

MỤC LỤC

CHƯƠNG 1:TỔNG QUAN VỀ XỬ LÝ ẢNH ,TIỀN XỬ LÝ VÀ PHÂN ĐOẠN ẢNH	5
1.1 Tổng Quan Về Xử Lý Ảnh	5
1.2 Tổng quan về phân đoạn ảnh.....	6
1.3 Tổng quan về tiền xử lý ảnh	7
CHƯƠNG 2:MỘT SỐ PHƯƠNG PHÁP TIỀN XỬ LÝ ẢNH.....	8
2.1. Nhị phân ảnh.....	8
2.1.1. Phân loại các phương pháp xác định ngưỡng T	10
2.1.2. Một số phương pháp xác định ngưỡng T	11
2.1.3. Nhận xét	15
2.2. Hiệu chỉnh độ nghiêng của trang văn bản	17
2.2.1. Phương pháp dựa trên biến đổi Hough	18
2.2.2. Phương pháp láng giềng gần nhất (nearest neighbours)	19
2.2.3. Phương pháp sử dụng chiếu nghiêng (project profile).....	21
2.2.4. Nhận xét	22
2.3. Các toán tử hình thái (Morphological operations).....	23
CHƯƠNG 3:PHƯƠNG PHÁP PHÂN ĐOẠN CHỨNG MINH NHÂN DÂN	27
3.1. Giới thiệu bài toán	27
3.2. Tách các trường thông tin ở mặt trước	29
3.2.1. Tiền xử lý ảnh.....	30
3.2.2. Tách trường Số CMND	33
3.2.3. Tách các trường thông tin còn lại.....	37
3.3. Tách các trường thông tin ở mặt sau	42
3.3.1. Tiền xử lý ảnh.....	43
3.3.2. Xác định cấu trúc bảng.....	43
3.3.3. Tách trường thông tin.....	45
CHƯƠNG 4:CÀI ĐẶT THỬ NGHIỆM.....	46
KẾT LUẬN	48
TÀI LIỆU THAM KHẢO.....	49

Danh mục thuật toán

Thuật toán 2.1. Nhị phân ảnh	8
Thuật toán 2.2. Phương pháp phân ngưỡng Niblack.....	11
Thuật toán 2.3. Phương pháp phân ngưỡng Otsu.....	13
Thuật toán 2.4. Hiệu chỉnh độ nghiêng của ảnh tài liệu.....	17
Thuật toán 2.5. Xoay ảnh	17
Thuật toán 2.6. Xác định góc nghiêng dựa vào biến đổi Hough.....	19
Thuật toán 2.7. Phương pháp lán giếng gần nhất	20
Thuật toán 2.8. Sử dụng chiếu nghiêng để xác định góc nghiêng	22
Thuật toán 3.1. Xác định các vùng có thể là Trường Số CMND.....	33
33.Thuật toán 3.2. Tìm và tách trường Số CMND.....	35
Thuật toán 3.3. Phân đoạn vùng Số CMND.....	35
Thuật toán 3.4. Ước lượng bề dày đường lượn sóng	36
Thuật toán 3.5. Tách các ký tự thuộc mỗi dòng	39
Thuật toán 3.6. Xoá phân tiêu đề	40
Thuật toán 3.7. Tìm các đường kẻ ngang trong ảnh.....	44

LỜI CẢM ƠN

Trước hết em xin chân thành cảm ơn các thầy giáo trong khoa công nghệ thông tin trường đại học dân lập Hải Phòng đã trang bị những cơ bản cần thiết để em có thể thực hiện đề tài của mình .

Đặc biệt em xin bày tỏ lòng kính trọng và biết ơn sâu sắc tới thầy giáo hướng dẫn PGS.TS Ngô Quốc Tạo người đã tận tình hướng dẫn ,chỉ bảo và tạo mọi điều kiện thuận lợi giúp em trong quá trình thực tập.

Mặc dù đã cố gắng hết sức cùng với sự tận tâm của thầy giáo hướng dẫn xong do trình độ có hạn ,nội dung đề tài còn quá mới mẻ với em nên khó tránh khỏi những sai sót trong quá trình tiếp nhận kiến thức.Em rất mong được sự chỉ dẫn của thầy cô và sự góp ý bạn bè để trong thời gian tới em có thể xây dựng đồ án một cách hoàn thiện nhất.

Sinh viên

Trần Văn Toàn

Mở Đầu

Xử lý ảnh là một trong những chuyên ngành quan trọng và lâu đời của ngành Công Nghệ Thông Tin. XLA được áp dụng cho nhiều lĩnh vực khác nhau như y học, vật lý, hóa học, truy tìm tội phạm... Mục đích chung của việc XLA thường là (1) xử lý ảnh ban đầu để có được một bức ảnh mới theo một yêu cầu cụ thể, (2) phân tích ảnh để thu được các thông tin đặc trưng trên ảnh nhằm hỗ trợ cho việc phân loại và nhận biết ảnh, (3) phân đoạn ảnh để nhận biết được các thành phần trong ảnh nhằm hiểu được kết cấu của bức ảnh có mức độ cao hơn. Để xử lý được một bức ảnh thì phải trải qua nhiều bước, nhưng trong phần này em xin trình bày 2 bước quan trọng trong xử lý ảnh là tiền xử lý ảnh và bước phân đoạn ảnh. Hiện nay có rất nhiều thuật toán được đề xuất để giải quyết bài toán về tiền xử lý và phân đoạn ảnh.

Phân đoạn ảnh thì hầu hết các thuật toán đều dựa vào hai thuộc tính quan trọng của mỗi điểm ảnh so với các điểm lân cận của nó đó là sự khác nhau và giống nhau. Các phương pháp dựa trên sự khác nhau của các điểm ảnh được gọi là phương pháp biên (boundary-based methods) còn các phương pháp dựa trên sự giống nhau của các điểm ảnh được gọi là phương pháp miền

Tiền xử lý ảnh là một bước quan trọng trong xử lý ảnh. ở bước này hình ảnh vẫn ở mức thấp nhất chưa được xử lý. Với mục đích cải thiện các dữ liệu hình ảnh và ngăn chặn các biến dạng không mong muốn hoặc tăng cường nội dung thông tin hình ảnh... nhiều phương pháp tiền xử lý hình ảnh đã được đề xuất. Dưới đây em xin trình bày một số phương pháp cho quá trình này.

CHƯƠNG 1:TỔNG QUAN VỀ XỬ LÝ ẢNH ,TIỀN XỬ LÝ VÀ PHÂN ĐOẠN ẢNH

1.1 Tổng Quan Về Xử Lý Ảnh

Trong xã hội loài người, ngôn ngữ là một phương tiện trao đổi thông tin phổ biến trong quá trình giao tiếp. Bên cạnh ngôn ngữ, hình ảnh cũng là một cách trao đổi thông tin mang tính chính xác biểu cảm khá cao và đặc biệt không bị cảm giác chủ quan của đối tượng giao tiếp chi phối. Thông tin trên hình ảnh rất phong phú, đa dạng và có thể xử lý bằng máy tính. Chính vì vậy, trong những năm gần đây sự kết hợp giữa ảnh và đồ họa đã trở lên chặt chẽ trong lĩnh vực xử lý thông tin.

Cũng như xử lý dữ liệu hình ảnh bằng đồ họa, việc xử lý ảnh số là một lĩnh vực của tin học ứng dụng. Việc xử lý dữ liệu bằng đồ họa đề cập đến những hình ảnh nhân tạo, các ảnh này được xem xét như là những cấu trúc dữ liệu và được tạo ra bởi các chương trình. XLA số thao tác trên các ảnh tự nhiên thông qua các phương pháp và kỹ thuật mã hóa. Ảnh sau khi được thu nhận bằng các thiết bị thu nhận ảnh sẽ được biến đổi thành ảnh số theo các phương pháp số hóa được nhúng trong các thiết bị kỹ thuật khác nhau và được biểu diễn trên máy tính dưới dạng ma trận 2 chiều hoặc 3 chiều

Mục đích của việc XLA được chia làm 2

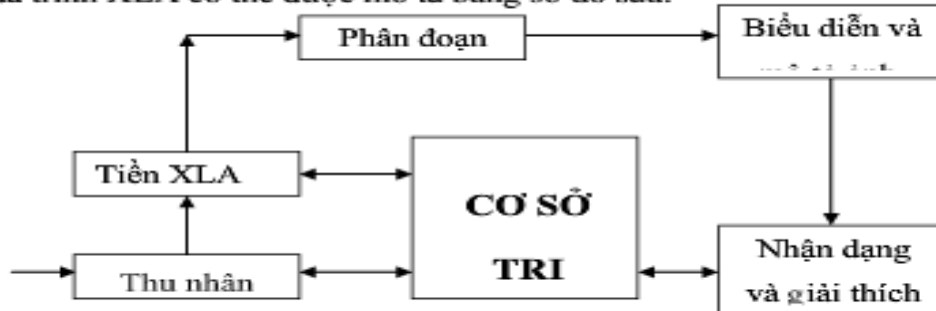
- Biến đổi ảnh làm tăng chất lượng ảnh
- Tự động nhận dạng, đoán ảnh, đánh giá nội dung ảnh

Phương pháp biến đổi các được sử dụng trong việc xử lý các ảnh chụp từ không trung. Một ứng dụng khác của việc biến đổi ảnh là mã hóa ảnh, trong đó các ảnh được xử lý để rồi lưu trữ hoặc truyền đi.

Các phương pháp nhận dạng ảnh được sử dụng khi xử lý tế bào, nhiễm sắc thể, nhận dạng chữ... Thực chất của công việc nhận dạng chính là sự phân loại đối tượng thành các lớp đối tượng chưa biết. Bài toán nhận dạng ảnh là một bài toán lớn, có rất nhiều ý nghĩa thực tiễn và ta cũng có thể thấy rằng để công việc nhận dạng trở lên dễ dàng thì ảnh phải được tách thành các đối tượng riêng biệt đây là mục đích chính của bài toán phân đoạn ảnh. Nếu phân đoạn ảnh không tốt sẽ dẫn đến sai lầm trong quá trình nhận dạng ảnh.

Quá trình XLA

Quá trình XLA có thể được mô tả bằng sơ đồ sau:



Hình 1. Quá trình xử lý ảnh

1.2 Tổng quan về phân đoạn ảnh

Để phân biệt các đối tượng trong ảnh, chúng ta cần phân biệt các đối tượng cần quan tâm với phần còn lại của ảnh, hay còn gọi là nền ảnh. Những đối tượng này có thể được tượng này có thể được tìm thấy nhờ kỹ thuật phân đoạn ảnh. Mỗi đối tượng trong ảnh được gọi là một vùng hay miền, đường bao quanh đối tượng gọi là đường biên. Mỗi một vùng ảnh phải có một đặc tính đồng nhất. Hình dáng của một đối tượng có thể được miêu tả hoặc bởi các tham số của đường biên hoặc các tham số của vùng mà nó chiếm giữ.

Có thể thấy kỹ thuật phát hiện biên và phân vùng ảnh là hai bài toán đối ngẫu của nhau. Dò biên để phân vùng được ảnh và ngược lại phân vùng được ảnh ta có thể phát hiện được biên.

Có rất nhiều kỹ thuật phân đoạn ảnh, nhìn chung ta có thể chia thành ba lớp khác nhau:

Các kỹ thuật cục bộ: dựa vào các thuộc tính cục bộ của điểm ảnh và láng giềng của nó.

Các kỹ thuật toàn thể: phân ảnh dựa trên thông tin chung của toàn bộ ảnh (vd sử dụng lược đồ xám của ảnh)

Các kỹ thuật tách (split), hợp (merge) và growing sử dụng các khái niệm đồng nhất và gắn về hình học.

1.3 Tổng quan về tiền xử lý ảnh

Hình ảnh tiền xử lý là hình ảnh chưa được chỉnh sửa ở bất kỳ phương diện nào. Ở bước này hình ảnh sẽ được cải thiện về độ tương phản, khử nhiễu, khử bóng, khử độ lệch... và với mục đích làm cho ảnh trở lên tốt hơn nữa và thường được thực hiện bởi những bộ lọc. Có rất nhiều phương pháp để xử lý ảnh ở giai đoạn này được trình bày và dưới đây em xin được trình bày một số phương pháp cụ thể của tiền xử lý ảnh.

CHƯƠNG 2: MỘT SỐ PHƯƠNG PHÁP TIỀN XỬ LÝ ẢNH

Đầu vào của các hệ thống xử lý ảnh thường là các tệp ảnh được thu nhận từ các thiết bị như: máy quét, máy ảnh, thiết bị ghi hình hay các thiết bị thu nhận hình ảnh khác. Các ảnh này thường có chất lượng thấp (bị lẫn các nhiễu, mất các chi tiết của đối tượng, hay bị lệch so với ảnh gốc một góc bất kỳ,...). Nguyên nhân là do: thiết bị thu nhận không đảm bảo, điều kiện thu nhận không tốt (độ sáng thay đổi, thu nhận trong khi di chuyển,...) hay quá trình sao lưu bị mất mát thông tin.

Để các bước xử lý tiếp theo thu được kết quả tốt cần phải có quá trình tiền xử lý để nâng cao chất lượng ảnh đầu vào. Quá trình này bao gồm các công đoạn khôi phục và tăng cường ảnh:

Khôi phục ảnh nhằm mục đích loại bỏ hay giảm thiểu các ảnh hưởng của môi trường tác động lên ảnh. Bao gồm các bước: lọc ảnh, khử nhiễu, xoay ảnh,... nhằm giảm bớt các biến dạng của ảnh và đưa ảnh về trạng thái gần như ban đầu.

Tăng cường ảnh không phải làm tăng lượng thông tin trong ảnh mà là làm nổi bật các đặc trưng của ảnh giúp cho công việc phía sau được hiệu quả hơn. Công đoạn này bao gồm các việc như: lọc độ tương phản, làm trơn ảnh, nhị phân ảnh,...

Trong đó các thao tác nhị phân ảnh, căn chỉnh độ nghiêng và xóa nhiễu là các thao tác cơ bản nhất và thường được áp dụng. Trong các phần tiếp theo của chương này sẽ trình bày một số thuật toán trong các thao tác đó.

2.1. Nhị phân ảnh

Ảnh nhận được từ các thiết bị thu nhận hình ảnh như máy ảnh hay camera thường là ảnh màu hay ảnh đa cấp xám, các thành phần trong ảnh là rất phức tạp (màu sắc, kết cấu...). Do đó muốn làm nổi bật các đặc trưng trong ảnh thì phải chuyển về dạng ảnh nhị phân, ảnh chỉ có hai màu (đen và trắng) – tương ứng với nền và tiền cảnh (đối tượng “quan tâm”). Nhị phân ảnh (hay còn gọi là phân ngưỡng) là thao tác chuyển từ ảnh đa cấp xám (hoặc ảnh màu) về ảnh nhị phân (Thuật toán 2.1).

Thuật toán 2.1. Nhị phân ảnh

INPUT: Ảnh màu hoặc ảnh đa cấp xám

OUTPUT: Ảnh nhị phân

1. Xác định ngưỡng T

2. Chuyển ảnh về dạng nhị phân

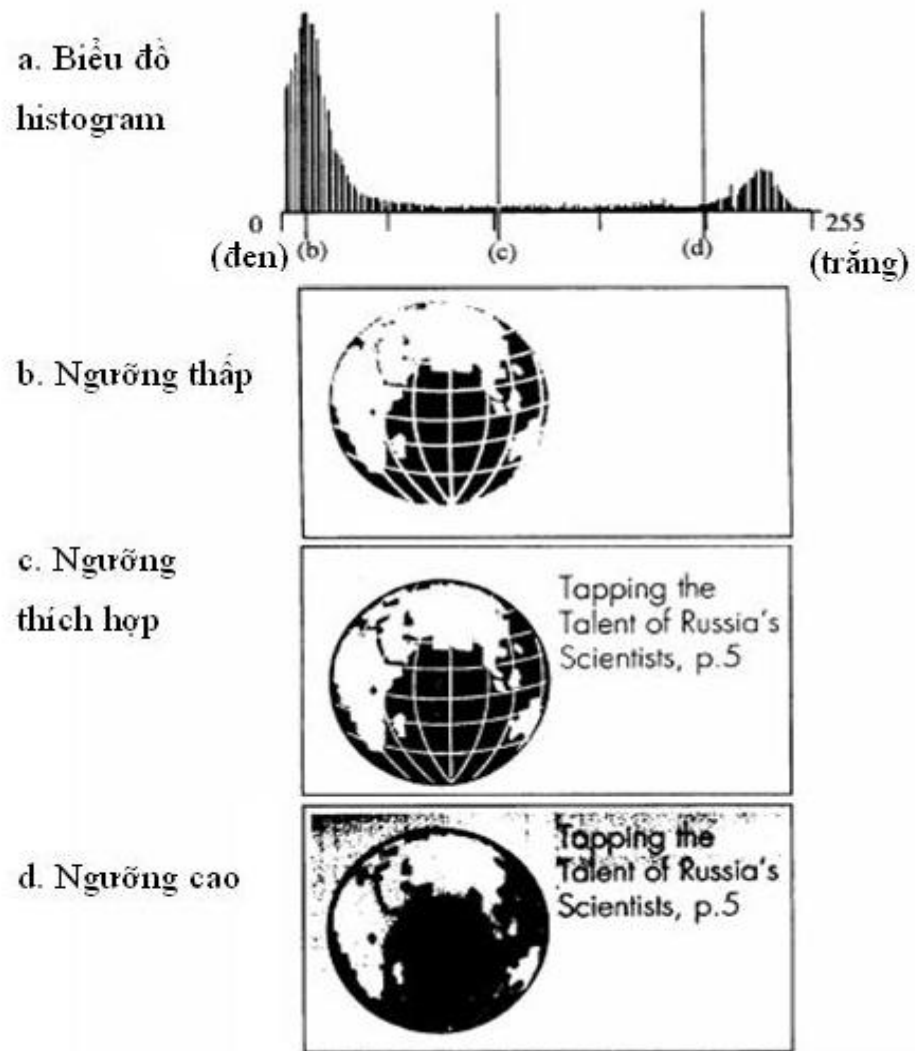
Như vậy, cơ bản của thuật toán nhị phân ảnh là xác định một ngưỡng T để phân tách giữa nền và đối tượng trong ảnh. Giả sử với ảnh đầu vào $I(x, y)$, có giá trị tại điểm (x, y) là $g(x, y)$ (đối với ảnh đa cấp xám: $g(x, y) \in [0, 255]$). Khi đó giá trị của điểm ảnh

(x, y) trong ảnh nhị phân $I'(x, y)$ sẽ được xác định như sau:

$$g'(x, y) = \begin{cases} 0 & \text{nếu } g(x, y) < T \\ 1 & \text{nếu } g(x, y) \geq T \end{cases}$$

Việc xác định một ngưỡng T thích hợp luôn là một quá trình khó khăn và dễ gây ra lỗi

(Hình 2. 1). Điều này sẽ đặc biệt khó khăn khi độ tương phản giữa các đối tượng và nền thấp hay khi ảnh có độ chiếu sáng không đồng đều khi thu nhận. Nếu ngưỡng T quá thấp thì các đối tượng thu được có thể bị xóa mất các chi tiết của ảnh, ngược lại nếu ngưỡng T quá cao thì có thể chứa các điểm ảnh nhiễu.



Hình 2. 1 Nhi phân ảnh

Có rất nhiều phương pháp để xác định ngưỡng phân tách T. Ngưỡng T có thể được xác định cho toàn bộ ảnh (ngưỡng tổng quát) hay được xác định cho mỗi điểm ảnh cụ thể (ngưỡng cục bộ). Trong phần tiếp theo sẽ phân loại và giới thiệu một số phương pháp xác định ngưỡng T.

2.1.1. Phân loại các phương pháp xác định ngưỡng T

Căn cứ vào phương pháp được áp dụng, có thể chia ra làm 6 nhóm sau [13]:

1. Các phương pháp dựa vào hình dạng của histogram (Histogram Shape-Based Thresholding Methods). Căn cứ vào hình dáng của histogram như: các đỉnh, các khe và độ cong (peaks, valleys and curvatures) để xác định ngưỡng. Vị trí lấy ngưỡng có thể là khe lõm nhất giữa hai đỉnh hay điểm cách xa đường thẳng nối hai đỉnh.

2. Các phương pháp dựa vào việc chia nhóm (Clustering-Based Thresholding Methods).

Các phương pháp loại này cố gắng chia ảnh ra làm hai nhóm tương ứng với nền và đối tượng dựa trên một số tiêu chí đánh giá “khoảng cách” giữa hai nhóm hay giữa các phần tử trong mỗi nhóm.

3. Các phương pháp dựa vào entropy (Entropy-Based Thresholding Methods).

Trong kỹ thuật này người ta chọn ngưỡng dựa vào entropy dựa trên một số cơ sở như: cực đại các entropy (nền và đối tượng), cực tiểu các entropy lai (giữa ảnh gốc và ảnh nhị phân) hay độ đo entropy mờ.

4. Các phương pháp dựa vào thuộc tính giống nhau (Thresholding Based on Attribute Similarity).

Ngưỡng được xác định dựa độ đo các thuộc tính giống nhau của ảnh gốc và ảnh nhị phân, chẳng hạn như căn cứ vào các cạnh thỏa mãn, độ chặt của hình dáng, momen mức xám, khả năng liên kết, kết cấu,...

5. Các phương pháp căn cứ vào không gian (Spatial Thresholding Methods).

Sử dụng sự tương liên hoặc/và phân phối thông kê bậc cao giữa các pixel để chọn ngưỡng.

6. Các phương pháp ngưỡng thích ứng cục bộ (Locally Adaptive Thresholding).

Kỹ thuật này sẽ xác định ngưỡng $t(x, y)$ cho từng điểm ảnh (x, y) riêng biệt căn cứ vào mối tương quan giữa điểm ảnh đó và các láng giềng của nó.

2.1.2. Một số phương pháp xác định ngưỡng T

Trong phần này sẽ trình bày hai phương pháp thường được sử dụng để xác định ngưỡng nhị phân T. Phương pháp Niblack xác định ngưỡng T cho mỗi điểm ảnh riêng biệt, trong khi phương pháp Otsu ngưỡng T được xác định cho toàn bộ ảnh.

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. Thuật toán được mô tả như sau:

Thuật toán 2.2. Phương pháp phân ngưỡng Niblack

INPUT: Giá trị mức xám của các điểm ảnh $g(x, y)$

OUTPUT: Ngưỡng nhị phân cho mỗi điểm ảnh $T(x, y)$

Duyệt tất cả các điểm ảnh:

1. Xác định cửa sổ ($w \times w$) bao quanh
2. Tính giá trị trung bình $m(x, y)$
3. Tính độ lệch chuẩn $\sigma(x, y)$
4. Xác định ngưỡng $T(x, y)$

Với mỗi điểm ảnh, xác định một cửa sổ kích thước ($w \times w$) bao quanh nó. Giá trị ngưỡng được tính toán dựa trên giá trị trung bình và độ lệch chuẩn trong cửa sổ đó.

Với điểm ảnh ở vị trí (x, y) giá trị ngưỡng được xác định như sau [14]:

$$T(x, y) = m(x, y) + k \cdot \sigma(x, y)$$

Trong đó: $m(x, y)$ và $\sigma(x, y)$ tương ứng là giá trị trung bình và độ lệch chuẩn cục bộ trong cửa sổ ($w \times w$) với tâm ở vị trí (x, y) , và được xác định như sau:

$$m(x, y) = \frac{1}{w^2} \sum_{i=x-\frac{w}{2}}^{x+\frac{w}{2}} \sum_{j=y-\frac{w}{2}}^{y+\frac{w}{2}} g(i, j)$$

$$\sigma^2(x, y) = \frac{1}{w^2} \sum_{i=x-\frac{w}{2}}^{x+\frac{w}{2}} \sum_{j=y-\frac{w}{2}}^{y+\frac{w}{2}} g^2(i, j) - m^2(x, y)$$

k là tham số dùng để xác định đường biên của đối tượng chiếm bao nhiêu phần trong đối tượng trả về. Kích thước của cửa sổ phải đủ nhỏ để giữ lại các chi tiết và cũng phải đủ lớn để khử các điểm nhiễu. Theo [13] thì tham số $k = -0.2$, kích thước của sổ $w = 15$.

Cơ bản dựa trên phương pháp của Niblack, Sauvola đưa ra công thức xác định ngưỡng như sau [15]:

$$T(x, y) = m(x, y) + \left[1 + k \left(\frac{\sigma(x, y)}{R} - 1 \right) \right]$$

Trong đó: R là giá trị lớn nhất của độ lệch chuẩn (với ảnh đa cấp xám: $R = 128$), k là tham số nằm trong khoảng $[0.2, 0.5]$, $m(x, y)$ và $\sigma(x, y)$ là giá trị đáp ứng các mức ngưỡng khác nhau tùy theo các điểm lân cận. Với một vài vùng ảnh có độ

tương phản cao thì $\sigma(x, y) \approx R$, khi đó $T(x, y) \approx m(x, y)$. Kết quả này giống như phương pháp Niblack. Trong trường hợp $T(x, y)$ nhỏ hơn giá trị trung bình thì sẽ xóa đi một vài vùng tối của nền. Tham số k dùng để điều chỉnh giá trị ngưỡng so với giá trị trung bình $m(x, y)$ (lớn hơn hay nhỏ hơn một tỷ lệ k).

Như vậy ngưỡng của mỗi điểm ảnh được xác định dựa trên việc đánh giá giá trị của các điểm ảnh lân cận với nó, do đó rất thích hợp cho những ảnh có độ sáng thay đổi (ví dụ như ảnh chụp từ camera). Nhưng thời gian tính toán là rất chậm, tùy thuộc vào kích thước của cửa sổ.

2/. Phương pháp Otsu

Đây là phương pháp xác định ngưỡng cho toàn bộ ảnh. Phương pháp này sẽ tìm một ngưỡng để phân chia các điểm ảnh vào hai lớp tiền cảnh (đối tượng) và nền. Giá trị ngưỡng được xác định sao cho “khoảng cách” giữa các điểm ảnh trong mỗi lớp là nhỏ nhất, điều này tương đương với khoảng giữa hai lớp là lớn nhất. Việc phân chia này dựa trên các giá trị trong histogram của ảnh. Các bước để xác định ngưỡng tOtsu của ảnh được tiến hành như sau:

Thuật toán 2.3. Phương pháp phân ngưỡng Otsu

INPUT: Ảnh đa cấp xám

OUTPUT: Ngưỡng nhị phân cho toàn bộ ảnh: tOtsu

1. Tính histogram của ảnh: $\{p_i\}$
2. Duyệt tất cả mức xám của ảnh: t

- a. Tính lũy tích cho nền P_B^t và tiền cảnh P_F^t
- b. Tính trị số trung bình của nền μ_B^t và tiền cảnh μ_F^t
- c. Tính độ lệch chuẩn của nền σ_B^t và tiền cảnh σ_F^t
- d. Xác định hàm khoảng cách: $\text{var}_{\text{between-class}}^t$ hoặc $\text{var}_{\text{within-class}}^t$
- e. Ngưỡng t_{Otsu} là đối số để hàm $\text{var}_{\text{between-class}}^t$ lớn nhất
Hoặc hàm $\text{var}_{\text{within-class}}^t$ đạt giá trị nhỏ nhất

Với ảnh đầu vào là ảnh đa cấp xám, mỗi điểm ảnh có giá trị cường độ nằm trong khoảng $[0, L]$ ($L=255$). Giả định rằng $\{p_i\}$ ($i = [0, L-1]$) là lược đồ mức xám của ảnh tỷ số giữa số lượng điểm ảnh có mức xám i so với toàn bộ ảnh và t là giá trị mức xám của ngưỡng lựa chọn. Sử dụng F và B để ký hiệu cho lớp tiền cảnh và

nền, khi đó việc tính toán xác suất nền và tiền cảnh bởi ngưỡng t được xác định bởi các hàm sau [16]:

Hàm lũy tích của tiền cảnh và nền:

$$P_B^t = \sum_{i=0}^t p_i \quad P_F^t = 1 - P_B^t = \sum_{i=t+1}^{L-1} p_i$$

Độ lệch chuẩn của tiền cảnh và nền:

$$\sigma_B^t = \left(\frac{1}{P_B^t} \right) \sum_{i=0}^t (i - \mu_B^t)^2 p_i \quad \sigma_F^t = \left(\frac{1}{P_F^t} \right) \sum_{i=t+1}^{L-1} (i - \mu_F^t)^2 p_i$$

Giá trị trung bình của tiền cảnh và nền:

$$\mu_B^t = \left(\frac{1}{P_B^t} \right) \sum_{i=0}^t i p_i \quad \mu_F^t = \left(\frac{1}{P_F^t} \right) \sum_{i=t+1}^{L-1} i p_i$$

Từ đó ta xác định được khoảng cách giữa hai lớp và kháng cách của lớp trong như sau:

$$\begin{aligned} \text{var}_{\text{between-class}}^t &= P_B^t (\mu_B^t - \mu)^2 + P_F^t (\mu_F^t - \mu)^2 \\ &= P_B^t P_F^t (\mu_B^t - \mu_F^t)^2 \end{aligned}$$

$$\text{var}_{\text{within-class}}^t = P_B^t \sigma_B^t + P_F^t \sigma_F^t$$

Trong đó: $\mu = \sum_{i=0}^{L-1} i p_i$ là giá trị trung bình của toàn ảnh.

Khi đó ngưỡng t_{Otsu} được xác định là đối số để hàm $\text{var}_{\text{between-class}}^t$ đạt giá trị lớn nhất hoặc hàm $\text{var}_{\text{within-class}}^t$ đạt giá trị nhỏ nhất:

$$\begin{aligned} t_{Otsu} &= \arg \left\{ \max_{1 \leq t \leq L} \left(\text{var}_{\text{between-class}}^t \right) \right\} \\ &= \arg \left\{ \min_{1 \leq t \leq L} \left(\text{var}_{\text{within-class}}^t \right) \right\} \end{aligned}$$

Đây là phương pháp được sử dụng phổ biến trong xử lý ảnh vì nó phân đoạn và làm nổi bật ảnh khá tốt. Tuy nhiên phương pháp này sẽ thất bại khi số điểm ảnh tiền cảnh nhỏ hơn 5%.

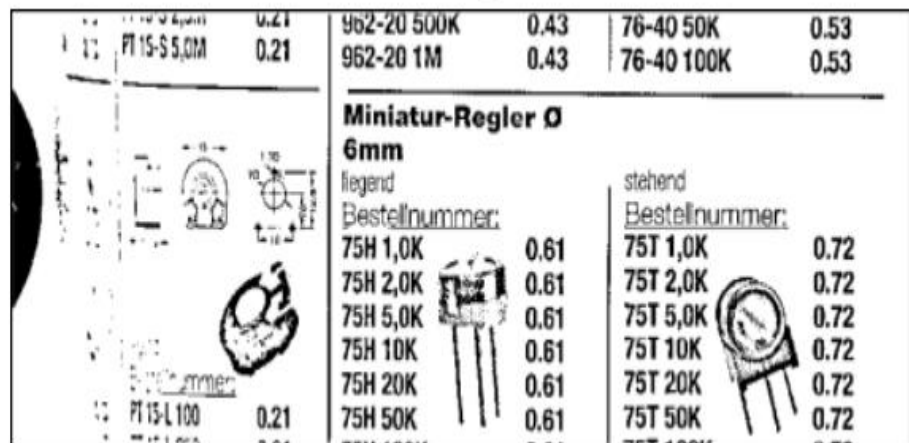
2.1.3. Nhận xét

Nhị phân ảnh là một thao tác cơ bản để phân tách giữa nền và đối tượng. Trong đó việc xác định ngưỡng nhị phân là thao tác quan trọng và khó khăn nhất. Có thể xác định ngưỡng cho toàn bộ ảnh (phương pháp Otsu) hay xác định ngưỡng cho từng điểm ảnh riêng biệt (phương pháp Niblack) tùy vào yêu cầu của bài toán và tích chất ảnh đầu vào.

Hình 2. 2 và Hình 2. 3 mô tả kết quả phân ngưỡng của hai phương pháp trên hai ảnh có tính chất khác nhau. Hình 2. 2 ảnh có độ chiếu sáng thay đổi, ta thấy phương pháp Niblack cho kết quả tốt hơn phương pháp Otsu. Trong khi Hình 2. 3 kết quả của phương pháp Otsu lại tốt hơn phương pháp Niblack. Kết quả so sánh giữa hai phương pháp này được chỉ ra trong Bảng 2. 1.



a. Ảnh gốc

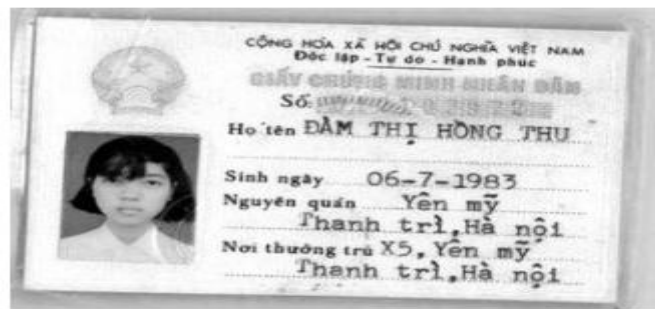


b. Phương pháp Niblack

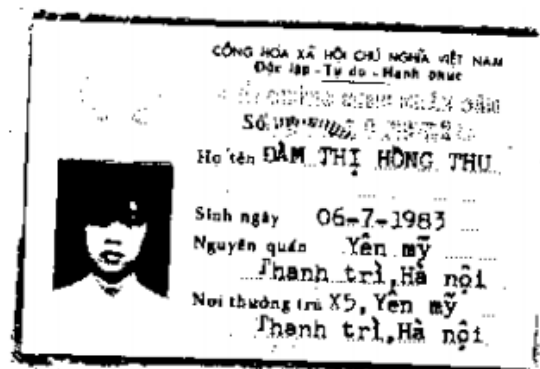
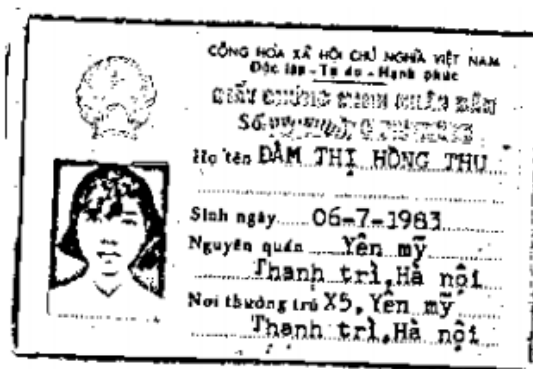
0.21	962-20 500K	0.43	76-40 50K	0.53
0.21	962-20 1M	0.43	76-40 100K	0.53
Miniatuur-Regler Ø				
6mm				
legend		stehend		
Bestellnummer:		Bestellnummer:		
75H 1,0K	0.61	75T 1,0K	0.72	
75H 2,0K	0.61	75T 2,0K	0.72	
75H 5,0K	0.61	75T 5,0K	0.72	
75H 10K	0.61	75T 10K	0.72	
75H 20K	0.61	75T 20K	0.72	
75H 50K	0.61	75T 50K	0.72	

c. Phương pháp Otsu (ngưỡng 148)

Hình 2. 2 So sánh các phương pháp nhị phân ảnh đối với có độ sáng thay đổi



A. ảnh gốc



b. Phương pháp Niblack

c. Phương pháp Otsu (ngưỡng 154)

Hình 2. 3 So sánh các phương pháp nhị phân ảnh đối với ảnh Chứng minh nhân dân

2.2. Hiệu chỉnh độ nghiêng của trang văn bản

Đối với những chương trình nhận dạng, có ảnh đầu vào là những trang tài liệu dạng văn bản thì các ảnh thu nhận được thường bị lệch so với ảnh gốc một góc bất kỳ. Nguyên nhân là do trong quá trình thu nhận: ảnh gốc bị đặt lệch, thiết bị ghi nhận hình ảnh đặt không đúng vị trí hay thu nhận ảnh bị xô dịch..., điều này là không thể tránh khỏi. Do đó, để cho các bước xử lý tiếp theo (phân tích và nhận dạng) được chính xác cần phải có thao tác hiệu chỉnh độ nghiêng của ảnh thu nhận được. Các bước hiệu chỉnh độ nghiêng của ảnh được mô tả trong Thuật toán 2.4.

Thuật toán 2.4. Hiệu chỉnh độ nghiêng của ảnh tài liệu

INPUT: Ảnh (nhị phân) bị nghiêng

OUTPUT: Ảnh đã chỉnh độ nghiêng

1. Xác định góc nghiêng α
2. Xoay ảnh với góc nghiêng α

Trong đó, xác định góc nghiêng là thao tác quan trọng nhất và khó khăn nhất. Có rất nhiều phương pháp khác nhau để xác định góc nghiêng: có thể trực tiếp dựa vào các thống kê, đánh giá góc nghiêng của các đối tượng trong ảnh hay phân tích, đánh giá trên ảnh đã được biến đổi. Trong đó có 3 phương pháp thường được sử dụng: phương pháp dựa trên biến đổi Hough, phương pháp láng giềng gần nhất (nearest neighbours) và phương pháp sử dụng chiếu nghiêng (project profile). Các phương pháp này sẽ được trình bày ở các phần tiếp theo.

Sau khi xác định được góc nghiêng của ảnh sẽ thực hiện thao tác xoay ảnh với góc nghiêng đã xác định được quanh một vị trí gốc (tâm xoay). Tâm xoay thường lấy là điểm chính giữa của ảnh ($w/2, h/2$). Các bước để xoay ảnh được thực hiện như sau:

Thuật toán 2.5. Xoay ảnh

INPUT:

Ảnh (nhị phân) bị nghiêng I

Góc nghiêng α

Tâm xoay (x_0, y_0)

OUTPUT: Ảnh đã chỉnh độ nghiêng I'

Duyệt tất cả các điểm ảnh $g(x, y)$ trong ảnh I

1. Xác định vị trí mới $g'(x', y')$ trong ảnh I'

$$x' = x_0 + (x-x_0).\cos(\alpha) - (y-y_0).\sin(\alpha)$$

$$y' = y_0 + (x-x_0).\sin(\alpha) + (y-y_0).\cos(\alpha)$$

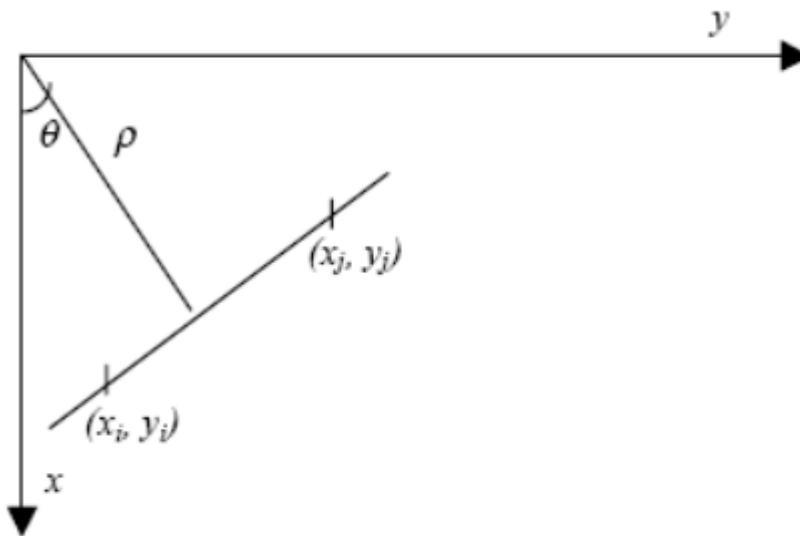
2. Chuyển giá trị điểm ảnh: $g'(x', y') = g(x, y)$

2.2.1. Phương pháp dựa trên biến đổi Hough

Biến đổi Hough là phép biến đổi điểm ảnh từ hệ tọa độ đề các Oxy sang hệ tọa độ cực ρ - θ . Thay vì biểu diễn một tập các điểm (x_i, y_i) (thuộc đường thẳng) trong mặt phẳng x - y thì ta có thể biểu diễn bằng một cặp (ρ, θ) trong mặt phẳng ρ - θ . Công thức của phép chuyển đổi là:

$$x.\cos\theta + y.\sin\theta = \rho$$

Trong đó: x và y được thay thế bởi x_i và y_i , θ là góc giữa vector khoảng cách (tính từ gốc tọa độ đến điểm gần nhất thuộc đường thẳng) và trục x , ρ là khoảng cách từ gốc tọa độ tới đường thẳng (Hình 2. 4).



Hình 2. 4 Biến đổi Hough

Như vậy, biến đổi Hough 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. Việc xác định góc nghiêng của ảnh dựa vào biến đổi Hough gồm hai bước chính:

- Thực hiện phép biến đổi Hough.
- Tính toán lũy tích để tìm góc nghiêng.

Thuật toán 2.6 thể hiện một các xác định góc nghiêng dựa vào biến đổi Hough. Trong đó bước 1 và bước 2 là thực hiện phép biến đổi Hough, bước 3 và bước 4 thể hiện một cách thống kê để tìm góc nghiêng.

Thuật toán 2.6. Xác định góc nghiêng dựa vào biến đổi Hough

INPUT: Ảnh (nhị phân) bị nghiêng I

OUTPUT: Góc nghiêng α

1. Khởi tạo mảng: $h[\rho][\theta] = 0$

(Đếm số lượng điểm thuộc đường thẳng)

2. Duyệt tất cả các điểm ảnh:

Duyệt tất cả các góc có thể θ_i

Tính: $\rho_i = x \cdot \cos\theta_i + y \cdot \sin\theta_i$

Tăng $h[\rho][\theta]$ lên 1

3. Tìm k phân tử trong mảng $h[\rho][\theta]$ có giá trị lớn nhất

(Tìm k đường thẳng trong ảnh)

4. $\bar{\theta}$ là tỉ số trung bình của các góc trong k phân tử trên

(Tính góc nghiêng chung bình của k đường thẳng)

Biến đổi Hough sử dụng rất nhiều tính toán do phải thao tác trên từng điểm ảnh riêng lẻ. Người ta đã cải tiến để tăng tốc độ thực hiện bằng cách thực hiện tính toán trên chùm điểm ảnh. Những chùm điểm ảnh này là các dải liên tục các điểm đen liên tiếp nhau theo chiều ngang hoặc chiều dọc. Mỗi chùm được mã hóa bởi độ dài của và vị trí kết thúc của nó. Với cải tiến này thì thuật toán này thích hợp với các góc nghiêng $\sim 15^\circ$ và cho độ chính xác rất cao. Tuy cải tiến này làm tăng tốc độ thuật toán nhưng vẫn rất chậm so với các phương pháp khác. Hơn nữa, trong trường hợp văn bản là thưa, thuật toán này tỏ ra không hiệu quả.

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 ký tự trong một từ và giữa các ký tự của từ trên cùng một dòng là nhỏ hơn khoảng cách giữa hai dòng văn bản, vì thế đối với mỗi ký tự, láng giềng gần nhất của nó sẽ là các ký tự liền kề trên cùng một dòng văn bản. Các bước chính của thuật toán

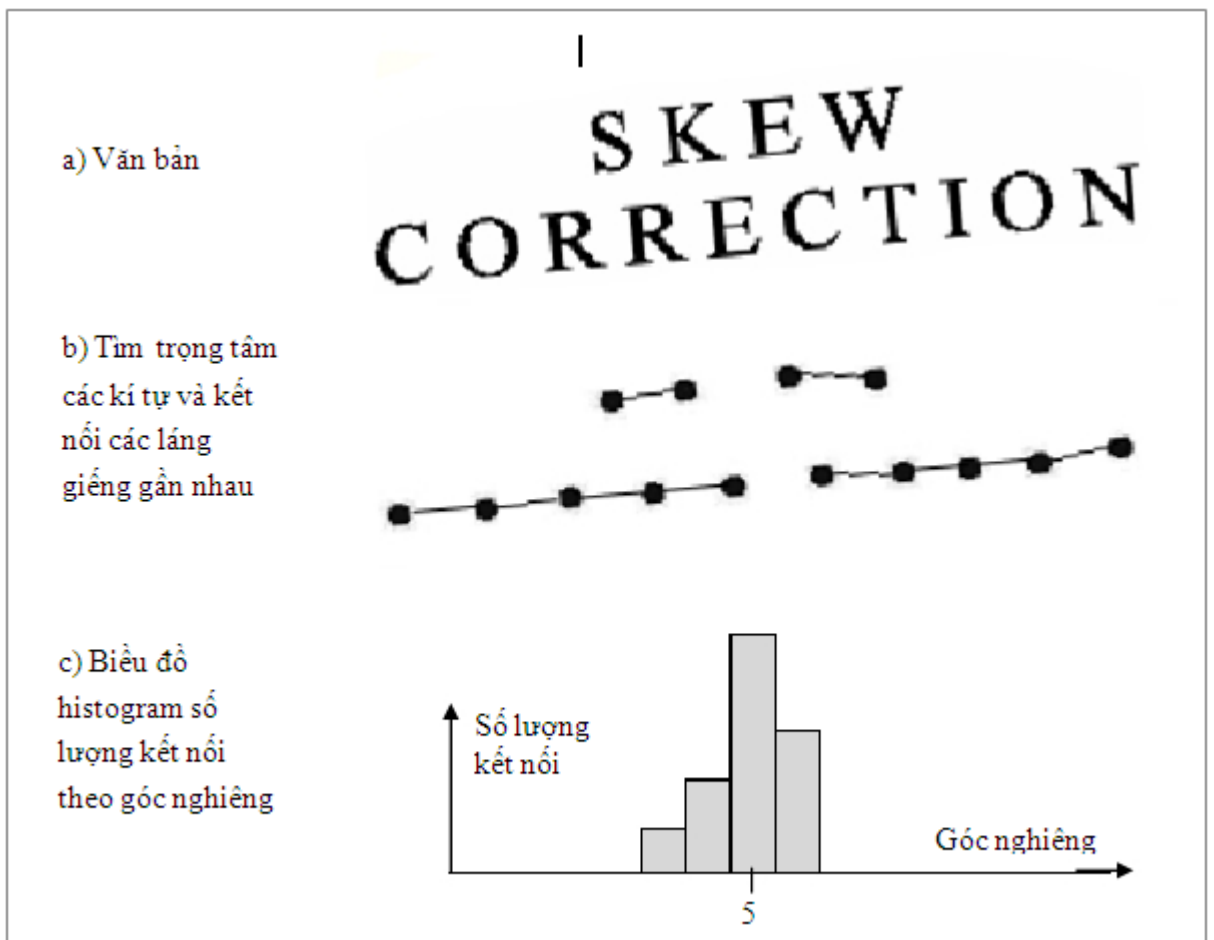
được mô tả như sau:

Thuật toán 2.7. Phương pháp lán giềng gần nhất

INPUT: Ảnh bị nghiêng I (ảnh nhị phân)

OUTPUT: Góc nghiêng α

1. Xác định các thành phần liên thông
2. Tìm lán giềng gần nhất của mỗi thành phần liên thông dựa vào khoảng cách Öclit giữa tâm của hai miền liên thông, nối hai tâm đó lại thành một vector
3. Tính góc của các vector
4. Xây dựng biểu đồ thể hiện số lượng các vector cùng phương
5. Góc nghiêng α tương ứng với vị trí có nhiều vector cùng phương nhất trên biểu đồ.



Hình 2. 5 Phương pháp lán giềng gần nhất

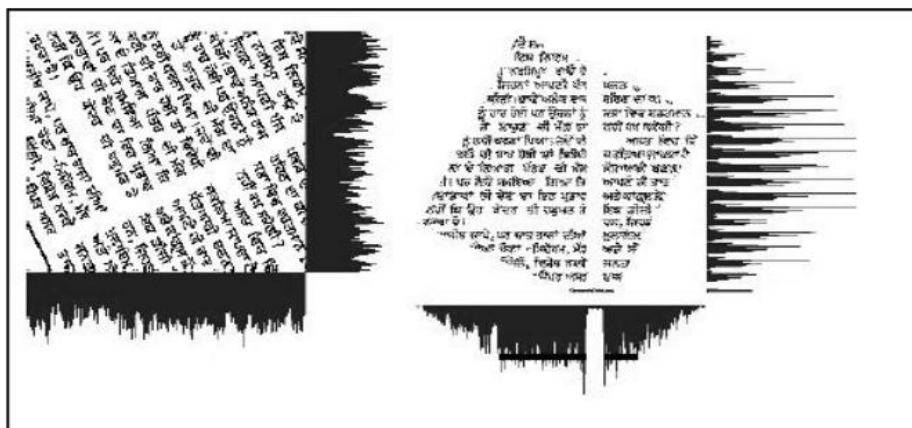
Chi phí tính toán của phương pháp này đã giảm đi nhiều so với phương pháp sử dụng biến đổi Hough tuy nhiên vẫn cao. Độ chính xác của phương pháp này phụ thuộc rất nhiều vào số thành phần của một kí tự trong văn bản. Đối với các kí tự có nhiều thành phần ví dụ như chữ ă có 3 thành phần gồm thân, mũ và dấu hỏi. Khi đó, láng giềng gần nhất của mỗi phần sẽ là một trong hai thành phần còn lại chứ không phải là kí tự liền kề với nó. Điều đó làm giảm đi độ chính xác của thuật toán, đồng thời khiến cho phương pháp này không thích hợp với chữ Tiếng Việt.

L’O Gorman đã phát triển thuật toán với ý tưởng là với mỗi thành phần lấy k láng giềng (k có thể là 4 hoặc 5) thay vì lấy một láng giềng duy nhất. Và góc thu được được sử dụng như là góc nghiêng ước lượng. Góc nghiêng ước lượng này được sử dụng để loại bỏ các liên kết mà góc của nó không gần với góc ước lượng. Sau đó, thực hiện xác định lại góc nghiêng theo các liên kết được giữ lại. Ý tưởng này đã cải thiện được độ chính xác của thuật toán nhưng lại làm tăng thời gian tính toán. Trên thực tế, phương pháp này được gọi là phương pháp docstrum – thực hiện cả việc xác định góc nghiêng và phân tích ảnh tài liệu.

2.2.3. Phương pháp sử dụng chiếu nghiêng (project profile)

Đây là phương pháp thường được sử dụng trong các hệ thống thương mại. Project profile là biểu đồ các giá trị điểm đen được tích lũy theo các dòng quét song song với một phương xác định trên toàn bộ ảnh. Biểu đồ này thường được ghi theo phương ngang và dọc của ảnh, gọi là các histogram chiếu ngang và histogram chiếu dọc (Hình

2. 6).



Hình 2. 6 Phương pháp chiếu nghiêng

Từ histogram theo chiều ngang ta có thể thấy, những đường thuộc vào dòng văn bản nằm ngang sẽ có giá trị cao hơn những dòng thuộc vào khoảng giữa các dòng với nhau. Dựa vào tính chất đó để xác định góc nghiêng của ảnh như sau:

Thuật toán 2.8. Sử dụng chiếu nghiêng để xác định góc nghiêng

INPUT: Ảnh bị nghiêng I (ảnh nhị phân)

OUTPUT: Góc nghiêng α

1. Duyệt một số góc nghiêng:

- Xoay ảnh với góc nghiêng đó
- Tính biểu đồ histogram theo phương ngang
- Tính độ biến thiên của histogram

2. Góc nghiêng α là góc tương ứng với độ biến thiên lớn nhất

H.S.Braid đã tiến hành cải biến phương pháp project profile. Đầu tiên, tìm các thành phần liên thông, mỗi thành phần liên thông được đại diện bởi tâm ở đáy của hình chữ nhật bao quanh nó. Tiếp theo, các thành phần liên thông được nối với nhau. Sau đó, tiến hành xoay văn bản ở một số vị trí và tính biểu đồ histogram. Vị trí có độ biến thiên lớn nhất của histogram chính là vị trí góc nghiêng cần tìm. Cải biến này đã làm tăng đáng kể tốc độ của phương pháp Project profile, đồng thời độ chính xác của nó đạt khá cao (~ 0.50), thích hợp với các trang có độ nghiêng trong khoảng ~ 100 .

2.2.4. Nhận xét

Việc chỉnh góc nghiêng của các ảnh thu được trong các hệ nhận dạng văn bản là rất cần thiết, bởi vì ảnh nhận được từ các thiết bị ghi hình thường bị lệch so với phương ban đầu. Trong đó, xác định góc lệch là việc quan trọng và khó khăn nhất. Trong phần trên đã giới thiệu ba phương pháp thường được sử dụng:

+ 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.

2.3. Các toán tử hình thái (Morphological operations)

Trong ảnh nhị phân, mỗi một điểm ảnh chỉ có hai mức xám (0 và 1). Do đó có thể coi mỗi phần tử ảnh như một phần tử logic và có thể áp dụng các toán tử hình thái đối với nó. Đầu vào của các toán tử hình thái thường là ảnh nhị phân (một số trường hợp là ảnh đa cấp xám) và phần tử cấu trúc (structuring element), kết hợp với việc sử dụng các toán tử tập hợp: hợp, giao, trừ và lấy phần bù. Các thao tác xử lý (trên ảnh đầu vào) cơ bản dựa trên những đặc trưng hình dáng của đối tượng như: hình bao, xương ảnh, bao lồi... dưới sự giám sát của phần tử cấu trúc.

Phần tử cấu trúc là một mặt nạ dạng bất kỳ, chỉ chứa thành phần đối tượng (thiết lập là 1) và thành phần “không quan tâm” (được để trống). Trong một số trường hợp, phần tử cấu trúc có thể chứa thành phần là nền (mang trị số 0). Có thể hiểu phần tử cấu trúc như là một tập tọa độ các điểm (kích thước nhỏ) chứa một gốc tọa độ (thường ở vị trí giữa). Trong Hình 2. 7 là ví dụ một phần tử cấu trúc kích thước 3x3.

1	1	1
1	1	1
1	1	1

Tập tọa độ điểm =
 { (-1, -1), (0, -1), (1, -1),
 (-1, 0), (0, 0), (1, 0),
 (-1, 1), (0, 1), (1, 1) }

Hình 2. 7 Phần tử cấu trúc 3x3

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). Có nhiều cách khác nhau để định nghĩa các toán tử hình thái (giãn ảnh và co ảnh). Giả sử $g(x, y)$ là ảnh nhị phân và $H(x, y)$ là phần tử cấu trúc.

Khi đó toán tử giãn ảnh $g_{out} = g_{in} \oplus H$ được định nghĩa như sau:

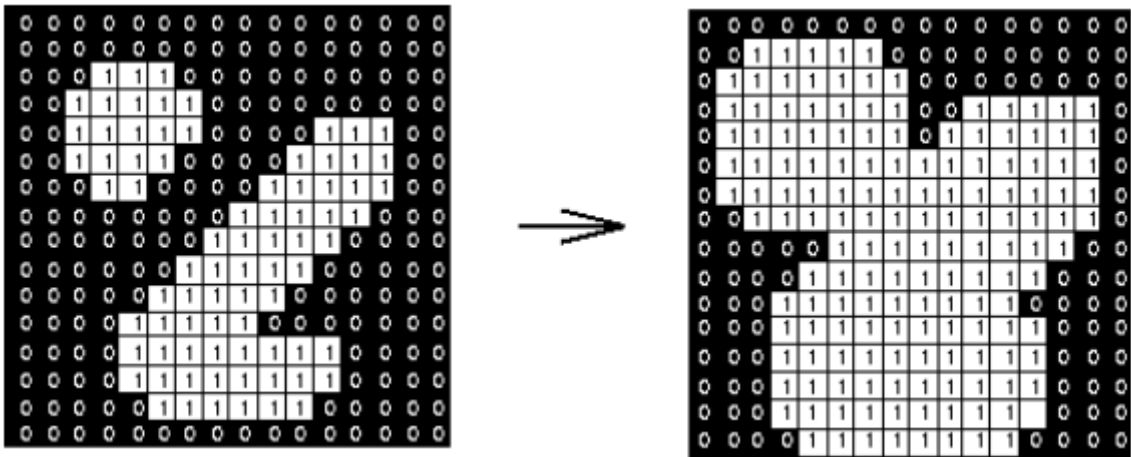
$$g_{out}(x, y) = \bigcup_{m, n} g_{in}(x + m, y + n) H(m, n)$$

Và định nghĩa của toán tử co ảnh là:

$$g_{out}(x, y) = \bigcap_m \bigcap_n g_{in}(x + m, y + n) H(m, n)$$

Trong đó: α có nghĩa là một dãy các toán tử OR, \square là một dãy các toán tử AND và (m, n) là tọa độ các điểm trong phần tử cấu trúc.

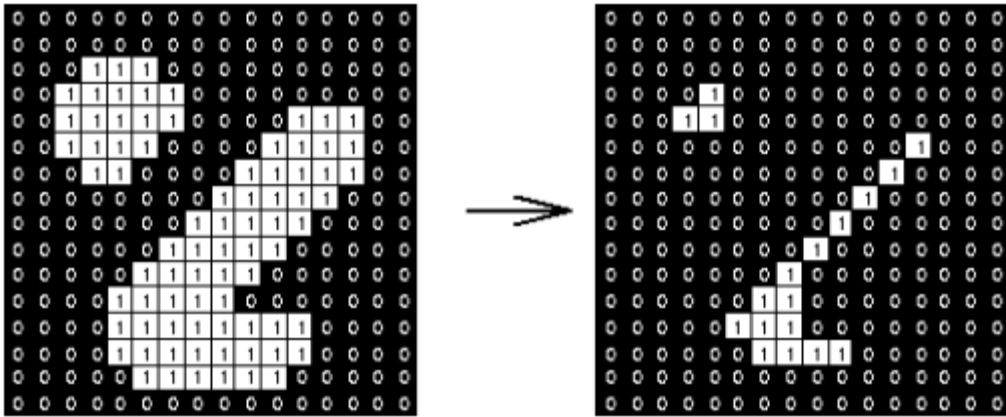
Hiệu ứng cơ bản của toán tử giãn ảnh trên ảnh nhị phân là sự mở rộng dần dần đường biên của các đối tượng ảnh (thường là các điểm ảnh màu trắng). Do đó kích thước của các đối tượng ảnh tăng lên trong khi lỗ hổng bên trong đối tượng và khoảng cách giữa các đối tượng thì giảm xuống. Mức độ giãn nở được quy định bởi tích chất của phần tử cấu trúc. Hình 2. 8 là ví dụ phép giãn ảnh với phần tử cấu trúc kích thước 3x3 (như mô tả trong Hình 2. 7).



Hình 2. 8 Phép giãn ảnh với phần tử cấu trúc 3x3

Trong khi đó phép co có hiệu ứng đối ngược lại, phép co ảnh làm cho các đường biên của đối tượng bị “xói 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. Mức độ bào mòn của các đối tượng cũng được quy định bởi tính chất của phần tử cấu trúc. Hình 2. 9 là một ví dụ của phép co ảnh với phần tử cấu trúc kích thước 3x3.

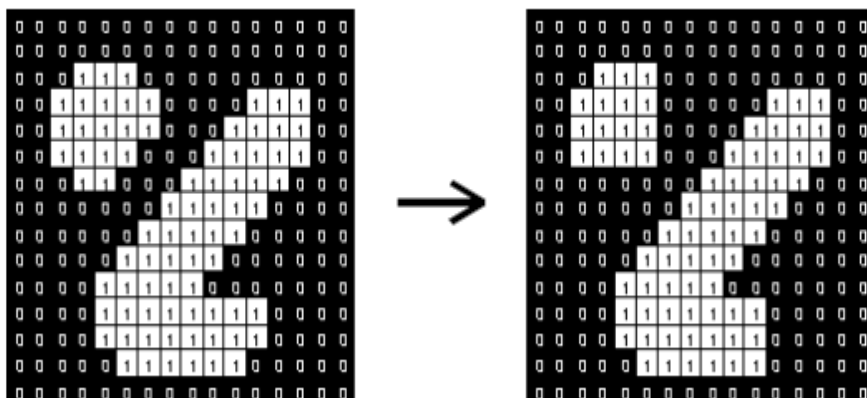


Hình 2. 9 Phép co ảnh với phần tử cấu trúc 3x3

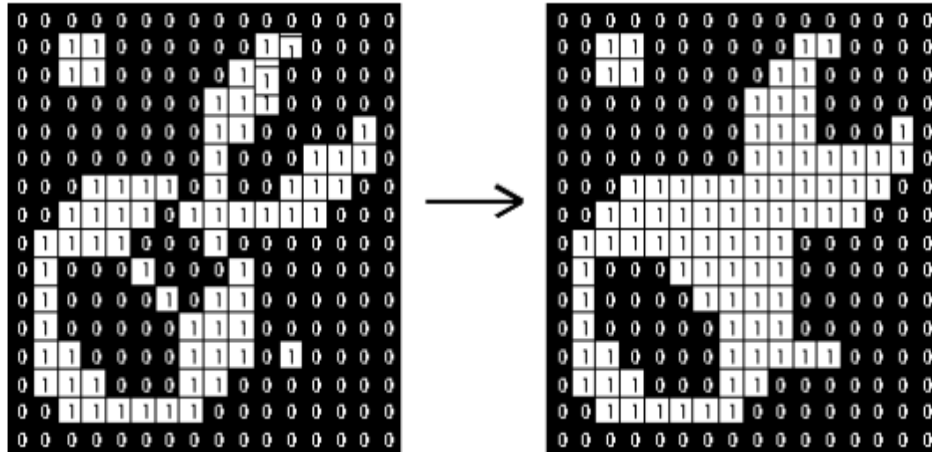
Nếu sử dụng các phép co ảnh và giãn ảnh một cách riêng lẻ thì sẽ làm mất đi các đặc trưng (hình dạng, kích thước) của ảnh. Do đó người ta thường kết hợp hai phép toán này với nhau, bằng cách: co bao nhiêu lần thì giãn bấy nhiêu lần và ngược lại. Một trong số những cách kết hợp đó được gọi là phép opening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

Còn closing được định nghĩa (giãn rồi mới co):

Hiệu ứng của hai phương pháp này tương tự như hiệu ứng của phương pháp co ảnh và giãn ảnh (opening tương đương với co ảnh còn closing tương đương với giãn ảnh) nhưng mức độ co/giãn thấp hơn. Hình 2. 10 và Hình 2. 11 minh họa điều này.



Hình 2. 10 Phép toán opening



Hình 2. 11 Phép toán closing

Toán tử opening sẽ xóa các điểm ảnh có kích thước nhỏ (nhỏ hơn hoặc bằng kích thước phần tử cấu trúc) trong khi vẫn giữ được các đặc trưng của các đối tượng trong ảnh. Không làm giảm kích thước của các đối tượng, chỉ xóa điểm ảnh là gai xung quanh viền đối tượng. Do đó nó thường được sử dụng để xóa nhiễu trong ảnh (các nhiễu hạt tiêu). Trong một số trường hợp nó sẽ xóa đi các liên kết “mảnh” giữa các đối tượng, ví dụ như chỗ dính nhau giữa các ký tự.

Toán tử closing thường dùng để nối các nét bị đứt trong đối tượng và lấp đầy các lỗ hổng bên trong đối tượng trong khi vẫn giữ được hình dạng và kích thước của đối tượng. Khi kích thước của phần tử cấu trúc lớn hơn khoảng cách giữa hai đối tượng, thì hai đối tượng này được nối với nhau nhưng hình dạng chung của khối (chứa hai đối tượng) không thay đổi. Dựa vào tính chất này để nối các ký tự trên cùng một dòng văn bản trong ảnh với nhau.

CHƯƠNG 3: PHƯƠNG PHÁP PHÂN ĐOẠN CHỨNG MINH NHÂN DÂN

3.1. Giới thiệu bài toán

Giấy chứng minh nhân dân hay chứng minh thư (CMND) 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. CMND được ban hành lần đầu tiên vào năm 1957 để thay thế cho thẻ căn cước [1], và cho đến nay đã qua nhiều lần thay đổi về nội dung. Theo quy định mới nhất năm 1999 [2] và được sửa đổi năm 2007 [3] thì CMND là hình chữ nhật (kích thước: 85,6 × 53,98 mm), hai mặt CMND in hoa văn màu xanh trắng nhạt và được ép nhựa trong. Các thông tin liên quan đến người được cấp CMND được dập hoặc in trên mẫu đã in sẵn. Tuy nhiên trên thực tế, những CMND được cấp từ trước tới nay lại có nội dung hoàn toàn khác với quy định ở trên, cụ thể như sau (Hình 3. 1):



a) Mặt trước

b) Mặt sau

Hình 3. 1 Mẫu “Giấy Chứng minh nhân dân”

Mặt trước: Có in hoa văn màu xanh nhạt. Bên trái, từ trên xuống: hình Quốc huy nước Cộng hòa xã hội chủ nghĩa Việt Nam; ảnh của người được cấp CMND. Bên phải, từ trên xuống: Cộng hòa xã hội chủ nghĩa Việt Nam; Độc lập – Tự do – Hạnh phúc; Giấy chứng minh nhân dân (màu đỏ); Số CMND (màu đỏ); Họ tên;

Ngày sinh; Nguyên quán; Nơi ĐKHK thường trú (hoặc Nơi thường trú). Mặt sau: Màu trắng, có cấu trúc dạng bảng. Ô trên cùng là Dân tộc và Tôn giáo. Bên trái

có hai ô: ô trên, vân tay ngón trỏ trái; ô dưới, vân tay ngón trỏ phải. Bên phải có hai ô: ô trên, dấu vết riêng và dị hình; ô dưới, Ngày cấp CMND; Nơi cấp

CMND; Họ tên người cấp; Đóng dấu và ký tên. CMND 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 CMND, 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 CMND 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:

+ Các trường thông tin có thể bị lệch so với dòng chuẩn hoặc chồm lên phần tiêu đề đã được in trước.

+ Nét chữ không đều nhau giữa các CMND, thậm chí là trong cùng một CMND: chữ quá đậm hoặc quá mờ.

+ CMND có thể bị ố, mốc, nhàu, nát,...

+ 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.

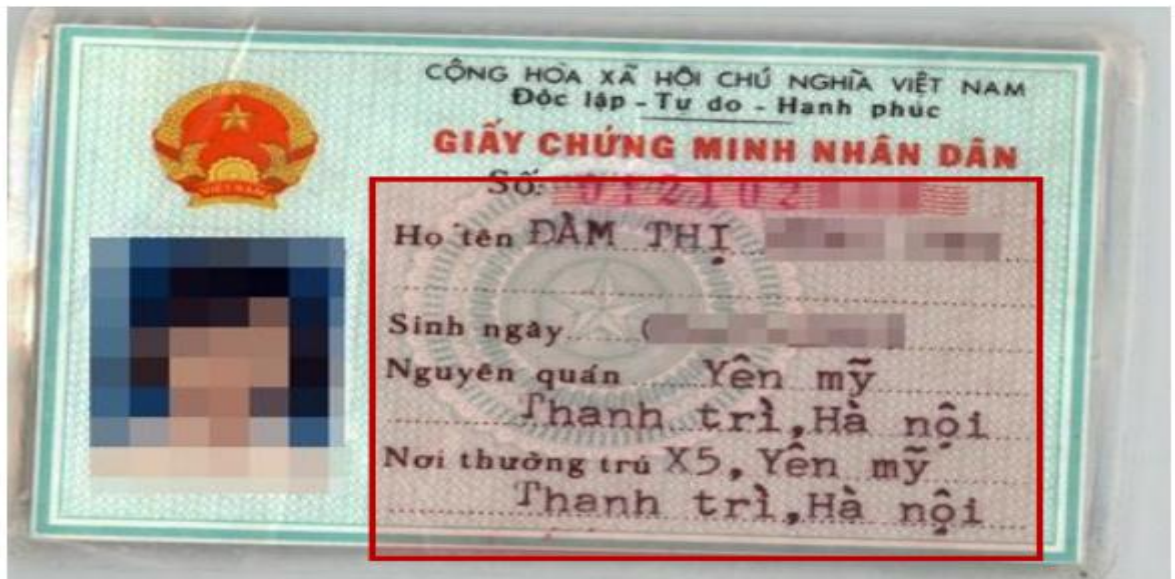
+ Ở mặt sau, trong khá nhiều trường hợp dấu màu đỏ và chữ ký đè cả lên phần

+ Ngày cấp và Nơi cấp, làm mờ đi một phần thông tin trên hai trường này.

Yêu cầu của bài toán: Từ ảnh CMND màu được quét bằng máy quét với độ phân giải 300 dpi, tách lấy các trường thông tin: Số CMND, 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 CMND, 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.

3.2. Tách các trường thông tin ở mặt trước



Hình 3. 2 Vùng thông tin cần tách ở mặt trước CMND

Các trường thông tin cần tách ở mặt trước bao gồm: Số CMND, Họ tên, Ngày sinh, Nguyên quán, Nơi thường trú (Hình 3. 2). Các trường này được đặt tuần tự từ trên xuống, ở bên phải của CMND, ngay dưới dòng “Giấy chứng minh nhân dân” (GCMND), có đặc điểm được mô tả trong Bảng 3. 1.

Bảng 3. 1 Đặc trưng các trường thông tin ở mặt trước CMND

Tên trường	Số dòng	Đặc trưng
Số CMND	1	Màu đỏ dập/in trên đường lượn sóng màu đỏ
Họ tên	2	Màu đen
Ngày sinh	1	Màu đen
Nguyên quán	2	Màu đen
Nơi thường trú	2	Màu đen

Để thấy rằng, trường Số CMND có màu đỏ nổi bật so với các trường khác nên có thể dựa vào đó để tách trường Số CMND trước, sau mới đó tách các trường còn lại dựa vào vị trí tương đối của nó so với trường Số CMND. Mặt khác, do ở mặt trước CMND 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 CMND: Dựa vào đặc điểm nổi bật riêng để tách, tiếp theo là loại bỏ các đường lượn sóng xong 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.

3.2.1. Tiền xử lý ảnh

Ảnh đầu vào thường có chất lượng rất khác nhau: có thể bị ố, mốc, nhàu, nát,... trong quá trình sử dụng hoặc bị lệch, nghiêng khi thu thập. Mặt khác, mặt trước có hoa văn nền khá rõ nét. Do đó, để đảm bảo cho việc tách các trường thông tin được chính xác (cũng như kết quả nhận dạng sau này), cần áp dụng các kỹ thuật xử lý ảnh để nâng cao chất lượng ảnh đầu vào. Các kỹ thuật đó bao gồm:

- . Chuyển từ ảnh màu về ảnh đa cấp xám
- . Làm trơn ảnh
- . Nhị phân ảnh
- . Căn chỉnh độ nghiêng

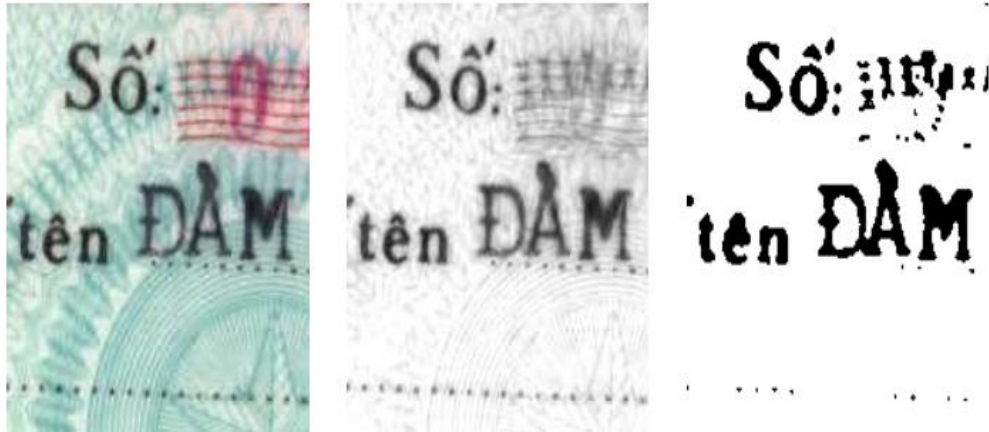
1/. Chuyển ảnh màu về ảnh đa cấp xám

Các phương pháp nhị phân ảnh thường được áp dụng trên ảnh đa cấp xám, do đó cần chuyển ảnh màu đầu vào về ảnh đa cấp xám trước khi chuyển sang bước nhị phân ảnh. Trong không gian màu RGB, mỗi điểm ảnh được tổ hợp từ ba thành phần màu R (Red), G (Green) và B (Blue) (mỗi thành phần màu có giá trị từ 0 đến 255), còn trong ảnh đa cấp xám, mỗi điểm ảnh mang một trị số mức xám từ 0 đến 255. Để chuyển đổi giữa hai không gian màu này, nếu áp dụng công thức tính độ sáng theo chuẩn NTSC (dựa trên thị giác của con người) thì ảnh thu được có hoa văn nền khá rõ nét (khá tương đồng với nét chữ).

Nhận thấy rằng, với những điểm sẫm 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 CMND lên trường Họ tên.



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

Hình 3. 3 Tiền xử lý ảnh mặt trước CMND

2/. Làm trơn ảnh

Ảnh đa cấp xám thu được ở bước trên thường không được “mịn” (vẫn còn “rấp”) 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. Áp dụng phương pháp làm trơn thích ứng:

$$F(x, y) = \frac{\sum_{i=-1, j=-1}^{i=1, j=1} \text{weight}(x+i, y+j) * I(x+i, y+j)}{\sum_{i=-1, j=-1}^{i=1, j=1} \text{weight}(x+i, y+j)}$$

Trong đó:

$$weight(x, y) = e^{-\frac{G_x^2 + G_y^2}{2 * factor}}$$

Với:

$$G_x(x, y) = \frac{I(x + 1, y) - I(x - 1, y)}{2}$$

$$G_y(x, y) = \frac{I(x, y + 1) - I(x, y - 1)}{2}$$

3/. Nhị phân ảnh

Bước tiếp theo là tìm ngưỡng để phân đoạn ảnh đa cấp xám ở trên về ảnh nhị phân (phân biệt rõ đâu là nền đâu là đối tượng). 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 IO.
- . Nhị phân ảnh bằng phương pháp Sauvola được ảnh IS.
- . Áp dụng toán tử AND đối với hai ảnh IO và IS để được ảnh nhị phân cần tìm:

$$IB = IO \text{ AND } IS$$

4/. Căn chỉnh độ nghiêng

Mặc dù công đoạn thu nhận ảnh được thực hiện bằng thiết bị chuyên dụng nhưng không thể tránh khỏi việc ảnh thu nhận được bị nghiêng. Do đó, để cho bước phân tích ảnh được chính xác thì ảnh sau khi phân đoạn phải được căn chỉnh độ nghiêng. Trong các phương pháp trình bày ở phần 2.2, nhận thấy rằng: Phương pháp dựa vào biến đổi Hough có thời gian tính toán lâu, nhất là khi áp dụng cho ảnh

có kích thước lớn do phải tính toán cho từng điểm ảnh một; Phương pháp láng giềng gần nhất không phù hợp đối với ảnh có chứa các ký tự Tiếng Việt có dấu. Do đó phương pháp được áp dụng ở đây là phương pháp sử dụng chiếu nghiêng. Phương pháp này cho kết quả tốt và thời gian thực hiện tương đối nhanh. Ảnh màu đầu vào còn được sử dụng để tách trường Số CMND, do đó việc hiệu chỉnh độ nghiêng được thực hiện trên cả ảnh nhị phân và ảnh màu dựa vào góc nghiêng xác định được trên ảnh nhị phân.

3.2.2. Tách trường Số CMND

Trong ảnh màu đã căn chỉnh độ nghiêng ở phần trên, trường Số CMND 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 CMND, trước tiên cần làm nổi bật các vùng này, sau đó tìm và tách vùng Số CMND, cuối cùng là loại bỏ các đường lượn sóng.

1/. Xác định vùng có thể là trường Số CMND

Để xác định các vùng có thể là trường Số CMND, trước tiên cần làm nổi bật những vùng có màu đỏ trong ảnh, sau đó tách lấy các vùng này. Những vùng này có giá trị của thành phần màu Red lớn hơn hẳn so với các thành phần màu khác, ngược lại ở vùng nền thì thành phần màu Green lại có giá trị lớn nhất. Dựa vào đặc trưng này để khử đi thành phần nền màu Green và làm nổi bật các vùng màu Red. Các bước của thuật toán như trong (Thuật toán 3.1).

Thuật toán 3.1. Xác định các vùng có thể là Trường Số CMND

Input: Ảnh màu đã chỉnh độ nghiêng: $I_{Color} = U(I_{Red}, I_{Green}, I_{Blue})$

Output: Các vùng có thể là Số CMND: $L = \{b_i\}$, với $b_i = (x_{b_i}, y_{b_i}, w_{b_i}, h_{b_i})$

Process:

- 1) Chuyển về ảnh đa cấp xám: $I_{Gray} = \text{Max}(I_{Red} - I_{Green}, 0)$
- 2) Áp dụng toán tử Closing 31×7
- 3) Nhị phân ảnh bằng phương pháp Otsu
- 4) Đảo ngược mức xám
- 5) Tách lấy các thành phần liên thông: L

Với mỗi điểm ảnh trong ảnh màu 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 màu (Hình 3. 4a). 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 CMND (Hình 3. 4b). Cuối cùng, tách lấy các thành phần liên thông và miêu tả chúng bằng hình bao chữ nhật của nó.



a) IRed – IGreen

b) Các vùng có thể là Số CMND

Hình 3. 4 Xác định vùng Số CMND

2/. Tìm và tách trường Số CMND

Trường Số CMND 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 CMND có thể bị nhoè lên dòng “GCMND”. Mặt khác, trong quá trình sử dụng trường Số CMND 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 CMND. Như vậy, khó có thể tìm trực tiếp trường Số CMND 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 CMND nên có thể dựa vào dòng này để xác định trường Số CMND (Hình 3. 4b). Các bước cụ thể của thuật toán được mô tả trong Thuật toán 3.2.

Thuật toán 3.2. Tìm và tách trường Số CMND**Input:**

Các vùng có thể là Số CMND: $L = \{b_i\}$, với $b_i = (x_{bi}, y_{bi}, w_{bi}, h_{bi})$

Ảnh màu thu được từ bước tiền xử lý: IColor

Output: Ảnh màu vùng Số CMND: INumber

Process:

1. Tìm dòng “GCMND”: $b_{Title} = \{b_i \mid x_{bi} \rightarrow \max\}$

2. \rightarrow vùng Số CMND: bNumber

a. Nếu $h_{b_{Title}} > h_{b_{MaxTitle}} \rightarrow$ tách bNumber từ bTitle

b. Ngược lại, Tìm bNumber trong L (nằm ngay dưới bTitle)

3. Tách lấy vùng Số CMND: INumber = Clip(IColor, bNumber)

Trong đó, $h_{b_{MaxTitle}}$ là chiều cao lớn nhất có thể của dòng “GCMND”, được xác định dựa vào chiều cao trung bình của các ký tự. Với trường hợp hai dòng dính nhau, dòng Số CMND có chiều dài ngắn hơn dòng “GCMND”, dựa vào hình chiếu dọc và ngang để tách riêng hai dòng này. Còn trường hợp không tìm thấy vùng nào nằm ngay dưới vùng có kích thước lớn nhất, có nghĩa là vùng Số CMND đã bị mờ mất. Sau khi xác định được vùng Số CMND, tách lấy vùng ảnh tương ứng trên ảnh màu đầu vào (đã được căn chỉnh độ nghiêng) để chuyển sang bước tiếp theo Hình 3. 5a.

3/. Phân đoạn vùng Số CMND

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, thuật toán cụ thể như trong (Thuật toán 3.3).

Thuật toán 3.3. Phân đoạn vùng Số CMND

Input: Ảnh màu vùng Số CMND: IColor = IRed \times IGreen \times IBlue

Output: Ảnh nhị phân: IBin2

Process:

1. Chuyển về ảnh đa cấp xám: $I_{\text{Gray1}} = I_{\text{Green}}$
2. Nhị phân ảnh: $I_{\text{Bin1}} = \text{Sauvola}(I_{\text{Gray1}}, w=15, k=0.0125)$
3. Ước lượng bề dày đường lượn sóng: *thick*
4. Làm mờ đường lượn sóng: $I_{\text{Gray2}} = \text{Median}(I_{\text{Gray1}}, \text{thick}+1, \text{thick}+3)$
5. Nhị phân ảnh: $I_{\text{Bin2}} = \text{Sauvola}(I_{\text{Gray2}}, w=15, k=0.1)$



Hình 3. 5 Phân đoạn vùng Số CMND

Để ước lượng được bề dày đường lượn sóng cần phân đoạn “thô” sao cho vẫn giữa 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 (Thuật toán 3.4). Từ đó làm mờ đường lượn sóng bằng phép lọc trung vị (median filter), cuối cùng là phân đoạn “tinh” để khử đi các thành phần nền (Hình 3. 5).

Thuật toán 3.4. Ước lượng bề dày đường lượn sóng

Input: Ảnh nhị phân vùng Số CMND: $I_{\text{Bin1}}(w, h)$

Output: Bề dày đường lượn sóng: *thick*

Process:

1. Khởi tạo mảng: $count[1..N]$
(N : bề dày lớn nhất có thể của đường lượn sóng, $N=5$)
2. “Đo” bề dày đường lượn sóng:

```

for x=1 to w
  t←0
  for y=1 to h
    if ( $I_{Bin1}(x, y)$ ) t++
    else count[t]++, t←0
  count[t]++

```
3. $thick = arg(count[t] > 100 \mid t \rightarrow max)$

3.2.3. 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 CMND, 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 cửa sổ “mặt nạ” của các dòng sau đó cố gắng lọc lấy các đối tượng (ký tự) thuộc mặt nạ dòng. Cụ thể, thuật toán bao gồm các bước sau:

- Tìm mặt nạ dòng: Xác định vị trí của các dòng.
- Tách các đối tượng thuộc mỗi dòng: Phân tích các thành phần liên thông để tìm các đối tượng thuộc mặt nạ dòng.
- Xoá phần tiêu đề và nhiễu: Loại bỏ phần tiêu đề của từng trường thông tin và các đối tượng là nhiễu, dòng không có ký tự.
- Lấy lại các ký tự bị mất: thuộc dòng nhưng không được xét thuộc mặt nạ dòng.

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, biểu diễn 7 dòng này bằng 7 hình chữ nhật – gọi là mặt nạ dòng – (xem Hình 3. 6), trong đó:

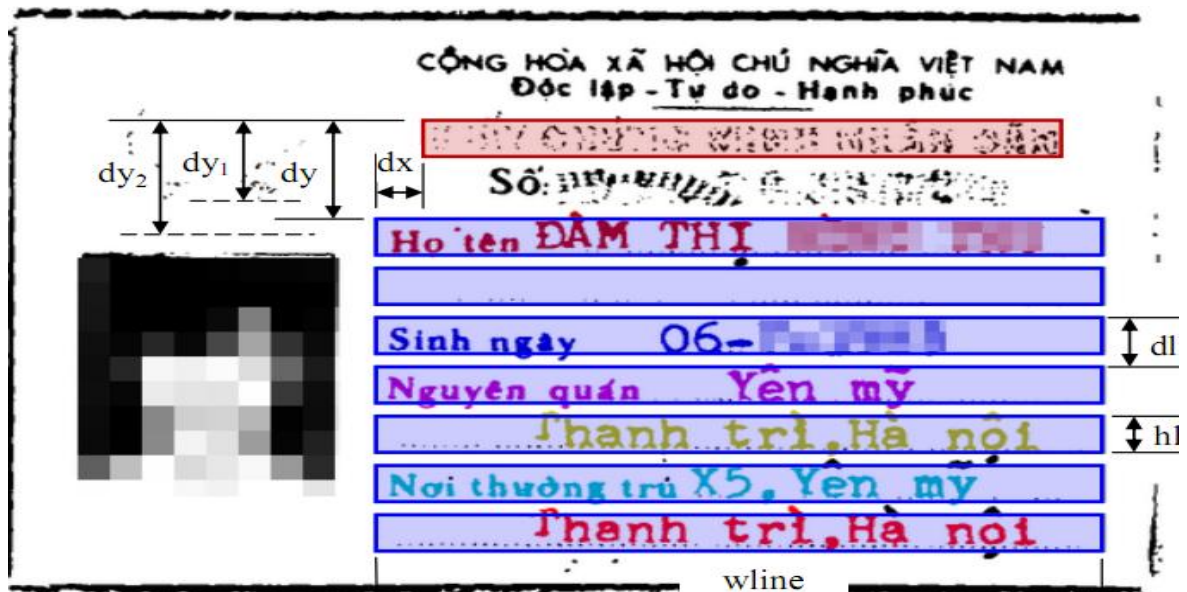
+ **dx, dy**: khoảng cách giữa vùng mặt nạ và dòng tiêu đề “GCMND” theo phương ngang và phương dọc, tương ứng. Trong đó, dx là cố định (vì được in theo mẫu), chỉ có dy là thay đổi do dòng thông tin đập/in vào có thể thay đổi so với dòng cơ sở (dịch lên hoặc dịch xuống): $dy \in [dy1, dy2]$.

+ **wline, hline, dline**: kích thước và khoảng cách giữa các hình chữ nhật là cố định.

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

$$dy = \arg\{\max_{y \in [dy_1, dy_2]} (\text{count}(y))\}$$

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



Hình 3. 6 Mật nạ dòng mặt trước

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), mỗi đối tượng được biểu diễn bằng hình chữ nhật bao quanh nó, trong đó:

xb, yb: Toạ độ điểm trái trên của đối tượng

wb, hb: Kích thước của đối tượng

Để việc lọc các đối tượng được chính xác, chia các đối tượng 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, nhiều,... ($wb < wmin$ hoặc $hb < hmin$).

Lớn: Lớp các đối tượng là ảnh chân dung hoặc các đường viền ($w_b > w_{max}$ hoặc $h_b > h_{max}$).

Vừa: Gồm các đối tượng còn lại, lớp các đối tượng có thể là ký tự.

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. Tuy nhiên, trong danh sách này vẫn có các ký tự là tiêu đề của các trường, các nhiễu, hoặc thiếu các ký tự của phần thông tin, do đó cần các bước hậu xử lý tiếp theo.

Thuật toán 3.5. Tách các ký tự thuộc mỗi dòng

Input:

1. Mặt nạ dòng: $LMask = \{bMaskk\}$
2. Ảnh nhị phân: $IBin$

Output: Danh sách các ký tự thuộc các dòng: $LLine = \{LChark\}$

Process:

1. Tách lấy các thành phần liên thông từ $IBin$: $LCCs = \{bCCsi\}$
2. Chia $LCCs$ là 3 lớp:
 - a. Nhỏ: $LSmall = \{bCCsi \mid wbCCsi < w_{min} \text{ OR } hbCCsi < h_{min}\}$
 - b. Lớn: $LLarge = \{bCCsi \mid wbCCsi > w_{max} \text{ OR } hbCCsi > h_{max}\}$
 - c. Vừa: $LMean = \{bMeani\} = \{bCCsi \mid bCCsi \square LSmall \text{ AND } bCCsi \square LLarge\}$

3. Lọc lấy các ký tự thuộc mỗi dòng:

$$LChark = \{bMeani \mid (xMeani, yMeani) \square bMaskk\}$$

Trong đó:

$$xMeani = xbMeani + wbMeani / 2$$

$$yMeani = ybMeani + hbMeani / 2$$

3/. Xoá tiêu đề

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.

Từ các đặc điểm trên có thể tính được hàm “khoảng cách” giữa phần tiêu đề và phần thông tin:

$$\text{distance}() = dh + dy + dd + dc$$

Trong đó:

- . +dh: Chênh lệch độ cao trung bình của phần tiêu đề và phần thông tin.
- . + dy: Tổng chênh lệch của đường baseline và đường mean line giữa phần tiêu đề và phần thông tin.
- . + dd: Khoảng cách giữa phần tiêu đề và phần thông tin.
- . + dc: Chênh lệch độ cao của hai ký tự tiếp giáp giữa phần tiêu đề và phần thông tin. Như vậy, có thể dựa vào hàm khoảng cách này để xác định vị trí phân tách giữa phần tiêu đề và phần thông tin (Thuật toán 3.6).

Thuật toán 3.6. Xoá phần tiêu đề

Input:

1. Mặt nạ dòng: LChar = {bChark}
2. Độ dài có thể của phần tiêu đề: [dTitle1, dTitle2]

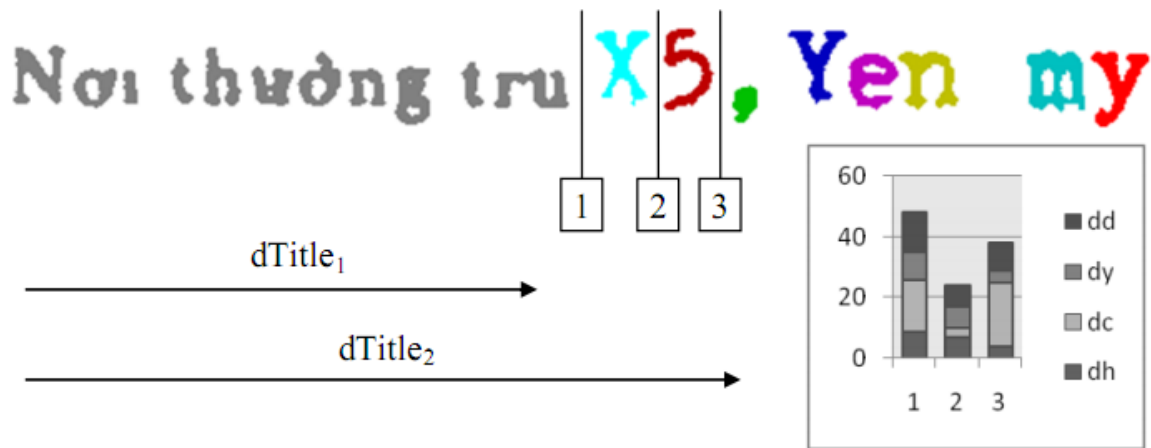
Output: Danh sách các ký tự thuộc phần thông tin: LInfo = {bInfok}

Process:

1. Sắp xếp LChar theo chiều tăng dần của xbChari
2. Tìm các vị trí có thể phân tách:

$$L_{Split} = \{xSplit_i \mid xSplit_i \in [dTitle_1, dTitle_2] \text{ AND} \\ xSplit_i \in [xbChar_k + wbChar_k, xbChar_{k+1}]\}$$

3. Tính hàm khoảng cách: $distance(xSplit_i)$
4. Chọn vị trí phân tách: $split = \arg\{\max(distance(xSplit_i))\}$
5. Xoá phần tiêu đề: $LInfo = \{bChark \mid xbChark > split\}$



Hình 3. 7 Xoá phần tiêu đề

4/. 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. 8a).

. + 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. 8b).

Nơi ĐKKH thường trú: 31 Nam Kỳ,
 Nam Ngạn, TP T-Hoa, Thanh Hoa

a) Năm ngoài vùng mặt nạ

Nơi thường trú Gia Lập
 Gia Viên, Ninh Bình

b) Hai ký tự dính nhau

Hình 3. 8 Kết quả lấy lại các ký tự bị mất

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.

3.3. Tách các trường thông tin ở mặt sau



Hình 3. 9 Các vùng thông tin cần tách ở mặt sau

Mặt sau của CMND 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 (Hình 3. 9). 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.

+Xác định cấu trúc bảng: Tìm các đường kẻ ngang/dọc và tạo lại cấu trúc bảng.

+Tách các trường thông tin: Tách lấy các dòng thông tin và loại bỏ phần tiêu đề của mỗi dòng.

3.3.1. Tiền xử lý ảnh

Mặt sau của CMND là đơn giản hơn mặt trước, không có hoa văn nền phức tạp mà chỉ là chữ đen trên nền trắng. Tuy nhiên lại có dấu màu đỏ, dấu này có thể đề lên trường Ngày cấp và Nơi cấp. Mặt khác, trong quá trình sử dụng ảnh cũng bị suy thoái chất lượng: bị ố, mốc,... hay bị nghiêng giống như mặt trước. Do đó, cần các thao tác tiền xử lý ảnh để khử đi các thành phần nền và dấu màu đỏ cũng như căn chỉnh độ nghiêng

(Hình 3. 10), cụ thể gồm các thao tác sau:

+Chuyển ảnh màu 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.

3.3.2. Xác định cấu trúc bảng

Trong phần này sẽ tách lấy các vùng thông tin yêu cầu từ ảnh nhị phân thu được ở bước trước thông qua việc xác định cấu trúc của 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 3.7. Tìm các đường kẻ ngang trong ảnh

Input: Ảnh nhị phân: IBin

Output: Danh sách các đường kẻ ngang trong ảnh: LLine = {yLinek}

Process:

1. Áp dụng các phép biến đổi hình thái đối với IBin:

a. erosion 15x1: Xoá đường kẻ dọc và nét bút của ký tự.

b. dilation 1x3: Làm dày đường kẻ ngang.

c. closing 51x1: Nối đường ngang bị đứt.

2. Tách lấy các thành phần liên thông: LCCs = {bCCSi}

3. □ Các đường kẻ ngang trong ảnh:

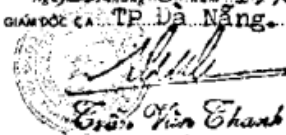
$$L_{Line} = \{y_{bCCS_i} \mid bCCS_i \in L_{CCs} \text{ AND } w_{bCCS_i} > w_{image} / 4\}$$

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 (Thuật toán 3.7). 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 gần nhau của đường cơ sở hoặc đường gạch chân của chữ ký,... (Hình 3. 11b)

+ 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 kẻ ngang còn thiếu (Hình 3. 11c).

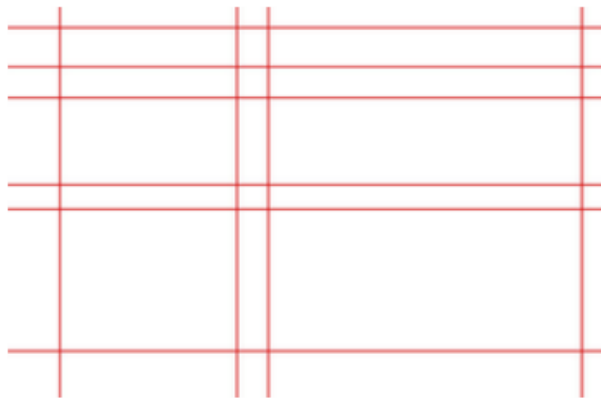
+Sau khi xác định được các đường kẻ ngang và dọc của bảng (tức là xác định được cấu trúc của bảng), tách lấy các vùng chứa các trường thông tin cần tìm từ ảnh: vùng Dân tộc nằm giữa đường kẻ ngang thứ nhất và thứ hai; vùng Ngày cấp và Nơi cấp nằm giữa đường kẻ ngang thứ tư và thứ sáu, và đường kẻ dọc thứ ba và thứ bốn (Hình 3. 11d). Từ các vùng này, bước tiếp theo sẽ tách lấy các trường thông tin cần tìm.

Dân tộc: kinh		Tôn giáo: không	
DẤU VẾT RIÊNG HOẶC DỊ HÌNH			
NGÓN TRÁI		Nốt ruồi C. 2cm, trên trước mép trái.	
		Ngày 13 tháng 8 năm 1998 GIÁM ĐỐC CA TP. Đà Nẵng.	
NGÓN PHẢI			
		Trần Văn Chánh	

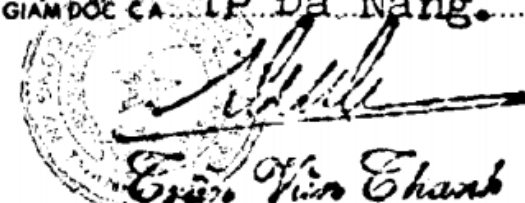


a) Ảnh nhị phân đầu vào

b) Các đường kẻ ngang trong ảnh



c) Các đường kẻ của bảng

Dân tộc: kinh	
Ngày 13 tháng 8 năm 1998 GIÁM ĐỐC CA TP. Đà Nẵng.	
	
Trần Văn Chánh	

d) Các vùng thông tin

Hình 3. 11 Xác định cấu trúc bảng

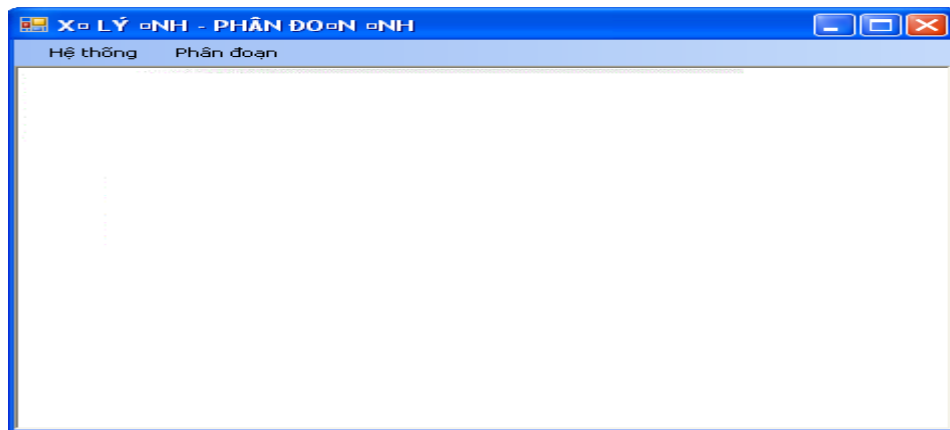
3.3.3. Tách trường thông tin

Việc tách cách trường thông tin ở mặt sau cũng tương tự như đối với mặt trước.

CHƯƠNG 4: CÀI ĐẶT THỬ NGHIỆM

4.1. Hệ thống chương trình gồm có các phần.

- Giao diện chính của chương trình.
- Load ảnh
- Phân đoạn ảnh Tách Cây Tứ Phân.
- Phân đoạn ảnh theo Single-Link.
- Phân đoạn ảnh theo K-Means.



4.2. Giao diện phân đoạn ảnh dựa vào phương pháp tách cây tứ phân.



4.3. Giao diện phân đoạn ảnh dựa vào phương pháp Single-Link



4.4. Giao diện phân đoạn ảnh dựa vào phương pháp K-Means



KẾT LUẬN

5.1 Nội dung của đề án

5.1.1 Các kết quả đạt được

Trong quá trình nghiên cứu tài liệu và thực hiện đề án dưới sự định hướng của thầy hướng dẫn em thấy bản thân mình đã đạt được một số kết quả sau:

Tìm hiểu được một cách tổng quan các vấn đề về XLA và phân đoạn ảnh hay tiền xử lý ảnh .Em đã có một cách nhìn có hệ thống về các phương pháp trong phần phân đoạn ảnh và tiền xử lý ảnh .Đồng thời biết được điểm mạnh ,yếu của từng phương pháp và có thể đưa ra cách lựa chọn phương pháp phù hợp với từng loại ảnh.

Ngoài ra,trong quá trình nghiên cứu em cũng tự tích lũy thêm cho mình các kiến thức về toán học ,về kỹ thuật lập trình ..Và quan trọng hơn là rèn luyện kỹ năng để thực hiện một nghiên cứu khoa học.

5.1.2 Một số hạn chế cần khắc phục

Bên cạnh những kết quả đạt được em tự thấy trong đề án vẫn còn một số hạn chế .

Chưa đưa ra được một phương pháp phân đoạn mới hoàn toàn .Trong khuôn khổ một đề án tốt nghiệp ,em mới chỉ trình bày lại các kiến thức tìm hiểu được chứ chưa đề xuất một phương pháp hoàn toàn mới.

Do thời gian có hạn nên trình bày các thuật toán phân đoạn cũng chưa được hệ thống và khoa học.có nhiều thuật toán được trình bày sơ lược .

Đề án cũng chưa chỉ ra được các ứng dụng thực tế của các thuật toán phân đoạn.

5.2 Công việc tiếp theo

Dựa trên những kết quả bước đầu đạt được trong đề án ,em có đề xuất một số cải tiến thuật toán Trong tương lai.

Hầu hết các thuật toán được xử lý trên ảnh tĩnh em hi vọng có thể tìm hiểu được và phát triển thêm các thuật toán ứng dụng trên ảnh động hay video

Bên cạnh đó em cũng sẽ tiếp tục tìm hiểu thêm những thuật toán mới hơn trong lĩnh vực phân đoạn ảnh hay tiền xử lý ảnh.

TÀI LIỆU THAM KHẢO

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 Geometric 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.