|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  **VIỆN ĐIỆN TỬ - VIỄN THÔNG**  logo_128  ĐỒ ÁN  **TỐT NGHIỆP ĐẠI HỌC**  **Đề tài:**  **THIẾT KẾ ỨNG DỤNG XỬ LÝ ẢNH ĐỂ TRÍCH**  **XUẤT THÔNG TIN TỪ CĂN CƯỚC CÔNG DÂN**   |  |  | | --- | --- | | Sinh viên thực hiện: | PHẠM ĐỨC HẢI | |  | Lớp ĐT09 – K60 | | Giảng viên hướng dẫn: | TS. PHẠM DOÃN TĨNH |     Hà Nội, 7-2020 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  **VIỆN ĐIỆN TỬ - VIỄN THÔNG**  logo_128  ĐỒ ÁN  **TỐT NGHIỆP ĐẠI HỌC**  **Đề tài:**  **THIẾT KẾ ỨNG DỤNG XỬ LÝ ẢNH ĐỂ TRÍCH**  **XUẤT THÔNG TIN TỪ CĂN CƯỚC CÔNG DÂN**     |  |  | | --- | --- | | Sinh viên thực hiện: | PHẠM ĐỨC HẢI | |  | Lớp ĐT09 – K60 | | Giảng viên hướng dẫn: | TS. PHẠM DOÃN TĨNH | | Cán bộ phản biện: |  |     Hà Nội, 7-2020 |

**ĐÁNH GIÁ QUYỂN ĐỒ ÁN TỐT NGHIỆP**

(Dùng cho giảng viên hướng dẫn)

Tên giảng viên đánh giá: TS. Phạm Doãn Tĩnh

Họ và tên sinh viên:Phạm Đức Hải MSSV:20151215

Tên đồ án: **Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân**

**Chọn các mức điểm phù hợp cho sinh viên trình bày theo các tiêu chí dưới đây:**

Rất kém (1); Kém (2); Đạt (3); Giỏi (4); Xuất sắc (5)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Có sự kết hợp giữa lý thuyết và thực hành (20)** | | | | | | |
| 1 | Nêu rõ tính cấp thiết và quan trọng của đề tài, các vấn đề và các giả thuyết (bao gồm mục đích và tính phù hợp) cũng như phạm vi ứng dụng của đồ án | 1 | 2 | 3 | 4 | 5 |
| 2 | Cập nhật kết quả nghiên cứu gần đây nhất (trong nước/quốc tế) | 1 | 2 | 3 | 4 | 5 |
| 3 | Nêu rõ và chi tiết phương pháp nghiên cứu/giải quyết vấn đề | 1 | 2 | 3 | 4 | 5 |
| 4 | Có kết quả mô phỏng/thực nghiệm và trình bày rõ ràng kết quả đạt được | 1 | 2 | 3 | 4 | 5 |
| **Có khả năng phân tích và đánh giá kết quả (15)** | | | | | | |
| 5 | Kế hoạch làm việc rõ ràng bao gồm mục tiêu và phương pháp thực hiện dựa trên kết quả nghiên cứu lý thuyết một cách có hệ thống | 1 | 2 | 3 | 4 | 5 |
| 6 | Kết quả được trình bày một cách logic và dễ hiểu, tất cả kết quả đều được phân tích và đánh giá thỏa đáng | 1 | 2 | 3 | 4 | 5 |
| 7 | Trong phần kết luận, tác giả chỉ rõ sự khác biệt (nếu có) giữa kết quả đạt được và mục tiêu ban đầu đề ra đồng thời cung cấp lập luận để đề xuất hướng giải quyết có thể thực hiện trong tương lai | 1 | 2 | 3 | 4 | 5 |
| **Kỹ năng viết quyển đồ án (10)** | | | | | | |
| 8 | Đồ án trình bày đúng mẫu quy định với cấu trúc các chương logic và đẹp mắt (bảng biểu, hình ảnh rõ ràng, có tiêu đề, được đánh số thứ tự và được giải thích hay đề cập đến; căn lề thống nhất, có dấu cách sau dấu chấm, dấu phảy v.v.), có mở đầu chương và kết luận chương, có liệt kê tài liệu tham khảo và có trích dẫn đúng quy định | 1 | 2 | 3 | 4 | 5 |
| 9 | Kỹ năng viết xuất sắc (cấu trúc câu chuẩn, văn phong khoa học, lập luận logic và có cơ sở, từ vựng sử dụng phù hợp v.v.) | 1 | 2 | 3 | 4 | 5 |
| **Thành tựu nghiên cứu khoa học (5)** *(chọn 1 trong 3 trường hợp)* | | | | | | |
| 10a | Có bài báo khoa học được đăng hoặc chấp nhận đăng/Đạt giải SVNCKH giải 3 cấp Viện trở lên/Có giải thưởng khoa học (quốc tế hoặc trong nước) từ giải 3 trở lên/Có đăng ký bằng phát minh, sáng chế | 5 | | | | |
| 10b | Được báo cáo tại hội đồng cấp Viện trong hội nghị SVNCKH nhưng không đạt giải từ giải 3 trở lên/Đạt giải khuyến khích trong các kỳ thi quốc gia và quốc tế khác về chuyên ngành (VD: TI contest) | 2 | | | | |
| 10c | Không có thành tích về nghiên cứu khoa học | 0 | | | | |
| **Điểm tổng** | | **/50** | | | | |
| **Điểm tổng quy đổi về thang 10** | |  | | | | |

***Nhận xét khác*** *(về thái độ và tinh thần làm việc của sinh viên)*

|  |  |
| --- | --- |
|  | Ngày: … / … / 20…  **Người nhận xét**  (Ký và ghi rõ họ tên) |

**ĐÁNH GIÁ QUYỂN ĐỒ ÁN TỐT NGHIỆP**

(Dùng cho cán bộ phản biện)

Giảng viên đánh giá:

Họ và tên sinh viên:Phạm Đức Hải MSSV:20151215

Tên đồ án: **Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân**

**Chọn các mức điểm phù hợp cho sinh viên trình bày theo các tiêu chí dưới đây:**

Rất kém (1); Kém (2); Đạt (3); Giỏi (4); Xuất sắc (5)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Có sự kết hợp giữa lý thuyết và thực hành (20)** | | | | | | |
| 1 | Nêu rõ tính cấp thiết và quan trọng của đề tài, các vấn đề và các giả thuyết (bao gồm mục đích và tính phù hợp) cũng như phạm vi ứng dụng của đồ án | 1 | 2 | 3 | 4 | 5 |
| 2 | Cập nhật kết quả nghiên cứu gần đây nhất (trong nước/quốc tế) | 1 | 2 | 3 | 4 | 5 |
| 3 | Nêu rõ và chi tiết phương pháp nghiên cứu/giải quyết vấn đề | 1 | 2 | 3 | 4 | 5 |
| 4 | Có kết quả mô phỏng/thực nghiệm và trình bày rõ ràng kết quả đạt được | 1 | 2 | 3 | 4 | 5 |
| **Có khả năng phân tích và đánh giá kết quả (15)** | | | | | | |
| 5 | Kế hoạch làm việc rõ ràng bao gồm mục tiêu và phương pháp thực hiện dựa trên kết quả nghiên cứu lý thuyết một cách có hệ thống | 1 | 2 | 3 | 4 | 5 |
| 6 | Kết quả được trình bày một cách logic và dễ hiểu, tất cả kết quả đều được phân tích và đánh giá thỏa đáng | 1 | 2 | 3 | 4 | 5 |
| 7 | Trong phần kết luận, tác giả chỉ rõ sự khác biệt (nếu có) giữa kết quả đạt được và mục tiêu ban đầu đề ra đồng thời cung cấp lập luận để đề xuất hướng giải quyết có thể thực hiện trong tương lai | 1 | 2 | 3 | 4 | 5 |
| **Kỹ năng viết quyển đồ án (10)** | | | | | | |
| 8 | Đồ án trình bày đúng mẫu quy định với cấu trúc các chương logic và đẹp mắt (bảng biểu, hình ảnh rõ ràng, có tiêu đề, được đánh số thứ tự và được giải thích hay đề cập đến; căn lề thống nhất, có dấu cách sau dấu chấm, dấu phảy v.v.), có mở đầu chương và kết luận chương, có liệt kê tài liệu tham khảo và có trích dẫn đúng quy định | 1 | 2 | 3 | 4 | 5 |
| 9 | Kỹ năng viết xuất sắc (cấu trúc câu chuẩn, văn phong khoa học, lập luận logic và có cơ sở, từ vựng sử dụng phù hợp v.v.) | 1 | 2 | 3 | 4 | 5 |
| **Thành tựu nghiên cứu khoa học (5)** *(chọn 1 trong 3 trường hợp)* | | | | | | |
| 10a | Có bài báo khoa học được đăng hoặc chấp nhận đăng/Đạt giải SVNCKH giải 3 cấp Viện trở lên/Có giải thưởng khoa học (quốc tế hoặc trong nước) từ giải 3 trở lên/Có đăng ký bằng phát minh, sáng chế | 5 | | | | |
| 10b | Được báo cáo tại hội đồng cấp Viện trong hội nghị SVNCKH nhưng không đạt giải từ giải 3 trở lên/Đạt giải khuyến khích trong các kỳ thi quốc gia và quốc tế khác về chuyên ngành (VD: TI contest) | 2 | | | | |
| 10c | Không có thành tích về nghiên cứu khoa học | 0 | | | | |
| **Điểm tổng** | | **/50** | | | | |
| **Điểm tổng quy đổi về thang 10** | |  | | | | |

***Nhận xét khác của cán bộ phản biện***

|  |  |
| --- | --- |
|  | Ngày: … / … / 20…  **Người nhận xét**  (Ký và ghi rõ họ tên) |

**LỜI NÓI ĐẦU**

Hiện nay, với sự phát triển của công nghệ thông tin và điện tử truyền thông, điện thoại thông minh đã không còn xa lạ với mỗi cá nhân và đã trở thành một vật dụng thiết yếu trong cuộc sống của chúng ta. Cùng nhu cầu về chức năng, dịch vụ nhanh chóng, tin cậy như thanh toán trực tuyến, tìm kiếm trực tuyến, giải trí, bản đồ..., nhiều hệ thống, ứng dụng đã và đang được phát triển để phục vụ cho cuộc sống. Hơn thế nữa, việc phát triển các ứng dụng trên điện thoại thông minh còn đem lại nguồn lợi nhuận khổng lồ nhờ việc đầu tư vốn nhỏ nhưng lại có số lượng người dùng lớn.

Ứng dụng trích xuất thông tin từ căn cước công dân được xây dựng dành cho đối tượng doanh nghiệp với nhân viên sở hữu các thiết bị di động thông minh để kiểm tra và lấy được thông tin khách hàng một cách nhanh chóng và tiện lợi. Dựa vào các ưu điểm của điện thoại thông minh, ứng dụng này có tiềm năng phát triển và mở rộng tốt trong tương lai.

Để thực hiện dự án này, em đã nhận được được sự hướng dẫn nhiệt tình và tận tâm từ TS. Phạm Doãn Tĩnh, em xin chân thành cảm ơn thầy. Do kiến thức hạn hẹp và thời gian nghiên cứu chưa có nhiều nên trong quá trình phân tích, thiết kế ứng dụng nếu có gì sai sót mong thầy cô đóng góp ý kiến để em có thể làm đồ án tốt nghiệp của mình hoàn chỉnh hơn nữa.

Hà Nội, ngày 28 tháng 6 năm 2020

Sinh viên thực hiện

Phạm Đức Hải

LỜI CAM ĐOAN

Tôi là Phạm Đức Hải, mã số sinh viên 20151215, sinh viên lớp Điện tử 09. Người hướng dẫn là TS. Phạm Doãn Tĩnh. Tôi cam đoan toàn bộ nội dung được trình bày trong đồ án “*Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân”* là kết quả của quá trình tìm hiểu và nghiên cứu của tôi. Các dữ liệu được nêu trong đồ án là hoàn toàn trung thực, phản ánh đúng kết quả đo đạc thực tế. Mọi thông tin trích dẫn đều tuân thủ các quy định về sở hữu trí tuệ; các tài liệu tham khảo được liệt kê rõ ràng. Tôi xin chịu hoàn toàn trách nhiệm với những nội dung được viết trong đồ án này.

Hà Nội, ngày 1 tháng 7 năm 2020

Người cam đoan

Phạm Đức Hải

MỤC LỤC

[MỤC LỤC 10](#_Toc44584583)

[DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT i](#_Toc44584584)

[DANH MỤC HÌNH VẼ ii](#_Toc44584585)

[DANH MỤC BẢNG BIỂU iii](#_Toc44584586)

[TÓM TẮT ĐỒ ÁN 1](#_Toc44584587)

[ABSTRACT 2](#_Toc44584588)

[CHƯƠNG 1. GIỚI THIỆU CHUNG 3](#_Toc44584589)

[1.1 Đặt vấn đề và lý do chọn đề tài 3](#_Toc44584590)

[1.2 Các sản phẩm đã có hiện nay 3](#_Toc44584591)

[1.3 Mục tiêu của dự án 5](#_Toc44584592)

[1.4 Bố cục của đồ án 6](#_Toc44584593)

[1.5 Kết luận 6](#_Toc44584594)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 7](#_Toc44584595)

[2.1 OCR là gì? 7](#_Toc44584596)

[2.2 Ưu điểm vượt trội của OCR 8](#_Toc44584597)

[2.3 Quy trình chung của một hệ thống nhân dạng chữ 8](#_Toc44584598)

[2.4 Giới thiệu về bộ nhân dạng ký tự quang học Tesseract 9](#_Toc44584599)

[2.4.1 Lịch sử 9](#_Toc44584600)

[2.4.2 Kiến trúc hoạt động 11](#_Toc44584601)

[2.5 Giới thiệu về Google Cloud Vision API và Firebase ML Kit 11](#_Toc44584602)

[2.5.1 Nền tảng Firebase 12](#_Toc44584603)

[2.5.2 Firebase Cloud Function 12](#_Toc44584604)

[2.5.3 ML Kit 13](#_Toc44584605)

[2.6 Thư viện OpenCV 14](#_Toc44584606)

[2.6.1 Giới thiệu về thư viện OpenCV 14](#_Toc44584607)

[2.6.2 Các phiên bản OpenCV 14](#_Toc44584608)

[2.6.3 Cách tổ chức OpenCV 16](#_Toc44584609)

[2.7 Tiến trinh nhận dạng sử dụng Tesseract OCR 17](#_Toc44584610)

[2.7.1 Thiết lập dòng cơ sở 17](#_Toc44584611)

[2.7.2 Cắt nhỏ từ 18](#_Toc44584612)

[2.7.3 Nhận dạng khoảng cách giữa số hoặc chữ 18](#_Toc44584613)

[2.7.4 Nhận dạng từ 18](#_Toc44584614)

[2.8 Đánh giá độ chính xác nhận dạng Tessract OCR 18](#_Toc44584615)

[2.9 Công cụ và ngôn ngữ sử dụng 19](#_Toc44584616)

[2.9.1 Công cụ 19](#_Toc44584617)

[2.9.2 Ngôn ngữ 20](#_Toc44584618)

[2.10 Kết luận 22](#_Toc44584619)

[CHƯƠNG 3. PHÂN TÍCH THIẾT KẾ HỆ THỐNG 23](#_Toc44584620)

[3.1 Khảo sát về thẻ căn cước công dân 23](#_Toc44584621)

[3.2 Phân tích và thiết kế tổng thể ứng dụng 25](#_Toc44584622)

[3.2.1 Yêu cầu chức năng của hệ thống 25](#_Toc44584623)

[3.2.2 Yêu cầu phi chức năng của hệ thống 25](#_Toc44584624)

[3.2.3 Mô tả chi tiết các chức năng của ứng dụng 25](#_Toc44584625)

[3.3 Biểu đồ use case 27](#_Toc44584626)

[3.4 Biểu đồ trạng thái 30](#_Toc44584627)

[3.5 Biểu đồ hoạt động 32](#_Toc44584628)

[3.6 Biểu đồ tuần tự 36](#_Toc44584629)

[3.7 Kết luận 39](#_Toc44584630)

[KẾT QUẢ MÔ PHỎNG 40](#_Toc44584631)

[3.8 Hình chụp từ máy ảnh 40](#_Toc44584632)

[3.9 Hình chọn từ thư viện trong điện thoại 41](#_Toc44584633)

[KẾT LUẬN 44](#_Toc44584634)

[Kết luận chung 44](#_Toc44584635)

[Hướng phát triển 44](#_Toc44584636)

[TÀI LIỆU THAM KHẢO 45](#_Toc44584637)

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| Ký hiệu và chữ viết tắt | Cụm từ đầy đủ và ý nghĩa | Tiếng việt |
| ML | Machine Learning | Học máy |
| OCR | **Optical Character Recognition** | **Nhận diện ký tự quang học** |
| OPENCV | Open Source Computer Vision | Môi trường thị giác máy tính mở |
| HP | Hewlett Packet |  |
| API | Application Programming Interface | Giao diện lập trình ứng dụng |
| GUI | Graphical User Interface | Giao diện đồ họa người dùng |
| IDE | Integrated Development Environment | Môi trường phát triển tích hợp |
| Java SE | Java Standard Edition | Phiên bản Java chuẩn |
| Java ME | Java Micro Edition | Phiên bản Java Micro |
| Java EE | Java Enterprise Edition | Phiên bản Java doanh nghiệp |
| OMG | Object Management Group | Nhóm quản lý đối tượng |
| RIAs | Rich Internet Applications | Ứng dụng Internet phong phú |

# DANH MỤC HÌNH VẼ

[Hình 2. 1 Ví dụ về một đường cơ sở dạng cong 17](#_Toc44582663)

[Hình 2. 2 Ví dụ về xác định ký tự 18](#_Toc44582664)

[Hình 2. 3 Quá trình nhận dạng từ 18](#_Toc44582665)

[Hình 3. 1 Thẻ căn cước công dân 24](#_Toc44582673)

[Hình 3. 2 Thông tin trên thẻ căn cước công dân 24](#_Toc44582674)

[Hình 3. 3 Ví dụ về crop ảnh 26](#_Toc44582675)

[Hình 3. 4 Biểu đồ use case tổng quát 28](#_Toc44582676)

[Hình 3. 5 Biểu đồ trạng thái chọn ảnh 31](#_Toc44582677)

[Hình 3. 6 Biểu đồ trạng thái xử lý ảnh 32](#_Toc44582678)

[Hình 3. 7 Biểu đồ hoạt động xác định dòng 34](#_Toc44582679)

[Hình 3. 8 Biểu đồ hoạt động cắt nhỏ từ 35](#_Toc44582680)

[Hình 3. 9 Biểu đồ hoạt động nhận dạng từ 36](#_Toc44582681)

[Hình 3. 10 Biểu đồ trình tự chọn ảnh 37](#_Toc44582682)

[Hình 3. 11 Biểu đồ trình tự lưu kết quả 38](#_Toc44582683)

[Hình 4. 1 Ảnh chụp từ điện thoại 40](#_Toc44582684)

[Hình 4. 2 Kết quả của ảnh chụp từ điện thoại 41](#_Toc44582685)

[Hình 4. 3 Ảnh chọn từ thư viện của điện thoại 42](#_Toc44582686)

[Hình 4. 4 Kết quả thu được từ ảnh chọn từ thư viện điện thoại 42](#_Toc44582687)

# DANH MỤC BẢNG BIỂU

[Bảng 2. 1 So sánh phần mềm thương mại và Tesseract 10](#_Toc44582694)

[Bảng 2. 2 Độ chính xác của Tesseract trên một số ngôn ngữ 10](#_Toc44582695)

[Bảng 3. 1 Các chức năng chính của hệ thống 26](#_Toc44582698)

[Bảng 3. 2 Mô tả các use case 28](#_Toc44582699)

# 

# TÓM TẮT ĐỒ ÁN

Trong đề tài này, dựa trên những phương thức, công cụ mới và phổ biến em đã xây dựng được phần mềm trích xuất thông tin từ ảnh căn cước khá hoàn chỉnh. Thêm vào đó, dựa trên việc tìm hiểu các công nghệ khả dụng hiện tại, em đã áp dụng các công nghệ phù hợp với yêu cầu và đặc trưng của đề tài.

Bài toán đặt ra là xây dựng phần mềm xử lý ảnh để trích xuất thông tin từ căn cước công dân, bắt đầu từ khâu phân tích và thiết kế đến khâu triển khai phần mềm. Với việc sử dụng UML (sơ đồ use-case, sơ đồ tuần tự, sơ đồ trạng thái, sơ đồ lớp, sơ đồ thực thế liên kết), mô hình thác nước và một số công cụ cần thiết cho việc lập trình, em đã có được một kết quả khả quan cho bài toán này.

Đầu tiên, em tìm hiểu về các phương pháp trích xuất chữ từ ảnh hiện nay và tìm ra điểm mạnh, điểm yếu của từng phương pháp. Tiếp đó, em lên ý tưởng, đặt ra các mục tiêu, yêu cầu, giới hạn của đề tài, tìm hiểu các quy trình thiết kế và các công nghệ phục vụ cho yêu cầu của bài toán. Sau đó, em tập trung vào việc phân tích thiết kế và triển khai phần mềm.

Trong thời gian ba tháng, em đã hoàn thành việc tìm hiểu các phương pháp và công cụ xử lý ảnh, trích xuất chữ từ ảnh, phân tích thiết và triển khai phần mềm.

ABSTRACT

In this project, based on new and popular methods, tools, I have built a software to extract information from the image of a fairly complete ID. In addition, based on understanding the available technologies, I have applied the technologies in accordance with the requirements and characteristics of the topic.

The problem is to build image processing software to extract information from citizen identity, starting from the analysis and design stage to the software deployment stage. With the use of UML (use-case diagrams, sequence diagrams, state diagrams, class diagrams, link fact diagrams), waterfall models and some necessary tools for programming, I have had a satisfactory result for this problem.

First, I learn about the methods of extracting text from the current image and find the strengths and weaknesses of each method. Next, I come up with ideas, set goals, requirements, limits of the topic, learn about design processes and technologies that serve the requirements of the problem. Later, I focused on analyzing the design and implementation of the software.

In a three-month period, I completed my study of image processing methods and tools, extracted text from images, set analysis and implemented the software.

# GIỚI THIỆU CHUNG

Chương 1 trình bày những vấn đề hiện tại trong thời buổi thế giới ngày càng phát triền cùng những thách thức và cơ hội đối với nền công nghệ 4.0 hiện nay.

## Đặt vấn đề và lý do chọn đề tài

Trong giai đoạn bùng nổ công nghệ hiện nay, yêu cầu về nguồn nhân lực chất lượng cao để đáp ứng cho nhu cầu cho các nhà tuyển dụng là vấn đề được xã hội quan tâm. Các kỹ năng về phân tích thiết kế và triển khai hệ thống cần được mỗi sinh viên rèn luyện một cách hoàn chỉnh và nghiểm túc mới có thể dáp ứng yêu cầu của các nhà tuyển dụng cũng như các doanh nghiệp. Ngoài ra, việc tiếp cận với các công nghệ mới cũng là một yếu tố cần thiết do thị trường công nghệ thông tin đang từng bước phát triển và đổi mới từng ngày.

Ngoài lý do trên, hiện nay việc lấy thông tin nhân sự hay khách hàng đang mất khá nhiều thời gian khi đa phần mọi thứ vẫn còn thực hiện một cách thủ công. Bên cạnh đó, sự phổ biến của điện thoại thông minh là không thể bàn cãi. Vậy nên câu hỏi đặt ra là tại sao ta không tận dụng điều này để khiến việc lấy thông tin trở nên dễ dàng hơn?

Để trả lời câu hỏi trên, em quyết định chọn đề tài “Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân”. Kết quả của đề tài có thể phục vụ cho các doanh nghiệp trong việc quản lý thông tin nhân viên cũng như chăm sóc khách hàng một cách thuận lợi. Hơn nữa, đây cũng là cơ hội đối với bản thân em để có thể hoàn thiện những kỹ năng về nghiệp vụ cũng như phân tích thiết kế để có thể đáp ứng yêu cầu của các nhà tuyển dụng trong tương lai.

## Các sản phẩm đã có hiện nay

Trên thị trường hiện tại cũng đã có nhiều sản phẩm trích xuất thông tin từ căn cước công dân, điển hình như công nghệ nhận dang hình ảnh FPT.AI Vision. Công nghệ nhận dạng hình ảnh FPT.AI Vision giúp giúp nhận dạng và trích xuất thông tin từ ảnh chụp chứng minh nhân dân một cách dễ dàng theo các trường như tên, ngày tháng năm sinh, quê quán, số chứng minh thư, ngày cấp… giúp việc khai thác thông tin nhan chóng, chính xác, tiết kiệm thời gian. Nhờ việc có thể trích xuất thông tin qua ảnh chụp từ điện thoại di động nên công nghệ của FPT có thể sử dụng linh động mọi lúc, mọi nơi. Đây là điểm mạnh của FPT.AI Vision mà nhiều đối thủ khác không có được.

[FPT.AI Vision](https://fpt.ai/vi/vision-vi/) nhận dạng chính xác đến 94.6% các trường thông tin đối với cả chứng minh thư kiểu cũ và trả kết quả nhận dạng tốt hơn với thẻ căn cước công dân mới. Đối với những chứng minh thư nhân đã cũ bị mờ số, FPT.AI Vision cho ra kết quả sai sót không đáng kể. Với những thông tin có khả năng sai sót, hệ thống của FPT sẽ gửi cảnh báo đến cho người dùng. Công nghệ nhận dạng chứng minh thư nhân dân của FPT nhận dạng và trích xuất chính xác các trường thông tin ở cả mặt trước và mặt sau của chứng minh thư. Ứng dụng công nghệ này giúp tiết kiệm thời gian gấp 60 lần so với việc nhập liệu theo cách thủ công. Thông thường để nhập thông tin từ CMT phải mất 3-4 phút nhưng với FPT.AI Vision, thông tin được nhân dạng và trích xuất vào máy tính chỉ mất 2-3 giây kể cả thời gian truyền ảnh qua mạng và có thể xử lý được hàng trăm ảnh cùng một lúc tùy theo nhu cầu của khách hàng.

FPT.AI Vision là một trong 4 module của [FPT.AI](https://techinsight.com.vn/fpt-ai-nen-tang-tri-tue-nhan-tao-toan-dien/?highlight=fpt.ai) được xây dựng và phát triển bởi Ban công nghệ FPT cùng các chuyên gia, nhà khoa học trong lĩnh vực trí tuệ nhân tạo, học máy,… FPT.AI Vision là sự kết hợp các các kỹ thuật xử lý ảnh nâng cao, xử lý ngôn ngữ tự nhiên, được phát triển trên cơ sở kế thừa các tính năng ưu việt của công nghệ OCR. FPT.AI Vision giúp các ứng dụng, dịch vụ có thể xác định, phân tích chính xác nội dung trong hình ảnh hoặc văn bản. Các chức năng chính của FPT.AI Vision tập trung vào việc tự động hoá các quy trình xử lý (Cognitive RPA) cho doanh nghiệp, nhận dạng hình ảnh, nhận dạng biểu mẫu và OCR (Nhận dạng ký tự quang học). Như vậy việc ứng dụng FPT.AI Vison giúp Chính phủ cũng như Doanh nghiệp số hóa toàn bộ các giấy tờ cơ bản như: Thẻ căn cước công dân, Giấy phép lái xe thành mẫu chung. FPT.AI Vision đảm bảo rút ngắn quy trình xử lý, giảm số lượng và đơn giản hóa, chuẩn hóa nội dung hồ sơ, giảm thời gian và chi phí thực hiện thủ tục hành chính.

Hiện nay, giải pháp nhận dạng ký tự quang học đang phát triển mạnh về mặt ứng dụng và liên tục có nhiều cải tiến mới để tăng tính ứng dụng và hiệu quả của sản phẩm đầu ra. Tuy nhiên, việc nhận dạng ký tự tiếng Việt (loại hình ngôn ngữ có dấu) vẫn là thách thức đối với sự phát triển của công nghệ [OCR](https://techinsight.com.vn/goc-nhin-cong-nghe-ve-fvi-he-thong-ocr-nhan-dien-chung-minh-thu-nhan-dan-viet-nam/) trên thế giới. Một hạn chế rất lớn mà tất cả các phần mềm OCR hiện nay gặp phải là thiếu hẳn tính năng rút trích thông tin từ văn bản, biểu mẫu. Hầu hết các phần mềm mới chỉ dừng lại việc nhận dạng toàn văn mà không thể rút trích thông tin theo các mẫu/định dạng, các trường theo nhu cầu của người sử dụng – một nhu cầu rất thiết yếu đối với những người làm công tác văn thư, lưu trữ. FPT.AI Vision giải quyết được các hạn chế của các chương trình hiện có và đặc biệt phù hợp với việc nhận dạng ký tự và rút trích thông tin từ chứng minh thư nhân dân ở Việt Nam.

Bằng việc giảm chi phí về nhân sự và thời gian nhập liệu, FPT.AI Vision giúp doanh nghiệp tăng hiệu quả vận hành đồng thời thay đổi quy mô hệ thống một cách linh hoạt. Doanh nghiệp không cần duy trì đội ngũ nhân lực cồng kềnh khi hệ thống hoàn toàn có thể tăng/giảm khối lượng xử lý dữ liệu bất cứ lúc nào. Với APIs, FPT.AI Vision có thể được tích hợp vào bất kỳ hệ thống nào của doanh nghiệp trong thời gian ngắn nhất. Việc số hoá tài liệu mà cụ thể là việc tự động hóa quy trình nhập liệu nhờ trích xuất các thông tin từ hình ảnh hoặc văn bản giúp cho việc cất trữ gọn gàng, bảo quản và duy trì tài liệu; nâng cao sự tiện lợi trong việc truy xuất, tìm kiếm thông tin, nhờ đó tiết kiệm tối đa sức người, sức của cho việc quản lý. FPT.AI Vision là một giải pháp tốt cho nhận dạng tiếng Việt vì nó có khả năng: tích hợp kết quả về chính tả, xử lý ngôn ngữ tự nhiên dành cho tiếng Việt; mô hình học Deep Learning cho tiếng Việt.

Bên cạnh đó, một số ứng dụng Android nhận dạng ký tự quang học (OCR) thực tiễn như:

* Google Keep: Ứng dụng ghi chú tuyệt vời từ Google có một số thủ thuật và nhiều ứng dụng sáng tạo. Nó cũng hỗ trợ công cụ OCR tích hợp sẵn. Google Keep trích xuất văn bản với nhiều định dạng khác nhau từ đơn giản đến phức tạp. Nó cũng giữ định dạng văn bản gốc ở mức độ tối đa nhất.
* Text Scanner [OCR]: Text Scanner [OCR] hỗ trợ hơn 50 ngôn ngữ bao gồm tiếng Trung, tiếng Nhật, tiếng Pháp và nhiều thứ tiếng khác nữa. Nó thậm chí còn hỗ trợ trích xuất văn bản từ bản viết tay. Giao diện ứng dụng có tính năng quét cơ bản như độ phóng đại và thanh trượt độ sáng để chụp văn bản rõ ràng nhất có thể.
* Text Fairy: Text Fairy là một công cụ trích xuất ảnh khác cho Android có khả năng nhận dạng văn bản từ hơn 50 ngôn ngữ bao gồm tiếng Trung, tiếng Nhật, tiếng Hà Lan, tiếng Pháp và nhiều thứ tiếng khác nữa. Nó hỗ trợ nhiều ngôn ngữ Ấn Độ như Hindi, Bengali, Marathi, Telugu, v.v... Bạn sẽ được hỏi để tải các ngôn ngữ yêu cầu trên ứng dụng khi chạy lần đầu. Nó làm việc tốt với các văn bản giấy nhưng gặp khó khăn khi nhận dạng văn bản với nhiều màu sắc.
* Office Lens: Office Lens là ứng dụng protable scan tài liệu của Microsoft đưa đến cho các thiết bị Android. Tính năng nổi bật của nó là khả năng quét và số hoá tài liệu, nhưng nó cũng đi kèm với một tùy chọn OCR tiện dụng. Nó có một phiên bản trả phí, nhưng bạn có thể sử dụng miễn phí bằng cách đăng ký một tài khoản của Microsoft. Với việc đăng ký tài khoản miễn phí này, bạn cũng sẽ được dùng các tính năng khác như 5GB bộ nhớ OneDrive miễn phí và khả năng lưu nhiều định dạng. Ứng dụng có thể nhận dạng văn bản với các phông chữ đầy màu sắc. Ngoài ra, nó có thể xác định văn bản từ ghi chú viết tay, tích hợp chặt chẽ với các sản phẩm khác của Microsoft như OneNote và Office 365.
* OCR Text Scanner: OCR Text Scanner có một giao diện đơn giản và hỗ trợ trên 55 ngôn ngữ bao gồm tiếng Anh, Pháp, Ý, Thụy Điển và nhiều ngôn ngữ khác. Ứng dụng không thể lấy văn bản từ bản viết tay. Ứng dụng có nhiều quảng cáo nên bạn sẽ phải đợi khoảng năm giây giữa mỗi lần quét tài liệu.

## Mục tiêu của dự án

Từ những vấn đề đã nêu trên, em xin chọn mục tiêu chính của đề tài “Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân” là:

* Tìm hiểu về các phương pháp xử lý và trích xuất chữ từ ảnh
* Khảo sát, thu thập dữ liệu về tính thiết yếu của bài toán
* Phân tích thiết kế hệ thống
* Triển khai hệ thống

## Bố cục của đồ án

Với đề tài này, bố cục đồ án sẽ bao gồm những phần sau:

* Chương 1: Giới thiệu chung
* Chương 2: Cơ sở lý thuyết
* Chương 3: Phân tích thiết kế hệ thống
* Chương 4: Kết quả mô phỏng

## Kết luận

Trên đây là những giới thiệu của em về ứng dụng xử lý ảnh và trích xuất thông tin từ căn cước công dân. Chương một đã trình bày được lý do em chọn đề tài, mục tiêu thực hiện phạm vi thực hiện của đề tài

# CƠ SỞ LÝ THUYẾT

Chương này sẽ giới thiệu tổng quan về các công nghệ được trong dự án thiết kế ứng dụng trích xuất thông tin từ căn cước công dân.

## OCR là gì?

Nhận dạng kí tự quang học (OCR – Optical Charater Recognition) là phần mềm máy tính xử lí và cho phép chuyển đổi tài liệu dạng ảnh (các ảnh đầu ra của máy scanner, máy ảnh, file PDF…) thành tài liệu có thể biên tập được (file word…). OCR thực hiện được điều này nhờ công nghệ xử lí ngôn ngữ tự nhiên (NLP), giúp nhận dạng các chữ cái và kí tự, sau đó sao chép chúng theo định dạng và thứ tự được viết. Đối với việc trích xuất các trường thông tin cần thiết trong giấy tờ tuy thân như chứng minh nhân dân, giấy phép lái xe,… khi đưa tài liệu lên hệ thống, người dùng chọn lựa và xác định các vùng cần bóc tách. Sau đó, hệ thống được OCR nhận dạng để chuyển sang dạng text và tự động trích xuất các trường thông tin cần thiết.

Công nghệ OCR đã mang lại vô số tiện ích cho ngành ngân hàng, nhưng đáng chú ý nhất là số hoá tài liệu. Đại diện của ngân hàng Rabobank (Hà Lan) cho biết, công ty họ ứng dụng công nghệ OCR vào rất nhiều hoạt động kinh doanh. Bởi họ cần tổng hợp và tích luỹ thông tin khách hàng cho chính mình, để tận dụng nguồn dữ liệu lớn này cho nhiều dự án của ngân hàng trong tương lai.

Những năm gần đây, với sự kết hợp của Trí tuệ nhân tạo (AI) và công nghệ xử lí ngôn ngữ tự nhiên (NLP), OCR được mở rộng thành công nghệ thu thập và quản lí dữ liệu, đóng vai trò vô cùng quan trọng trong hệ thống tự động hoá quy trình kinh doanh của ngân hàng. Nhờ tích hợp AI, OCR hiểu được những gì nó trích xuất và tự động cải thiện đầu ra bằng Học Máy, học hỏi từ các dữ liệu sẵn có và liên tục bổ sung kiến thức còn thiếu, giúp xử lí và kiểm tra các lỗi tài liệu vật lí một cách nhanh chóng, tiết kiệm thời gian và chi phí.

Đối với văn bản đánh máy, OCR cho độ chính xác trên 98%. Tỉ lệ chính xác là một vấn đề rất lớn vì những lỗi sai nhỏ có thể dẫn đến việc mất dữ liệu quan trọng. Các ngân hàng sử dụng OCR như một phương tiện bảo mật giao dịch và quản lí rủi ro. Bởi khi dùng OCR truyền thống, người dùng chỉ có thể kiểm tra tài liệu một cách thủ công. Nhưng khi được tích hợp AI và công nghệ xử lí ngôn ngữ tự nhiên (NLP), OCR có thêm khả năng tự động đánh giá rủi ro cho bất kì tài liệu giấy tờ nào. AI có thể học hiểu để phân tích và xử lí nội dung tài liệu kĩ lưỡng, phát hiện điều bất thường từ thông tin thanh toán tín dụng trong dữ liệu trích xuất của khách hàng.

Ngoài các loại giấy tờ phổ biến nhất mà ngân hàng thường xuyên phải xử lí là chứng minh nhân dân (để lấy thông tin khách hàng), hợp đồng, biên lai, giấy vay tiền, báo cáo tài chính,… thì các dịch vụ ngân hàng như cho vay thế chấp và thẻ ghi nợ cũng có 1 khối lượng lớn tài liệu giấy cần số hoá. Ngân hàng sử dụng phần mềm OCR để quét các giấy đảm bảo của khách hàng được viết bằng tay như giấy vay vốn, giấy uỷ quyền… Phần mềm có khả năng nhận ra sự khác biệt trong sắp xếp văn bản mới. Từ đó gửi đi cảnh báo cho người giám sát về sự thay đổi trong định dạng nếu có sự sửa chữa.

Trên thế giới, nhiều ngân hàng sử dụng OCR kết hợp phần mềm nhận dạng khuôn mặt để cung cấp bảo mật 2 lớp tại các cây ATM. Nhiều ứng dụng di động đã ra đời dựa trên công nghệ OCR mang lại tiện ích cho người dùng như quét Séc gửi tiền bằng điện thoại di động. Khách hàng chỉ cần chụp lại số tài khoản, số tiền và chữ kí trên Séc ở độ phân giải cao bằng điện thoại, sau đó OCR sẽ diễn ra trong ứng dụng di động và dữ liệu được gửi đến ngân hàng để xử lí một cách nhanh chóng.

## Ưu điểm vượt trội của OCR

Các tài liệu sau khi số hoá sẽ được định tuyến thông qua 1 luồng dữ liệu trung tâm, nơi tất cả các chuyên gia, nhà phân tích và nhân viên ở các bộ phân liên quan đều có thể truy cập, tra cứu thông tin nhanh chóng. Việc áp dụng công nghệ OCR vào hệ thống giúp ngân hàng quản lí công việc hiệu quả, chuẩn hoá hoạt động của từng bộ phận, đơn giản hoá các quy trình, giảm gánh nặng hành chính, tăng năng suất lao động và tối ưu hoá chi phí doanh nghiệp.

Số hoá tài liệu giúp ngân hàng tiết kiệm không gian lưu trữ cần thiết cho cùng 1 thông tin, giảm từ vài mét vuông xuống chỉ còn vài byte trên máy tính, giải phóng không gian, mang lại nhiều diện tích sử dụng hơn. Các tài liệu được số hoá có thể lưu trữ ở nhiều định dạng, thuận tiện cho việc tìm kiếm, chỉnh sửa dễ dàng, sao lưu đơn giản và bền vững theo thời gian. Ngoài ra, việc giảm thiểu sử dụng giấy giúp bảo vệ môi trường, tiết kiệm chi phí sản xuất giấy.

Hiện nay, ngoài chữ Latinh, công nghệ OCR có thể xử lí được rất nhiều tập lệnh với đa dạng ngôn ngữ tượng hình như Ả Rập, Ấn Độ, Trung Quốc… Được kết hợp với các tiêu chuẩn Unicode và phầm mềm dịch mchine (google Dịch), tài liệu bằng ngôn ngữ nào cũng có thể được quét, số hoá và dịch sang ngôn ngữ khác một cách nhanh chóng, giúp cắt giảm nhu cầu dịch thuật của con người, giảm thời gian quay vòng hoạt động của doanh nghiệp. Ngân hàng là 1 trong những ngành luôn tiên phong song hành cùng với sự phát triển của khoa học công nghệ. Việc ứng dụng OCR cùng nhiều nền tảng khoa học tiên tiến đã và đang giúp ngân hàng trở nên thông minh và thân thiện với người dùng hơn.

## Quy trình chung của một hệ thống nhân dạng chữ

Có nhiều phương pháp để tạo ra một phần mềm dạng OCR, độ chính xác của các phương pháp này phụ thuộc vào công nghệ tạo nên phần mềm. Các phương pháp này đạt được độ tin cậy trong các hình ảnh có chất lượng tốt và vừa. Độ chính xác của việc rút trích văn bản là điều quan trọng nhất. Nhóm tác giả Kirill Safronov cho rằng một số sai sót trong quá trình chuyển đổi thường không quá quan trọng trừ các trường hợp như rút trích số serial từ ảnh chụp,...

Để khắc phục tình trạng kết quả xuất ra không chính xác của công nghệ OCR, nhiều công nghệ khác đã ra đời, tác giả A. Vinutha M H đã ứng dụng định hướng robot (Optical Character Recognition Based Auto Navigation of Robot). Việc định hướng của robot dựa vào bảng tín hiệu như là một cột mốc đánh dấu đường đi tiếp theo của robot. Định hướng tự động của các robot trong một vùng lớn đòi hỏi nhiều bảng tín hiệu khác nhau với mô hình nhận dạng duy nhất. Ngoài ra, hệ thống này còn cho phép nhận diện vị trí tên riêng.

Bên cạnh việc cải thiện độ chính xác, cần có sự thay đổi kích thước của thiết bị nhận dạng, tác giả Ali Ahmadi đã đề cập trong nghiên cứu của mình, tốc độ xử lý và độ chính xác cao là yêu cầu lớn hiện nay của các thiết bị nhận dạng ký tự dạng nhỏ, ví dụ như bút biết nhận dạng. Nhưng dù có nhiều mặt hàng loại này được chào bán trên thị trường nhưng nó vẫn không đáp ứng nhu cầu sử dụng và kích thước thiết bị. Ngoài sự đa dạng trong cách thức nhận dạng, OCR còn đa dạng về cách dùng, nó được chia thành hai cách, dùng online và dùng offline, có nhận xét về hai cách dùng này như sau:

Cách 1 nhận dạng offline: nhận dạng các văn bản in ra giấy hoặc các bản viết tay và nó đòi hỏi quá trình scan trên mặt giấy hoặc mặt vật liệu có chữ. Cách này thường đòi hỏi con người phải thực hiện một số thao tác như phân loại, lưu trữ và chỉnh sửa văn bản trước khi scan.

Cách 2 nhận dạng online: thường chỉ được dùng cho nhận dạng chữ viết tay được lưu trữ ở dạng kỹ thuật số, thông thường để scan dạng này chúng ta thường dùng một loại bút đặc biệt nhưng do sự thành công của các nghiên cứu gần đây mà giờ đã có các thiết bị khác thay thế. Việc nhận dạng online nhằm giúp con người giao tiếp với máy tính tốt hơn bằng cách viết tay thay vì gõ phím.

Trong đề tài này, em sẽ tìm hiểu về một công cụ OCR điển hình và là một trong những nền tảng quan trọng, đó là Tesseract. Bên cạnh đó, em cũng nghiên cứu và tìm hiểu thêm về Google Cloud Vision API trong việc trích xuất thông tin chữ từ ảnh đầu vào.

## Giới thiệu về bộ nhân dạng ký tự quang học Tesseract

### Lịch sử

Tesseract là một phần mềm mã nguồn mở và ban đầu nó được nghiên cứu và phát triển tại hãng Hewlett Packet (HP) trong khoảng từ năm 1984 đến 1994. Vào năm 1995, Tesseract nằm trong nhóm ba bộ nhận dạng OCR đứng đầu về độ chính xác khi tham gia trong hội nghị thường niên của tổ chức UNLV.

Lúc mới khởi động thì Tesseract là một dự án nghiên cứu tiến sĩ tại phòng thí nghiệm HP ở Bristol và đã được tích hợp vào trong các dòng máy quét dạng phẳng của hãng dưới dạng các add-on phần cứng hoặc phần mềm. Nhưng thực tế dự án này đã thất bại ngay từ trong trứng nước vì nó chỉ làm việc hiệu quả trên các tài liệu in có chất lượng tốt.

Sau đó, dự án này cùng với sự cộng tác của bộ phận máy quét HP ở bang Colorado đã đạt được một bước tiến quan trọng về độ chuẩn xác khi nhận dạng và vượt lên nhiều bộ nhận dạng OCR thời đó nhưng dự án đã không thể trở thành sản phẩm hoàn chỉnh vì độ cồng kềnh và phức tạp. Sau đó, dự án được đưa về phòng thí nghiệm của HP để nghiên cứu về cách thức nén và tối ưu mã nguồn. Dự án tập trung cải thiện hiệu năng làm việc của Tesseract dựa trên độ chính xác đã có. Dự án này được hoàn tất vào cuối năm 1994 và sau đó vào năm 1995 bộ Tesseract được gửi đi tham dự hội nghị UNLV thường niên về độ chính xác của OCR, vượt trội hơn hẳn so với các phần mềm OCR lúc bấy giờ. Tuy nhiên, Tesseract đã không thể trở thành một sản phẩm thương mại hoàn chỉnh được và vào năm 2005, HP đã chuyển Tesseract sang mã nguồn mở và được hãng Google tài trợ. Tesseract cho đến nay vẫn được nhiều nhà phát triển cộng tác và tiếp tục hoàn thiện. Phiên bản mới nhất của bộ nhận dạng Tesseract là phiên bản 3.2.2.

Bảng 2. So sánh phần mềm thương mại và Tesseract

|  |  |
| --- | --- |
| Phần mềm thương mại | Bộ nhận dạng Tesseract |
| Hỗ trợ hơn 100 ngôn ngữ | Hỗ trợ trên 40 ngôn ngữ và đang tăng dần |
| Có giao diện đồ họa | Không hỗ trợ giao diện đồ họa (dùng Command Line để gõ lệnh) |
| Hầu hết chỉ hỗ trợ trên nền tảng Windows | Hỗ trợ trên Windows, Linux, Mac OS |
| Độ chính xác cao mới đây | Độ chính xác cao từ năm 1995 |
| Chi phí khá cao 130$ - 500 $ | Hoàn toàn miễn phí (mã nguồn mở) |

Vì Tesseract hiện nay là bộ thư viện mã nguồn mở hoàn toàn miễn phí nên trên thế giới đã có nhiều phần mềm nhận dạng ký tự quang học ra đời dựa trên bộ Tesseract với giao diện và các tính năng dễ sử dụng hơn so với giao diện đơn giản của Tesseract ban đầu như: VietOCR cho nhận dạng tiếng Việt, Tessenet2 bộ nhận diện Tesseract trên nền .Net của Microsoft, giao diện Java (Java GUI frontend) cho Tesseract…

Bảng 2. Độ chính xác của Tesseract trên một số ngôn ngữ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ngôn ngữ | Tổng số ký tự (triệu) | Tổng số từ (triệu) | Lỗi ký tự (%) | Lỗi từ (%) |
| Tiếng Anh | 39 | 4 | 0.5 | 3.72 |
| Tiếng Nga | 213 | 26 | 0.75 | 5.78 |
| Tiếng Hoa giản thể | 0.25 | không xác định | 3.77 | không xác định |
| Tiếng Hindi | 1.4 | 0.33 | 15.41 | 69.44 |

### Kiến trúc hoạt động

Đầu tiên, bộ nhận diện Tesseract sẽ nhận đầu vào là ảnh màu hoặc ảnh mức xám. Ảnh này sẽ được chuyển đến bộ phận phân tích ngưỡng thích ứng (adaptive thresholding) để cho ra ảnh nhị phân. Vì trước kia HP cũng đã phát triển bộ phận phân tích bố cục trang nên Tesseract không cần phải có thành phần đó và được thừa hưởng từ HP. Vì thế mà Tesseract nhận đầu vào là một ảnh nhị phân với các vùng đa giác tùy chọn đã được xác định.

Ban đầu, Tesseract được thiết kế làm việc trên ảnh nhị phân sau đó chương trình được cải tiến để có thể nhận dạng cả ảnh màu và ảnh mức xám. Chính vì thế mà cần bộ phận phân tích ngưỡng thích ứng để chuyển đổi ảnh màu / ảnh mức xám sang ảnh nhị phân.

Sau đó quá trình nhận dạng sẽ được thực hiện tuần tự theo từng bước.

* Bước đầu tiên là phân tích các thành phần liên thông. Kết quả của bước này sẽ là tạo ra các đường bao quanh các ký tự.
* Bước thứ hai là tìm hàng và tìm từ, kết quả của bước này cũng giống như bước trên sẽ tạo ra các vùng bao quanh các hàng chữ và ký tự chứa trong vùng văn bản.
* Bước tiếp theo sẽ là nhận dạng từ. Công đoạn nhận dạng từ sẽ được xử lý qua 2 giai đoạn. Giai đoạn đầu sẽ là nhận dạng các từ theo lượt. Các từ thỏa yêu cầu trong giai đoạn này sẽ được chuyển sang bộ phân loại thích ứng (adaptive classifier) để làm dữ liệu huấn luyện. Chính nhờ đó mà bộ phân loại thích ứng sẽ có khả năng nhận diện được chính xác hơn ở phần sau của trang. Sau khi bộ phân loại thích ứng đã học được các thông tin có ích từ giai đoạn đầu khi nhận dạng phần trên của trang thì giai đoạn thứ 2 của việc nhận dạng sẽ được thực hiện. Giai đoạn này sẽ quét hết toàn bộ trang, các từ không được nhận diện chính xác ở giai đoạn đầu sẽ được nhận diện lại lần nữa. Cuối cùng bộ nhận diện sẽ tổng hợp lại các thông tin ở trên và cho ra kết quả nhận diện hoàn chỉnh.

## Giới thiệu về Google Cloud Vision API và Firebase ML Kit

Google Cloud Vision API là một công cụ rất mạnh có thể mang đến cho cuộc sống các khả năng ứng dụng vô tận khi kết hợp với thư viện Python. Vision API là mô hình được đào tạo trước của Google, giúp phát hiện các đối tượng, nhận dạng khuôn mặt, nhận dạng hình ảnh, phân loại, gán nhãn và trích xuất văn bản của văn bản in hoặc hình ảnh chữ viết tay. Ngoài các tính năng trên, Vision API còn một tính năng rất thực tế và thú vị đó là cho phép bạn phát hiện nội dung không lành mạnh của hình ảnh. Vision API cho phép các nhà phát triển tích hợp các tính năng thú vị vào các ứng dụng dễ dàng.

Firebase ML Kit là một thư viện cho phép bạn sử dụng dễ dàng và với mã rút gọn, sử dụng một loạt các mô hình linh hoạt, chính xác cao trong các ứng dụng Android. Hầu hết các models mà nó cung cấp đều có thể sử dụng ở cả local lẫn trên Google Cloud.

Hiện tại, các mô hình trong Firebase ML Kit chỉ giới hạn cho các tác vụ liên quan đến tầm nhìn của máy tính, chẳng hạn như nhận dạng ký tự quang học, quét mã vạch và phát hiện đối tượng.

### Nền tảng Firebase

Firebase là một nền tảng cho phép xây dựng các ứng dụng di động một cách nhanh chóng, nó cung cấp các dịch vụ có sẵn như: Realtime Database (cơ sở dữ liệu thời gian thực), Firebase Authentication (xác thực người dùng), Cloud Function (chức năng), Cloud Store (lưu trữ file), Google Analytic (phân tích dữ liệu), Admob (quảng cáo)... Firebase cung cấp một hệ thống khá toàn diện các chức năng để phục vụ phát triển ứng dụng di động, cung cấp cho nhà phát triển các API để phát triển ứng dụng, nhà phát triển sẽ chỉ cần tập trung phát triển ứng dụng về logic, trải nghiệm người dùng, nội dung... Còn các việc xử lý phức tạp bên phía máy chủ (duy trì máy chủ, duy trì cơ sở dữ liệu, xác thực người dùng...) sẽ được xử lý một cách tự động.

### Firebase Cloud Function

Cloud Function cho phép chạy mã nguồn phía máy chủ một cách tự động để phản hồi lại các sự kiện được kích hoạt bởi các thành phần khác của Firebase hoặc yêu cầu HTTPS. Mã nguồn ứng dụng phía máy chủ sẽ được lưu trên đám mây của Google và được chạy trên một môi trường được quản lý, nhà phát triển sẽ không phải quan tâm đến việc duy trì và cân bằng máy chủ.

Các khả năng chính của Cloud Function:

* Hợp nhất với nền tảng Firebase: các chức năng được nhà phát triển xây dựng có thể phản hồi lại các sự kiện sinh ra bởi các thành phần khác của nền tảng Firebase như Cloud Store, Realtime Database, Firebase Authentication, Google Analytics, HTTP Triggers...
* Nhà phát triển không cần duy trì máy chủ: nhà phát triển sẽ triển khai mã nguồn Javascript lên máy chủ sử dụng của sổ dòng lệnh, sau đó Firebase sẽ tự động cân bằng tài nguyên phù hợp với số lượng người dùng. Nhà phát triển không cần quan tâm về cài đặt máy chủ, nâng cấp, bảo trì, giất phép...
* Bảo mật: Cloud Function hoàn toàn tách biệt với phía máy trạm, mã nguồn sẽ được lưu trữ an toàn, bảo mật và đảm bảo thực hiện đúng các chức năng mà nhà phát triển đã thiết kế.

đời của một chức năng (function):

* Nhà phát triển viết mã nguồn cho một function mới, lựa chọn nguồn phát sinh sự kiện (event provider) như HTTPS, Realtime Database..., và định nghĩa điều kiện để function được thực hiện
* Nhà phát triển triển khai function và Firebase sẽ kết nối chức năng đó với nguồn phát sinh sự kiện được lựa chọn
* Khi nguồn phát sinh sự kiện tạo ra một sự kiện khớp với điều kiện của function, mã nguồn sẽ được thực hiện
* Nếu function đang bận xử lý nhiều sự kiện, Google sẽ tạo ra thêm các instance để xử lý công việc nhanh hơn. Nếu function rảnh rỗi thì các instance sẽ được xóa bỏ
* Khi nhà phát triển cập nhật function bằng cách triển khai thêm mã nguồn mới, tất cả các instance của phiên bản cũ sẽ bị loại bỏ và thay thế bằng các instance mới
* Khi nhà phát triển xóa bỏ function, tất cả các instance sẽ được xóa bỏ, và kết nối giữa function và nguồn phát sinh sự kiện cũng bị xóa bỏ theo

### ML Kit

ML Kit là một SDK di động mang lại kiến ​​thức chuyên môn về học máy của Google cho các ứng dụng Android và iOS trong một gói mạnh mẽ nhưng dễ sử dụng. Cho dù bạn là người mới hoặc có kinh nghiệm về học máy, bạn cũng có thể thực hiện chức năng mà minh mong muốn chỉ trong một vài dòng code. Không cần phải có kiến ​​thức sâu về Neural Network hoặc mô hình tối ưu hóa để bắt đầu. Mặt khác, nếu bạn là nhà phát triển ML có kinh nghiệm, ML Kit cung cấp các API tiện lợi giúp bạn sử dụng các mô hình TensorFlow Lite linh hoạt trong các ứng dụng di động của bạn.

ML Kit giúp dễ dàng áp dụng các kỹ thuật ML trong ứng dụng của bạn bằng cách mang đến cho bạn các công nghệ ML của Google, chẳng hạn như Google Cloud Vision API, TensorFlow Lite và Android Neural Network API cùng trong một SDK. Cho dù bạn cần đến khả năng xử lý dựa trên cloud, khả năng tối ưu hóa thời gian thực của các mô hình trên thiết bị di động hay tính linh hoạt của việc tùy chỉnh các mô hình TensorFlow Lite, ML Kit đều có thể thực hiện được chỉ với một vài dòng code.

## Thư viện OpenCV

### Giới thiệu về thư viện OpenCV

OpenCV (Open Source Computer Vision) là một thư viện mã nguồn mở về thị giác máy với hơn 500 hàm và hơn 2500 các thuật toán đã được tối ưu về xử lý ảnh, và các vấn đề liên quan tới thị giác máy. OpenCV được thiết kế một cách tối ưu, sử dụng tối đa sức mạnh của các dòng chip… để thực hiện các phép tính toán trong thời gian thực, nghĩa là tốc độ đáp ứng của nó có thể đủ nhanh cho các ứng dụng thông thường. OpenCV là thư viện được thiết kế để chạy trên nhiều nền tảng khác nhau (cross-patform), nghĩa là nó có thể chạy trên hệ điều hành Window, Linux, Mac, iOS … Việc sử dụng thư viện OpenCV tuân theo các quy định về sử dụng phần mềm mã nguồn mở BSDdo đó bạn có thể sử dụng thư viện này một cách miễn phí cho cả mục đích phi thương mại lẫn thương mại.

Dự án về OpenCV được khởi động từ những năm 1999, đến năm 2000 nó được giới thiệu trong một hội nghị của IEEE về các vấn đề trong thị giác máy và nhận dạng, tuy nhiên bản OpenCV 1.0 mãi tới tận năm 2006 mới chính thức được công bố và năm 2008 bản 1.1 (pre-release) mới được ra đời. Tháng 10 năm 2009, bản OpenCV thế hệ thứ hai ra đời (thường gọi là phiên bản 2.x), phiên bản này có giao diện của C++ (khác với phiên bản trước có giao diện của C) và có khá nhiều điểm khác biệt so với phiện bản thứ nhất.

Thư viện OpenCV ban đầu được sự hỗ trợ từ Intel, sau đó được hỗ trợ bở Willow Garage, một phòng thí nghiệm chuyên nghiên cứu về công nghệ robot. Cho đến nay, OpenCV vẫn là thư viện mở, được phát triển bởi nguồn quỹ không lợi nhuận (none -profit foundation) và được sự hưởng ứng rất lớn của cộng đồng.

### Các phiên bản OpenCV

Cho tới nay, trải qua hơn nhiều năm từ lúc phiên bản OpenCV đầu tiên được công bố, đã có lần lượt nhiều phiên bản OpenCV ra đời, tuy nhiên có thể chia thư viện này thành hai bản chính dựa trên những đặc điểm khác biệt lớn nhất của chúng: phiên bản OpenCV thế hệ thứ nhất (hay còn gọi là phiên bản OpenCV 1.x) và phiên bản OpenCV thứ hai (hay còn gọi là phiên bản OpenCV 2.x). Sau đây ta sẽ chỉ ra một số điểm khác biệt cơ bản giữa hai phiên bản này.

* OpenCV 1.x (bao gồm bản 1.0 và bản pre-release 1.1) dựa trên giao diện C, cấu trúc của một ảnh số dựa trên cấu trúc của IplImage, trong khi thư OpenCV 2.x dựa trên giao diện C++, cấu trúc của ảnh số, ma trận dựa trên cấu trúc của cv::Mat.
* Trong OpenCV 1.x, người sử dụng phải hoàn toàn quản lý bộ nhớ của các đối tượng, nghĩa là khi một đối tượng mới được tạo ra, ta phải luôn chú . để giải phóng nó khi không còn sử dụng nữa (trong nhiều trường hợp có thể sẽ bị tràn bộ nhớ nếu không chú . đều này), trong khi thư viện OpenCV 2.x việc quản lý bộ nhớ trở nên dễ dàng hơn nhờ các hàm hủy các các lớp đối tượng trong OpenCV 2.x đã thực hiện điều này khi một đối tượng không còn được sử dụng nữa.
* Việc viết các dòng lệnh để thực hiện cùng một chức năng trong OpenCV 2.x là dễ dàng hơn nhiều so với OpenCV 1.x, một phần là là giao diện C++ có phần dễ hiểu hơn so với C, một phần là các hàm trong OpenCV 2.x đã được tối ưu hóa nhiều bước trung gian không cần thiết về mặt giao diện người sử dụng.
* Thư viện OpenCV 1.x tuy chứa một lượng lớn hàm xử lý và thuật toán, tuy nhiên nó vẫn ở dạng sơ khai. Thư viện OpenCV 2.x đã được bổ xung khá nhiều hàm, thuật toán và được tối ưu khá nhiều đặc biệt trong các khía cạnh về phát hiện đối tượng (detection), nhận dạng đối tượng (partten regconition) và theo dỗi đối tượng (tracking). Hơn thế nữa, tuy có giao diện là C++ nhưng OpenCV 2.x vẫn dữ một phần giao diện C để tương thích với các phiên bản của OpenCV 1.x …

Từ một số đặc điểm trên ta có thể thấy rằng thư viện OpenCV phiên bản 2.x là có nhiều điểm nổi trội hơn so với phiên bản 1.x, Tuy nhiên trong một số trường hợp như ở các hệ thống nhúng khi mà trình dịch chỉ đơn thuần chấp nhận ngôn ngữ C thì phiển bản 1.x vẫn còn giá trị. Trong cuốn sách này, các nội dung cài đặt, thuật toán, ứng dụng … chỉ dành cho OpenCV phiên bản 2.x trên nền tảng hệ điều hành Window. OpenCV  rất đa dạng, nó hỗ trợ rất nhiểu hệ điểu hành như: Window, Linux và MacOSX.

OpenCV có những điểm đặc trưng sau:

* OpenCV có rất nhiều chức năng. Bạn có thể bất ngờ khi lần đầu tiếp xúc với nó. Tuy nhiên, bạn sẽ chỉ cần một vài lần đầu để khởi động nó
* Sau đây là những tóm tắt cơ bản về hệ thống các về chức năng của các hàm trong OpenCV

Image and Video I/O: Những giao diện này sẽ giúp bạn đọc được dự liệu ảnh từ file hoặc trực tiếp từ video. Bạn cũng có thể tạo các file ảnh và video với giao diện này

Thị giác máy và các thuật toán xử lý ảnh ( General computer-vision and image-processing algorithms(mid – and low level APIs)): Sử dụng những giao diện này, bạn có thể thực hành với rất nhiều chuẩn thị giác máy mà không cần phải có mã nguồn của chúng.

Modul thị giác máy ở cấp độ cao: OpenCV gồm một vài tác dụng ở cấp độ cao. Thêm vào nhận dạng mặt, dò tìm, theo dõi. Nó bao gồm luồng thị giác (sử dụng camera di động để xác định cấu trúc 3D), kích cỡ camera và âm thanh nổi.

AI and machine-learning: Các ứng dụng của thị giác máy thường yêu cầu máy móc phải học (machine learning) hoặc các hình thức trí tuệ nhân tạo khác. Một vài trong số chúng là có sẵn trong gói OpenCV

Lấy mẫu ảnh và phép biến đổi: Nó thường rất tốt cho quá trình xử lý một nhóm phần tử ảnh như là một đơn vị. OpenCV bao gồm  lấy tách ra, lấy mẫu ngẫu nhiên, phục chế, xoay ảnh, làm cong ảnh ( warping), thay đổi hiệu ứng của ảnh.

Cách thức tạo và phân tích ảnh nhị phân: Ảnh nhị phân thường xuyên được dùng trong các hệ thống kiểm tra có khuyết điểm hình dạng hoặc các bộ phận quan trọng. Sự biểu diễn ảnh  cũng rất thuận tiện khi chúng ta biết rõ vật thể cần bắt.

Cách thức cho tính toán thông tin 3D (methods for computing 3D information): Những hàm này  rất có ích khi cần sắp xếp và xác định với một khối lập thể (with a stereo rig) hoặc với không gian nhìn phưc tạp (multiple views) từ một camera riêng.

Các phép toán cho xử lý ảnh, thị giác máy và biểu diễn ảnh( image interpretation): OpenCV sử dụng các phép toán phổ biến như: đại số học, thống kê và tính toán hình học

Đồ họa: Những giao diện này giúp bạn viết chữ và vẽ trên hình ảnh. Thêm vào đó những chức năng này được sử dụng nhiểu trong ghi nhãn và đánh dấu. Ví dụ nếu bạn viết một chương trình cần nhận dạng nhiểu đối tượng thì nó sẽ rất có ích cho tạo nhãn ảnh ( label image) với kích thước và vị trí.

Phương thức GUI: OpenCV bao gồm cửa sổ  giao diện của chính bản thân nó. Trong khi đó những giao diện này được so sánh giới hạn với khả năng có thể thực hiện  trong mỗi môi trường. Chúng cung cấp những môi trường API đa phương tiện và đơn giản để hiển thị hình ảnh, cho phép người dùng nhập dữ liệu thông qua chuột, bàn phím và điều khiển quá trính.

Cấu trúc dữ liệu và giải thuật: Với những giao diện này bạn có thể giữ lại, tìm kiếm, lưu và cách danh mục điều khiển, các tuyển tập (cũng  như các tập hợp lệnh được gọi), đồ họa và sơ đồ nhánh một cách hiệu quả.

Khả năng tồn tại lâu dài của dữ liệu (Data persistence): Những phương pháp này cung cấp các giao diện một cách thuận lợi để lưu trữ các dạng khác nhau của dữ liệu vào đĩa để có thể khôi phục khi cần thiết.

### Cách tổ chức OpenCV

Cấu trúc OpenCV gồm nhiều module. CXCORE bao gồm các dạng dữ liệu cơ bản rõ ràng. Ví dụ cấu trúc dự liệu về ảnh, điểm, hình chữ nhật được xác định trong file cxtypes.h. CXCORE chứa các phép toán đại số tuyến tính và thống kê, các hàm lưu trữ lâu dài (persistence fun…) và các lỗi thao tác. Có điều lạ lung thay là các hàm đồ họa được cho việc vẽ ảnh cũng được lưu trữ tại đây.

CV chứa đựng quá trình xử lý ảnh và các phương pháp đánh giá sơ bộ kích thước ảnh. Những hàm tính toán hình học cũng được lưu trữ tại đây. CVAUX được mô tả trong văn bản của OpenCV như là modul cũ và chỉ dùng để thí nghiệm. Tuy nhiên, giao diện đơn giản nhất cho nhận dạng mặt được nằm trong modul này. Những mã nguồn nằm trong module  này rất phù hợp cho việc nhận dạng mặt và chúng được sử dụng rộng rãi cho mục đích này.

Những hàm còn lại được nằm trong HighGUI và CVCAM. Cả hai đều nẳm ở thư mục có tên “otherlibs”, sử dụng chúng rất dễ gặp lỗi. Vì rằng HighGUI chứa các thư viện vào ra cơ bản , bạn sẽ muốn chắc chắn hơn, đừng bỏ sót nó.Nó cũng chứa đựng nhiều cửa sổ đa chức năng.

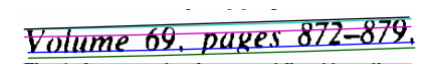
CVCAM là thư viện chứa các hàm truy nhập video thông qua DirectX trên môi trường Window 32 bit. Tuy nhiên, HighGUI cũng có các giao diện video.Trong bài báo này, tôi chỉ xem xét HighGUI. Chúng đơn giản, làm việc trên nhiều môi trường. Nếu bạn sử dụng Window XP hoặc 2000, bạn có thể làm tăng hiểu quả của nó bằng cách chuyển qua giao diện CVCAM, nhưng học OpenCV sẽ đơn giản hơn khi dùng HighGUI.

## Tiến trinh nhận dạng sử dụng Tesseract OCR

Mục đích của bước này là nhận dạng các dòng của các hình ảnh bị nghiêng, giúp giảm sự mất thông tin khi nhận dạng ảnh nghiêng. Các bộ phận quan trọng của quá trình này là lọc dãy màu (còn được gọi là blobs) và xây dựng dòng. Bước này cũng giúp loại bỏ các văn bản có dropcap.

### Thiết lập dòng cơ sở

Khi dòng văn bản được tìm thấy, các dòng cơ sở được thiết lập chính xác hơn bằng cách sử dụng một đường có tên là spline toàn phương (là dòng mà được kết hợp từ nhiều đoạn). Nó giúp Tesseract xử lý các trang có đường cơ sở là đường cong. Các dòng cơ sở được thiết lập bằng cách phân vùng các blobs thành các nhóm có thể thay thế thích hợp liên tục trong đường cơ sở thẳng ban đầu. Một spline toàn phương được thiết lập cho phân vùng dày đặc nhất, (giả định là đường cơ sở) của một hình có phương ít nhất. Spline có lợi thế là tính toán ổn định, nhược điểm là sự gián đoạn có thể xảy ra khi nhiều phân đoạn spline được yêu cầu.



Hình 2. Ví dụ về một đường cơ sở dạng cong

### Cắt nhỏ từ

Tesseract sẽ xác định xem có các ký tự dính với nhau trong một từ hay không. Nếu có nó sẽ cắt nhỏ các ký tự ra thành các ký tự riêng lẻ.



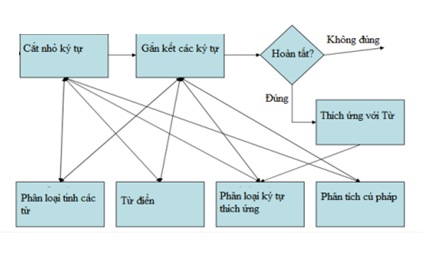
Hình 2. Ví dụ về xác định ký tự

### Nhận dạng khoảng cách giữa số hoặc chữ

Xác định khoảng cách giữa các số hoặc giữa các chữ là một vấn đề khá phức tạp. Tesseract giải quyết những vấn đề này bằng cách đo khoảng cách trong một phạm vi hạn chế theo chiều dọc giữa dòng cơ sở và dòng trung bình.

### Nhận dạng từ

Quá trình nhận dạng một từ là quá trình phân tích một từ được chia ra thành các ký tự như thế nào.



Hình 2. Quá trình nhận dạng từ

Khi kết quả xuất ra một từ mà nó không thỏa mãn nhu cầu thì Teseract cố gắng cải thiện kết quả này bằng cách cắt nhỏ các từ có nghĩa không tốt nhất. Nếu việc cắt nhỏ không làm tăng chất lượng từ thì nó sẽ phục hồi lại từ trước đó.

## Đánh giá độ chính xác nhận dạng Tessract OCR

Hình ảnh muốn được OCR cần quét ở độ phân giải từ 200 DPI (dot per inch) trở lên tới 400 trong trắng đen hoặc grayscale. Quét ảnh với độ phân giải cao hơn nữa chưa hẳn sẽ tăng sự chính xác của kết quả nhận dạng. Hiện tại, mức chính xác có thể lên trên 97% cho Tiếng Việt ([ảnh chuẩn](http://downloads.sourceforge.net/vietocr/vietsample.zip)), và phiên bản tới của Tesseract có thể nâng cao độ nhận dạng hơn nữa. Dầu vậy, độ chính xác thực thụ vẫn còn tùy thuộc rất lớn vào phẩm chất của ảnh quét.

Thông số tiêu biểu cho quét ảnh là 300 DPI và 1 bpp (bit per pixel) trắng đen hoặc 8 bpp grayscale dạng không nén (uncompressed) TIFF hay PNG. PNG nhỏ gọn hơn những dạng ảnh khác mà vẫn giữ được chất lượng cao nhờ sử dụng thuật toán lossless data compression; TIFF có lợi điểm ở khả năng chứa nhiều trang ảnh (multi-page) trong một file.

Chế độ Screenshot Mode cung cấp độ nhận dạng tốt hơn cho những hình ảnh có độ phân giải thấp, chẳng hạn như ảnh in màn hình, bằng cách rescaling chúng tới 300 DPI.

## Công cụ và ngôn ngữ sử dụng

### Công cụ

IDE (Integrated Development Environment): là một phần mềm hỗ trợ lập trình viên viết phần mềm. Nó bao gồm trình soạn thảo mã lệnh (code), môi trường chạy, trình gỡ lỗi (Debug)… Một số IDE nổi tiếng: Netbeans, Eclipse (Dùng để lập trình Java), Android Studio (Dùng để lập trình Android), Micrsoft Visual Studio (dùng để lập trình C#, .Net), Dev C++ (dùng để lập trình C/C++)…

Eclipse là một trình IDE dùng để lập trình Java (Eclipse cũng có thể dùng để lập trình các ngôn ngữ khác nhau C/C++, PHP… nhưng mạnh nhất là về Java). Một số IDE khác cũng dùng để lập trình Java khác như Netbeans, Intellij IDEA. Trong đó Netbeans mình thấy chủ yếu dùng trong giảng dạy, còn Intellij IDEA thì khá mới nhưng rất mạnh mẽ và hỗ trợ nhiều tính năng tuy nhiên nó hơi khó dùng với người mới làm quen.

Có nhiều công cụ để phát triển Android nhưng đến nay công cụ chính thức và mạnh mẽ nhất là Android Studio. Đây là IDE (Môi trường phát triển tích hợp) chính thức cho nền tảng Android, được phát triển bởi Google và được sử dụng để tạo phần lớn các ứng dụng mà mỗi người có thể sử dụng hàng ngày. Android Studio lần đầu tiên được công bố tại hội nghị Google I/O vào năm 2013 và được phát hành cho công chúng vào năm 2014 sau nhiều phiên bản beta khác nhau. Trước khi được phát hành, các nhà phát triển Android thường sử dụng các công cụ như Eclipse IDE, một IDE Java chung cũng hỗ trợ nhiều ngôn ngữ lập trình khác.

Chức năng của Android Studio là cung cấp giao diện để tạo các ứng dụng và xử lý phần lớn các công cụ quản lý file phức tạp đằng sau hậu trường. Ngôn ngữ lập trình được sử dụng ở đây là [Java](https://quantrimang.com/java-la-gi-tai-sao-nen-chon-java-143130) và được cài đặt riêng trên thiết bị của người dùng. Android Studio rất đơn giản, lập trình viên chỉ cần viết, chỉnh sửa và lưu các dự án của mình và các file trong dự án đó. Đồng thời, Android Studio sẽ cấp quyền truy cập vào Android SDK. Hãy coi đây là đuôi cho code Java cho phép nó chạy trơn tru trên các thiết bị Android và tận dụng lợi thế của phần cứng gốc. Lập trình viên cần sử dụng ngôn ngữ lập trình Java để viết các chương trình, Android SDK có nhiệm vụ kết nối các phần này lại với nhau. Cùng lúc đó Android Studio kích hoạt để chạy code, thông qua trình giả lập hoặc qua một phần cứng kết nối với thiết bị. Sau đó, bạn cũng có thể “gỡ rối” chương trình khi nó chạy và nhận phản hồi giải thích sự cố, v.v… để lập trình viên có thể nhanh chóng giải quyết vấn đề.

Google đã nỗ lực rất nhiều để làm cho Android Studio trở nên mạnh mẽ và hữu ích nhất có thể. Nó cung cấp những gợi ý trực tiếp trong khi viết code và thường đề xuất những thay đổi cần thiết để sửa lỗi hoặc làm code hiệu quả hơn. Ví dụ, nếu không sử dụng biến, biến đó sẽ được tô đậm bằng màu xám. Và khi bắt đầu gõ một dòng code, Android Studio sẽ cung cấp danh sách gợi ý tự hoàn thành để giúp lập trình viên hoàn thiện dòng code đó. Chức năng này rất hữu ích khi bạn không nhớ được chính xác cú pháp hoặc để tiết kiệm thời gian.

Với đề tài “Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân”, em sử dụng công cụ Eclipse và Android Studio để hoàn thành dự án.

### Ngôn ngữ

**Java** là một **một ngôn ngữ lập trình hiện đại, bậc cao, hướng đối tượng, bảo mật và mạnh mẽ.** và là một **Platform**. Bất cứ môi trường phần cứng hoặc phần mềm nào mà trong đó có một chương trình chạy, thì được hiểu như là một Platform. Với môi trường runtime riêng cho mình (JRE) và API, Java được gọi là Platform.

Ngôn ngữ lập trình Java ban đầu được phát triển bởi **Sun Microsystems** do **James Gosling** khởi xướng và phát hành vào năm 1995. Phiên bản mới nhất của Java Standard Edition là Java SE 8. Với sự tiến bộ của Java và sự phổ biến rộng rãi của nó, nhiều cấu hình đã được xây dựng để phù hợp với nhiều loại nền tảng khác nhau. Ví dụ: J2EE cho các ứng dụng doanh nghiệp, J2ME cho các ứng dụng di động.

Các phiên bản J2 mới đã được đổi tên thành Java SE, Java EE và Java ME. Phương châm của java là **"Write Once, Run Anywhere"** - viết một lần chạy nhiều nơi, nghĩa là bạn chỉ cần viết một lần trên window chẳng hạn, sau đó vẫn chương trình đó bạn có thể chạy trên Linux, Android, các thiết bị J2ME...

Ngôn ngữ lập trình java có các tính năng sau:

* **Hướng đối tượng** - Trong Java, mọi thứ đều là một Object. Java có thể dễ dàng mở rộng và bảo trì vì nó được xây dựng dựa trên mô hình Object.
* **Nền tảng độc lập** - Không giống nhiều ngôn ngữ lập trình khác bao gồm cả C và C ++, khi Java được biên dịch, nó không được biên dịch thành ngôn ngữ máy nền tảng cụ thể, thay vào mã byte - nền tảng độc lập. Mã byte này được thông dịch bởi máy ảo (JVM) trên nền tảng nào đó mà nó đang chạy.
* **Đơn giản** - Java được thiết kế để dễ học. Nếu bạn hiểu khái niệm cơ bản về OOP Java, sẽ rất dễ để trở thành master về java.
* **Bảo mật** - Với tính năng an toàn của Java, nó cho phép phát triển các hệ thống không có virut, giả mạo. Các kỹ thuật xác thực dựa trên mã hoá khóa công khai.
* **Kiến trúc - trung lập** - Trình biên dịch Java tạo ra định dạng tệp đối tượng kiến trúc trung lập, làm cho mã biên dịch được thực thi trên nhiều bộ vi xử lý, với sự hiện diện của hệ điều hành Java.
* **Portable** - Là kiến trúc tập trung và không có khía cạnh thực hiện phụ thuộc của đặc tả này làm cho Java khả chuyển. Trình biên dịch trong Java được viết bằng ANSI C, đó là một tập con POSIX.
* **Mạnh mẽ** - Java làm nỗ lực để loại trừ các tình huống dễ bị lỗi bằng cách kiểm tra lỗi tại thời gian biên dịch và kiểm tra lỗi tại runtime.
* **Đa luồng** - Với tính năng đa luồng của Java có thể viết các chương trình có thể thực hiện nhiều tác vụ đồng thời. Tính năng thiết kế này cho phép các nhà phát triển xây dựng các ứng dụng tương tác có thể chạy trơn tru hơn.
* **Thông dịch** - Mã byte Java được dịch trực tiếp tới các máy tính gốc và không được lưu trữ ở bất cứ đâu.
* **Hiệu năng cao** - Với việc sử dụng trình biên dịch Just-In-Time, Java cho phép thực hiện hiệu năng cao.
* **Phân tán** - Java được thiết kế cho môi trường phân tán của Internet.
* **Năng động** - Java là năng động hơn C hoặc C++ vì nó được thiết kế để thích nghi với môi trường đang phát triển. Các chương trình Java có thể mang một lượng lớn thông tin tại runtime mà có thể được sử dụng để xác minh và giải quyết các truy cập vào các đối tượng tại runtime.

Có 4 kiểu ứng dụng chính của java app:

* **Standalone App**: Standalone App cũng được biết đến như Desktop App hoặc Window-based App. Để tạo ra ứng dụng kiểu này người ta thường sử dụng AWT, Swing hoặc JavaFX framework.
* **Web App**: Web App là ứng dụng chạy trên server và tạo được các trang động. Hiện nay, servlet, jsp, struts, jsf, spring... là những công nghệ được sử dụng để tạo Web App trong java.
* **Enterprise App**: Một ứng dụng dạng như Banking App, có lợi thế là tính bảo mật cao, cân bằng tải (load balancing) và clustering. Trong java, EJB được sử dụng để tạo các Enterprise App.
* **Mobile App**: Mobile App là ứng dụng được tạo ra cho các thiết bị di động. Hiện nay Android và Java ME được sử dụng để chạy các ứng dụng này.

Có 4 nền tảng Java:

* **Java SE (Java Standard Edition)**: Java SE là một nền tảng lập trình Java. Nó bao gồm các API lập trình Java như java.lang, java.io, java.net, java.util, java.sql, java.math, v.v. Nó bao gồm các chủ đề cốt lỗi như OOPs, String, Regex, Exception, Inner classes, Multithreading, I/O Stream, Networking, AWT, Swing, Reflection, Collection, v.v.
* **Java EE (Java Enterprise Edition)**: Đây là một nền tảng doanh nghiệp chủ yếu được sử dụng để phát triển các ứng dụng web và doanh nghiệp. Nó được xây dựng trên nền tảng Java SE. Nó bao gồm các chủ đề như Servlet, JSP, Web Services, EJB, JPA , v.v.
* **Java ME (Java Micro Edition)**: Đây là một nền tảng vi mô chủ yếu được sử dụng để phát triển các ứng dụng di động.
* **JavaFX**: JavaFX là một nền tảng phần mềm phát triển các ứng dụng Internet phong phú (Rich Internet Applications – RIAs) có thể chạy trên nhiều loại thiết bị, nhiều hệ điều hành khác nhau. JavaFX là một giải pháp công nghệ cho GUI trên nền tảng Java nhằm tạo giao diện đồ họa người dùng dựa trên Swing và Java2D.

Để thực hiện đề tài “Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân”, ngôn ngữ mà em sử dụng là Java.

## Kết luận

Trong chương này, em đã giưới thiệu về những kiến thức đã tìm hiểu cùng những công cụ để phục vụ cho đề tài “Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân”.

# PHÂN TÍCH THIẾT KẾ HỆ THỐNG

Chương này bao gồm các bước về phân tích và thiết kế ứng dụng từ tổng thể đến từng chức năng của ứng dụng.

Mô hình tổng thể hệ thống xác định người sử dụng của hệ thống và các nhiệm vụ mà họ phải thực hiện với hệ thống:

* Người sử dụng (Actor) gồm: Người dùng.
* Nhiệm vụ (Use case) gồm: Chọn ảnh/chụp ảnh, crop ảnh, xử lý ảnh, thoát.

## Khảo sát về thẻ căn cước công dân

Thẻ căn cước công dân hay còn được gọi với cái tên thông dụng khác như: Thẻ nhận dạng cá nhân, chứng mình thư nhân dân. Loại thẻ này là một trong những tài liệu để có thể nhận dạng từng công dân, được sử dụng để xác minh tất cả các chi tiết quan trọng mang bản sắc cá nhân của mỗi người. Tất cả những thông tin nhận dạng theo quy định của Nhà nước được trình bày dưới hình thức của một chiếc thẻ nhỏ, có kích thước theo tiêu chuẩn mà Nhà nước đề ra.

Theo đó, chúng ta có thể hiểu cụ thể về thẻ căn cước công dân như sau:

* Căn cước công dân: Căn cước công dân chính là các thông tin cơ bản về lai lịch, đặc điểm nhận dạng của công dân từ đủ 14 tuổi trở lên. Trong đó, đặc điểm nhận dạng công dân chính là đặc điểm cá biệt và vô cùng ổn định về vẻ bề ngoài của công dân đó, nhằm phân biệt giữa công dân này với công dân khác.
* Thẻ căn cước công dân: Thẻ căn cước công dân chính là một loại thẻ thể hiện tất cả những thông tin căn cước công dân. Đây là loại thẻ thể hiện được lao lịch, đặc điểm nhận dạng của mỗi công dân. Thẻ căn cước công dân dần thay thế cho Chứng minh thư nhân dân, loại thẻ này có kích cỡ nhất định, được làm bằng nhựa Plastic (được làm từ phôi thẻ từ giống như thẻ ATM mà các ngân hàng sử dụng)

A close up of text on a white background

Description automatically generated

Hình 3. Thẻ căn cước công dân

Thẻ căn cước công dân có vai trò quan trọng trong việc chúng minh công dân của người được cấp thẻ, giúp các công dân thực hiện các giao dịch trên cả nước. Loại thẻ này có thể được sử dụng thay cho hộ chiếu (khi giữa Việt Nam và các nước khác có sự ký kết cho phép thẻ căn cước công dân có thể thay cho hộ chiếu)

Nội dung cần có theo quy định trong thẻ căn cước công dân bao gồm: Họ và tên công dân, ngày/tháng/năm sinh của người công dân, đặc điểm nhận dạng, dấu vân tay, ảnh. Đó là những thông tin cơ bản trong thẻ căn cước. Tuy nhiên, một số quốc gia trên thế giới có yêu cầu về một số chi tiết khác như là: Thông tin của cha mẹ (họ tên), dân tộc, quê quán, giới tính…

Cụ thể:

* Mặt trước thẻ căn cước: Số thẻ căn cước công dân, họ và tên người công dân, ngày/tháng/năm sinh của công dân, giới tính, quốc tịch, quê quán và chức danh của người cấp thẻ, dấu xác nhận của cơ quan cấp thẻ.
* Mặt sau: Mã vạch của từng thẻ, vân tay ngón trỏ 2 bên phải và trái, đặc điểm nhận dạng, ngày cấp, nơi cấp, họ tên và chức danh của người cấp thẻ, chữ ký của người cấp thẻ, dấu xác nhận của cơ quan cấp thẻ.

A picture containing cat, sitting, laying, bird

Description automatically generated

Hình 3. Thông tin trên thẻ căn cước công dân

## Phân tích và thiết kế tổng thể ứng dụng

### Yêu cầu chức năng của hệ thống

Yêu cầu chức năng của hệ thống bao gồm:

* Cho phép truy cập vào thư viện ảnh của điện thoại
* Cho phép sử dụng camera để chụp ảnh
* Crop ảnh vào phần muốn xử lý
* Xử lý ảnh để trích xuất chữ từ ảnh
* Xác định được chữ tiếng Việt và các ký tự đặc biệt
* Hiện thị thông tin đã thông qua xử lý trên màn hình

### Yêu cầu phi chức năng của hệ thống

Yêu cầu phi chức năng của hệ thống gồm:

* Thời gian xử lý hình ảnh nhanh
* Giao diện thân thiện dẽ sử dụng
* Ứng dụng chạy không ổn định, đảm bảo yêu cầu của người dùng

### Mô tả chi tiết các chức năng của ứng dụng

Ứng dụng bao gồm 5 chức năng chính: chọn ảnh, chụp ảnh, crop ảnh, xử lý ảnh và thoát chương trình. Những chức năng này được thể hiện trong bảng … Cụ thể, chức năng “Chọn/chụp ảnh”, là chức năng tương tác với người dùng. Chức năng này hỗ trợ người dùng có thể chọn ảnh trong thư viện có sẵn hoặc chụp ảnh mà mình mong muốn để đưa vào chương trình. Thao tác của người dùng trong chức năng “Chọn/chụp ảnh”: Đầu tiên người sử dụng chọn icon chọn ảnh có trên thanh công cụ. Giao diện sẽ hiển thị cho người dùng chọn 1 trong 2 chức năng là Chọn hoặc Chụp ảnh. Khi người dùng nhấn vào phần Chọn ảnh, ứng dụng sẽ di chuyển vào thư viện ảnh có trong điện thoại để người dùng có thể thỏa thích chọn ảnh mà mình mong muốn. Ngược lại, khi người dùng chọn phần Chụp ảnh, người dùng có thể tự mình chụp thẻ căn cước công dân để để lên hê thống xử lý. Sau khi đã chọn/chụp xong ảnh, tiếp theo sẽ đến phần crop ảnh đầu vào – bước tiền xử lý ảnh đầu vào.

Trước khi tiến vào bước “crop ảnh”, tôi sẽ trình bày về thuật ngữ crop ảnh và công dụng của crop ảnh trong AI nói riêng và xử lý ảnh nói chung. Crop ảnh là một cách đơn giản để tái bố cục và tạo điểm nhân cho bức ảnh. Với việc sử dụng cách này, ảnh sẽ được loại bỏ các khuyết điểm và trập trung vào nhưng phần nổi bật. Đối với AI nói riêng và xử lý ảnh nói chung, kỹ thuật crop ảnh thường xuyên được xử dụng trong bước tiền xử lý ảnh đầu vào để tập trung vào những điểm ảnh quan trong nhất và thu thập được một bộ dữ liệu hoàn chỉnh.

A close up of text on a white background

Description automatically generated

Hình 3. Ví dụ về crop ảnh

Trong bước này, người dùng sẽ di chuyển các đường biên của hình ảnh để giới hạn lại hình ảnh và tạo ra hình ảnh đầu vào tốt nhất cho hệ thống xử lý. Khi đã căn chỉnh phù hợp, người dùng nhân vào núy Cắt để xác nhận đã kết thúc quá trình crop. Khi đó, ứng dụng sẽ nhận được tín hiệu của người dùng và tiến hành crop ảnh, chỉ để lại phần ảnh mà người dùng đã chọn và kết thúc quá trình tiền xử lý ảnh. Kết quả thu được ảnh đầu vào hoàn hảo và được hệ thống tiến hành phân tích và xử lý đề trích xuất được thông tin có trong ảnh.

Áp dụng các thuật toán về ML và sử dụng thư viện Tesseract OCR, hệ thống sẽ tiến hành phân tích dữ liệu ở trên ảnh, đóng khối lại những phần có chữ và phân vùng chúng. Tiếp đến, hệ thống sẽ xử lý từng khối đã thu thập được và tiến hành lấy dòng cơ sở cho từng khối này. Sau đó, hệ thống sẽ tiến hành phân biệt từ và số và tiến hành vẽ lại các ký tự để hiển thị lên trên màn hình. Cuối cùng ta thu được kết quả là chuỗi ký tự mang thông tin trong hình ảnh căn cước công dân từ đầu vào. Bảng 3.1 thể hiện các chức năng chính của hệ thống.

Bảng 3. Các chức năng chính của hệ thống

|  |  |
| --- | --- |
| Chọn/chụp ảnh | |
| Tác nhân | Người dùng |
| Mô tả | Người dùng chọn/chụp ảnh trong thư viện để load vào chương trình. |
| Sự kiện | Nhấn vào chọn ảnh trên menu giao diện  Chọn ảnh trong thư viện  Nhấn Chọn  Hoặc nhấn chụp ảnh để chụp |
| Kết quả | Ảnh đầu vào được load lên |
| Crop ảnh | |
| Tác nhân | Người dùng |
| Mô tả | Người dùng sau khi chọn ảnh đầu vào xong sẽ tiến hành thao tác crop ảnh |
| Sự kiện | Di chuyển các đường viền để crop ảnh |
| Kết quả | Trả ra ảnh đã được crop |
| Xử lý hình ảnh | |
| Tác nhân | Người dùng |
| Mô tả | Ứng dụng xử lý ảnh đầu vào đã thông qua crop |
| Sự kiện | Ấn OK để tiến hành xử lý |
| Kết quả | Hiển thị thông tin được trích xuất từ ảnh |
| Thoát chương trình | |
| Tác nhân | Người dùng |
| Mô tả | Người dùng nhấn vào thoát trong giao diện |
| Sự kiện | Nhấn vào “Exit” |
| Kết quả | Ra khỏi chương trình |

## Biểu đồ use case

Use case là kỹ thuật được dùng trong kỹ thuật phần mềm và hệ thống nhằm nắm bắt những yêu cầu chức năng của hệ thống. Use case mô tả sự tương tác đặc trưng giữa người dùng bên ngoài (actor) và hệ thống. Use case cũng mô tả các yêu cầu đối với hệ thống.

Mỗi Use case sẽ mô tả cách thức người dùng bên ngoài tương tác với hệ thống để đạt được mục tiêu nào đó. Một hoặc nhiều kịch bản (scenario) có thể được tạo ra từ mỗi use case, tương ứng với chi tiết về mỗi cách thức đạt được mục tiêu nào đó. Khi mô tả Use case, người ta thường tránh dùng thuật ngữ kỹ thuật, thay vào đó họ sử dụng ngôn ngữ của người dùng cuối hoặc chuyên gia về lĩnh vực đó. Để tạo ra use case, cần phải có sự hợp tác chặt chẽ giữa người phân tích hệ thống và người dùng cuối. Một trong những cách biểu diễn trực quan phổ biến hiện nay là lược đồ use case của UML.

[Use case](https://phongvu.vn/thiet-bi-tin-hoc/case-thung-may) thường được đặt tên giống động từ hoặc động từ + cụm danh từ. Đặc điểm của tên Use case là ngắn gọn, cụ thể và miêu tả đầy đủ nghĩa của đối tượng người dùng. Những động từ như “do”, “perform”, các danh từ như ”data”, “information” nên tránh nếu có thể. Người dùng sử dụng Use case để đại diện cho các nghiệp vụ trong hệ thống.

Một mô hình Use case là một loại của ngôn ngữ mô hình hóa thống nhất(UML) được định nghĩa bởi Object Management Group(OMG). Mô hình Use case cung cấp cái nhìn trực quan đồ họa về nghiệp vụ của hệ thống(đại diện là Use case), người dùng hệ thống. Mô hình hóa nhằm tạo ra một tài liệu mô tả tổng thể về hệ thống bằng cách xác định mục đích của người dùng. Các [Use case](https://phongvu.vn/thiet-bi-tin-hoc/case-thung-may) trong mô hình Use case có thể được tổ chức, sắp xếp theo logic nghiệp vụ trong hệ thống và mức độ trìu tượng, ảnh hưởng đến người dùng. Chúng có thể được kết nối để hiển thị những tác động, những quan hệ phụ thuộc và mở rộng. Mục đích chính của việc mô hình hóa Use case là để tạo ra một tài liệu mô tả tổng thể về hệ thống bằng cách xác định mục đích người dùng.

Hình 3.4 dưới đây mô tả biểu đồ use case tổng quát.

A close up of a map

Description automatically generated

Hình 3. Biểu đồ use case tổng quát

Bảng 3.2 sau để sẽ mô tả các use case được sư dụng trong đề tài “Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân”

Bảng 3. Mô tả các use case

|  |  |  |
| --- | --- | --- |
| STT | TÊN USE CASE | Ý NGHĨA/GHI CHÚ |
| 1 |  | Người dùng chương trình |
| 2 |  | Chọn ảnh đầu vào |
| 3 |  | Chụp ảnh đầu vào |
| 4 |  | Crop ảnh đầu vào |
| 5 |  | Bắt đầu xử lý ảnh |
| 6 |  | Thoát chương trình |

Đặc tả use case “Chọn ảnh”: Khi người dung nhấn “Chọn ảnh” trên giao diện, hê thống bắt đầu truy cập đến phần thư viện ảnh của điện thoại để người dùng chọn ảnh. Dòng sự kiện chính của use case “Chọn ảnh” là người dùng chọn ảnh và kết thúc use case.Trạng thái của hệ thống khi bắt đầu thực hiện use case là truy cập vào thư viện để chọn ảnh. Trang thái của hệ thống sau khi thực hiện xong này là chọn xong được ảnh đầu vào để tiến hành tiền xử lý.

Đặc tả use case “Chụp ảnh”: Khi người dùng chọn phần chụp ảnh trên giao diện, ứng dụng sẽ truy cập vào camera của điện thoại để người dùng tiến hành thao tác chụp ảnh đầu vào. Dòng sự kiện chính của use case “Chụp ảnh” là người dùng chọn phần chụp ảnh, tiến hành chụp ảnh đầu và và kết thúc use case. Trạng thái của hệ thống khi bắt đầu thực hiện use case là truy cập vào camera để chụp ảnh đầu vào. Trạng thái của hệ thống sau khi thực hiện use case là chụp được ảnh đầu và để tiến hành tiền xử lý.

Đặc tả use case “Crop ảnh”: Khi người dùng di chuyển các đường biên của ảnh đầu vào cho đến khi tập trung vào hình ảnh mong muốn, người dùng thực hiện thao tác nhấn nút “Cắt” trên giao diện để tiến hành crop hình ảnh – tiền xử lý ảnh. Dòng sự kiện chính của use case này là người dùng di chuyển các đường biên của ảnh đầu vào, nhấn nút “Cắt” để thực hiện thao tác crop và kết thúc use case. Trạng thái của hệ thống khi bắt đầu use case là khi ứng dụng đang trong giao diện crop. Trạng thái của hệ thống sau khi thực hiện use case là ảnh đầu vào đã xong bước tiền xử lý.

Đặc tả use case “Xử lý ảnh”: Sau khi ảnh đã được crop xong, người dùng ấn nút “OK” để xác thực hình ảnh đầu vào đã qua bước tiền xử lý và hệ thống sẽ tiến hành xử lý, trích xuất thông tin từ ảnh căn cước công dân đầu vào. Dòng sự kiện chính của use case này là người dùng ấn nút “OK” để tiến hành xử lý trích xuất thông tin từ ảnh căn cước công dân đầu vào và kết thúc use case. Trạng thái của hệ thống khi bắt đầu use case là khi ảnh đầu vào bắt đầu bước vào quá trình xử lý trích xuất thông tin. Trạng thái của hệ thống sau khi thực hiện use case là hiển thị kết quả trích xuất thông tin trên màn hình điện thoại.

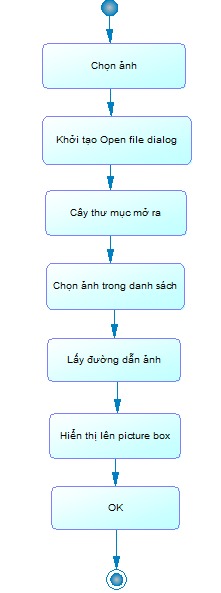
Đặc tả use case “Thoát”: Khi người dùng muốn thoát chương trình, người dùng chọn nút “Thoát” trên ứng dụng để thoát khỏi chương trình. Dòng sự kiện chính của use case này là người dùng chọn nút “Thoát” để kết thúc chương trình và kết thúc use case. Trạng thái của hệ thống khi bắt đầu use case này là bất cứ khi nào người dùng muốn thoát chương trình. Trạng thái của hệ thống sau khi thực hiện use case là chương trình đã kết thúc.

## Biểu đồ trạng thái

Trạng thái là một trong các điều kiện có thể để đối tượng tồn tại, là kết quả của một hoạt động trước đó của đối tượng. Trạng thái của đối tượng thường được mô tả trong hình chữ nhật góc tròn và được xác định bởi: Tên gọi trạng thái, thường bắt đầu bằng động từ, Biến trạng thái mô tả các giá trị hiện thời của trạng thái, Hoạt động là hành vi mà đối tượng sẽ thực hiện khi nó ở vào trạng thái đó.

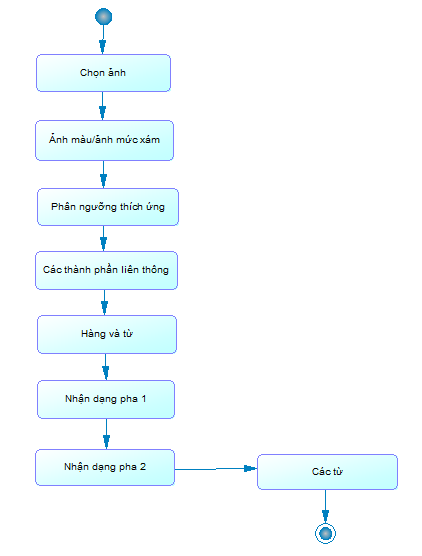
Biểu đồ trạng thái là dạng biểu đồ mô tả trạng thái có thể có và sự chuyển đổi giữa các trạng thái đó khi có các sự kiện tác động của một đối tượng. Đối với các đối tượng có nhiều trạng thái thì biểu đồ trạng thái là sự lựa chọn tốt nhất giúp chúng ta có thể hiểu rõ hơn về hệ thống. Ứng dụng hiện tại có 2 trạng thái tiêu biểu là trạng thái chọn ảnh và trạng thái xử lý ảnh.

Hình 3.5 mô tả trạng thái chọn ảnh của ứng dụng. Bao gồm các bước chọn ảnh trong thư mục đến khi lấy được đường dẫn và hiển thị hình ảnh lên picture box.



Hình 3. Biểu đồ trạng thái chọn ảnh

Hình 3.6 mô tả từng bước xử lý ảnh. Từ việc xác định ảnh màu/mức xám, phân ngưỡng thích ứng cho ảnh cho đến xác định hàng và từ. Cuối cùng là nhận diện ra các từ thông qua việc nhận dạng ký tự.



Hình 3. Biểu đồ trạng thái xử lý ảnh

## Biểu đồ hoạt động

Biểu đồ hoạt động nắm bắt hành động và các kết quả của chúng. Biểu đồ hoạt động tập trung vào công việc được thực hiện trong khi thực thi một thủ tục (hàm), các hoạt động trong một lần thực thi một trường hợp sử dụng hoặc trong một đối tượng. Biểu đồ hoạt động là một biến thể của biểu đồ trạng thái và có một mục tiêu tương đối khác, đó là nắm bắt hành động (công việc và những hoạt động phải được thực hiện) cũng như kết quả của chúng theo sự biến đổi trạng thái. Các trạng thái trong biểu đồ hoạt động (được gọi là các trạng thái hành động) sẽ chuyển sang giai đoạn kế tiếp khi hành động trong trạng thái này đã được thực hiện xong (mà không xác định bất kỳ một sự kiện nào theo như nội dung của biểu đồ trạng thái). Một sự điểm phân biệt khác giữa biểu đồ hoạt động và biểu đồ trạng thái là các hành động của nó được định vị trong các luồng (swimlane). Một luồng sẽ gom nhóm các hoạt động, chú ý tới khái niệm người chịu trách nhiệm cho chúng hoặc chúng nằm ở đâu trong một tổ chức. Một biểu đồ hoạt động là một phương pháp bổ sung cho việc miêu tả tương tác, đi kèm với trách nhiệm thể hiện rõ các hành động xảy ra như thế nào, chúng làm gì (thay đổi trạng thái đối tượng), chúng xảy ra khi nào (chuỗi hành động), và chúng xảy ra ở đâu (luồng hành động).

Biểu đồ hoạt động có thể được sử dụng cho nhiều mục đích khác nhau, ví dụ như:

- Để nắm bắt công việc (hành động) sẽ phải được thực thi khi một thủ tục được thực hiện. Đây là tác dụng thường gặp nhất và quan trọng nhất của biểu đồ hoạt động.

- Để nắm bắt công việc nội bộ trong một đối tượng.

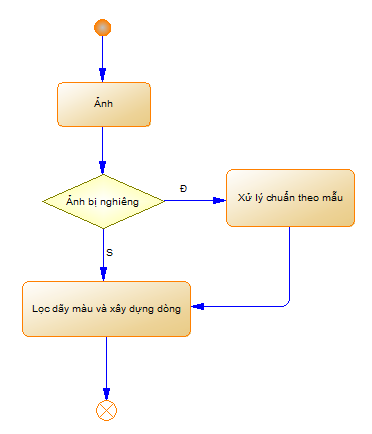
- Để chỉ ra một nhóm hành động liên quan có thể được thực thi ra sao, và chúng sẽ ảnh hưởng đến những đối tượng nằm xung quanh chúng như thế nào.

- Để chỉ ra một trường hợp sử dụng có thể được thực thể hóa như thế nào, theo khái niệm hành động và các sự biến đổi trạng thái của đối tượng.

- Để chỉ ra một doanh nghiệp hoạt động như thế nào theo các khái niệm công nhân (tác nhân), qui trình nghiệp vụ (workflow), hoặc tổ chức và đối tượng (các khía cạnh vật lý cũng như tri thức được sử dụng trong doanh nghiệp).

Biểu đồ hoạt động có thể được coi là một loại Flow chart. Điểm khác biệt là Flow Chart bình thường ra chỉ được áp dụng đối với các qui trình tuần tự, biểu đồ hoạt động có thể xử lý cả các các qui trình song song.

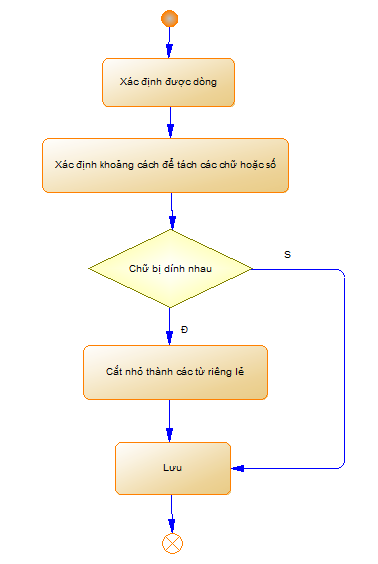
Hình 3.7 thể hiện hoạt động xác định dòng của ứng dụng. Ứng dụng sẽ xác định ảnh có bị nghiêng hay không. Nếu bị nghiêng sẽ xử lý theo mẫu rồi mới lọc dãy màu và xây dựng dòng.



Hình 3. Biểu đồ hoạt động xác định dòng

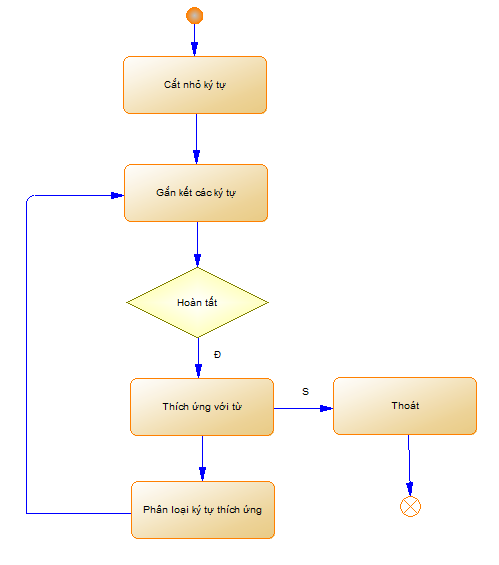
Ảnh bị nghiêng có thể ảnh hương rất lớn tới kết quả đầu ra nên việc xử lý ảnh đầu vào bị nghiêng cũng vô cùng quan trọng. Ảnh sau khi đã qua xử lý phần bị nghiêng mới bắt đầu tiến vào quá trình lọc dãy màu và xây dựng dòng.  
Tiếp đó, ứng dụng sẽ dựa trên dòng đã được xây dựng và bắt đầu xác định cách giữa các ký tự để tách lấy chữ hoặc số. Nếu các ký tự bị dính nhau, ứng dụng sẽ cắt nhỏ các ký tự ra và lưu lại để chuẩn bị cho bước tiếp theo.

Hình 3.8 mô tả hoạt động cắt nhỏ từ. Sau khi xác định được dòng, ứng dụng sẽ tiến hành xác định khoảng cách để tách các chữ hoặc số. Nếu các chữ không bị dính nhau, ứng dụng sẽ tiến hành cắt nhỏ thành các từ riêng lẻ rồi mới lưu những ký tự đó lại.



Hình 3. Biểu đồ hoạt động cắt nhỏ từ

Cuối cùng, sau khi đã cắt nhỏ được các ký tự, ứng dụng sẽ tiến hành gắn kết các ký tự lại với nhau để tạp thành 1 câu hoàn chỉnh (Hình 3.9).

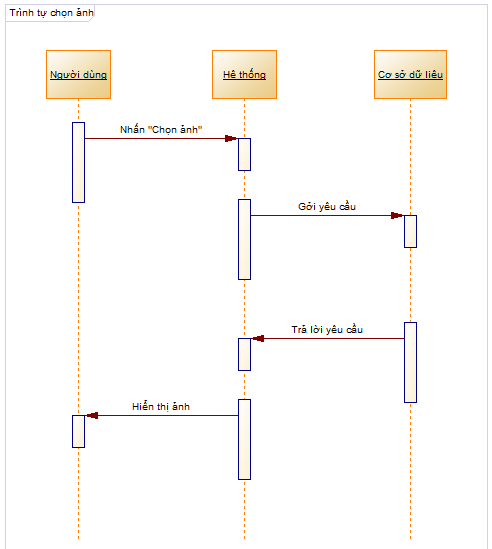


Hình 3. Biểu đồ hoạt động nhận dạng từ

## Biểu đồ tuần tự

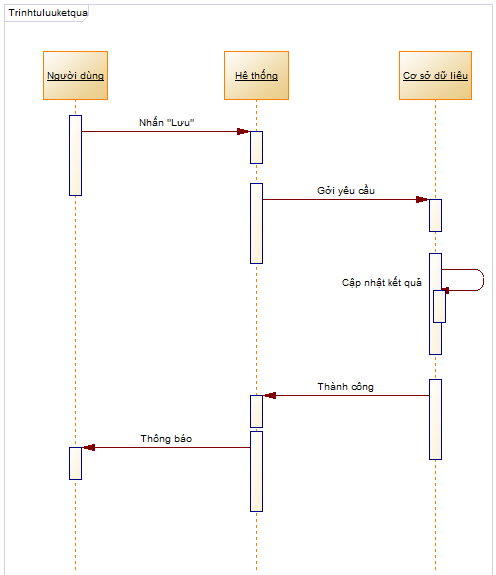
Biểu đồ tuần tự là biểu đồ dùng để xác định các trình tự diễn ra sự kiện của một nhóm đối tượng nào đó. Nó miêu tả chi tiết các thông điệp được gửi và nhận giữa các đối tượng đồng thời cũng chú trọng đến việc trình tự về mặt thời gian gửi và nhận các thông điệp đó.

Hình 3.10 biểu diễn trình tự chọn ảnh . Người dùng thao tác với hệ thống để chọn ảnh. Hệ thống sẽ gửi yêu cầu xử lý về cơ sở dữ liệu và cơ sở dữ liệu sẽ xử lý rồi trả lời yêu cầu của hệ thống. Cuối cùng, hệ thống sẽ hiển thị ảnh mà người dùng đã chọn.



Hình 3. Biểu đồ trình tự chọn ảnh

Hình 3.11 biểu diễn trình tự lưu kết quả. Trình tự này sẽ thể hiện quá trình mà hệ thống gửi yêu cầu đến cơ sở dữ liệu để lưu lại kết quả đã xử lý.



Hình 3. Biểu đồ trình tự lưu kết quả

## Kết luận

Trong chương 3 này, em đã đi từng bước khảo sát thông tin có trên thẻ căn cước công dân, phân tích yêu cầu chức năng và phi chức năng, cho đến phân tích thiết kế hệ thống. Kết quả mô phỏng đạt được sẽ được thể hiện trong Chương 4.

# KẾT QUẢ MÔ PHỎNG

Chương này tiến hành thử nghiệm trên hai loại hình ảnh: Hình chụp từ máy ảnh và hình chọn từ thư viện trong điện thoại.

## Hình chụp từ máy ảnh

Ảnh chụp đầu vào sau khi crop có dạng như sau:

A screenshot of a cell phone

Description automatically generated

Hình 4. Ảnh chụp từ điện thoại

Sau khi đã sau bước tiền xử lý ảnh đầu vào, ứng dụng sẽ tiến hành xử lý ảnh để lấy thông tin từ trong ảnh ra. Cụ thể kết quả đạt được sẽ như hình dưới đây:

A screenshot of a cell phone

Description automatically generated

Hình 4. Kết quả của ảnh chụp từ điện thoại

## Hình chọn từ thư viện trong điện thoại

Tiến hành chọn ảnh căn cước công dân có trong thư viện của điện thoại:

A close up of text on a white background

Description automatically generated

Hình 4. Ảnh chọn từ thư viện của điện thoại

Kết quả thu được sẽ như sau:

A screenshot of a cell phone

Description automatically generated

Hình 4. Kết quả thu được từ ảnh chọn từ thư viện điện thoại

Từ Hình 4.2 và Hình 4.3, ta thấy được độ chính xác trong việc nhận diện thẻ căn cước công dân và hoàn toàn có thể phát triển để phục vụ cho dự án mang tính xác thực cao.

# KẾT LUẬN

## Kết luận chung

Sau khoảng thời gian tích cực và nghiêm túc trong việc nghiên cứu đề tài “*Thiết kế ứng dụng xử lý ảnh để trích xuất thông tin từ căn cước công dân*”, em đã đạt được một số kết quả sau:

* Hiểu tổng quan về hệ thống OCR và OpenCV cho việc xử lý ảnh
* Năm bắt thêm về ML Kit và Firebase
* Hiểu được cách trích xuất chữ từ ảnh đầu vào
* Biết được cách sử dụng mô hình được huấn luyện sẵn.
* Kết quả được đánh giá trên tập dữ liệu ảnh chữ đánh máy, chữ từ bản pdf.
* Thực hiện triển khai trên ngôn ngữ Java
* Có thêm kiến thức về ML

Kết quả của em chưa thực sự hoàn hảo nhưng cũng đạt được đến một số chất lượng nhất đinh để có thể đủ để phục vụ các hướng tìm hiểu cũng như nghiên cứu tiếp theo.

## Hướng phát triển

Dựa vào những kết quả trên, em xin đề ra những định hướng phát triển trong tương lai của ứng dụng trích xuất thông tin từ ảnh như sau:

* Phát triển các ứng dụng cho điện thoại chạy hệ điều hành IOS dựa trên những phân tích thiết kế đã được thực hiện
* Phát triển thêm các chức năng mới cho hệ thống như scan hình ảnh, lưu trữ vào cơ sở dữ liệu,…
* Phát triển thêm tinh năng về chữ ký điện tử.
* Triển khai thực tiễn để đưa tới tay người sử dụng.
* Cải thiện giao diện để giúp cho người dùng có trải nghiệm tốt hơn.
* Phát triển ứng dụng phục vu nhu cầu cho các doanh nghiệp, tập đoàn trên khắp cả nước và toan thế giới.

# TÀI LIỆU THAM KHẢO

1. Tesseract OCR Engine – Ray Smith, Google Inc OSCON2007
2. <https://code.google.com/p/tesseract-ocr>, truy nhập cuối cùng ngày 17/04/2020
3. <https://github.com/tesseract-ocr>, truy nhập cuối cùng ngày 17/04/2020
4. Machine learning, “<https://en.wikipedia.org/wiki/Machine_learning>”, truy nhập cuối cùng ngày 17/04/2020
5. <https://www.geeksforgeeks.org/tesseract-ocr-with-java-with-examples/>, truy nhập cuối cùng ngày 30/04/2020
6. <https://gist.github.com/berak/788da80d1dd5bade3f878210f45d6742>, truy nhập cuối cùng ngày 05/05/2020
7. <https://opencv.org/opencv-4-0/>, truy nhập cuối cùng ngày 20/04/2020
8. <https://medium.com/fnplus/ml-kit-android-implementing-text-recognition-firebase-8cb225a23c38>, truy nhập cuối cùng ngày 01/05/2020
9. <https://firebase.google.com/docs/ml-kit/android/recognize-text>, truy nhập cuối cùng ngày 02/05/2020
10. Vũ Hữu Tiệp, “Machine learning cơ bản”
11. <https://machinelearningcoban.com/2016/12/27/categories/>, truy nhập cuối cùng ngày 17/04/2020
12. GOOGLE. Google Code. google code. [Online] 2012. <http://code.google.com/p/tesseract-ocr>
13. Y. WEN, Y. L. 2011.,An Algorithm for License Plate Recognition Applied to Intelligent Transportation System. IEEE Transactions on Intelligent Systems, pp. 1-16
14. A. Veit, T. Matera, L. Neumann, J. Matas, and S. Belongie, “Coco-text: Dataset and benchmark for text detection and recognition in natural images,” arXiv preprint arXiv:1601.07140, 2016.
15. M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, “Reading text in the wild with convolutional neural networks,” International Journal of Computer Vision, vol. 116, no. 1, pp. 1–20.