**TRƯỜNG ĐẠI HỌC VINH**

**VIỆN KỸ THUẬT VÀ CÔNG NGHỆ**

****

**ĐỒ ÁN HỌC PHẦN**

**THỊ GIÁC MÁY TÍNH**

**ĐỌC THÔNG TIN TỪ**

**CĂN CƯỚC CÔNG DÂN**

**NHÓM: 03**

|  |  |
| --- | --- |
| GVHD: | ThS. Nguyễn Thị Minh Tâm |
| SVTH: | Lê Xuân Lộc, 215748010310023 |
|  | Lê Việt Bách, 215748010310010  Nguyễn Văn Trường, 215748010310015 |
|  |  |

**Nghệ An, 11/202****4**

# MỞ ĐẦU

Không có sự thành công nào mà không gắn liền với những nỗ lực, cố gắng, sự hỗ trợ, giúp đỡ dù ít hay nhiều, dù trực tiếp hay gián tiếp. Là cả quá trình phấn đấu không ngừng nghỉ bên cạnh những hướng dẫn, chỉ bảo tận tình, gắn bó và tình cảm sâu sắc.

Để hoàn thành tốt để tài và bài báo cáo này, chúng em xin gửi lời cảm ơn chân thành đến giảng viên, cô Nguyễn Thị Minh Tâm đã trực tiếp hỗ trợ chúng em trong suốt quá trình làm đề tài. Chúng em cảm ơn cô đã đưa ra những lời khuyên từ kinh nghiệm thực tiễn của mình để định hướng cho chúng em đi đúng với yêu cầu của đề tài đã chọn, luôn giải đáp thắc mắc và đưa ra những góp ý, chỉnh sửa kịp thời giúp chúng em khắc phục nhược điểm và hoàn thành tốt cũng như đúng thời hạn đã đề ra.

Đề tài và bài báo cáo được chúng em thực hiện với những hạn chế về mặt kiến thức, kĩ thuật và cũng như kinh nghiệm thực tế. Do đó, trong quá trình làm nên đề tài có những thiếu sót là điều không thể tránh khỏi nên chúng em rất mong nhận được những ý kiến đóng góp quý báu của cô để kiến thức của chúng em được hoàn thiện hơn và chúng em có thể làm tốt hơn nữa trong những lần sau.

# LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành đến ThS. Nguyễn Thị Minh Tâm đã dành thời gian và công sức để hướng dẫn, giảng dạy và truyền đạt kiến thức cho chúng em trong suốt quá trình học tập. Sự am hiểu, sự tận tâm và sự chia sẻ của thầy đã giúp chúng em tiếp thu và hiểu bài học một cách dễ dàng hơn.

Chúng em rất biết ơn cô vì cách thức giảng dạy tận tâm và sự chia sẻ kiến thức thực tế của mình. Cô đã tạo ra một môi trường học tập tích cực, nơi chúng em có thể thảo luận, trao đổi ý kiến và phát triển khả năng của mình. Cô cung cấp cho chúng tôi những bài giảng thú vị và đa dạng, giúp chúng em hiểu rõ và ứng dụng kiến thức vào thực tế.

Chúng em cũng nhận thức rằng nhóm còn nhiều hạn chế và khó khăn trong việc tiếp thu và áp dụng kiến thức. Tuy nhiên, sự hướng dẫn và những lời khuyên của thầy đã giúp chúng em vượt qua những thách thức đó. Chúng em cam kết sẽ tiếp tục nỗ lực hơn nữa để phát triển và áp dụng kiến thức một cách hiệu quả.

Cuối cùng, chúng em xin chân thành cảm ơn thầy vì sự tận tâm và mong muốn giúp đỡ chúng em trong quá trình học tập. Chúng em hứa sẽ tiếp thu và áp dụng những kiến thức đã học để đóng góp vào sự phát triển của xã hội và đạt được thành công trong sự nghiệp sau này.

Chúng em xin chân thành cảm ơn thầy!

*Nghệ An, ngày 6 tháng 11 năm 2024*

**Nhóm Sinh Viên**

Lê Xuân Lộc

Lê Việt Bách

Nguyễn Văn Trường

**CHƯƠNG 1: MÔ TẢ BÀI TOÁN ĐỌC THÔNG TIN TỪ**

**CĂN CƯỚC CÔNG DÂN**

## 1.1. Bài toán xây dựng đọc thông tin từ căn cước công dân

Thực tế cho ta thấy rằng chứng minh thư hay căn cước công dân là một phần không thể thiếu trong xã hội ngày nay. Chứng minh thư hay căn cước công dân có giá trị chứng minh về căn cước, lai lịch của công dân của người được cấp để thực hiện các giao dịch hành chính trên toàn lãnh thổ Việt Nam, có thể được sử dụng thay hộ chiếu trong trường hợp Việt Nam và một nước khác có điều ước hoặc thỏa thuận quốc tế cho phép công dân hai bên sử dụng thẻ Căn cước công dân thay cho việc sử dụng hộ chiếu.

Có thể thấy được rằng căn cước công dân rất quan trọng nhưng để xử lý một số lượng thông tin lớn từ căng cước nếu như chúng ta làm thủ công thì sẽ mất rất nhiều thời gian và nhân lực. Ví dụ như năm 2021 Nước Cộng hoà xã hội chủ nghĩa Việt Nam thực hiện chuyển đổi căn cước công dân từ không có chip thành căn cước công dân có chip mất rất nhiều thời gian. Dẫn đến xảy ra rất nhiều việc không đáng có. Điển hình như việc người dân xếp hàng dài từ sáng đến đêm nhưng chưa được làm. Các cán bộ chiến sĩ công an làm tăng ca qua giờ hành chính đến đêm nhưng vẫn không thể xử lý được số lượng thông tin người dân đến đổi thẻ. Dẫn đến hậu quả là có những người làm căn cước cả năm vẫn chưa được nhận.

Thứ 2 các ngân hàng tốn rất nhiều nhân lực để thực hiện việc lấy thông tin khách hàng từ chứng minh thư. Việc này tốn rấn nhiều thời gian của cả khách hàng lẫn nhân viên ngân hàng dẫn đến việc làm chậm hệ thống của cả ngân hàng.

Tổng quan lại thì vấn đề chung ở đây là mất quá nhiều thời gian để lấy thông tin cá nhân từ chứng minh thư.

Ngoài ra thì một số công ty hay một số cá nhân cần thông tin từ chứng minh thư cho công việc mà họ đang làm.

Hoặc đơn giản hơn nữa thì là cá nhân họ chỉ muốn lấy ra thông tin của mình để tiện cho công việc.

### CHƯƠNG 2: CÁC THUẬT TOÁN SỰ DỤNG

### 1.2. Các thuật toán hiện có

### 1.2.1. Video Capture

Sử dụng cv2.VideoCapture() để truyền phát video từ một URL đã chỉ định. Các khung video được hiển thị bằng cv2.imshow().

**1.2.2. Document Detection**

***1.2.2.1. Tiền xử lý***

Chuyển đổi khung sang thang độ xám bằng cv2.cvtColor(). Áp dụng lọc song phương bằng cv2.bilateralFilter() để giảm nhiễu trong khi vẫn giữ nguyên các cạnh.

***1.2.2.2. Phát hiện cạnh***

Sử dụng thuật toán phát hiện cạnh Canny (cv2.Canny()) để phát hiện các cạnh trong ảnh.

***1.2.2.3. Phát hiện đường viền***

cv2.findContours() được sử dụng để tìm các đường viền trong ảnh có cạnh. Các đường viền được sắp xếp theo diện tích để tìm các đường viền lớn nhất có thể biểu diễn cho tài liệu.

***1.2.2.4. Xấp xỉ đường viền***

Sử dụng cv2.approxPolyDP() để xấp xỉ các đường viền và xác định các đường viền có bốn góc (có thể là tài liệu).

***1.2.3. Sắp xếp các điểm***

Hàm order\_points() sắp xếp bốn góc theo một trình tự cụ thể (trên cùng bên trái, trên cùng bên phải, dưới cùng bên phải, dưới cùng bên trái) để đơn giản hóa quá trình biến đổi phối cảnh.

***1.2.4. Chuyển đổi phối cảnh***

Sử dụng cv2.getPerspectiveTransform() và cv2.warpPerspective() để áp dụng phép chuyển đổi "làm phẳng" vùng tài liệu được phát hiện thành chế độ xem chuẩn hóa.

***1.2.5. Nhận dạng kí tự quang học***

Phương thức pytesseract.image\_to\_string() từ thư viện pytesseract trích xuất văn bản từ hình ảnh đã chuyển đổi. Được cấu hình với --psm 6 để đặt chế độ phân đoạn OCR để xử lý hình ảnh như một khối văn bản.

***1.2.6. Phân tích văn bản***

Hàm parse\_ocr\_text() sử dụng biểu thức chính quy (re.search()) để trích xuất các trường dữ liệu cụ thể từ đầu ra OCR: Số ID: Tìm kiếm mẫu 12 chữ số. Họ và tên: Trích xuất tên viết hoa. Ngày sinh: Phù hợp với định dạng ngày (ví dụ: dd/mm/yyyy). Dữ liệu được lưu vào tệp văn bản bằng các phương thức open() và write() của Python.

## CHƯƠNG 3: GIAO DIỆN KẾT QUẢ

**3.1. Ảnh sau khi được cắt 4 góc**

**A close-up of a id card

Description automatically generated**

**3.2. Sau khi đọc cho kết quả**

A computer screen with white text

Description automatically generated

**3.3. Sau khi cắt thông tin thêm vào file txt**

A screenshot of a computer

Description automatically generated

# TÀI LIỆU THAM KHẢO

[1] Lê Văn Phùng, *Kỹ nghệ phần mềm*, NXB Thông tin và Truyền thông, 2014.

[2] Lê Văn Phùng, Lê Hương Giang, *Kỹ nghệ phần mềm nâng cao*, NXB Thông tin và truyền thông, 2015.

[3] Đoàn Văn Ban, Nguyễn Thị Tĩnh, *Giáo trình phân tích thiết kế hệ thống hướng đối tượng bằng UML*, NXB Đại học sư phạm, 2011.

[4] Ian Sommerville*, Software Engineering*, Ninth Edition, Addison-Wesley, 2011.

[5] Scott Tilley, Harry J. Rosenblatt, *Systems Analys and Design*, Shelly Cashman Series, 11th Edition, 2016.