

An Application of Image Processing in Optical Mark Recognition

Tran Vu Ha¹ & Nguyen Thi Thu²

¹Faculty of Information Technology, Vietnam National University of Agriculture, Hanoi 131000, Vietnam

²Faculty of Animal Science, Vietnam National University of Agriculture, Hanoi 131000, Vietnam

Abstract

The Optical Mark Recognition (OMR) is very popular with universities for the reading of multiple-choice questions. In this article, we presented a software system for processing surveys at the Vietnam National University of Agriculture based on digital image processing. This software was built using MATLAB and easy to use. The surveys were digitized using a scanner and sent to the software tool. In this study, we tested more than 170 surveys of nine different types. The software tool correctly detected all the valid answers. It was also able to detect all questions with no or multiple marks.

Keywords

Image processing, optical mark recognition, survey

Introduction

Optical mark recognition (OMR) is a form of automated data processing. Questions with multiple choices are printed on paper. Respondents then mark their answers using pens. In the next step, the sheets are scanned and sent to a computer for processing. There are many applications of OMR including multiple-choice examinations (for students and pupils) and feedback collection (from customers, students, and users, etc.). In universities (i.e., Vietnam National University of Agriculture), collecting feedback from students plays an important role in evaluating and improving the quality of education.

Nowadays, many commercial solutions for OMR are available (e.g., OpScan Series Product from SCANTRON). In common, these products require a dedicated scanner and answer sheets, which motivates the finding of cheaper solutions. Hong Duc University created a software named TickREC for this purpose (Hong-Duc University, 2014). The Vietnam Forestry University also has its software solutions (Mai Ha An, 2014). Increasingly more methods for mark detection have been published. Gaikwad (2015) applied a

Received: April 17, 2020
Accepted: December 5, 2020

Correspondence to
ntthu@vnua.edu.vn

ORCID
Le Thanh Ha
<https://orcid.org/0000-0001-5090-5491>

template matching algorithm after finding the region of interest to find the answers marked (Gaikwad, 2015). Loke *et al.* (2018)*et al.* proposed a method based on pixel counting and simple thresholding that can be used under a variety of conditions. Another method by Belag *et al.* was developed based on the creation of template answer sheets and key points detection algorithms (Belag *et al.*, 2018). Each of these methods (and corresponding software tools) has its own advantages and disadvantages. For example, Belag's tool used a dedicated sheet for answers, this sheet also had checkmarks that helped in case the scanned image was rotated. This kind of sheet is suitable for tests but is not good for surveys. In cases of TickREC and the tool of Mai Ha An (2014), they could process the sheets that contained both questions and answers (Mai Ha An, 2014). Because each software works with a certain type of answer sheet, which was designed as needed by the authors, it is not possible to apply these softwares instantly for the surveys at the Vietnam National University of Agriculture.

Hence, in this work, we created a software for processing surveys at the Vietnam National University of Agriculture. The surveys were scanned by an ordinary scanner and sent to the software to process. This software was designed in such a manner that it was easy to use and no special training was required. This system was cost-effective because no dedicated machine or answer sheets were required.

Materials and Methods

Materials

In this project, we used nine different types of questionnaires. All of these were used by the Center for Quality Assurance, Vietnam National University of Agriculture:

- (i) Employee feedback about the operation of a number of divisions
- (ii) Member feedback about the support of the Ho Chi Minh Communist Youth Union
- (iii) Student feedback about the support of a number of divisions

- (iv) Student feedback about an advanced education program

- (v) Master student feedback about a specific course

- (vi) Graduate student feedback about an educational program

- (vii) Student feedback about a theoretical course of an ordinary education program

- (viii) Student feedback about a practical course of an ordinary education program

- (ix) Student feedback about a theoretical course of a Professional Oriented to Higher Education (POHE) program

For each type of questionnaire, there were more than 30 sheets that were randomly filled. All of the sheets were scanned with an HP scanner (ScanJet Pro 3000 s3). The output file format was normally JPEG but could also be PNG, BMP, or some other formats supported by MATLAB (see method section for more details). The width and the height of the images were 1655 and 2338 pixels, respectively (these dimensions of images could be slightly different depending on the scanner). The examples of surveys are shown in **Figures 1 and 2**.

Methods

MATLAB - Environment for software development

MATLAB (short name for matrix laboratory) was developed in the 1970s by Cleve Moler (Haigh, 2008). Most of the codes of MATLAB was written by Cleve Moler using FORTRAN. Jack Little and Steve Bangert then reprogrammed MATLAB in C. Together with Cleve Moler, three of them founded the MathWorks in California in 1984. MathWorks then develops, maintains, and distributes MATLAB as a commercial product (Sandeep, 2017). Nowadays, MATLAB supports various platforms such as LINUX, Windows, and MacOS. With MATLAB, users write a few lines of code to acquire instant results without involving a compiler. MATLAB is used for data analysis and visualization. It supports multiple types of data (audios, images, videos, CSV, and

MẪU 1

HỌC VIỆN NÔNG NGHIỆP VIỆT NAM
TEN_DON_VI

PHIẾU LẤY Ý KIẾN PHẢN HỒI TỪ CÁN BỘ
(Về hoạt động hỗ trợ của(Tên đơn vị))

Để cải tiến, nâng cao chất lượng các hoạt động hỗ trợ của(Tên đơn vị) và phục vụ cho công tác kiến định Học viện, Khoa/Trung tâm/Ban/Trạm y tế/Nhà xuất bản rất mong nhận được những ý kiến đóng góp từ ông/bà bằng cách trả lời các câu hỏi dưới đây.

Trân trọng cảm ơn!

I. Nội dung lấy ý kiến phản hồi:
Ông/bà vui lòng tô kín vào ô tròn về mức độ hài lòng đối với từng tiêu chí theo thang điểm từ 1 đến 5:
1. Rất không hài lòng 2. Không hài lòng 3. Tạm hài lòng 4. Hài lòng 5. Rất hài lòng

STT	NỘI DUNG LẤY Ý KIẾN	MỨC ĐỘ ĐÁNH GIÁ				
		1	2	3	4	5
A	Thái độ phục vụ					
1	Lịch sự, hòa nhã trong giao tiếp	<input type="radio"/>				
2	Thái độ hỗ trợ tích cực, nhiệt tình	<input type="radio"/>				
3	Phục vụ với tinh thần trách nhiệm cao	<input type="radio"/>				
4	Tuân thủ thời gian làm việc	<input type="radio"/>				
B	Chất lượng hỗ trợ	1	2	3	4	5
5	Hướng dẫn, giải đáp chi tiết, rõ ràng	<input type="radio"/>				
6	Giải quyết công việc thỏa đáng và hiệu quả	<input type="radio"/>				
7	Giải quyết công việc kịp thời	<input type="radio"/>				
8	Tuân thủ đúng các quy định và quy trình giải quyết công việc	<input type="radio"/>				
C	Nội dung lấy ý kiến riêng của đơn vị (Đơn vị bổ sung thêm nội dung khác muốn lấy ý kiến phù hợp với đặc điểm riêng của đơn vị mình)	1	2	3	4	5
9	<input type="radio"/>				
...	Mức độ hài lòng chung của ông/bà về Khoa/Trung tâm/Ban/Trạm y tế/Nhà xuất bản	<input type="radio"/>				

Ý kiến khác:

Để xuất ý kiến nâng cao chất lượng hỗ trợ của đơn vị:

Cảm ơn ông/bà đã tham gia trả lời!

(a) A survey for employees

MẪU 1

HỌC VIỆN NÔNG NGHIỆP VIỆT NAM
ĐOÀN THANH NIÊN

PHIẾU LẤY Ý KIẾN PHẢN HỒI TỪ ĐOÀN VIÊN
(Về hoạt động hỗ trợ của Đoàn thanh niên)

Để cải tiến, nâng cao chất lượng các hoạt động hỗ trợ của Đoàn Thanh niên và phục vụ cho công tác kiểm định Học viện, Đoàn Thanh niên rất mong nhận được những ý kiến đóng góp từ anh/chị bằng cách trả lời các câu hỏi dưới đây.

Trân trọng cảm ơn!

I. Nội dung lấy ý kiến phản hồi:
Anh/chị vui lòng tô kín vào ô tròn về mức độ hài lòng đối với từng tiêu chí theo thang điểm từ 1 đến 5:
1. Rất không hài lòng 2. Không hài lòng 3. Tạm hài lòng 4. Hài lòng 5. Rất hài lòng

STT	NỘI DUNG LẤY Ý KIẾN	MỨC ĐỘ ĐÁNH GIÁ				
		1	2	3	4	5
A	Thái độ phục vụ					
1	Lịch sự, hòa nhã trong giao tiếp	<input type="radio"/>				
2	Thái độ hỗ trợ tích cực, nhiệt tình	<input type="radio"/>				
3	Phục vụ với tinh thần trách nhiệm cao	<input type="radio"/>				
4	Tuân thủ thời gian làm việc	<input type="radio"/>				
B	Chất lượng hỗ trợ	1	2	3	4	5
5	Hướng dẫn, giải đáp chi tiết, rõ ràng	<input type="radio"/>				
6	Giải quyết công việc thỏa đáng và hiệu quả	<input type="radio"/>				
7	Giải quyết công việc kịp thời	<input type="radio"/>				
8	Tuân thủ đúng các quy định và quy trình giải quyết công việc	<input type="radio"/>				
C	Nội dung lấy ý kiến riêng của Đoàn thanh niên (Đoàn thanh niên bổ sung thêm nội dung khác muốn lấy ý kiến phù hợp với đặc điểm riêng của mình)	1	2	3	4	5
9	<input type="radio"/>				
...	Mức độ hài lòng chung của anh/chị về Đoàn Thanh niên	<input type="radio"/>				

Ý kiến khác:

Để xuất ý kiến nâng cao chất lượng hỗ trợ của Đoàn thanh niên:

Cảm ơn anh/chị đã tham gia trả lời!

(b) A survey for students

Figure 1. Example of surveys with one page

BỘ NÔNG NGHIỆP VÀ PHÁT TRIỂN NÔNG THÔN
HỌC VIỆN NÔNG NGHIỆP VIỆT NAM

PHIẾU LẤY Ý KIẾN PHẢN HỒI TỪ SV TỐT NGHIỆP
(Về Chương trình đào tạo)

I. THÔNG TIN CÁ NHÂN NGƯỜI HỌC
1. Khoa: Ngành: Chuyên ngành:
Lớp: Tốt nghiệp đợt: tháng năm
2. Xếp loại tốt nghiệp của Anh (Chị): Xuất sắc Giỏi Khá Trung bình

3. Loại hình đào tạo: DH chính quy DH chính quy VB2 DH hệ liên thông DH hệ VLXH Cao đẳng

II. PHẢN NHẬN XÉT CHƯƠNG TRÌNH ĐÀO TẠO
Tô kín vào ô tròn về mức độ đồng ý đối với từng tiêu chí mà Anh (Chị) cho là đúng nhất theo thang điểm từ 1 đến 5:
1. Rất không hài lòng 2. Không hài lòng 3. Tạm hài lòng 4. Hài lòng 5. Rất hài lòng

MỨC TIÊU VÀ CHƯƠNG TRÌNH ĐÀO TẠO	1	2	3	4	5
1. Mục tiêu đào tạo của ngành học được phổ biến tới người học	<input type="radio"/>				
2. Nội dung chương trình phù hợp với chuẩn đầu ra	<input type="radio"/>				
3. Cấu trúc chương trình mềm dẻo, thuận lợi cho người học lựa chọn	<input type="radio"/>				
4. Cấu trúc chương trình logic, các học phần có liên kết, ít trùng lặp	<input type="radio"/>				
5. Khối lượng kiến thức trong chương trình đào tạo hợp lý	<input type="radio"/>				
6. Chương trình đảm bảo tính thực tế và ứng dụng	<input type="radio"/>				
7. Phân bố số giờ lý thuyết, thực hành, rèn nghề hợp lý	<input type="radio"/>				
8. Các học phần tự chọn đáp ứng nhu cầu học tập đa dạng của người học	<input type="radio"/>				
9. Sự cân đối giữa kiến thức đại cương, cơ sở và chuyên ngành	<input type="radio"/>				
10. Tinh cấp nhất, đổi mới trong nội dung chương trình	<input type="radio"/>				
11. Các học phần trung cấp kiến thức và kỹ năng	<input type="radio"/>				
12. Nội dung kiểm tra, đánh giá sát với chương trình học	<input type="radio"/>				
ĐỘI NGŨ GIẢNG VIÊN	1	2	3	4	5
13. Kiến thức chuyên môn và sự cập nhật	<input type="radio"/>				
14. Phương pháp sư phạm	<input type="radio"/>				
15. Nhiệt tình, sẵn sàng hướng dẫn, giúp đỡ người học	<input type="radio"/>				
16. Thông báo đầy đủ về tiêu chí đánh giá cho từng môn học	<input type="radio"/>				
17. Đảm bảo giờ lên lớp và kế hoạch giảng dạy	<input type="radio"/>				
18. Công bằng, khách quan và phân định đúng năng lực của người học trong kiểm tra, đánh giá	<input type="radio"/>				
KẾT QUẢ HỌC TẬP	1	2	3	4	5
19. Người học được phát triển kỹ năng chuyên môn, nghiệp vụ nghề nghiệp	<input type="radio"/>				
20. Người học có đủ kiến thức cần thiết, cập nhật	<input type="radio"/>				

(a) The first page of a student survey

21	Người học được phát triển những kỹ năng mềm cần thiết cho nghề nghiệp (giao tiếp, trình bày, tổ chức, quản lý, làm việc nhóm...)	<input type="radio"/>				
22	Người học được phát triển năng lực ngoại ngữ, tin học	<input type="radio"/>				
23	Người học tự tin về khả năng đáp ứng các yêu cầu của công việc	<input type="radio"/>				
24	Người học được phát triển kỹ năng tự học, tự nghiên cứu	<input type="radio"/>				
25	Người học được phát triển đạo đức, nhân cách, ý thức kỷ luật	<input type="radio"/>				
QUẢN LÝ VÀ PHỤC VỤ TỔ CHỨC ĐÀO TẠO		1	2	3	4	5
26	Thông báo rõ kế hoạch giảng dạy từng môn học	<input type="radio"/>				
27	Quy trình giải quyết những khó khăn, thắc mắc của người học rõ ràng, kịp thời	<input type="radio"/>				
28	Người học được tư vấn, tạo điều kiện thuận lợi để đăng ký học phần	<input type="radio"/>				
29	Kết quả học tập được thông báo đến người học đúng thời gian quy định	<input type="radio"/>				
30	Thái độ phục vụ người học của cán bộ, nhân viên các phòng ban, khoa	<input type="radio"/>				
31	Người học được tham gia đối thoại trực tiếp với lãnh đạo Học viện để đưa ra ý kiến phản hồi về các hoạt động của Học viện	<input type="radio"/>				
THỰC HIỆN, TRANG THIẾT BỊ VÀ CƠ SỞ HẠ TẦNG		1	2	3	4	5
32	Giáo trình, tài liệu tham khảo tại thư viện	<input type="radio"/>				
33	Trang thiết bị phục vụ giảng dạy và học tập	<input type="radio"/>				
34	Trang thiết bị phục vụ thí nghiệm, thực hành, thực tập, rèn nghề	<input type="radio"/>				
35	Điện tích, ánh sáng, nhiệt độ, âm thanh và độ thông thoáng của phòng học	<input type="radio"/>				
DỊCH VỤ VÀ SINH HOẠT		1	2	3	4	5
36	Đáp ứng nhu cầu ăn, ở của người học	<input type="radio"/>				
37	Người học được đảm bảo chế độ chính sách xã hội	<input type="radio"/>				
38	Người học được chăm lo khám sức khỏe theo quy định y tế học đường	<input type="radio"/>				
39	Người học được tạo điều kiện hoạt động Đoàn, Hội, tập luyện văn nghệ, thể dục thể thao	<input type="radio"/>				
40	Các hoạt động Đoàn, Hội bổ ích và có ý nghĩa thiết thực	<input type="radio"/>				
41	Người học được tư vấn hướng nghiệp, giao lưu với doanh nghiệp và tham gia hội chợ việc làm	<input type="radio"/>				
42	Mức độ hài lòng chung của Anh (Chị) về chương trình đào tạo	<input type="radio"/>				
43	Mức độ hài lòng chung của Anh (Chị) về môi trường sống và học tập tại Trường	<input type="radio"/>				

III. Các ý kiến khác:
Theo anh/chị, để sinh viên Học viện Nông nghiệp Việt Nam được học tập, rèn luyện và sinh hoạt tốt hơn thì Học viện cần làm những gì?
.....
.....
.....
.....

Học viện chân thành cảm ơn sự hợp tác của Anh/Chị.

(b) The second page of a student survey

Figure 2. Example of surveys with two pages

different databases). MATLAB also provides App Designer tool which allows the users to build GUI (Graphical User Interface) for their programs (Educba, 2020). For these reasons, we used MATLAB to develop our software tool for data processing.

Processing workflow

Figure 3 shows the basic steps needed for the processing of one scanned page of questionnaires. For the first step, the selected machine (ScanJet Pro 3000 s3) scanned multiple pages in a single run. After that, our software tool then came into play.

Because our questionnaires were printed in monochrome and then filled using black or blue (the colors of most ballpoint pens), converting images to binary would save us memory and time for processing. With the support from MATLAB, converting images to binary was straightforward. We only needed to call the `im2bw` function with the original image as a parameter, the function then returned a binary image.

To extract the region of interest (ROI), the region in which people filled in the options, we used a special image called a mask. As shown in **Figure 4a**, a mask contained only filled options. Our program would then find the ROI. The position and size of ROI (the region inside the red rectangle, **Figure 4b**) was then used to crop the other scanned images.

With the function `imfindcircles` from MATLAB, we were able to locate all the options on the cropped images. The number of black pixels in each circle helped us to indicate the selected one.

Our software tool then outputted the selected options for every question on the sheet. The output was eventually stored in a plain text file.

Results and Discussion

The software tool

Figure 5 shows the main graphical user interface (GUI) of the program. The user first needed to specify the directory of scanned images by clicking *Select image folder* button

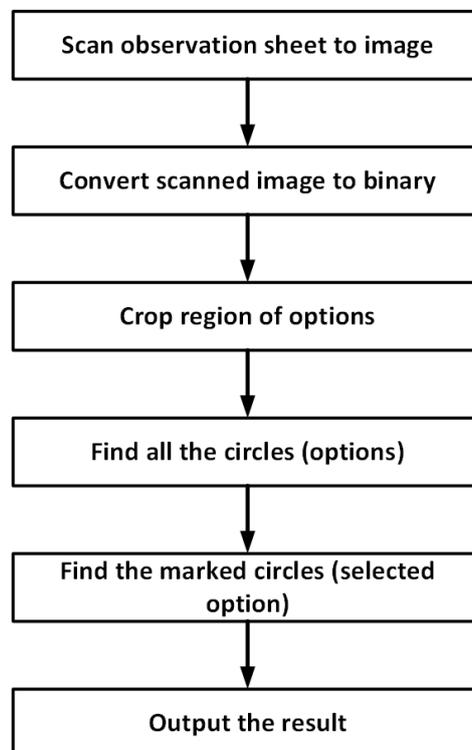
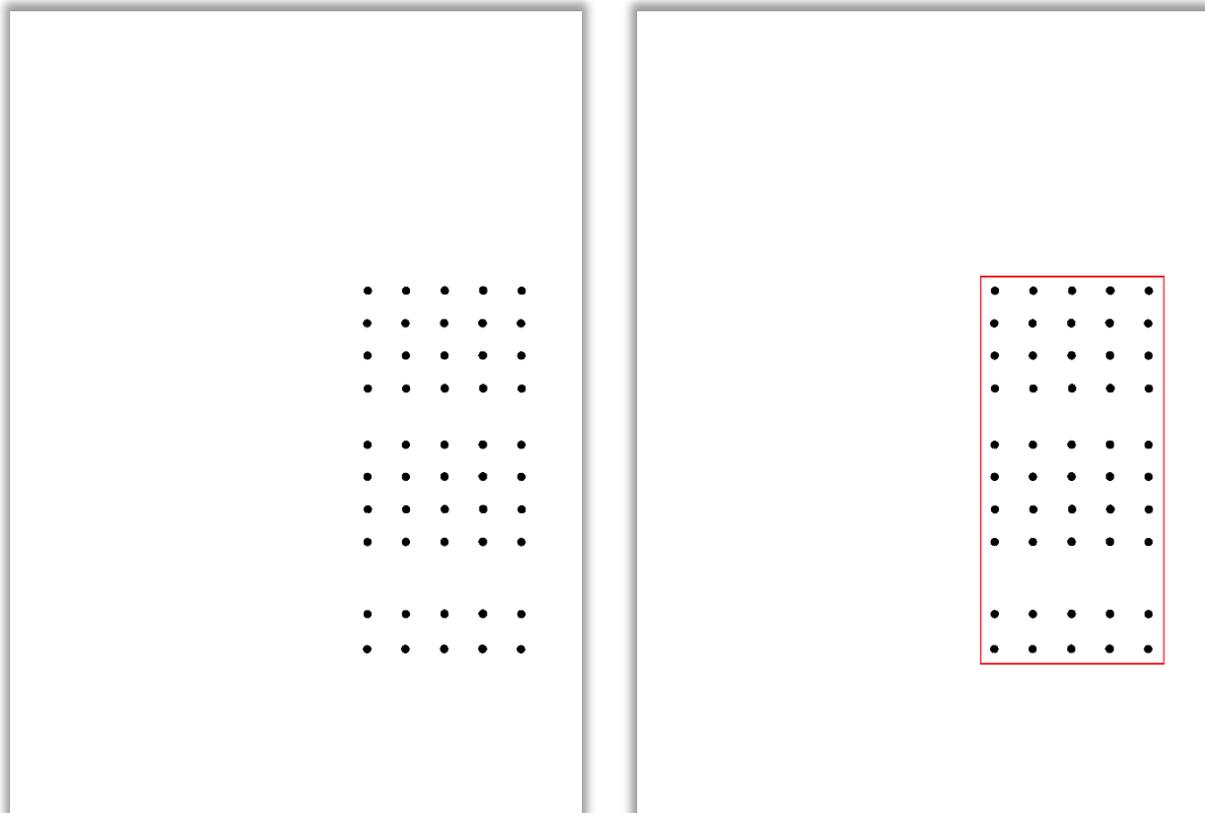


Figure 3. The proposed stages for data processing

An application of image processing in Optical Mark Recognition



(a) An example of mask image

(b) ROI on mask image (the area inside the red rectangle)

Figure 4. Mask image

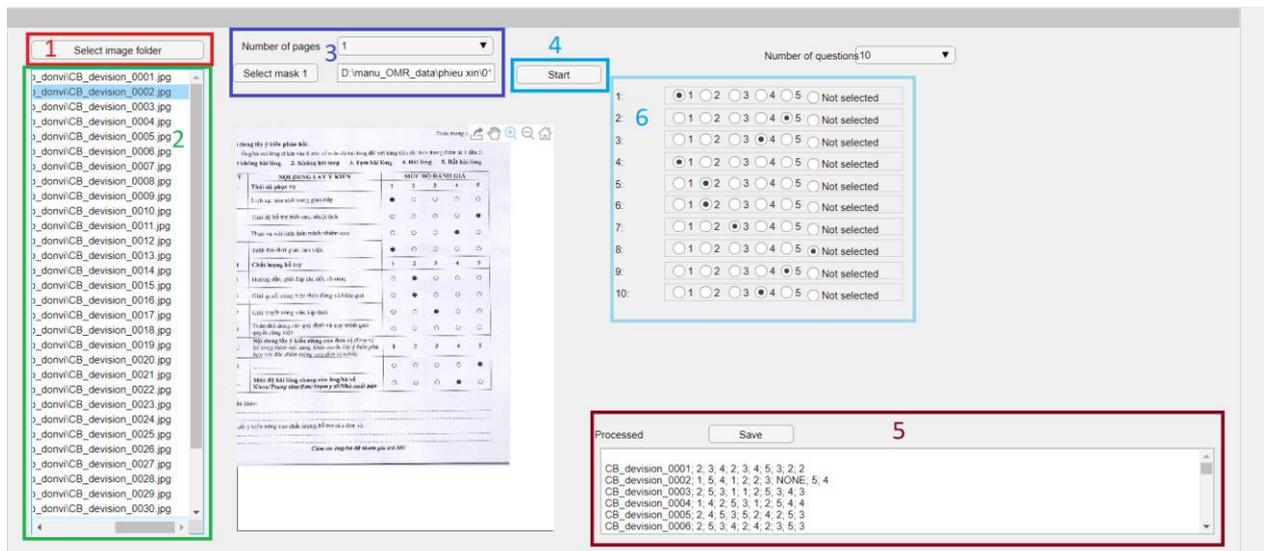


Figure 5. The main user interface of the program

(area 1). All images in the selected directory would be listed in the area below the button (area 2). The user then selected the mask file by clicking *Select mask* button (area 3). Depend on the type of questionnaire, we might need to select two masks if the questionnaire contained two pages. To start processing images, the user clicked on *Start* button (area 4). The result would be displayed at the bottom right of the window (area 5).

Processing questionnaires

Table 1 shows a summary of the analysis of

179 questionnaires belonging to nine different types. Our tool correctly detected all valid questions (questions having one option filled). It correctly identified all questions that were not filled (not evaluated by students, as shown in **Figure 6a**). The tool could also detect the question that had multiple options filled (the students changed their mind and chose another option) (**Figure 6b**).

Because the number of black pixels in each option was used to identify which options were filled, our tool might not work correctly in some cases as follows:

Table 1. Results of data processing

Type of questionnaires	Number of questions in the questionnaires	Number of questionnaires	Total number of questions	Number of correctly detected questions	Number of unfilled questions detected	Number of multiple filled questions detected
Employee feedback about the operation of a number of divisions	10	35	350	339	11	0
Member feedback about the support of Ho Chi Minh Communist Youth Union	10	34	340	338	1	1
Student feedback about the support of a number of divisions	10	35	350	342	5	3
Student feedback about an advanced education programs	25	35	875	866	2	7
Master student feedback about a specific course	23	35	805	800	3	2
Graduate student feedback about an educational program	43	35	1505	1498	2	5
Student feedback about a theoretical course of an ordinary education program	22	35	770	769	0	1
Student feedback about a practical course of an ordinary education program	18	35	630	628	1	1
Student feedback about a theoretical course of a POHE program	18	35	630	629	0	1

Instead of filling in the option, the user used a checkmark (tick) or x a mark (cross) to mark the selected option (**Figure 6c**). The number of black pixels inside a checked option might not be enough for a valid filled option.

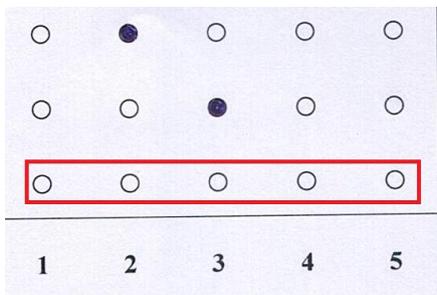
Options were not completely filled (**Figure 6d**). Similar to the previous case, the option might not be bold enough to be a marked one.

The user used light colors to mark the selected option. In this case, filled areas might become unfilled because of the conversion from color images to binary images.

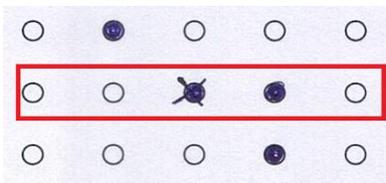
Apparently, our tool marked this question as NULL in the result area. The user could easily see this and check the answer sheet manually by

selecting the corresponding image from the list of images. After checking the images, the user was able to make direct modifications in the result area before exporting the final result to the output file.

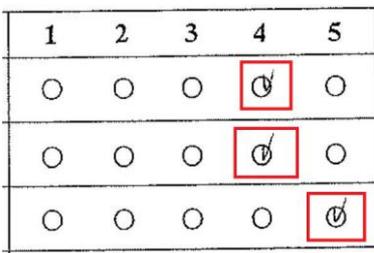
If the scanned images were rotated, our tool might encounter a problem due to the scanning or copying process. Especially, when the crop area did not contain all the options, the program could not obtain enough data for analysis (**Figure 6e**). In the future update, we will give a warning for this kind of sheet. One possible solution to this problem is using checkmarks. Checkmarks are black-filled rectangles or squares located at the corners and the margins of



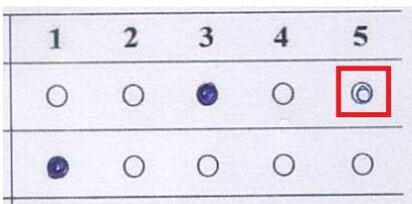
(a) No options filled



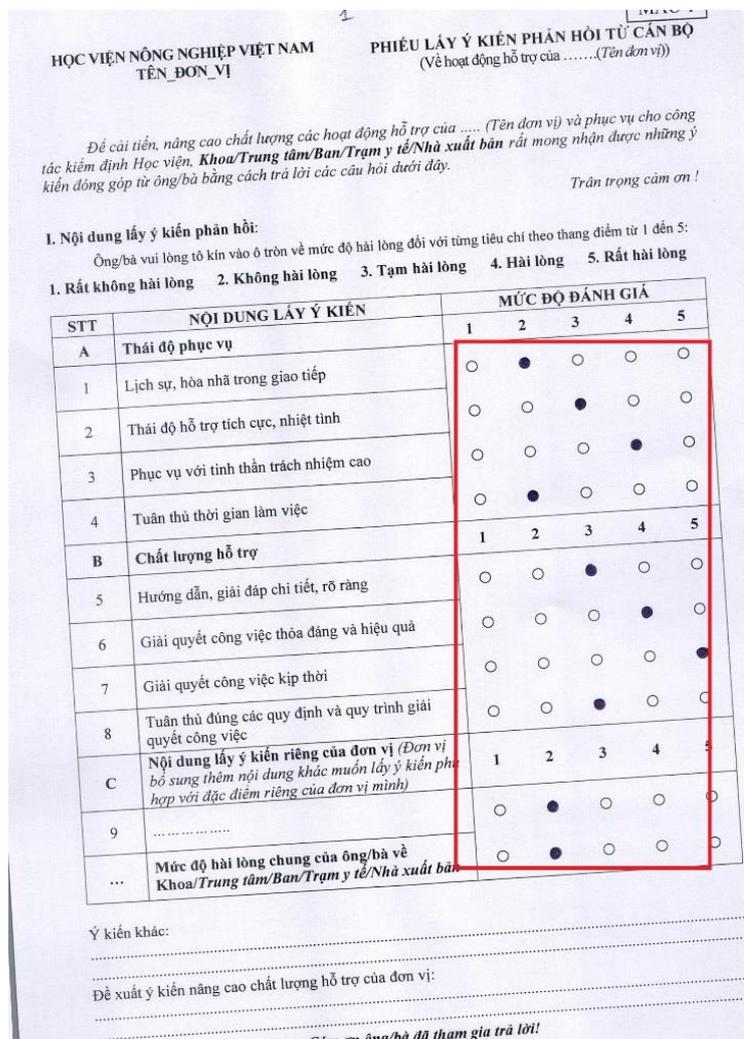
(b) Multiple options filled



(c) Checkmarks used



(d) Options not completely filled



(e) Cropping the wrong area due to image rotation

Figure 6. Problems with questionnaires and scanned images

the sheet. By first detecting checkmarks, it is possible to identify whether the sheet is rotated too much if one or more checkmarks at the corners are absent. If all of the checkmarks at four corners are detected, then we can calculate the rotate angle of the sheet. We can eventually rotate the scanned sheet in the reverse angle before detecting the options.

Conclusions

In this study, we have proposed a solution for optical mark recognition problems that do not require a dedicated machine or answer sheet. Instead, we used ordinary scanners and printers with A4 paper. We have built a software program that works with different image formats. It can detect filled options and questions with no/multiple filled options. The output of the program is in plain text and can be easily opened in various softwares, including Microsoft Excel. While other tools only work with one-page questionnaires, our tool can work with surveys that contain two pages. The first result looks promising, but still has room for improvement. Most of the questionnaires contain an area for other ideas (and comments) which may contain handwriting text. In the next version, it is our intention that our software tool will utilize the latest achievements of artificial intelligence to solve this problem or at least give users a warning about having handwriting text on questionnaires. We also want to solve the problem with rotated images. This can be done by detecting rectangles on the questionnaires. The problem now becomes selecting the right one (the rectangle that has options inside), but there are multiple and overlapping rectangles on a single sheet. Another solution for the rotating problem that we

plan to apply is using checkmarks (bold rectangles located at the corners and the margins of the questionnaires).

Acknowledgments

We would like to thank the Vietnam National University of Agriculture for funding this project.

References

- Belag I. A., Gulpete Y. & Elmanti T. M. (2018). An Image Processing Based Optical Mark Recognition with the Help of Scanner. *International Journal of Engineering Innovation and Research*. 7(2): 5.
- Educba W. (2020). Matlab Features [Online]. Retrieved from <https://www.educba.com/matlab-features/> on April 19, 2020.
- Gaikwad S. B. (2015). Image Processing Based OMR Sheet Scanning. *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*. 4.
- Haigh T. (2008). Cleve Moler: Mathematical Software Pioneer and Creator of Matlab. *IEEE Annals of the History of Computing*. 30(1): 87-91.
- Hong-Duc University. (2014). An introduction to TickREC - an automatic survey processing tool [Online]. Hong-Duc University. Retrieved from <http://hdu.edu.vn/vi-vn/4/3030/Gioi-thieu-phan-mem-xu-ly-phieu-dieu-tra-tu-dong-TickREC.html> on April 22.
- Loke S. C., Kasmiran K. A. & Haron S. A. (2018). A new method of mark detection for software-based optical mark recognition. *PLOS ONE*. 13(11): e0206420.
- Mai Ha An (2014). Research and applying image processing techniques to process the survey questionnaire on training of Vietnam forestry university. *Journal of Forestry Science and Technology*. 1(1): 6.
- Sandeep N. (2017). Introduction to MATLAB for Engineers and Scientists: Solutions for Numerical Computation and Modeling. Apress. 222 pages.