

XÂY DỰNG PHẦN MỀM CHẤM ĐIỂM THI TRẮC NGHIỆM CÓ KHẢ NĂNG HỌC MẪU PHIẾU TRẢ LỜI TRẮC NGHIỆM

DEVELOPMENT A SOFTWARE TO SCORE MULTIPLE CHOICE TEST WITH THE CAPACITY OF LEARNING MULTIPLE-CHOICE QUESTIONNAIRE

CHÂU QUANG VŨ^(*) và LÊ NHẬT TÙNG^(**)

TÓM TẮT: Hiện nay việc thi, kiểm tra và đánh giá kết quả học tập được thực hiện bởi nhiều hình thức khác nhau như thi tự luận, vấn đáp, trắc nghiệm nhằm mục đích đánh giá và phân loại người học. Trong các hình thức trên, hình thức thi trắc nghiệm dùng bút chì tô trên giấy được nhiều cơ quan và đơn vị sử dụng cụ thể như kỳ thi Trung học phổ thông quốc gia, kỳ thi đầu vào ở các trường học, thi kết thúc học kỳ,... Việc thi trắc nghiệm trên giấy cần phải có người chấm bài thủ công hoặc phần mềm chuyên dụng để chấm thi. Hiện nay, các phần mềm chấm thi thường được sử dụng với một số mẫu phiếu trả lời trắc nghiệm cố định. Trong bài viết này, tác giả đề xuất một phương pháp để xây dựng phần mềm chấm điểm thi trắc nghiệm có khả năng học mẫu phiếu trả lời trắc nghiệm bằng cách sử dụng phương pháp xử lý ảnh số nhằm tạo tính linh động và đa dạng mẫu phiếu làm bài, phù hợp với nhu cầu thực tế sử dụng của các kỳ thi khác nhau.

Từ khóa: hình thức thi trắc nghiệm, phiếu trả lời, xử lý và nhận diện hình ảnh.

ABSTRACT: The evaluation of learning outcomes is currently conducted by various methods such as assignment, oral quizzes and multiple choice tests.... Among the listed above methods, the multiple choice test is applied for many examinations, such as national high school exams, school entrance exams, and final examinations... The paper test requires a person or specialized software to mark the test. Currently, the marking software is often used with some form of fixed answer questionnaire. In this paper, the author proposes a method for developing test scoring software capable of learning multiple-choice questionnaire form using a digital image processing method for flexibility and diversity use in accordance with the actual use of the various examinations.

Key words: multiple choice test, answer sheet, processing and image recognition.

1. ĐẶT VẤN ĐỀ

Ngày nay việc thi, kiểm tra và đánh giá kết quả học tập được thực hiện trên nhiều hình thức khác nhau như thi tự luận, vấn đáp, trắc nghiệm nhằm mục đích đánh giá

và phân loại người học. Với số lượng thí sinh dự kỳ thi Trung học phổ thông Quốc gia, dự thi tuyển vào các trường đại học, các kỳ thi kết thúc khóa học ngày càng đông dẫn đến hình thức thi trắc nghiệm

(*) ThS. Trường Đại học Văn Lang, Email: chauquangvu@vanlanguni.edu.vn

(**) ThS. Trường Đại học Giao thông Vận tải, Email: lntung@utc2.edu.vn

khách quan được nhiều trường và đơn vị sử dụng.

Trắc nghiệm khách quan giúp cho các trường khảo sát được số lượng lớn thí sinh, chấm thi nhanh chóng, kết quả thi đáng tin cậy và nội dung thi có thể bao trùm toàn bộ kiến thức cần kiểm tra, tránh được tình trạng “học tủ”. Trắc nghiệm khách quan được chia ra nhiều hình thức khác nhau như: lựa chọn đáp án đúng nhất, ghép nối, điền khuyết,... trong đó trắc nghiệm bằng phương pháp đưa ra nhiều lựa chọn và yêu cầu thí sinh chọn đáp án đúng là phổ biến.

Hình thức thi trắc nghiệm chọn đáp án đúng hiện nay đang được áp dụng cho kỳ thi đầu vào các trường đại học, thi tốt nghiệp,... Hình thức này yêu cầu thí sinh sử dụng bút chì có độ đậm phù hợp tô kín các ô tròn bao gồm mã đề thi, mã thí sinh và các ô trả lời tương ứng với các câu trắc nghiệm.

Để thực hiện việc chấm bài thi trắc nghiệm các trường thường sử dụng các máy quét bài thi trắc nghiệm rồi sử dụng phần mềm chuyên dụng để xử lý. Phần mềm cần phải nhận dạng được số báo danh, mã đề,

cũng như các lựa chọn của thí sinh cho các câu trả lời. Việc sử dụng một số mẫu phiếu cố định giúp cho phần mềm xử lý nhanh chóng, tuy nhiên trong thực tế tùy theo kỳ thi, đối tượng tham gia thi, các mẫu phiếu trả lời trắc nghiệm sẽ rất đa dạng.

Trong nghiên cứu này, tác giả tập trung đề xuất một phương pháp xây dựng phần mềm chấm điểm thi có khả năng học mẫu phiếu trắc nghiệm với các thành phần chính như: Học khu vực mã đề, học khu vực số báo danh, khu vực lựa chọn đáp án của các câu hỏi. Để xây dựng và phát triển, chúng tôi đã sử dụng các kỹ thuật xử lý ảnh với thư viện OpenCV và ngôn ngữ lập trình Java.

2. PHÂN TÍCH VÀ ĐỀ XUẤT THUẬT TOÁN

2.1. Dữ liệu đầu vào

Dữ liệu đầu vào là các tập tin ảnh bài thi trắc nghiệm, các tập tin này được quét bằng các máy quét, thông thường các mẫu phiếu trả lời trắc nghiệm sẽ có bốn ô hình chữ nhật ở bốn góc hỗ trợ việc cho việc tiền xử lý cũng như nhận dạng các câu trả lời của thí sinh:

Hình 1. Ảnh phiếu bài thi trắc nghiệm

Trong một phiếu trả lời trắc nghiệm thường sẽ có các thành phần chính sau:

Phần ghi thông tin thí sinh, chữ ký của cán bộ coi thi: đây là phần không cần xử lý vì các thông tin trên không ảnh hưởng trực tiếp đến kết quả của việc chấm thi.

Phần số báo danh: đây là phần mà sinh viên phải tô chọn số báo danh dự thi của mình, tùy theo kỳ thi và mẫu phiếu mà số lượng lựa chọn, độ dài của số báo danh sẽ khác nhau.

Phần mã đề thi: đây là phần mà sinh viên phải tô chọn, dựa trên mã đề thi nhận dạng phần mềm sẽ lựa chọn bộ đáp án tương ứng từ cơ sở dữ liệu để chấm điểm bài thi.

Phần trả lời các câu hỏi: đây là phần mà sinh viên phải tô chọn, tùy theo mẫu phiếu trả lời, số lượng câu hỏi sẽ khác nhau, thậm chí tùy theo yêu cầu của đề thi, mẫu phiếu sẽ có số lượng sự lựa chọn trong một câu hỏi khác nhau. Ví dụ như: mẫu phiếu có 2 lựa chọn, có 4 lựa chọn, có 5 lựa chọn,... trong một câu hỏi.

2.2. Tiền xử lý ảnh mẫu phiếu trả lời trắc nghiệm

Hình ảnh bài thi trắc nghiệm tiền xử lý là hình ảnh chưa được chỉnh sửa ở bất kỳ phương diện nào. Việc tiền xử lý ảnh bài thi trắc nghiệm nhằm làm cho ảnh tốt hơn, giúp quá trình học mẫu phiếu hoặc chấm điểm thi được chính xác hơn.

Bước 1: Nhị phân ảnh bài thi trắc nghiệm trên giấy

Ảnh bài thi nhận được từ thiết bị thu nhận hình ảnh là máy scan thường là ảnh màu hay ảnh đa mức xám, các thành phần trong ảnh rất phức tạp về màu sắc và kết cấu. Do đó, muốn làm nổi bật các đặc trưng

trong ảnh thì phải chuyển về dạng ảnh nhị phân, ảnh chỉ có hai màu (đen và trắng).

Đối với ảnh màu hệ RGB, ta cần chuyển ảnh về dạng đa mức xám, việc chuyển ảnh màu về ảnh xám được thực hiện theo công thức sau [1]:

$$g(x,y) = fR(x,y) * 0.287 + fG(x,y) * 0.599 + fB(x,y) * 0.114$$

Trong đó: x,y : tọa độ của điểm ảnh; $fR,G,B(x,y)$: lần lượt là thành phần màu R, G, B tại điểm (x,y) ; $g(x,y)$: ảnh đầu ra (ảnh đa mức xám); $f(x,y)$: là ảnh đầu vào (ảnh màu).

Trong nghiên cứu này, chúng tôi thực hiện việc nhị phân ảnh dựa trên phương pháp xác định ngưỡng tự động Otsu, được tác giả Nobuyuki Otsu giới thiệu năm 1979 [2].

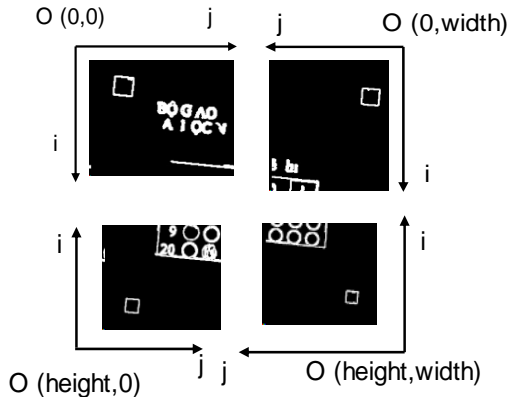
Một điểm trong ảnh nhị phân $B(x, y)$ tại vị trí (x, y) được xác định thông qua cường độ ảnh xám và công thức dưới đây với k là ngưỡng tự động được chọn thông qua phương pháp Otsu:

$$B(x,y) = \begin{cases} 0 & \text{nếu } I(x,y) < k \\ 1 & \text{nếu } I(x,y) \geq k \end{cases} \quad (2)$$

Bước 2: Hiệu chỉnh góc nghiêng

Đối với những chương trình nhận dạng, có ảnh đầu vào là những trang tài liệu dạng văn bản thì các ảnh thu nhận được thường bị lệch so với ảnh gốc một góc bất kỳ. Nguyên nhân là do trong quá trình thu nhận: ảnh gốc bị đặt lệch, thiết bị ghi nhận hình ảnh đặt không đúng vị trí hay thu nhận ảnh bị xô dịch,... Các yếu tố gây lệch ảnh này là không thể tránh khỏi trong quá trình xử lý ảnh. Do đó, để cho các bước học mẫu phiếu trắc nghiệm được chính xác cần phải có thao tác hiệu chỉnh độ nghiêng của ảnh bài thi trắc nghiệm thu nhận được.

Ảnh bài thi trắc nghiệm của hầu hết các mẫu phiếu thường được đánh dấu bởi bốn hình chữ nhật ở các góc, thông qua việc xác định tọa độ của các hình chữ nhật này ta có thể xác định được góc nghiêng của ảnh bài thi trắc nghiệm.



Hình 2. Ảnh xác định các hình chữ nhật nhận dạng ở các góc

Sau các bước nhận dạng các ô hình chữ nhật ở trên, giả sử ta có tọa độ tâm của hình chữ nhật góc trái dưới là $A(x_A, y_A)$ và tọa độ tâm của hình chữ nhật góc phải dưới là $B(x_B, y_B)$. Để tìm được góc nghiêng của ảnh, chúng ta cần phải tìm đường góc nghiêng giữa đường thẳng AB và đường thẳng song song với trục tọa độ Ox trên hệ trục tọa độ chuẩn Oxy.

Đường thẳng song song với hệ trục tọa độ Ox:

$$y = 0 \quad (3)$$

Xây dựng đường thẳng AB:

$$\frac{(x_A - x)}{(x_A - x_B)} = \frac{(y_A - y)}{(y_A - y_B)} \quad (4);$$

$$\text{Đặt: } \begin{cases} a = (x_A - x_B) \\ b = (y_A - y_B) \end{cases} \quad (5)$$

Từ (4) và (5) ta có:

$$-bx + ay + (bx_A - ay_A) = 0 \quad (6)$$

Từ (3) ta có vector pháp tuyến n_1 là $(0,1)$ (7)

Từ (6) ta có vector pháp tuyến n_2 là $(-b,a)$ (8)

Việc tìm góc của hai đường thẳng trở nên đơn giản bằng cách tìm góc giữa 2 vector pháp tuyến.

Gọi α là góc giữa vector (7) và (8), ta có:

$$\cos(\alpha) = \frac{|0(-b) + 1a|}{\sqrt{0^2 + 1^2} \sqrt{(-b)^2 + a^2}} = \frac{|a|}{\sqrt{(-b)^2 + a^2}} \quad (9)$$

Từ (9) ta suy ra góc giữa 2 đường thẳng cũng là góc nghiêng:

$$\alpha = \arccos\left(\frac{|a|}{\sqrt{(-b)^2 + a^2}}\right) \quad (10)$$

Nếu $y_A < y_B$, ảnh bị nghiêng về bên trái, ta cần xoay ảnh theo chiều bên phải một góc có giá trị là α . Nếu $y_A > y_B$, ảnh bị nghiêng về bên phải, ta cần xoay ảnh theo chiều bên phải một góc có giá trị là $-\alpha$. Trục xoay là trọng tâm của ảnh bài thi trắc nghiệm trên giấy.

Bước 3: Cắt xén, loại bỏ vùng ngoài ảnh bài thi trắc nghiệm

Tại bước này, chúng ta sẽ loại bỏ thành phần nằm bên ngoài các hình chữ nhật nhận dạng, mục tiêu giúp phần mềm chỉ tập trung vùng liên quan đến các nội dung của bài thi. Hình ảnh nội dung liên quan đến phiếu trả lời trắc nghiệm nằm bên trong khu vực các hình chữ nhật đánh dấu nhận dạng.

Bước 4: Tách biên

Biên là một phần quan trọng trong xử lý ảnh số. Đối với ảnh bài thi trắc nghiệm trên giấy, việc tách biên giúp cho công đoạn nhận dạng các đặc trưng của bài thi được thuận lợi hơn. Cơ sở của phép toán phát hiện biên là quá trình biến đổi về giá trị độ sáng của các điểm ảnh. Trong nghiên cứu này, chúng tôi sử dụng phương pháp tách biên Canny [3], một phương pháp tốt và thường được sử dụng hiện nay.

PHIẾU TRẢ LỜI CÂU HỎI TRẮC NGHIỆM

Họ và tên: Lên Văn Cường.....

Mã số sinh viên: 565.467.408.6.....

Môn thi: Công nghệ Java.....

Ngày thi: 19-6-2019.....

Chữ ký Giám thị 1: [Chữ ký]

Chữ ký Giám thị 2: [Chữ ký]

Lưu ý:

- Sinh viên ghi và tô đầy đủ mã đề thi, MSV.
- Sinh viên tô các ô đáp án lựa chọn, không đánh x hoặc khoanh tròn.
- Giám thị ký và ghi tên vào đúng vị trí, không ký tên ở vị trí khác.

MÃ ĐỀ THI				MÃ SINH VIÊN			
0	1	2	3	4	5	6	7
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	

Bán kính trung bình của các đường tròn phát hiện được xem là bán kính chuẩn của các đường tròn lựa chọn trong bài thi. Tìm bán kính trung bình của các đường tròn bằng công thức sau:

$$avR = \frac{\sum_{i=1}^n R(i)}{n} \quad (11)$$

Trong đó: avR là bán kính trung bình của các đường tròn, $R(i)$ là bán kính đường tròn thứ i , n là tổng số đường tròn được phát hiện.

Thông tin tọa độ tâm (x, y) , bán kính R , số thứ tự, nhãn và giá trị lựa chọn (trong quá trình học mẫu phiếu, ta không quan tâm đến nhãn và giá trị lựa chọn). Sau khi xác định được các thông số của một đường tròn, chúng ta sắp xếp thứ tự của các đường tròn tăng dần từ trái sang phải và từ trên xuống dưới theo phương thẳng đứng của bài thi.

Bước 2: Học các khu vực số báo danh, mã đề và khu vực bài làm

Đối với từng khu vực riêng biệt, người sử dụng có thể đánh dấu vị trí bắt đầu và kết thúc của từng khu vực dựa trên nhãn của các hình tròn.

Một khu vực bao gồm: Nhãn bắt đầu; Nhãn kết thúc; Số lượng câu hỏi; Số lượng lựa chọn của một câu hỏi.

Quá trình nhận dạng các đường tròn ban đầu có thể dư thừa hoặc thiếu đi một số lựa chọn. Tại bước này ta cần thực hiện việc tái tạo nhằm mục đích loại bỏ các đường tròn bị nhận dạng sai và bổ sung các đường tròn bị sót khi thực hiện nhận dạng.

Trong một khu vực học mẫu phiếu cụ thể, từ tập hợp các đường tròn theo chiều ngang của mẫu phiếu, ta tính khoảng cách trung bình theo chiều ngang giữa các tâm đường tròn liên tiếp bằng công thức sau:

$$avD_x = \frac{\sum_{i=1}^n D_x}{n} = \frac{\sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2}}{n} \quad (12)$$

Trong đó: n là số lượng đường tròn được phát hiện theo chiều ngang; x_i là tọa độ x của đường tròn thứ i ; avD_x là khoảng cách trung bình của các đường tròn theo chiều ngang.

Trong một khu vực học mẫu phiếu cụ thể, từ tập hợp các đường tròn theo chiều dọc của mẫu phiếu, ta tính khoảng cách trung bình theo chiều dọc giữa các tâm đường tròn liên tiếp bằng công thức sau:

$$avD_y = \frac{\sum_{i=1}^m D_y}{m} = \frac{\sum_{i=1}^{m-1} \sqrt{(y_{i+1} - y_i)^2}}{m} \quad (13)$$

Trong đó: m là số lượng đường tròn được phát hiện theo chiều dọc; y_i là tọa độ y của đường tròn thứ i ; avD_y là khoảng cách trung bình của các đường tròn theo chiều dọc.

Từ (11), (12), (13) thực hiện việc bố trí lại các đường tròn thành lưới như sau:

for i=1 to So_Luong_Cau_Hoi

for j=1 So_Luong_Lua_Chon

$$x[i,j] = x[1,1] + i*avD_x$$

$$y[i,j] = y[1,1] + j*avD_y$$

$$r[i,j] = avR$$

Trong đó: $x[i,j]$: là tọa độ x của tâm đường tròn thứ i, j ; $y[i,j]$: là tọa độ y của tâm đường tròn thứ i, j ; $r[i,j]$: là bán kính đường tròn thứ i, j .

Toàn bộ các khu vực đã học được sẽ được lưu thành một tập tin, từ tập tin này ta sẽ tiến hành chấm thi cho hàng loạt các bài thi sử dụng mẫu phiếu tương tự.

2.4. Xử lý bài thi và chấm điểm

Dữ liệu đầu vào: thư mục ảnh bài thi, tập tin chứa đáp án, tập tin chứa thông tin mẫu phiếu đã học tại mục 2.3.

Bước 1: Thực hiện các bước tiền xử lý tương tự mục 2.2 đối với từng ảnh bài thi.

Bước 2: Thực hiện xác định sự lựa chọn của thí sinh dựa trên việc xác định số pixel có trên mỗi ô tròn, nếu xác định khu vực được tô nhiều hơn $\frac{1}{2}$ diện tích ô tròn thì ta xem xét giá trị tại ô tròn đã được lựa chọn.



Hình 7. Ảnh mẫu phiếu đã học

Bước 3: Ta đã có chuỗi các đáp án mà thí sinh đã lựa chọn, mã đề thi. Việc tính điểm của câu được thực hiện bằng việc so khớp lựa chọn của thí sinh với tập tin chứa đáp án với mã đề thi tương ứng.

Bước 4: Lưu trữ toàn bộ kết quả chấm điểm.

STĐ	tID	Correct	Score	choose
5651071041	02	15	2.5	C C A B B B
5651071046	01	40	6.67	B A B D C C
5651071044	03	27	4.5	C B A B B B
5651071035	06	29	4.83	A B C D B A
5651071006	02	30	5.0	B D C C C B
5651071016	02	33	5.5	B D A A B C
5651071029	09	31	5.17	A C A B B A
5651071023	08	31	5.17	B B A C C B
5651071047	04	32	5.33	C B A C C B
5651071040	03	26	4.33	C B C B B A
5651071017	08	45	7.5	B D B B B C
5651071003	09	32	5.33	C B D D B D
5156071011	07	32	5.33	D B D C C A
5651071018	06	41	6.83	A D B D C C
5651071025	10	37	6.17	D B D C C C
5651071043	10	27	4.5	C B B B C C
5651071019	04	32	6.83	D B D C C A
5651071037	06	31	5.17	A B B D B A
5651071021	04	36	6.0	C B B C A D
5651071013	02	24	4.0	C B C A C B
5651071019	08	33	5.5	B A B C B A
5651071034	03	37	4.5	D B C C B D
5651071009	05	34	5.67	B A B C C A
5651071005	05	18	3.0	B C B A C A
5651071028	07	34	5.67	B B D D B A
5651071010	10	25	4.17	C D C C A C
5651071027	04	28	4.67	C C B B C B
5651071024	01	39	6.5	A A B D C C
5651071036	07	38	6.33	C A B D C A
5651071020	09	23	3.83	C C A D B B
5651071026	05	36	6.0	C A C B C A

Hình 8. Mẫu phiếu lưu kết quả chấm điểm

3. THỰC NGHIỆM

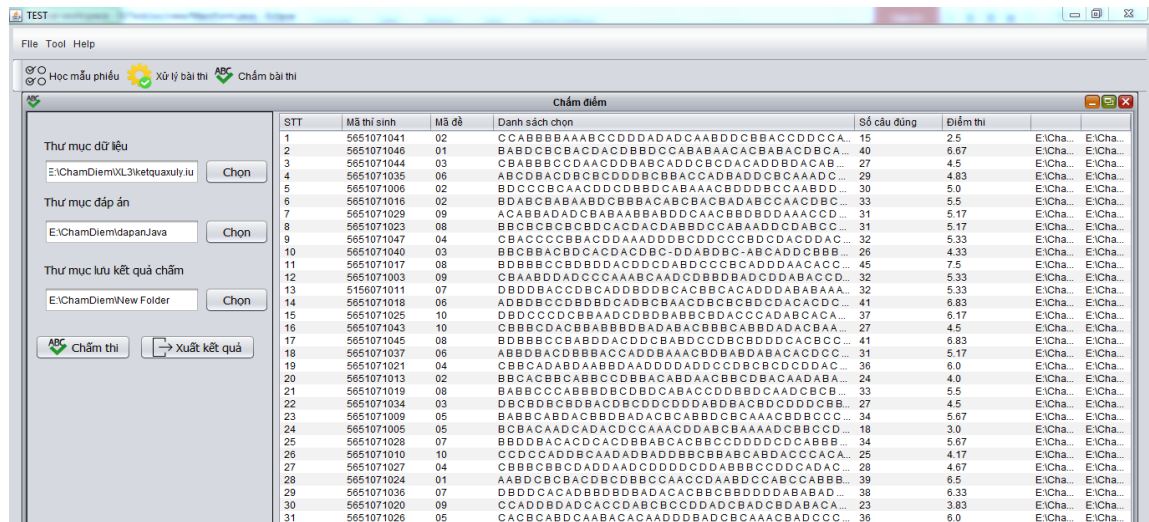
Để kiểm chứng hiệu quả của hệ thống và thuật toán đề xuất, chúng tôi tiến hành thực nghiệm dựa trên ngôn ngữ lập trình Java và thư viện xử lý ảnh OpenCV. Tập cơ sở dữ liệu gồm 5 loại mẫu phiếu trắc nghiệm khác nhau, mỗi loại gồm 30 phiếu

làm bài.

Kết quả học mẫu phiếu cho kết quả tốt đối với cả phiếu màu và phiếu trắng đen. Tỷ lệ học mẫu phiếu chính xác là 100%.

Tỷ lệ nhận dạng và chấm điểm bài thi với tỷ lệ chính xác là:

STT	Mẫu phiếu	Số phiếu làm bài	Tỷ lệ phiếu chấm chính xác
1	Mẫu phiếu 1	30	96.7%
2	Mẫu phiếu 2	30	100%
3	Mẫu phiếu 3	30	100%
4	Mẫu phiếu 4	30	100%
5	Mẫu phiếu 5	30	96.7%



Hình 10. Hình phần mềm tại bước chấm bài thi

Một số khó khăn trong quá trình xử lý: ảnh bài thi trắc nghiệm bị nhàu nát, khu vực nhận dạng bị nhiễu do thí sinh làm bài không cẩn trọng. Việc xử lý ảnh đối với các bài thi bị nhàu nát sẽ tạo ra các bóng trên bài thi dẫn đến việc hệ thống ghi nhận hình ảnh không chính xác. Các bài thi làm không cẩn thận, các đáp án hủy không được tẩy xóa kỹ lưỡng sẽ tạo nên những sự cố ghi nhận, nhiều hơn 01 đáp án cho 1 câu hỏi. Với các bài thi có ghi nhận nhiều hơn 1 đáp án cho 1 câu hỏi, hệ thống sẽ đánh dấu lại bài thi đó và thông báo cho giám khảo để có thể chấm lại bài thi đó.

4. KẾT LUẬN

Qua nghiên cứu này, chúng tôi đã đề xuất một hướng tiếp cận trong việc xây dựng phần mềm chấm điểm thi trắc nghiệm có khả năng học mẫu phiếu. Các bước tiền xử lý ảnh có thể hỗ trợ cho việc nghiên cứu xây dựng dựng các ứng dụng chấm điểm thi trắc nghiệm.

Ưu điểm của phần mềm là có khả năng ứng dụng trong thực tiễn, tỷ lệ chính xác cao, áp dụng được cho nhiều loại mẫu

phiếu trả lời trắc nghiệm phổ biến hiện nay. Các loại phiếu có kích cỡ, màu sắc khác nhau phần mềm vẫn xử lý và nhận dạng với kết quả tốt.

Tuy nhiên thuật toán đề xuất có phần hạn chế vì chỉ học 3 khu vực là số báo danh, mã đề, khu vực làm bài. Đối với một số bài thi quốc tế như TOEFL hoặc TOEIC thì phiếu trả lời trắc nghiệm còn có thêm các mục nhận dạng khác. Chưa giải quyết được tình huống mẫu phiếu không có điểm nhận dạng cũng như các hình ảnh phiếu trắc nghiệm bị nhiễu nặng.

Thuật toán đề xuất chưa giải quyết tốt trên các bài thi nhàu nát, thí sinh chưa xóa kỹ đáp án không sử dụng. Vấn đề về các bài nhàu nát có thể phân loại thủ công để có thể chấm điểm thủ công.

Nội dung nghiên cứu này có thể tiếp tục được mở rộng và đẩy nhanh tốc độ xử lý bằng cách xử lý nhận dạng các khối ảnh, ngoài ra hướng phát triển các ứng dụng chấm điểm thi trực tiếp trên thiết bị di động với máy ảnh được tích hợp cũng là hướng nghiên cứu mà chúng tôi quan tâm.

TÀI LIỆU THAM KHẢO

1. Charles Poynton (1996), *The magnitude of nonconstant luminance errors in Charles Poynton, A Technical Introduction to Digital Video*, New York: John Wiley & Sons.
2. Zhang, Jun & Hu, Jinglu (2008), *Image segmentation based on 2D Otsu method with histogram analysis*, *Computer Science and Software Engineering*, 2008 International Conference on. 6.
3. Green, B. (2002), *Canny Edge Detection Tutorial*, Retrieved December 3, 2014.

Ngày nhận bài: 08/08/2017. Ngày biên tập xong: 30/09/2017. Duyệt đăng: 18/10/2017