

Ứng dụng một số kỹ thuật xử lý ảnh trong phân tích chứng minh thư nhân dân

Lê Đức Hiếu

Trường Đại học Công nghệ

Luận văn ThS ngành: Công nghệ thông tin; Mã số: 60 48 10

Cán bộ hướng dẫn khoa học: PGS. TS. Ngô Quốc Tạo

Năm bảo vệ: 2012

Abstract. Tổng quan về phân tích ảnh tài liệu: giới thiệu về hệ thống nhận dạng ảnh tài liệu nói chung và hệ thống phân tích ảnh tài liệu nói riêng; giới thiệu các hướng tiếp cận trong phân tích bố cục và cấu trúc của ảnh tài liệu. Nghiên cứu một số thuật toán tiền xử lý ảnh: giới thiệu một số thuật toán tiền xử lý ảnh thường được áp dụng trong hệ thống phân tích và nhận dạng ảnh tài liệu như nhị phân ảnh, căn chỉnh độ nghiêng, các toán tử hình thái. Nghiên cứu phương pháp phân tích giấy chứng minh nhân dân: giới thiệu bài toán và các vấn đề cần giải quyết, từ đó áp dụng các kỹ thuật xử lý ảnh và phân tích tài liệu để phân tích ảnh chứng minh nhân dân. Tiến hành cài đặt thử nghiệm để đánh giá hiệu quả của giải pháp đưa ra trên tập ảnh chứng minh nhân dân.

Keywords: Công nghệ phần mềm; Xử lý ảnh; Thuật toán; Giấy chứng minh nhân dân

Content.

I. MỞ ĐẦU

Nhận dạng mẫu là một ngành khoa học của học máy (hay trí tuệ nhân tạo) nhằm phân loại dữ liệu (các mẫu) vào một số lớp. Một trong những ứng dụng phổ biến hiện nay của nhận dạng mẫu là phân tích và nhận dạng ảnh tài liệu (có nguồn gốc từ hệ thống nhận dạng ký tự quang học), nhằm số hoá các trang tài liệu giấy như sách, báo, tạp chí,... Bên cạnh lớp bài toán chung ở trên còn có lớp bài toán riêng biệt cho từng ngành, từng lĩnh vực cụ thể, như: phân tích và nhận dạng bảng biểu, phiếu điều tra, mẫu điền thông tin, danh thiếp, hộ chiếu,... Đối với lớp bài toán này thì việc phân tích cấu trúc của ảnh tài liệu là đặc biệt quan trọng, bởi vì nó sẽ quyết định đến việc tách và nhận dạng chính xác các trường thông tin cần thiết cho từng ứng dụng.

Trên thế giới đã có nhiều sản phẩm phần mềm phân tích và nhận dạng ảnh thẻ chứa thông tin cá nhân (như hộ chiếu, danh thiếp,...) và được ứng dụng trong nhiều lĩnh vực, như: làm thủ tục hải quan, các giao dịch ở các cửa hàng, khách sạn,... Ở Việt Nam loại thẻ chứa thông tin các nhân được sử dụng nhiều nhất là Chứng minh thư nhân dân (CMT), nhưng hiện nay chưa có một giải pháp nào được đưa ra cho việc phân tích và nhận dạng CMT. Do đó, trong luận văn này, tôi xin đề xuất một phương pháp

phân tích ảnh CMT dựa trên việc phân tích và nhận dạng biểu mẫu. Luận văn gồm bốn chương với các nội dung như sau:

Chương 1: Tổng quan về phân tích ảnh

Chương 2: Một số thuật toán tiền xử lý ảnh

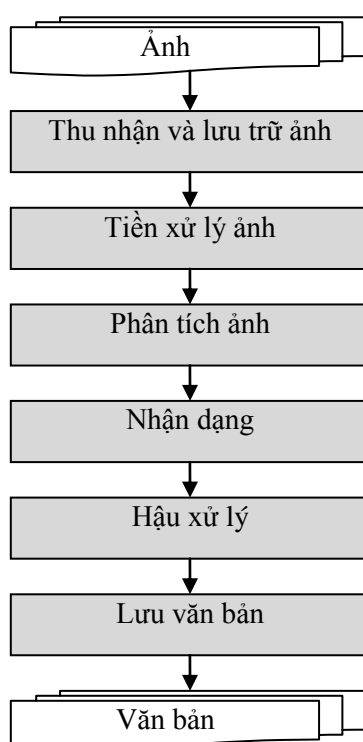
Chương 3: Thuật toán phân tích chứng minh thư nhân dân

Chương 4: Cài đặt thử nghiệm và đánh giá

II. CHƯƠNG 1: TỔNG QUAN VỀ PHÂN TÍCH ẢNH

A. 1.1. Tổng quan về hệ thống nhận dạng

Một hệ thống nhận dạng ảnh tài liệu gồm các bước sau (Hình 1. 1):



Hình 1. 1 Sơ đồ hệ thống nhận dạng

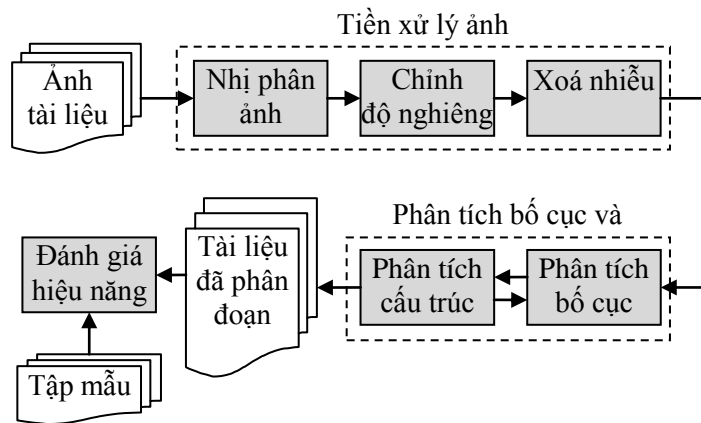
B. 1.2. Phân tích ảnh tài liệu

Quá trình **phân tích ảnh tài liệu** là việc cố gắng phân ly tài liệu thành các vùng và hiểu vai trò chức năng cũng như mối quan hệ giữa các vùng (Hình 1. 2).

1) 1.2.1. Bố cục tự nhiên và cấu trúc logic

Bố cục tự nhiên của ảnh tài liệu là đề cập đến hình bao và vị trí tự nhiên của các vùng khác nhau trong ảnh tài liệu. **Phân tích bố cục tự nhiên** là quá trình phân ly ảnh tài liệu thành các dãy các vùng đồng nhất, như: hình ảnh, khối văn bản, dòng văn bản, từ, ký tự,...

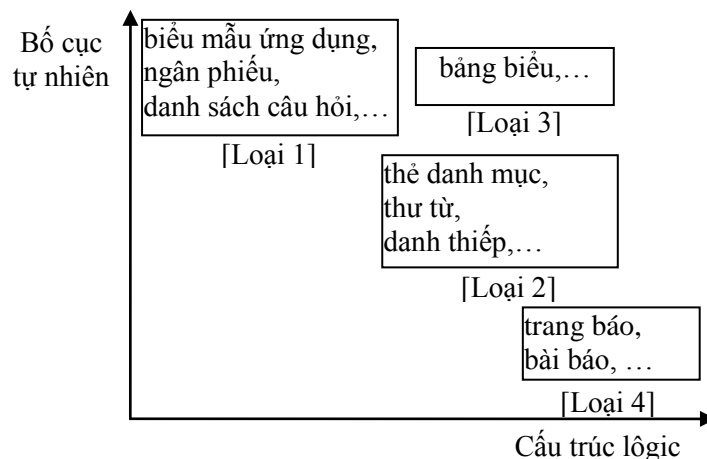
Ngoài bố cục tự nhiên, các tài liệu còn chứa các thông tin bổ xung về nội dung của nó như: nhan đề, tiêu đề,... Một tập các thực thể logic hoặc chức năng trong tài liệu, với các quan hệ bên trong của nó được hiểu là cấu trúc logic của tài liệu. **Phân tích cấu trúc logic** là việc biểu diễn kết quả của phân tích cấu trúc bố cục.



Hình 1. 2 Sơ đồ hệ thống phân tích bố cục và cấu trúc tài liệu

2) 1.2.2. Phân loại ảnh tài liệu

Dựa trên mức độ phụ thuộc của tài liệu vào bố cục tự nhiên hay cấu trúc logic, người ta chia làm 4 loại ảnh tài liệu (Hình 1. 3):



Hình 1. 3 Các loại tài liệu

C. 1.3. Phân tích bố cục

Mục đích chính của phân tích bố cục là xác định các vùng tự nhiên khác nhau trong tài liệu và các đặc trưng của nó, bao gồm:

- + **Phân đoạn vùng:** Là việc chia một hình ảnh tài liệu thành vùng thành phần, khi mà thuộc tính của tính đồng nhất không còn thoả mãn nữa.
- + **Phân loại vùng:** Là việc quyết định xem các vùng đã phân đoạn được thuộc loại nào.

1) 1.3.1. Hướng tiếp cận trên-xuống

Bắt đầu từ cả trang ảnh tài liệu, cố gắng chia nó thành các vùng nhỏ hơn dựa vào hình chiếu ngang/dọc (như X-Y Cut) hay phân tích cấu trúc nền của ảnh (như Whitespace).

2) 1.3.2. Hướng tiếp cận dưới-lên

Bắt đầu từ các thành phần nhỏ nhất của tài liệu (như điểm ảnh, thành phần liên thông), lặp đi lặp lại quá trình gom nhóm chúng thành các khối đồng nhất lớn hơn, như thuật toán Docstrum và Voronoi.

D. 1.4. Phân tích cấu trúc

Phân tích cấu trúc là quá trình gán các nhãn logic cho các vùng tự nhiên đã xác định ở bước phân tích bố cục dựa vào phân tích cú pháp hoặc kết quả nhận dạng.

III. CHƯƠNG 2: MỘT SỐ THUẬT TOÁN TIỀN XỬ LÝ ẢNH

A. 2.1. Nhị phân ảnh

Nhị phân ảnh (hay còn gọi là *phân ngưỡng*) là thao tác chuyển từ ảnh màu, ảnh đa cấp xám về ảnh nhị phân bằng cách tìm một ngưỡng: tổng quát hoặc cục bộ.

1) 2.1.1. Phân loại

Căn cứ vào phương pháp được áp dụng, có thể chia ra làm 6 nhóm sau:

- + Các phương pháp dựa vào hình dạng của histogram (*Histogram Shape-Based Thresholding Methods*).
- + Các phương pháp dựa vào việc chia nhóm (*Clustering-Based Thresholding Methods*).
- + Các phương pháp dựa vào entropy (*Entropy-Based Thresholding Methods*).
- + Các phương pháp dựa vào thuộc tính giống nhau (*Thresholding Based on Attribute Similarity*).
- + Các phương pháp căn cứ vào không gian (*Spatial Thresholding Methods*).
- + Các phương pháp ngưỡng thích ứng cục bộ (*Locally Adaptive Thresholding*).

2) 2.1.2. Một số phương pháp

a) 1/. Phương pháp Niblack

Đây là phương pháp xác định ngưỡng cục bộ dựa trên việc tính toán giá trị trung bình và độ lệch chuẩn cục bộ.

b) 2/. Phương pháp Otsu

Đây là phương pháp xác định ngưỡng toàn cục dựa vào histogram để phân chia các điểm ảnh vào hai lớp tiền cảnh (đối tượng) và nền sao cho “khoảng cách” giữa các điểm ảnh trong mỗi lớp là nhỏ nhất.

3) 2.1.3. Kết luận

Bảng 2. 1 So sánh phương pháp Niblack và Otsu

| Tiêu chí | Niblack | Otsu |
|--|----------------|---------------------------------------|
| Thời gian thực hiện | Chậm | Nhanh |
| Độ phức tạp của thuật toán | $O(N^2)O(W^2)$ | $O(N^2)$ |
| Ảnh có độ sáng thay đổi | Tốt | Không tốt (Mất một số vùng thông tin) |
| Phụ thuộc vào kích thước của các đối tượng trong ảnh | Có | Không |
| Phụ thuộc vào tham số | Có | Không |

B. 2.2. *Chỉnh độ nghiêng của trang văn bản*

Trong quá trình thu thập, ảnh có thể bị nghiêng do đó cần phải căn chỉnh độ nghiêng của ảnh bằng các xác định góc nghiêng và xoay ảnh trở lại như ảnh gốc.

1) 2.2.1. *Phương pháp dựa trên biến đổi Hough*

Biến đổi Hough là phép biến đổi điểm ảnh từ hệ tọa độ x-y sang hệ tọa độ ρ - θ do đó rất hữu ích cho việc dò tìm đường thẳng trong ảnh vì thế rất thích hợp cho việc xác định góc nghiêng của ảnh có chứa các thành phần là các dòng văn bản.

2) 2.2.2. *Phương pháp láng giềng gần nhất (nearest neighbours)*

Phương pháp này dựa trên một nhận xét rằng trong một trang văn bản, khoảng cách giữa các kí tự trong một từ và giữa các kí tự của từ trên cùng một dòng là nhỏ hơn khoảng cách giữa hai dòng văn bản, vì thế đối với mỗi kí tự, láng giềng gần nhất của nó sẽ là các kí tự liền kề trên cùng một dòng văn bản.

3) 2.2.3. *Phương pháp sử dụng tia quay (project profile)*

Dựa vào hình chiếu ngang/dọc của ảnh để tìm góc nghiêng.

4) 2.2.4. *Kết luận*

- + Phương pháp dựa vào biến đổi Hough rất tốt cho việc xác định góc nghiêng của các đường thẳng trong ảnh. Nhưng lại thất bại khi ảnh có quá ít các đối tượng (sự thừa thớt của các ký tự) và thời gian thực hiện lâu do phải thao tác trực tiếp trên các điểm ảnh.
- + Phương pháp láng giềng gần nhất cho kết quả tốt hơn nhưng thời gian tính toán cũng vẫn còn chậm. Mặt khác đối trang văn bản là Tiếng Việt có dấu, phương pháp này thường cho kết quả không chính xác.
- + Phương pháp sử dụng tia qua cho kết quả tương đối chính xác và thời gian thực hiện nhanh. Thường được sử dụng trong các sản phẩm thương mại.

C. 2.3. *Các toán tử hình thái (Morphological operations)*

Đầu vào của các toán tử hình thái thường là ảnh nhị phân (một số trường hợp là ảnh đa cấp xám) và phân tử cấu trúc (structuring element), kết hợp với việc sử dụng các toán tử tập hợp: hợp, giao, trừ và lấy phần bù. Tất cả các toán tử hình thái đều là sự phối hợp của hai toán tử cơ bản: **giãn ảnh** (*dilation*) và **co ảnh** (*erosion*). Hiệu ứng cơ bản của toán tử giãn ảnh là sự mở rộng dần dần đường biên của các đối tượng ảnh, do đó kích thước của các đối tượng ảnh tăng lên trong khi lỗ hổng bên trong đối tượng và khoảng cách giữa các đối tượng thì giảm xuống. Ngược lại, phép co ảnh làm cho các đường biên của đối tượng bị “bào mòn”, dẫn đến kích thước của các đối tượng trong ảnh giảm đi, khoảng cách giữa các đối tượng thì tăng lên và lỗ hổng trong mỗi đối tượng thì được mở rộng ra.

Nếu sử dụng các phép co ảnh và giãn ảnh một cách riêng lẻ thì sẽ làm mất đi các đặc trưng (hình dạng, kích thước) của ảnh. Do đó người ta thường kết hợp hai phép toán này với nhau, bằng cách: co bao nhiêu lần thì giãn bấy nhiêu lần và ngược lại. Một trong số những cách kết hợp đó được gọi là phép **opening** và **closing**, hai phép toán này đối xứng nhau: phép toán opening “co trước giãn sau” còn closing “giãn trước co sau”.

D. 2.4. Kết chương

Trong chương này, luận văn đã trình bày ba kỹ thuật tiền xử lý ảnh quan trọng, thường xuyên được sử dụng trong các hệ thống nhận dạng: nhị phân ảnh, căn chỉnh độ nghiêng của ảnh và một số phép toán hình thái.

IV. CHƯƠNG 3: THUẬT TOÁN PHÂN TÍCH CHỨNG MINH THƯ NHÂN DÂN

Giấy chứng minh nhân dân hay chứng minh thư (CMT) là một loại giấy tờ tùy thân của công dân Việt Nam, được xác nhận bởi cơ quan nhà nước có thẩm quyền về lý lịch của người được cấp.



a) Mặt trước

b) Mặt sau

Hình 3. 1 Mẫu “Giấy chứng minh thư nhân dân”

CMT là yêu cầu bắt buộc trong rất nhiều các giao dịch cũng như các thủ tục hành chính,... Ví dụ, khi chúng ta muốn mở một tài khoản ở ngân hàng hay đăng ký một thuê bao điện thoại, chúng ta cần photo CMT, đến các địa điểm giao dịch và điền thông tin cá nhân của mình vào các biểu mẫu có sẵn. Việc này mất rất nhiều thời gian và phiền phức cho những người đi làm thủ tục đó. Cùng với quá trình tin học hoá, đơn giản các thủ tục hành chính, một yêu cầu đặt ra là: cần có một hệ thống tự động tách và nhận dạng các trường thông tin trên CMT để điền vào các biểu mẫu có sẵn. Như vậy cần tách chính xác các trường thông tin yêu cầu trước khi chuyển cho phần nhận dạng, tuy nhiên việc này gặp một số vấn đề khó khăn sau:

- + Vì các trường thông tin được dập/in trên mẫu có sẵn nên có thể bị lệch so với dòng chuẩn hoặc chồm lên phần tiêu đề đã được in trước.
- + Đối với trường hợp CMT là chữ dập, nét chữ không đều nhau giữa các CMT, thậm chí là trong cùng một CMT. Có trường hợp nét chữ quá đậm, thậm chí là bị nhoè, có trường hợp nét chữ quá mờ.
- + Trong quá trình sử dụng, CMT có thể bị ố, mốc, nhàu, nát,... nét chữ bị mờ, nhoè,...
- + Ở mặt trước, trong một số trường hợp hoa văn nền khá rõ nét trong khi nét chữ lại quá mờ, khó có thể phân biệt đâu là nét chữ đâu là hoa văn nền.
- + Ở mặt sau, trong khá nhiều trường hợp dấu màu đỏ đè cả lên phần Ngày cấp và Nơi cấp, làm mờ đi một phần thông tin trên hai trường này.

Yêu cầu của bài toán: Từ ảnh CMT màu được quét bằng máy quét với độ phân giải 300 dpi, tách lấy các trường thông tin: *Số CMT, Họ tên, Ngày sinh, Nguyên quán, Nơi thường trú, Dân tộc, Ngày cấp và Nơi cấp*. Các trường thông tin này nằm trên cả hai mặt của CMT, mà mỗi mặt lại có các đặc điểm khác nhau nên tôi đề xuất hai thuật toán khác nhau để tách các trường thông tin trên mỗi mặt.

A. 3.1. Tách các trường thông tin ở mặt trước

Nhận thấy rằng, trường Số CMT có màu đỏ nổi bật so với các trường khác nên có thể dựa vào đó để tách trường Số CMT trước, sau mới đó tách các trường còn lại dựa vào vị trí tương đối của nó so với trường Số CMT. Mặt khác, do ở mặt trước CMT có các hoa văn nền (màu xanh – green) khá rõ nét nên cần các thao tác tiền xử lý ảnh hiệu quả để khử đi các thành phần nền phức tạp này. Cụ thể, thuật toán bao gồm các bước sau:

- + **Tiền xử lý ảnh:** Chuyển ảnh màu về ảnh đen trắng sao cho vừa khử được nền mà vẫn giữ được các đặc trưng của nét chữ, đồng thời cũng thực hiện căn chỉnh độ nghiêng của ảnh.
- + **Tách trường Số CMT:** Dựa vào đặc điểm nổi bật riêng để tách, tiếp theo là loại bỏ các đường lượn sóng xung vẫn phải giữ lại các đặc trưng của nét chữ.
- + **Tách các trường còn lại:** Tách lấy các dòng thông tin sau đó loại bỏ đi phần tiêu đề của các trường.

1) 3.1.1. Tiền xử lý ảnh



a) Ảnh đầu vào b) Ảnh đa cấp xám c) Ảnh nhị phân

Hình 3. 2 Tiền xử lý ảnh mặt trước CMT

a) 1/. Chuyển ảnh màu về ảnh đa cấp xám

Nhận thấy rằng, với những điểm sánh nền thì giá trị của thành phần màu Green chênh lệch lớn so với thành phần màu Red. Do đó, để chuyển đổi từ ảnh màu về ảnh đa cấp xám, ta lấy thành phần màu Green cộng với độ chênh lệch giữa hai thành phần màu Green và Red:

$$\text{Cường độ xám} = G + |G - R|$$

Điều này không những làm nổi bật được các nét chữ (hoa văn nền mờ đi, trong khi đặc trưng nét chữ không đổi) mà còn làm mờ đi mực nhoè màu đỏ ở trường Số CMT lên trường Họ tên.

b) 2/. Làm trơn ảnh

Ảnh đa cấp xám thu được ở bước trên thường không được “mịn”, vẫn còn “ráp” do một số thành phần nền quá rõ nét, để làm mịn ảnh và loại bỏ các nhiễu này cần áp dụng một phép lọc để làm trơn ảnh. Ở đây, tôi áp dụng thuật toán làm trơn thích ứng:

c) 3/. Nhị phân ảnh

Các phương pháp trình bày ở trên, trong một số trường hợp này thì cho kết quả tốt, trong một số trường hợp khác lại cho kết quả không tốt:

- + Phương pháp **Otsu**: Khử nền của ảnh khá tốt, nhưng không làm rõ các nét chữ.
- + Phương pháp **Sauvola**: Làm rõ được các nét chữ, nhưng vẫn còn có các nhiễu của nền.

Do đó, tôi đề xuất một phương thức kết hợp được ưu điểm của cả hai phương pháp trên. Các bước thực hiện như sau:

- + Nhị phân ảnh bằng phương pháp Otsu được ảnh I_O .
- + Nhị phân ảnh bằng phương pháp Sauvola được ảnh I_S .
- + Áp dụng toán tử AND đối với hai ảnh I_O và I_S để được ảnh nhị phân cần tìm:

$$I_B = I_O \text{ AND } I_S$$

d) 4/. *Căn chỉnh độ nghiêng*

Sử dụng phương pháp tia quay.

2) 3.1.2. *Tách trường Số CMT*

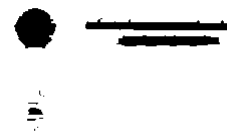
Trong ảnh mẫu đã căn chỉnh độ nghiêng ở phần trên, trường Số CMT có màu đỏ cùng màu với các vùng: dòng “GCMND”, hình quốc huy và đôi khi là vùng mặt người hay màu áo của ảnh chân dung. Để tách trường Số CMT, trước tiên cần làm nổi bật các vùng này, sau đó tìm và tách vùng Số CMT, cuối cùng là loại bỏ các đường lượn sóng.

a) 1/. *Xác định vùng có thể là trường Số CMT*

Với mỗi điểm ảnh trong ảnh mẫu đầu vào, lấy thành phần màu Red trừ đi thành phần màu Green (nếu nhỏ hơn không thì gán bằng không) sẽ được ảnh đa cấp xám, trong đó vùng sáng tương ứng với vùng màu đỏ còn vùng tối tương ứng với vùng nền trong ảnh mẫu (Hình 3. 3a). Tiếp theo, áp dụng phép biến đổi hình thái closing để “nối” các ký tự (các thành phần) liền kề nhau thành các dòng (các khối). Sau đó, sử dụng phương pháp phân ngưỡng Otsu và đảo ngược mức xám để chuyển ảnh đa cấp xám về ảnh nhị phân, những vùng màu đen là các vùng có thể là vùng Số CMT (Hình 3. 3b). Cuối cùng, tách lấy các thành phần liên thông và miêu tả chúng bằng hình bao chữ nhật của nó.



a) $I_{Red} - I_{Green}$



b) Các vùng có thể là Số CMT

Hình 3. 3 Xác định vùng Số CMT

b) 2/. *Tìm và tách trường Số CMT*

Trường Số CMT nằm ngay dưới dòng “GCMND”, khi đập/in nó có thể bị chồm lên (hoặc sát vào) dòng “GCMND” hoặc khi sử dụng trường Số CMT có thể bị nhoè lên dòng “GCMND”. Mặt khác, trong quá trình sử dụng trường Số CMT có thể bị mờ (thậm chí mất hẳn). Do đó trong các vùng tìm được ở trên, hai dòng này có thể bị dính vào nhau thành một vùng hoặc không chứa vùng Số CMT. Như vậy, khó có thể tìm trực tiếp trường Số CMT trong các vùng có thể ở trên, mà phải tìm thông qua một vùng khác. Nhận thấy rằng, dòng “GCMND” có đặc điểm nổi trội hơn hẳn: có độ dài lớn nhất, ít bị mờ hoặc nhoè trong quá trình sử dụng và nằm ngay bên trên trường Số CMT nên có thể dựa vào dòng này để xác định trường Số CMT.

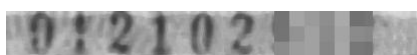
c) 3/. Phân đoạn vùng Số CMT

Trong phần này sẽ tiến hành phân đoạn để loại bỏ đi những đường lượn sóng nằm ngang trong khi vẫn phải giữ được các đặc trưng nét bút của chữ số. Việc này là rất khó khăn, bởi vì các chữ số có thể bị nhoè mờ và nét chữ có độ đậm nhạt thay đổi, trong khi các đường lượn sóng lại có bề dày khác nhau, thậm chí là tương đồng với bề dày nét bút. Do đó, cần ước lượng bề dày và làm mờ đường lượn sóng trước khi tiến hành phân đoạn.

Để ước lượng được bề dày đường lượn sóng cần phân đoạn “thô” sao cho vẫn giữ được các đường lượn sóng. Sau đó “đo” bề dày đường lượn sóng tại tất cả các lát dọc để thống kê chọn lấy bề dày thích hợp nhất. Từ đó làm mờ đường lượn sóng bằng phép lọc trung vị (median), cuối cùng là phân đoạn “tinh” để khử đi các thành phần nền [Hình 3. 4].



a) Vùng Số CMT



b) Làm mờ đường lượn sóng



c) Kết quả phân đoạn

Hình 3. 4 Phân đoạn vùng Số CMT

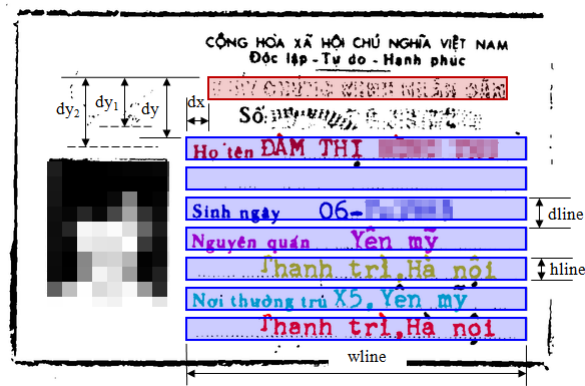
3) 3.1.3. Tách các trường thông tin còn lại

Các trường thông tin còn lại (bao gồm 4 trường: Họ tên, Ngày sinh, Nguyên quán, Nơi thường trú) được phân bố trên 7 dòng, các dòng này cách đều nhau và nằm bên dưới dòng “GCMND” và Số CMT, việc tách các trường thông tin còn lại thực chất là việc tách lấy 7 dòng thông tin này. Để tách các dòng này, từ ảnh nhị phân thu được ở bước tiền xử lý và vị trí của dòng “GCMND” đã xác định ở bước trên, xác định “mặt nạ” của các dòng sau đó cố gắng lọc lấy các đối tượng (ký tự) thuộc mặt nạ dòng.

a) 1/. Tìm mặt nạ dòng

Vị trí của các dòng thông tin còn lại được xác định thông qua vị trí tương đối của chúng so với dòng “GCMND” đã xác định được ở trên. Để thuận tiện, ta biểu diễn 7 dòng này bằng 7 hình chữ nhật – gọi là *mặt nạ dòng* – như [Hình 3. 5], trong đó:

- + dx, dy : khoảng cách giữa vùng mặt nạ và dòng tiêu đề “GCMND” theo phương ngang và phương dọc, tương ứng. Trong đó, dx là cố định (vì được in theo mẫu), chỉ có dy là thay đổi do dòng thông tin đập/in vào có thể thay đổi so với dòng cơ sở (dịch lên hoặc dịch xuống): $dy \in [dy_1, dy_2]$.
- + $wline, hline, dline$: kích thước và khoảng cách giữa các hình chữ nhật là cố định.



Hình 3. 5 Mặt nạ dòng mặt trước

Như vậy, để xác định mặt nạ dòng chỉ cần xác định dy , dy được xác định dựa vào sự phân bố của các điểm ảnh trong vùng mặt nạ:

$$dy = \arg\{\max_{y \in [dy1, dy2]} (\text{count}(y))\}$$

Trong đó: $\text{count}(y)$ là số lượng điểm ảnh trong vùng mặt nạ, tương ứng với $y = dy$.

b) 2/. Tách các đối tượng thuộc mỗi dòng

Từ ảnh nhị phân thu được ở bước tiền xử lý, tách lấy các đối tượng (các thành phần liên thông) và chia làm 3 lớp dựa trên kích thước của các đối tượng:

- + **Nhỏ**: Lớp các đối tượng là dấu của ký tự, dấu chấm của đường cơ sở, nhiễu,...
- + **Lớn**: Lớp các đối tượng là ảnh chân dung hoặc các đường viền.
- + **Vừa**: Gồm các đối tượng còn lại, lớp các đối tượng có thể là ký tự.

\updownarrow dline

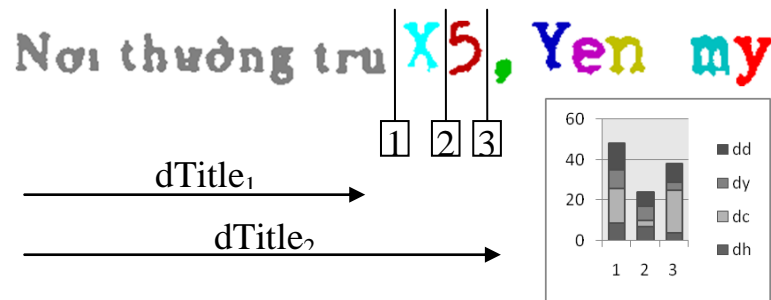
\updownarrow hline

Từ các đối tượng thuộc lớp **vừa**, lọc lấy các đối tượng thuộc mặt nạ của mỗi dòng. Một đối tượng được cho là thuộc mặt nạ của một dòng nếu tâm của nó nằm trong hình chữ nhật của mặt nạ dòng đó. Lúc này mỗi dòng sẽ được biểu diễn bằng một danh sách các đối tượng.

c) 3/. Xóa phần tiêu đề và nhiễu

Trong 7 dòng đã tìm được ở trên, có 4 dòng có chứa phần tiêu đề (tương ứng với 4 trường thông tin), đó là các dòng thứ 1, 2, 4, 6. Để loại bỏ các phần tiêu đề này cần xác định vị trí phân tách giữa phần tiêu đề và phần thông tin trong các dòng đó. Do phần thông tin được đập/in vào mẫu có sẵn (đã có phần tiêu đề) nên giữa phần hai phần này có những đặc điểm khác nhau sau:

- + Chiều cao trung bình của các ký tự trong phần tiêu đề thường nhỏ hơn chiều cao trung bình của các ký tự trong phần thông tin.
- + Theo phương dọc, phần thông tin có thể bị lệch so với phần tiêu đề (dịch lên hoặc dịch xuống).
- + Theo phương ngang, giữa phần thông tin và phần tiêu đề thường có khoảng cách lớn hơn khoảng cách giữa các ký tự (các từ) trong dòng.
- + Ký tự cuối cùng của phần tiêu đề là chữ in thường, trong khi ký tự đầu tiên của phần thông tin là chữ in hoa nên có sự khác nhau lớn về chiều cao.



Hình 3. 6 Xoá phần tiêu đề

d) 5/. Lấy lại các ký tự bị mất

Trong quá trình tách lấy các đối tượng thuộc dòng, có thể một số ký tự thuộc dòng nhưng không được chọn, nguyên nhân là do:

- + Các ký tự này nằm ngoài vùng mặt nạ dòng (không được đập/in vào vùng thông tin đã được định trước) do lượng thông tin quá nhiều nên “tràn” ra khỏi vùng đã được định trước [Hình 3. 7a].
- + Các ký tự này thuộc vùng mặt nạ dòng nhưng tâm của nó không nằm trong mặt nạ của dòng, do các ký tự trên hai dòng dính nhau hoặc dính vào các đường viền [Hình 3. 7b].

Từ các nguyên nhân trên, có các giải pháp khác nhau để lấy lại các ký tự bị mất:

- + Với nguyên nhân thứ nhất: Mở rộng mặt nạ dòng về hai phía, rồi lấy lại các ký tự như ở bước trên.
- + Với nguyên nhân thứ hai: Tìm phần giao nhau giữa hình bao của các dòng với các đối tượng, nếu phần giao này có chiều cao lớn hơn một nửa chiều cao của dòng thì đó là ký tự bị mất.



a) Nằm ngoài vùng mặt nạ



b) Hai ký tự dính nhau

Hình 3. 7 Lấy lại các ký tự bị mất

B. 3.2. Tách các trường thông tin ở mặt sau

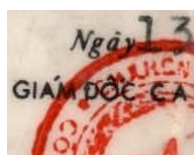
Mặt sau của CMT có cấu trúc dạng bảng, các trường thông tin được phân bổ vào các ô của bảng: trường Dân tộc ở ô trên cùng, trường Ngày cấp và Nơi cấp ở ô dưới cùng bên phải. Do vậy, để tách được các trường thông tin yêu cầu cần xác định được cấu trúc bảng sau đó mới tách các trường thông tin này, thuật toán gồm các bước:

- + **Tiền xử lý ảnh:** Khử các thành phần nền của trong ảnh, đặc biệt là dấu đỏ.
- + **Xác định cấu trúc bảng:** Tìm các đường kẻ ngang/dọc để xây dựng lại cấu trúc bảng
- + **Tách các trường thông tin:** Loại bỏ đi phần tiêu đề của các dòng.

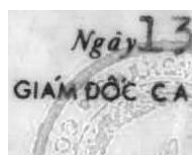
1) 3.2.1. Tiền xử lý ảnh

Các thao tác tiền xử lý ở mặt sau là:

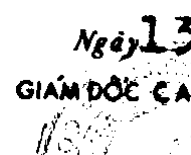
- + **Chuyển ảnh mẫu về ảnh đa cấp xám:** bằng cách tách lấy kênh màu Red. Điều này không những loại bỏ được dấu màu đỏ, mà còn vẫn giữ được các đặc trưng nét bút.
- + **Nhị phân ảnh:** Vì mặt sau có cấu trúc nền khá đơn giản nên chỉ cần áp dụng phương pháp phân ngưỡng Sauvola để khử nền.
- + **Căn chỉnh độ nghiêng:** Áp dụng phương pháp như đối với mặt trước.



a) Ảnh đầu vào



b) Ảnh đa cấp xám



c) Ảnh nhị phân

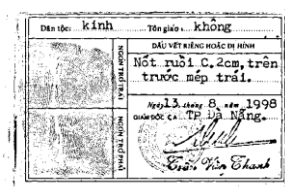
Hình 3. 8 Tiền xử lý mặt sau CMT

2) 3.2.2. Xác định cấu trúc bảng

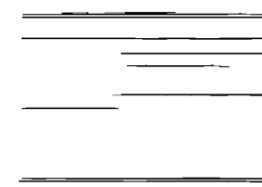
Cấu trúc bảng được tạo bởi các đường kẻ ngang và dọc liền nét nhưng trong quá trình sử dụng các đường kẻ này có thể bị mờ hoặc đứt nét. Mặt khác, trong lúc đập/in thông tin và lăn tay, các ký tự hoặc dấu vân tay có thể chồm lên các đường kẻ, gây khó khăn cho việc xác định cấu trúc bảng. Như vậy, để xác định được cấu trúc bảng cần xác định được các đường kẻ ngang và dọc của bảng. Vì các đường kẻ ngang và dọc là có tính chất như nhau nên trong phần này chỉ trình bày thuật toán xác định đường kẻ ngang của bảng, việc xác định đường kẻ dọc là tương tự.

Thuật toán xác định đường kẻ ngang của cấu trúc bảng được chia làm hai phần chính:

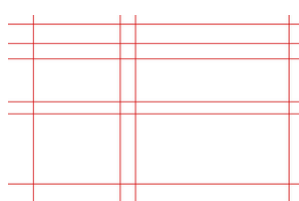
- + Đầu tiên áp dụng các phép biến đổi hình thái để làm nổi bật các đường kẻ ngang, sau đó phân tích các thành phần liên thông để tách lấy các **đường kẻ ngang trong ảnh**. Các đường kẻ này có thể là đường kẻ ngang của bảng hoặc không, bởi vì có cả các đường kẻ được tạo bởi các dấu chấm liền nhau của đường cơ sở hoặc đường gạch chân của chữ ký,... [Hình 3. 9b]
- + Dựa vào khoảng cách tương đối giữa các đường kẻ ngang, tìm các **đường kẻ ngang của bảng** từ các đường kẻ ngang trong ảnh, bằng cách loại bỏ đi các đường kẻ ngang không phải của bảng và bổ sung các đường của ngang còn thiếu [Hình 3. 9c].



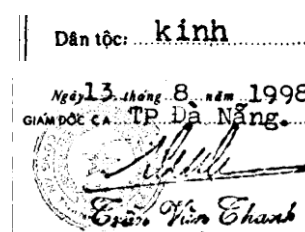
a) Ảnh nhị phân đầu vào



b) Các đường kẻ ngang



c) Các đường kẻ của bảng



b) Các đường kẻ ngang

d) Các vùng thông tin

Hình 3. 9 Xác định cấu trúc bảng

Sau khi xác định được các đường kẻ ngang và dọc của bảng (tức là xác định được cấu trúc của bảng), tách lấy các vùng chứa các trường thông tin cần tìm từ ảnh: vùng Dân tộc nằm giữa đường kẻ ngang thứ nhất và thứ hai; vùng Ngày cấp và Nơi cấp nằm giữa đường kẻ ngang thứ tư và thứ sáu, và đường kẻ dọc thứ ba và thứ bốn [Hình 3. 9d]. Từ các vùng này, bước tiếp theo sẽ tách lấy các trường thông tin cần tìm.

3) 3.2.3. Tách trường thông tin

Việc tách các trường trong tin ở mặt sau cũng tương tự như đối với mặt trước.

V. CHƯƠNG 4: CÀI ĐẶT THỬ NGHIỆM VÀ ĐÁNH GIÁ

Trong chương này của luận văn sẽ mô tả một cách chi tiết quá trình cài đặt thử nghiệm thuật toán, cũng như đánh giá các kết quả đạt được trên các bộ dữ liệu khác nhau.

A. 4.1. Môi trường cài đặt

Thuật toán thử nghiệm được cài đặt bằng ngôn ngữ lập trình C/C++ (Microsoft Visual Studio 2008), sử dụng thư viện xử lý ảnh Leptonica cho việc đọc/ghi ảnh và các thao tác xử lý ảnh cơ bản. Chương trình được thử nghiệm trên máy PC, tốc độ 3,4GHz, bộ nhớ RAM 1,0 GB.

B. 4.2. Dữ liệu kiểm thử

Chương trình được thử nghiệm trên tập 926 ảnh CMT được quét với độ phân giải 300dpi, kiểu ảnh màu. Các mẫu CMT được lấy từ nhiều tỉnh thành khác nhau, các đơn vị cấp CMT khác nhau. Mặc dù CMT được in theo mẫu chung nhưng vẫn có sự khác nhau giữa các đơn vị cấp về kích thước kiểu chữ, vị trí tương đối giữa các trường thông tin,...

C. 4.3. Kết quả thực nghiệm

Có rất nhiều phương pháp đánh giá độ chính xác của việc tách các trường thông tin, trong luận văn này sử dụng diện tích miền bao phủ để tính các tham số **Precision** và **Recall**. Kết quả đánh giá như trong Bảng 4. 1.

D. 4.4. Đánh giá

Thuật toán đưa ra trong luận văn tách được gần như 100% các trường thông tin yêu cầu. Một số trường hợp các có thể bị mất một phần thông tin hoặc coi nhiều như một phần thông tin của trường. Thuật toán chỉ thất bại trong trường hợp các trường thông tin in/dập vào CMT bị lệch một góc đánh kẻ so với các dòng in sẵn trong CMT.

Bảng 4. 1 Kết quả phân tích ảnh CMT

| | Precision (%) | Recall (%) |
|--------|----------------------|-------------------|
| Số CMT | 100,00 | 100,00 |
| Họ tên | 96,56 | 96,15 |

| | | |
|----------------|-------|-------|
| Ngày sinh | 99,41 | 99,21 |
| Nguyên quán | 97,91 | 96,58 |
| Nơi thường trú | 94,61 | 84,53 |
| Dân tộc | 95,99 | 95,82 |
| Ngày cấp | 97,83 | 97,46 |
| Nơi cấp | 78,56 | 86,08 |
| Cả CMT | 95,25 | 92,89 |

VI. KẾT LUẬN

Trong quá trình nghiên cứu và thực hiện luận văn, tôi nhận thấy đã đạt được một số kết quả chính sau:

- + Nắm bắt được các bước chính trong một hệ thống xử lý ảnh, hiểu được các khái niệm cơ bản trong một xử lý ảnh. Thấy được vai trò quan trọng của xử lý ảnh đối với một hệ nhận dạng, đó là bước tiền xử lý nhằm nâng cao chất lượng của nhận dạng.
- + Tìm hiểu được một số thuật toán xử lý ảnh hay được dùng trong bước tiền xử lý của một hệ thống nhận dạng hoàn chỉnh. Đó là các phương pháp nhị phân ảnh, căn chỉnh độ nghiêng của trang văn bản, các toán tử hình thái. Trên cơ sở hiểu biết đó có thể vận dụng được vào các bài toán khác.
- + Tìm hiểu và tổng quát hoá các phương pháp phân tích ảnh tài liệu, cũng như nắm được các ưu nhược điểm của từng phương pháp. Từ đó đưa ra được các giải pháp cho bài toán đặt ra trong luận văn.
- + Đã áp dụng thành công các kiến thức tìm hiểu được vào cài đặt thử nghiệm chương trình phân tích ảnh CMT. Kết quả của chương trình đạt được là tốt và có thể áp dụng vào thực tế. Đây là bài toán khó, hiện nay chưa có giải pháp nào khác được đưa gia để giải quyết bài toán này. Đó là đóng góp lớn nhất của luận văn.

Tuy nhiên, do thời gian làm luận văn hạn chế, trong khi khối lượng công việc lớn nên còn nhiều vấn đề tồn tại chưa được giải quyết:

- + Chương trình mới dừng lại ở bước thử nghiệm, chưa phải là một chương trình hoàn chỉnh, đầy đủ các tính năng
- + Thuật toán mới chỉ dừng lại ở việc phân tích ảnh CMT, chưa khái quát hoá cho các ảnh thẻ bất kỳ.

Hướng phát triển tiếp theo là, tiếp tục nghiên cứu hoàn thiện chương trình để có thể áp dụng vào thực tế. Mở rộng các tính năng của chương trình (như thêm phần nhận dạng, kiểm lỗi chính tả) để thành một chương trình hoàn chỉnh. Khái quát hoá thuật toán để có thể xử lý được ảnh thẻ bất kỳ.

References.

Tiếng Việt

- [1] Nghị định số: 577-TTg “Đặt giấy chứng minh và quy định thể lệ cấp phát giấy chứng minh”, Chính phủ, Hà Nội, ngày 27 tháng 11 năm 1957.
- [2] Nghị định số: 05/1999/NĐ-CP “Về chứng minh nhân dân”, Chính phủ, Hà Nội, ngày 03 tháng 02 năm 1999.
- [3] Nghị định số: 170/2007/NĐ-CP “Sửa đổi, bổ sung một số điều của nghị định số 05/1999/NĐ-CP ngày 03 tháng 02 năm 1999 của chính phủ về chứng minh nhân dân”, Hà Nội, ngày 19 tháng 11 năm 2007.

Tiếng Anh

- [4] K. Kpalma and J. Ronsin, “An Overview of Advances of Pattern Recognition Systems in Computer Vision”, 2007.
- [5] Anoop M Namboodiri and Anil Jain, “Document Structure and Layout Analysis”, in Digital Document Processing: Major Directions and Recent Advances B. B. Chaudhuri (ed.), Springer-Verlag, London, (ISBN:978-1-84628-501-1), Jan. 2007.
- [6] Toyohide WATANABE, “Document Analysis and Recognition”, IEICE TRANS. INF. & SYST., vol.e82-d, no.3.
- [7] G. Nagy, S. Seth and M. Viswanathan, "A Prototype Document Image-Analysis System for Technical Journals", Computer 25, (1992), 10–22.
- [8] Baird, U.S. and Jones, S.E. and Fortune, S.J.: Image Segmentation by Shape-Directed Covers. in Proceedings of International Conference on Pattern Recognition, Atlantic City, NJ (1990), 820-825.
- [9] Pavlidis, T. and Zhou, J.: Page Segmentation by White Streams. Proceedings of International Conference on Document Analysis and Recognition, Saint-Malo, France (1991), 945-953.
- [10] Breuel, T.M.: Two Geometrie Algorithms for Layout Analysis, in Proceedings of the Fifth International Workshop on Document Analysis Systems, Princeton, NY (2002), LNCS 2423, 188-199.
- [11] O’Gorman, L.: The Document Spectrum for Page Layout Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (1993), 1162-1173.
- [12] Kise, K. and Sato, A. and Iwata, M.: Segmentation of Page Images using the Area Voronoi Diagram. Computer Vision and Image Understanding 70 (1998), 370-382.
- [13] Mehmet Sezgin & Bülent Sankur, “Survey over image thresholding techniques and quantitative performance evaluation”, Journal of Electronic Imaging 13(1), 146– 165 (January 2004).
- [14] W. Niblack, “An Introduction to Digital Image Processing”, pp. 115-116, Prentice Hall, 1986.
- [15] Z. Zhang and C. L. Tan, “Restoration of images scanned from thick bound documents”, Proc. Int. conf. Image Processing., vol. 1, 2001, pp.1074-1077.

- [16] N. Otsu, “*A threshold selection method from gray-level histograms*” IEEE Trans. Syst. Man Cybern. 9(1), 62–66 (1979).
- [17] Yichao Ma, Chunheng Wang, Baihua Xiao, et Ruwei Dai. “Usage-oriented performance evaluation for text localization algorithms”. In Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, volume 2, page(s) 1033-1037, 2007.
- [18] Xian-Sheng Hua, Liu WenYin, and Hong-Jiang Zhang, “An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms”. In IEEE Transactions on circuits and systems for video technology, vol.14, no.4, avril 2004
- [19] Vasant Manohar , Padmanabhan Soundararajan , Matthew Boonstra, Harish Raju, Dmitry Goldgof, Rangachar Kasturi, and John Garofolo, “Performance Evaluation of Text Detection and Tracking in Video”. In Document analysis systems VII, vol. 3872, pages 576-587, 2006.