, which is task-driven [6]. We focus on pre-attentive bottom-up saliency detection techniques. These methods are extensions of expert-driven human saliency that tends to use cognitive psychological knowledge of the human visual system and to find image patches on edges and junctions as salient using local contrast or global unique frequencies. Local contrast methods are based on investigating rarity of an image region with respect to local neighborhoods [8]. Whereas, global contrast based methods evaluate saliency of an image region using its contrast with respect to the entire image [10].

In this paper, the authors propose an efficient based human vision computation method to detect automatically high informative regions based on regional contrast in order to determine which region contains meaningful keypoint candidates. This reduces redundant candidates for further processing steps.

3 Overview of Proposed System

3.1 Motivations: Advantages of Smart Environment

The main objective of our system is to create a smart interactive education environment to support real-time collaboration and active interaction between educators and learners. Via special prisms, i.e. mobile devices or AR glasses, both educators and learners are linked to the virtual learning environment with real-time communication and interactions. Our proposed system has the following main characteristics:

1. *Interactivity*: Learners and educators can interact with augmented content, including multimedia and social media, or interact with others via augmented activities, such as exercises or discussion.
2. *Personalization*: Augmented content and activities can be adapted to each learner to provide the learner with the most appropriate, individualized learning paradigm. The adaptation can be in active or passive modes. In active mode, each learner can customize which types of augmented content and activities that he or she wants to explore or participate. In passive mode, an educator can individualize teaching materials to meet the progress, knowledge level, personal skills and attitudes of each learner.
3. *Feedback*: Interactive feedbacks from learners can be used to help an educator redesign existing teaching materials or design future teaching activities. Besides, feedbacks of a learner can also be used to analyze his or her personal interests, knowledge level, personal skills and attitudes toward certain types of activities in learning.
4. *Tracking*: The progress of a learner or a group of learners can be monitored so that an educator can keep track of the performance of each individual or a group.

3.2 Typical Scenarios of Usage for an Educator

The proposed system provides an educator with the following main functions:

1. Design augmented content and activities for lectures
2. Personalize or customize augmented content and activities for each learner or a group of learners
3. Monitor feedbacks and progress of each learner or a group of learners

The first step to create an AR-supported lecture is to design augmented content and activities for lectures. Lecture documents in each course include textbooks, reference books, and other printed materials. An educator can now freely design lectures with attached augmented materials (including multimedia, social media, or activities) that can be revised and updated over terms/semesters, specialized for different classes in different programs such as regular or honors program, and adapted to different languages. Because of wide variety of attached augmented media and activities, an educator can customize a curriculum and teaching strategies to deliver a lecture.

An educator uses our system to design augmented content (including multimedia objects or activities) for a lecture and assigns such content to link with a specific region in a printed page of a lecture note/textbook (c.f. Figure 1). Augmented media are not only traditional multimedia contents, such as images, 3D models, videos, and audios, but also social media contents or activities, such as different types of exercises, an URL to a reference document, or a discussion thread in an online forum, etc.

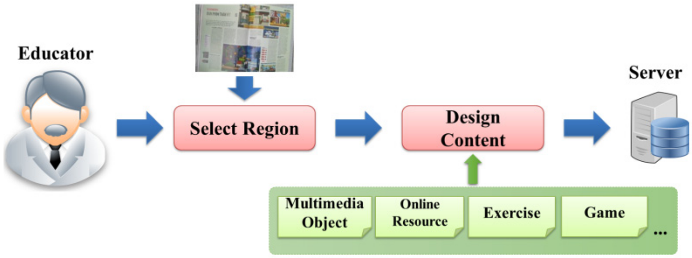


Fig. 1. Design augmented content for printed lecture notes and textbooks

For a specific page, an educator first selects a graphical region that can visually attract learners' attention, and links it to augmented contents, either resources or activities. The system automatically learns features to recognize the selected graphical region together with embedded resources to a remote server. An educator can also design different sets of augmented contents for the same printed teaching materials for different groups of learners in the same class, or for classes in classes in different programs, to utilize various teaching strategies and learning paradigms.

After designing AR-supported teaching materials, an educator can interact with learners via augmented activities during a course. Useful information on learners' activities and interactions are delivered to an educator so that the educator can keep track of the progress of a learner or a group of learners, update and customize augmented resources or activities to meet learners' expectation or level of knowledge, and redesign the teaching materials for future classes.

3.3 Typical Scenarios of Usage for a Learner

A learner can use a mobile device or smart glasses to see pages in a textbook, a reference book, or a printed lecture handout. Upon receiving the visual information that a learner is looking at, the server finds the best match (c.f. Section 4). Then the system transforms the reality in front of the learner's eyes into an augmented world with linked media or activities. Dynamic augmented contents that match a learner's personal profile and preferences are downloaded from the server and displayed on the learner's mobile device screen or glasses(c.f. Figure 2)..

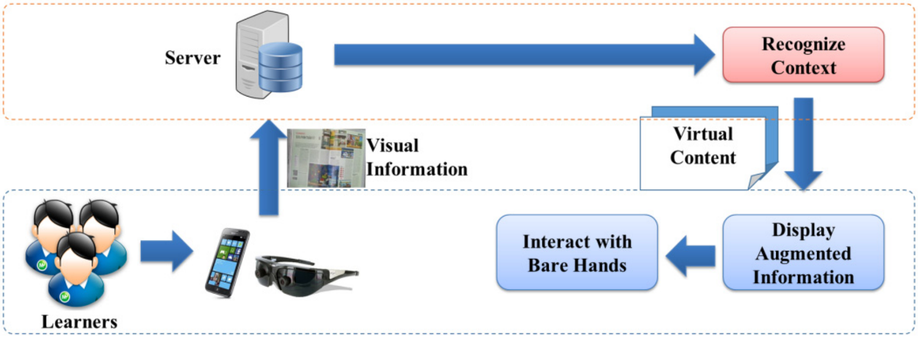


Fig. 2. Learners use the proposed system

Learners can interact with these virtual objects with their bare hands. Skin detection algorithm is used to enable learners use their bare hands to interact with virtual objects appearing in front of their eyes. An event corresponding to a virtual object is generated if that object is occluded by a human skin color object long enough.

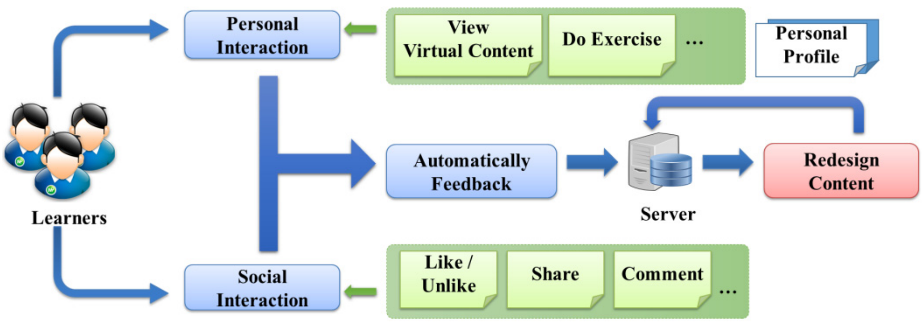


Fig. 3. Interaction and feedback

Learners can add a new virtual note or comment to a specific part of a printed lecture and share with others. They can also do exercises embedded virtually as an augmented content linked to printed lecture notes. When learners use this system, their behaviors are captured as implicit feedbacks to the educator (c.f. Figure 3). An educator can now analyze learners' behaviors and intention to adjust teaching materials to well adapt to each learner of a group. With collaborative filtering methods, the system can recommend to educators which types of augmented content are appropriate for a specific learner based on learners' profiles.

4 Visual Search Optimization with Saliency Based Metric

4.1 Overview

For mobile visual search (MVS) applications, most of existing methods use all keypoints detected from a given image, including those in unimportant regions such as small decoration or text areas. Different from state-of-the-art methods, our approach reduces the number of local features instead of reducing the size of each descriptor. Only keypoints with meaningful information are considered. As our method is independent of the choice of features, the combination of our idea with compact visual descriptors will give more efficiency.

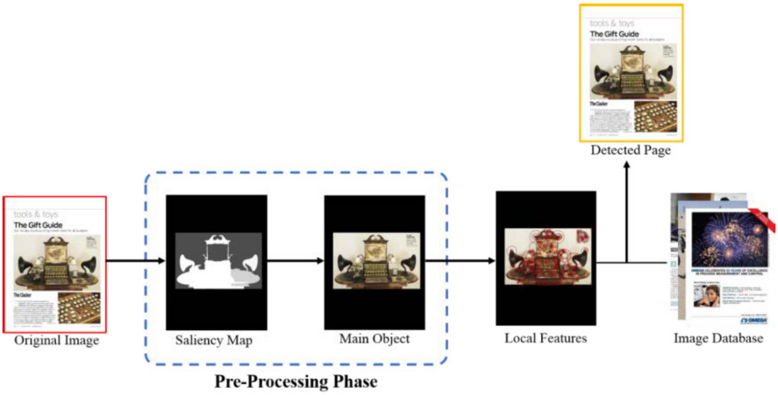


Fig. 4. Our approach to detect a page in a printed lecture note or textbook

We propose the idea to utilize the saliency map of an image to quickly discard keypoints in unimportant or insensitive regions of a template image as well as a query image (c.f. Figure 4). The visual sensitivity of each region is evaluated to determine keypoints to be preserved and those to be removed. This helps to reduce computational cost in local feature extraction of an image. As keypoints in unimportant regions can be removed, the accuracy of visual object recognition can also be improved.

Figure 5 shows our proposed method with two main steps. First, an arbitrary image is decomposed into perceptually homogeneous elements. Then, saliency maps are derived based on the contrast of those elements. The proposed saliency detection algorithm is inspired by the works in object segmentation with image saliency [10]. In our approach, regions of interest can be discrete and there is no need of merging.

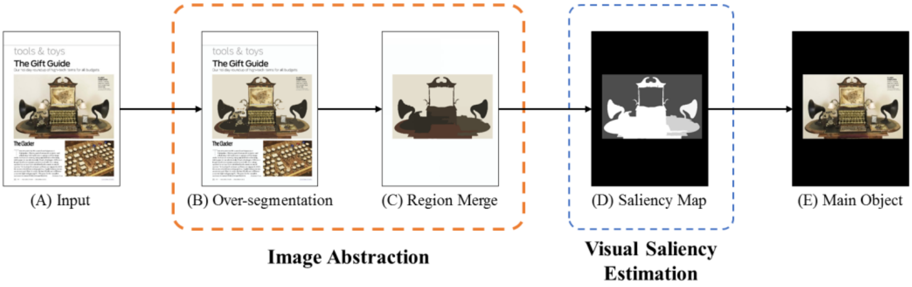


Fig. 5. Pre-processing phase

4.2 Image Abstraction

To simplify illustrations from color images, visual contents are abstracted by region based segmentation algorithms. A region grows by adding similar neighboring pixels according to certain homogeneity criteria, increasing size of region gradually. The proposed algorithm for this phase includes two steps: Over-Segmentation (c.f. Figure 5.B) and Region Growing (c.f. Figure 5.C).

Over-Segmentation: An image is over-segmentation by the watershed-like method. The regions are merged on the basis of a similarity color criterion afterwards: $\left\|c_i - c_j\right\|_2 \leq \theta$ where c_i and c_j are pixels in the same region.

Region Growing: Neighboring segments are merged based on their sizes, which are the number of pixels of each region. If a region whose size is below a threshold, it is merged to its nearest region, in terms of average Lab color distance. To speed up, we use Prim's algorithm [11] to optimize merging regions.

4.3 Visual Saliency Estimation

An image captured from a camera is intentionally focused on meaningful regions by human vision which reacts to regions with features such as unique colors, high contrast, or different orientation. Therefore, to estimate the attractiveness, the contrast metric is usually used to evaluate sensitivity of elements in image.

A region with high level of contrast with surrounding regions can attract human attention and is perceptually more important. Instead of evaluating the contrast difference between regions in an original image, the authors only calculate the contrast metric based on Lab color between regions in the corresponding segmented image. As the number of regions in the original image is much more than the number of regions in its corresponding segmented image, our approach not only simplifies the calculation cost but also exploits the meaningful regions in the captured image efficiently. The contrast C_i of a region $\mathcal{R}_i$ is calculated as the difference between Lab color of R_i and its surrounding regions:

$$C_i = \frac{\sum_{j=1}^{n} \omega(\mathcal{R}_j) \left\|c_i - c_j\right\|_2}{\sum_{j=1}^{n} \omega(\mathcal{R}_j)} \quad (1)$$

where c_j and c_i are Lab colors of regions $\mathcal{R}_j$ and $\mathcal{R}_i$ respectively, and $\omega(\mathcal{R}_j)$ is the number of pixels in region $\mathcal{R}_i$. Regions with more pixels contribute higher local-contrast weights than those containing only a few pixels. Finally, C_i is normalized to the range $[0,1]$. Figure 6 shows that our method can provide better results than existing saliency calculation techniques.

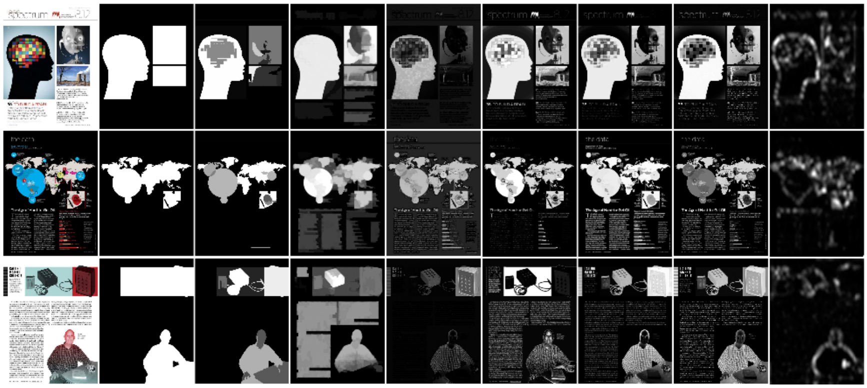


Image Ground truth Ours BMS [12] FT [7] GC [13] HC [14] LC [15] SR [16]

Fig. 6. Visual comparison between the proposed method and other state-of-the-art methods

5 Experiments and Evaluation

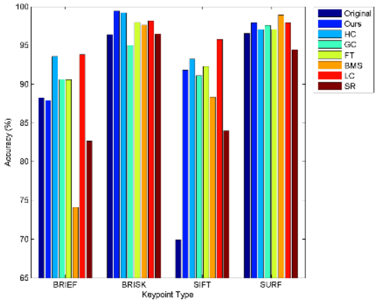
5.1 Page Detection Evaluation

We conduct the experiment to evaluate the efficiency of our proposed method by matching local features extracted from images in the dataset to compare the accuracy and performance of the proposed process with the original method which does not filter out keypoints and the other state-of-the-art saliency detection methods. Since the proposed process is independent of the keypoint extraction and recognition algorithms, experiments are conducted to evaluate our approach using four popular local features: BRIEF [17], BRISK [18], SIFT [19], and SURF [20].

Experiment is conducted in a system using CPU Core i3 3.3 GHz (with 4GB RAM). Our dataset consists of 200 pages (with resolution 566×750) of reference materials for students in Computer Science, including MSDN Magazine, ACM Transaction Magazine, and IEEE Transaction Magazine. Each typical page includes three types of regions: background, text region, and image.

All local features are extracted in two scenarios: extracting all keypoints and extracting only keypoints in important regions. Image matching is then performed with each pair of images. The accuracy of matching is computed as proportion of correctly matched pairs of images over the number of image pairs. The result of this experiment is shown in Figure 7(a).

On average, the proposed method outperforms conventional methods up to 7%. Especially, when using SIFT feature, the accuracy is boosted approximately 22%. Moreover, our saliency detection module is replaced by different existing state-of-the-art methods such as BMS [12], FT [7], GC [13], HC [14], LC [15], and SR [16] to evaluate efficiency of our approach. In most cases, our process can provide better results than others. Incorporating our pre-process stage can not only preserve the robustness of conventional methods but also boost up the accuracy.



A. Accuracy (%)

Saliency Methods	Local Features			
	BRIEF	BRISK	SIFT	SURF
Ours	**90.42**	**4.02**	**21.33**	**44.34**
FT	93.48	6.12	26.80	46.76
SR	444.35	43.83	35.17	263.76
HC	261.04	23.09	37.60	180.75
BMS	193.48	14.44	41.50	79.55
GC	328.97	43.87	88.19	220.26
LC	717.25	56.92	104.62	465.71
Original	1000.48	78.00	218.66	661.12

B. Performance (millisecond)

Fig. 7. Accuracy and performance of page detection of printed reference materials

In addition, the experiments also show that our method outperforms other algorithms with all common local features (c.f. Figure 7.B). On average, using SIFT, our method is 10.3 times faster than conventional method with no filtering out keypoints. Similarly, using BRIEF and SURF, our method is 11 and 15 times faster, and especially that of using BRISK features is more than 19.4 times.

Overall, our approach does not only boost up the running time up 19.4 times but also increases the accuracy of recognizing magazines to 22%. This is the crucial criteria for real-time AR system for magazines, books, and newspapers.

6 Potential Usage of Proposed System

For each course in a specific teaching environment, it is necessary to identify which types of augmented contents are required by end-users, i.e. educators and learners. Therefore, we conduct surveys to evaluate the practical need for our system in enhancing the enthusiasm and attractiveness for learners, including high school students and undergraduate students.

In the meeting with high school teachers and students in enhancing learning experience in Chemistry, we identify the first two main requirements for our system. The first is 3D visualization of chemical elements, substances, atoms, molecules, and stoichiometry. The second is to assist teachers in the visualization and simulation for chemical reactions. Although no activities have been set up for students, it is a new teaching activity with the assistance of our smart educational environment via AR.

In the meeting with instructors of the two courses on Introduction to Information Technology 1 and 2, we identify more interesting augmented contents including multimedia/social media data and augmented activities that can be established via our system. These two courses aim to provide the overview on different aspects in Information Technology for freshmen as a preparation and guidance for students following the teaching strategy of Conceive - Design - Implementation - Operation (CDIO). With the assistance of our system, we can deploy the trial teaching environment for freshmen volunteers to join the active learning activities with AR interactions. The volunteers use the proposed system in the two courses. Students are assigned to read printed materials with AR media, to discuss and do exercises with others via our system. We collect useful feedbacks from participants to evaluate the usefulness and convenience of our system as well as the satisfaction of volunteers with the system and favorite functions. Based on the qualitative interviews in this study, most students find that our system can provide a more interesting and attractive way to study than traditional approaches do. Moreover, the features of collaboration in our system successfully attract students' interest and trigger their motivation in reading documents.

7 Conclusion and Future Work

The authors propose a new method for organizing a collaborative class using AR and interaction. Via our proposed system, learners and educators can actively interact with others. Learners can do exercises embedded virtually as augmented contents linked to a printed lecture note. They can add a new virtual note or comment to a specific part of a printed lecture and share with others as well. Besides, educators get feedbacks from learners on the content and activities designed and linked to a specific page in a lecture note or textbook to improve the quality of lecture designs. Educators can also keep track of the learning progress of each individual or each group of learners.

In our proposed system, we focus on providing the natural means of interactions for users. The system can recognize the context, i.e. which section of a page in a lecture note or a textbook is being read, by natural images, not artificial markers. Users can also interact with related augmented contents with their bare hands.

We also propose a new method based on saliency metric to quickly eliminate irrelevant regions in a page of a book or printed material to enhance the accuracy and performance of the context aware process on mobile devices or AR glasses. Furthermore, our method works independently of the training and detecting stage. It is compatible to most well-known local features. Therefore, this stage can be incorporated into any existed system for printed material detection and recognition.

There are more saliency metrics for implementation in our visual search engine, thus requires further experiments. In addition, the authors are interested in applying psychology and neuroscience knowledge of human vision in further research. To enhance the system, we are doing classification by Neuron network algorithm to analysis the profiles and learn their behaviors in order to utilize better.

References

[1] Bonwell, C.C., Eison, J.A.: Active learning: Creating excitement in the classroom. School of Education and Human Development, George Washington University, Washington, DC, USA (1991)

[2] Kaufmann, H., Schmalstieg, D., Wagner, M.: Construct3D: A Virtual Reality Application for Mathematics and Geometry Education. Education and Information Technologies 5(4), 163–276 (2000)
[3] Winkler, T., Kritzenberger, H., Herczeg, M.: Mixed Reality Environments as Collaborative and Constructive Learning Spaces for Elementary School Children. In: The World Conference on Educational Multimedia, Hypermedia and Telecommunications (2002)
[4] Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook - Moving Seamlessly between Reality and Virtuality. IEEE Computer Graphics and Applications 21(3), 6–8 (2001)
[5] Woods, E., et al.: Augmenting the science centre and museum experience. In: The 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (2004)
[6] Goferman, S., Zelnik-Manor, L., Tal, A.: Context-Aware Saliency Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(10), 1915–1926 (2012)
[7] Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: 22nd IEEE Computer Society on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
[8] Ma, Y.-F., Zhang, H.-J.: Contrast-based image attention analysis by using fuzzy growing. In: 11th ACM International Conference on Multimedia, pp. 374–381 (2003)
[9] Siva, P., Russell, C., Xiang, T., Agapito, L.: Looking Beyond the Image: Unsupervised Learning for Object Saliency and Detection. In: 26th IEEE Conference on Computer Vision and Pattern Recognition (2013)
[10] Qiong, Y., Xu, L., Shi, J., Jia, J.: Hierarchical Saliency Detection. In: 26th IEEE Conference on Computer Vision and Pattern Recognition (2013)
[11] Prim, R.C.: Shortest connection networks and some generalizations. Bell System Technical Journal 36(6), 1389–1401 (1957)
[12] Zhang, J., Sclaroff, S.: Saliency detection: A boolean map approach. In: The IEEE International Conference on Computer Vision, ICCV (2013)
[13] Cheng, M.-M., et al.: Efficient Salient Region Detection with Soft Image Abstraction. In: IEEE International Conference on Computer Vision (2013)
[14] Cheng, M.-M., Zhang, G.-X., Mitra, N.J., Huang, X., Hu, S.-M.: Global contrast based salient region detection. In: 24th IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416 (2011)
[15] Zhai, Y., Shah, M.: Visual Attention Detection in Video Sequences Using Spatiotemporal Cues. In: The 14th Annual ACM International Conference on Multimedia, pp. 815–824 (2006)
[16] Hou, X., Zhang, L.: Saliency Detection: A Spectral Residual Approach. In: 20th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2007)
[17] Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary robust independent elementary features. In: 11th European Conference on Computer Vision, pp. 778–792 (2010)
[18] Leutenegger, S., Chli, M., Siegwart, R.Y.: BRISK: Binary Robust Invariant Scalable Keypoints. In: 13th IEEE International Conference on Computer Vision (ICCV), pp. 2548–2555 (2011)
[19] Lowe, D.: Distinctive Image Features from Scale Invariant Keypoints. International Journal of Computer Vision 20(2), 91–110 (2004)
[20] Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: SURF: Speeded Up Robust Features. In: 9th European Conference on Computer Vision, pp. 404–417 (2006)