

Ts. NGUYỄN KIM SÁCH

XỬ LÝ ẢNH VÀ VIDEO SỐ

(Digital Processing of picture and video
by D.Sc., Ph. D. Nguyen Kim Sach)



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT
HÀ NỘI - 1997

Ts. NGUYỄN KIM SÁCH

XỬ LÝ ẢNH VÀ VIDEO SỐ

(Digital Processing of picture and video
by D.Sc., Ph. D. Nguyen Kim Sach)



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT
HÀ NỘI - 1997

Lời nói đầu

Từ thập niên 20, xử lý ảnh (Image processing) đã được đề cập đến, nhằm nâng cao chất lượng ảnh và xử lý số liệu. Vấn đề xử lý ảnh thật sự mang lại ý nghĩa thực tế phải kể đến sự kiện ảnh chụp mặt trăng từ vệ tinh vào năm 1964. Các bức ảnh chụp này được truyền về quả đất, được xử lý bằng máy tính (sửa méo, lọc nhiễu, làm nổi đường biên, ...). Xử lý ảnh cho phép khôi phục chất lượng ảnh tốt và trung thực. Từ đó đến nay, xử lý ảnh được phát triển không ngừng trên cơ sở công nghệ tin học phát triển và được ứng dụng trong hầu hết các lĩnh vực (truyền ảnh, truyền chữ, truyền hình, nhận dạng chữ viết và vân tay, máy cắt lớp trong y học, viễn thám, quân sự, nội vụ, nghiên cứu khoa học, ...). Ngày nay xử lý ảnh giải quyết nhiều vấn đề như nén dữ liệu ảnh, nhằm tiết kiệm dung lượng bộ nhớ và tận dụng hiệu quả kênh truyền. Nổi bật nhất là nén video (nén ảnh động) trong hệ thống truyền hình số (một kênh thông thường có thể truyền được nhiều đường video), truyền hình có độ phân giải cao HDTV / ATV (từ 1,5 Gbit/s có thể nén xuống còn 17 - 24 Mbit/s để truyền trên một kênh truyền hình thông thường, mà chất lượng hình ảnh HDTV / ATV cao hơn rất nhiều so với truyền hình thông thường NTSC, PAN, SECAM).

"Xử lý ảnh và video số" đề cập đến vấn đề xử lý ảnh tĩnh (still picture) và ảnh động (moved picture hay video), là vấn đề rất rộng, bao gồm nhiều hướng nghiên cứu. Những năm gần đây ở Việt nam, nhiều người đã quan tâm đến vấn đề này (sinh viên, cao học, nghiên cứu sinh, v.v. ở Hà nội và TP. Hồ Chí Minh) và đã có một số công trình được công bố về xử lý ảnh. Tuy nhiên vấn đề xử lý ảnh và đặc biệt là xử lý video (xử lý ảnh động) còn là mới mẻ đối với một số đông người, mặc dù môn xử lý ảnh đã bắt đầu được đưa vào các trường đại học.

Tài liệu này có thể giúp bạn đọc tìm hiểu và nghiên cứu vấn đề xử lý ảnh (tĩnh và động), mở rộng kiến thức về xử lý ảnh. Tài liệu cũng thích hợp cho sinh viên, kỹ sư, cán bộ khoa học, nghiên cứu sinh làm việc và nghiên cứu trong lĩnh vực xử lý ảnh.

Tài liệu chắc không tránh khỏi thiếu sót. Chúng tôi mong được bạn đọc đóng góp ý kiến. Thu góp ý xin gửi về Nhà xuất bản KHKT, 70 Trần Hưng Đạo Hà nội.

Tác giả

TS. NGUYỄN KIM SÁCH

1. MỞ ĐẦU

1.1. Lịch sử về xử lý ảnh

Các phương pháp xử lý ảnh bắt nguồn từ hai ứng dụng : nâng cao chất lượng thông tin hình ảnh đối với mắt người và xử lý số liệu cho máy tự động. Một trong những ứng dụng đầu tiên xử lý ảnh là nâng cao chất lượng ảnh báo truyền qua cáp giữa London và New York vào những năm 1920. Thiết bị đặc biệt mã hóa hình ảnh (báo), truyền qua cáp và khôi phục lại ở phía thu.

Vấn đề nâng cao chất lượng hình ảnh (số) lúc đầu có liên quan đến việc lựa chọn quá trình in và phân bố các mức sáng (tông và độ phân giải của ảnh).

Hệ thống đầu tiên (Bartlane) có khả năng mã hóa hình ảnh với 5 mức sáng. Khả năng này tăng lên 15 mức vào 1929.

Việc nâng cao chất lượng ảnh bằng các phương pháp xử lý để truyền ảnh (số) được liên tục nghiên cứu 35 năm sau đó. Do kỹ thuật máy tính phát triển, nên xử lý hình ảnh ngày càng phát triển. Năm 1964, các bức ảnh chụp mặt trăng được vệ tinh Ranger 7 (Mỹ) truyền về trái đất, được xử lý bằng máy tính để sửa méo (gây ra do camera truyền hình đặt trên vệ tinh ở các góc độ khác nhau). Các kỹ thuật cơ bản cho phép nâng cao chất lượng hình ảnh như làm nổi đường biên và lưu hình ảnh.

Từ 1964 đến nay, phạm vi xử lý ảnh và video (ảnh động) lớn mạnh không ngừng. Các kỹ thuật xử lý ảnh số (digital image processing) hiện nay được sử dụng để giải quyết hàng loạt vấn đề, nhằm nâng cao chất lượng thông tin hình ảnh.

Trong y học, các thuật toán máy tính nâng cao độ tương phản, hoặc mã hóa các mức sáng thành các màu để nội suy ảnh x quang (x - rays) và các hình ảnh y sinh học dễ dàng. Các nhà địa vật lý sử dụng kỹ thuật tương tự để nghiên cứu các mẫu vật chất từ ảnh vệ tinh. Các thuật toán làm nổi đường biên (image enhancement) và khôi phục hình ảnh (image restoration) được sử dụng để xử lý hình ảnh bị giảm chất lượng. Trong thiên văn học, các phương pháp xử lý ảnh nhằm khôi phục hình ảnh bị nhiễu hoặc bị mất do bóng (artifacts) sau khi chụp. Trong vật lý và các lĩnh vực có liên quan, kỹ thuật máy tính nâng cao được chất lượng ảnh trong các lĩnh vực như plasmas (có năng lượng cao) và microscopy điện tử. Tương tự, người ta đã ứng dụng xử lý ảnh có kết quả tốt trong trong viễn thám, sinh học, y tế hạt nhân, quân sự, công nghiệp ... Nâng cao chất lượng (làm nổi đường biên) và khôi phục ảnh bị nhiễu là quá trình xử lý ảnh dùng cho mục đích nội suy của mắt người. Lĩnh vực ứng dụng quan trọng thứ hai là xử lý ảnh số gắn liền với việc cảm nhận của máy. Trong lĩnh vực thứ hai, các cố gắng đều tập trung vào

các quá trình trích thông tin ảnh và chuyển thành dạng thích hợp cho xử lý máy tính. Ví dụ như thông tin dùng cho máy tính là các moments thống kê, các hệ số biến đổi Fourier và đo khoảng cách nhiều chiều.

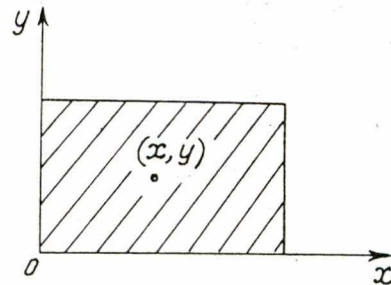
Những vấn đề tiêu biểu của kỹ thuật xử lý ảnh (tĩnh) và video (ảnh động) được ứng dụng nhiều trong thực tế, có thể kể như : tự động nhận dạng đặc trưng, máy nhìn công nghiệp để điều khiển và kiểm tra sản phẩm, nhận dạng mục tiêu quân sự, tự động xử lý vân tay, hiển thị lên màn hình ảnh x quang và các mẫu máu, xử lý bằng máy các hình ảnh chụp từ vệ tinh để dự báo thời tiết, nén ảnh (tĩnh) và video (ảnh động) để lưu và truyền được nhiều hơn tín hiệu ảnh và video số trong thông tin, máy tính, truyền hình thông thường và truyền hình có độ phân giải cao HDTV/ATV.

1.2. Biểu diễn ảnh số

Đối với ảnh đen trắng (gọi là ảnh đơn giản), thì ảnh được biểu diễn bằng một hàm cường độ sáng hai chiều $f(x, y)$, trong đó x, y là các giá trị tọa độ không gian và giá trị của f tại bất kỳ một điểm (x, y) sẽ tỉ lệ với độ sáng (hoặc mức xám) của ảnh tại điểm này (hình 1.1).

Đôi khi người ta còn biểu diễn hàm ảnh với một trục thứ 3 là cường độ sáng (trong hình 1.1, trục thứ 3 bằng 0).

Một ảnh số là một ảnh $f(x, y)$ được gián đoạn theo không gian và độ sáng. Một ảnh số được xem như một ma trận với hàng và cột biểu diễn một điểm trong ảnh và giá trị điểm ma trận tương ứng mức xám tại điểm đó. Các phần tử của một dãy số như thế (digital array) được gọi là các điểm ảnh (image elements, picture elements, pixels, pels).



Hình 1.1. Biểu diễn ảnh bằng hàm $f(x, y)$.

Kích thước của một ảnh số thay đổi theo mục đích sử dụng. Các chương sau đây sẽ trình bày ưu điểm của việc chọn kích thước hình vuông (Square array) và số lượng mức xám là số nguyên tố bậc 2 (integer). Ví dụ một ảnh có chất lượng ảnh truyền hình đen trắng sẽ có kích thước 512×512 array với 128 mức xám ($2^7 = 128$).

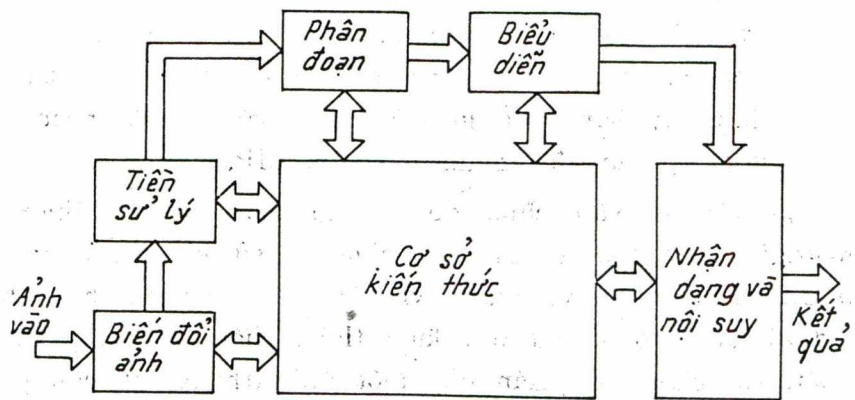
1.3. Cơ sở về xử lý ảnh

Xử lý ảnh số bao gồm một phạm vi rộng về phần cứng (thiết bị), phần mềm (chương trình máy tính) và các cơ sở lý thuyết. Trong mục này sẽ đề cập đến các bước cơ bản cần thiết cho xử lý ảnh.

Hình 1.2 trình bày các bước cơ bản trong xử lý ảnh số.

Phạm vi vấn đề ở đây bao gồm các mẫu thư tín và chủ đề là đọc địa chỉ trên mỗi mẫu thư. Cho nên, đầu ra ở đây là một dòng các ký tự (chữ cái).

Bước đầu tiên là thu ảnh (image acquisition), thu ảnh số bằng bộ cảm biến ảnh (imaging sensor) với khả năng số hóa tín hiệu của bộ cảm biến. Bộ cảm biến có thể là một camera truyền hình màu hoặc đen trắng ; nó tạo một ảnh đầy đủ trong thời gian 1/30 giây (một ảnh truyền hình hệ NTSC là 1/30 giây, một ảnh truyền hình hệ PAL là 1/25 giây). Bộ cảm biến ảnh có thể là một camera quét dòng, tạo một



Hình 1.2 Các bước cơ bản trong xử lý ảnh số.

dòng ảnh riêng trong một đơn vị thời gian. Trong trường hợp này, vật thể chuyển động qua bộ quét dòng, sẽ tạo một ảnh hai chiều. Nếu đầu ra của camera (hoặc bộ cảm biến ảnh) không cho ở dạng số, thì phải dùng bộ biến đổi tương tự - số (ADC) để biến đổi ảnh tương tự thành ảnh số.

Sau khi nhận được một ảnh số (digital image), bước tiếp theo là tiền xử lý ảnh (image preprocessing). Chức năng chủ yếu của tiền xử lý là nâng cao khả năng để các quá trình tiếp theo đạt kết quả tốt. Trong ví dụ trên, tiền xử lý là nâng cao độ tương phản, khử nhiễu và tách các vùng có đặc trưng giống thông tin chữ cái của bức điện (bức thư).

Bước tiếp theo là phân đoạn (segmentation). Phân đoạn ảnh là tách một ảnh đầu vào thành các thành phần hoặc các vật thể. Nói chung, tự động phân đoạn ảnh là một trong những nhiệm vụ khó nhất trong xử lý ảnh số. Một mặt thuật toán phân đoạn ảnh thô là quá trình nâng cao độ phân giải ảnh. Mặt khác, các thuật toán yếu (hoặc gián đoạn) sẽ gây ra sự mất mát. Trong nhận dạng ký tự, vai trò chính của phân đoạn ảnh là trích riêng các ký tự và các từ ra khỏi nền ảnh.

Đầu ra của quá trình phân đoạn ảnh thường là số liệu pixel chưa lọc, bao gồm cả các liên kết (lân cận) của một vùng hoặc tất cả các điểm ảnh trong vùng đó. Trong cả hai trường hợp, thì việc biến đổi số liệu thành dạng thích hợp cho xử lý bằng máy tính là cần thiết. Số liệu cần được biểu diễn như là một liên kết (boundary) hoặc là một vùng tổng hợp. Biểu diễn liên kết là một xấp xỉ, nếu tập trung vào các đặc trưng hình dạng bên ngoài, như là các góc hoặc đường cong. Còn biểu diễn vùng là xấp xỉ, nếu chú ý đến các đặc tính bên trong, ví dụ như đặc trưng hoặc dạng chủ yếu. Trong một vài ứng dụng, cả hai loại biểu diễn này cùng tồn tại. Đó là các trường hợp nhận dạng ký tự.

Việc lựa chọn một loại biểu diễn chỉ là phân kết quả của việc biến đổi số liệu thô thành dạng thích hợp cho xử lý bằng máy tính. Sự miêu tả (description), gọi là chọn đặc trưng (feature selection), gắn liền với việc tách (trích) đặc trưng, cho kết quả dưới dạng thông tin định lượng hoặc các đặc trưng mà nó là cơ sở để phân biệt một lớp các vật thể từ các lớp khác. Trong nhận dạng ký tự, việc miêu tả thiếu (hổng) các đặc trưng có hiệu lực giúp phân biệt một phân chữ cái so với các chữ cái khác.

Trường hợp sau trong *hình 1.2* phục vụ cho việc nhận dạng và nội suy. Nhận dạng (recognition) là quá trình phân chia một bảng (label) cho một vật thể dựa trên cơ sở mô tả. Nội suy (interpretation) phục vụ cho việc phân chia theo ý nghĩa, để hòa hợp các vật thể được nhận dạng. Ví dụ để nhận dạng một chữ c, cần kết hợp việc miêu tả cho chữ cái này với c. Nội suy là phân chia theo ý nghĩa cho một tập toàn bộ được lập bảng. Ví dụ một chuỗi gồm 5 số (hoặc 5 số theo một dấu ngang) có thể được nội suy để thành một mã ZIP.

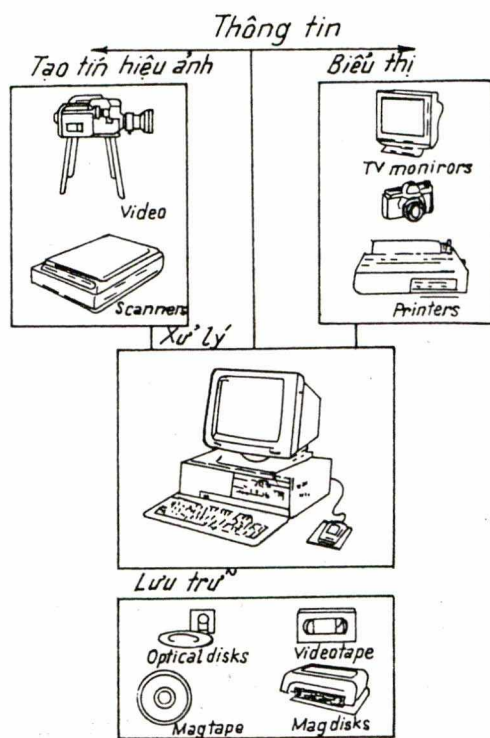
Như vậy ta vẫn chưa đề cập gì đến sự cần thiết về tiên kiến thức (prior knowledge) hoặc tốc độ qua lại giữa cơ sở kiến thức và các modules xử lý trong *hình 1.2*. Kiến thức về phạm vi vật thể được mã hóa thành một hệ thống xử lý ảnh dưới dạng cơ sở dữ liệu kiến thức. Kiến thức này có thể là về các vùng càng chi tiết, thì càng đơn giản của một ảnh (thông tin cần quan tâm được cô lập lại). Cơ sở kiến thức có thể hoàn toàn phức tạp như là một danh sách có liên quan với nhau về tất cả các hư hỏng chính có thể (trong kiểm tra vật liệu hoặc một cơ sở dữ liệu ảnh chứa các ảnh vệ tinh có độ nét cao của một vùng).

Ngoài ra để hướng dẫn hoạt động của từng module xử lý, cơ sở kiến thức kiểm tra sự tương tác giữa các modules. Trong *hình 1.2*, các đường có mũi tên hai đầu nối giữa các modules xử lý và cơ sở kiến thức, các đường có một mũi tên điều khiển các modules xử lý. Điều này thể hiện thông tin giữa các modules xử lý nói chung dựa trên cơ sở tiên kiến thức. Ví dụ, để cho một máy gồm một vòng các ký tự là 1 mã ZIP, thì hệ thống được cung cấp kiến thức (để nhận dạng ý nghĩa vị trí của vòng) phải chấp nhận các thành phần khác của vùng địa chỉ. Kiến thức này không những cần cho hoạt động của mỗi module, mà còn cần cho hoạt động phản hồi giữa các modules qua cơ sở kiến thức. Tất nhiên, một vòng các số trong một vùng cụ thể chỉ bao gồm 4 ký tự (một trong số đó có thể không được nhận dạng) có thể dẫn đến module nội suy để xác định 2 ký tự làm việc: Mạch phản hồi đòi hỏi cơ sở kiến thức cho trạng thái phân đoạn là một ví dụ sử dụng kiến thức để xử lý ảnh. Mặc dù ta chưa đề cập đến màn hiển thị ảnh (image display), nhưng nó rất quan trọng cho việc hiển thị kết quả xử lý ảnh (đặt ở đầu ra các bước trên *hình 1.2*). Ta nhận thấy là không phải tất cả các ứng dụng xử lý ảnh đều đòi hỏi sự phức tạp của các tương tác trong *hình 1.2*. Thực tế không phải tất cả các modules ở đây đều cần cho cùng một trường hợp. Nói chung, các chức năng xử lý ảnh (bao gồm nhận dạng và nội suy) có liên quan đến phân tích ảnh (trong đó vật thể được tự động hoặc từng phần tự động tách thông tin ảnh), ví dụ như nhận dạng ký tự.

1.4. Các phần tử của hệ thống xử lý ảnh số

Các phần tử của một hệ thống chung có khả năng biểu diễn các hoạt động xử lý ảnh (*mục 1.3*, xem *hình 1.3*). Loại hệ thống chung biểu diễn ảnh bao gồm:

1. cảm biến ảnh (acquisition),
2. lưu ảnh (storage),
3. xử lý ảnh (processing),
4. thông tin (communication),
5. hiển thị ảnh (display).



Hình 1.3. Các phần tử chức năng chính của 1 hệ thống xử lý ảnh.

1.4.1. Cảm biến ảnh

Hai phần tử cần cho công đoạn này là linh kiện nhạy với phổ năng lượng điện từ trường (như băng tần x-ray, cực tím, nhìn thấy và hồng ngoại) : Loại thứ nhất tạo tín hiệu điện ở đầu ra tỉ lệ với mức năng lượng mà nó cảm biến ; loại thứ hai là bộ số hóa (digitizer), là phần tử biến đổi tín hiệu tương tự thành tín hiệu số.

Ví dụ với hệ thống chụp ảnh x - quang (x - ray), nguồn x - quang chiếu vào vật thể và bộ cảm biến được đặt hướng vào vật thể. Môi trường vật liệu (ví dụ như xương và nước trong cơ thể) hấp thụ tia x với nhiều mức độ khác nhau. Máy chụp ảnh hoặc camera truyền hình có thể ghi lại hình ảnh trên (phim trong máy chụp ảnh, ống vidicon trong camera truyền hình là bộ cảm biến). Có nhiều loại cảm biến (sensor) làm việc với ánh sáng nhìn thấy và hồng ngoại, ví dụ như *microdensitometers*, *image dissectors*, *camera vidicon*, *linh kiện quang điện* bằng bán dẫn. Loại linh kiện thứ nhất đòi hỏi hình ảnh cần phải số hóa dưới dạng phim âm bản hoặc chụp ảnh. Camera vidicon và linh kiện bán dẫn quang điện có thể cho ảnh ghi trên băng từ và nó có thể số hóa để cường độ sáng có hiệu quả hơn.

Trong *microdensitometers*, phim và ảnh chụp được gắn trên một mặt phẳng hoặc cuộn quanh một trống. Quét ảnh được thực hiện nhờ một tia sáng (có thể là tia laser) trên ảnh và dịch chuyển mặt phim hoặc quay trống tương đối theo tia sáng. Trong trường hợp dùng phim thì tia sáng đi qua phim, đối với ảnh chụp thì phản xạ từ bề mặt của ảnh. Trong cả hai trường hợp, tia sáng được hội tụ trên phần tử tách sóng ánh sáng (photodetector) và mức xám tại một điểm bất kỳ trong ảnh được ghi lại trên cơ sở cường độ sáng của tia. Một ảnh số nhận được bằng cách chỉ cho phép các giá trị rời rạc về cường độ sáng xuất hiện ở vị trí đầu ra. Mặc dù *microdensitometers* là loại linh kiện chậm, nhưng nó cho độ chính xác cao.

Hoạt động của camera vidicon (truyền hình) dựa trên nguyên tắc điện dung cảm

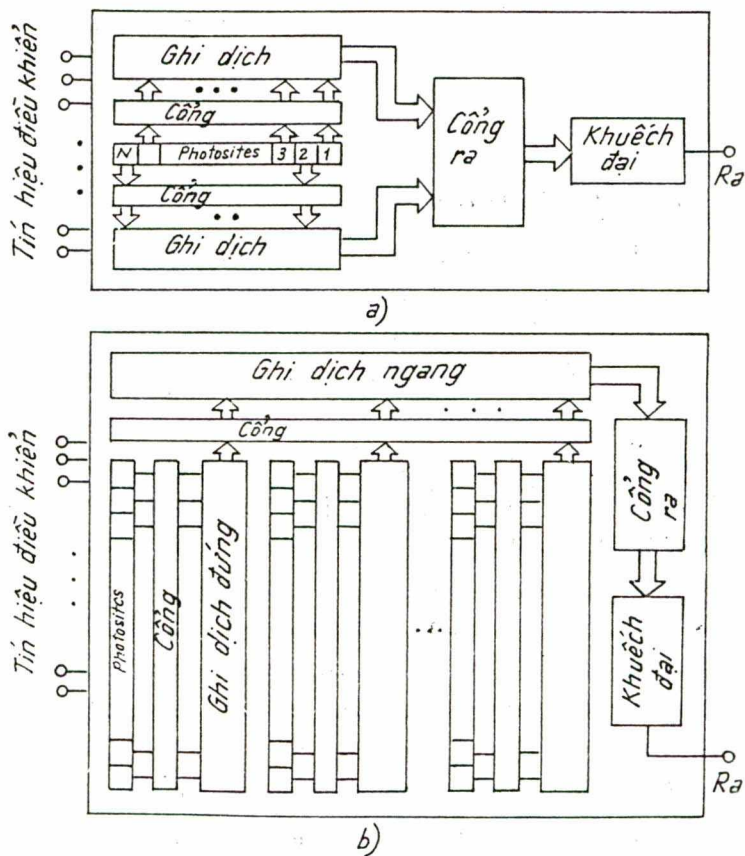
quang (photoconductivity). Hình ảnh hội tụ trên bề mặt đèn vidicon, tạo ra một ảnh (pattern) làm thay đổi các tụ phân bố độ sáng trong ảnh quang học. Một tia điện tử độc lập và hội tụ quét lên bề mặt của màng target (photoconductive target) và phóng nạp để trung hòa; Tia này mang hiệu thế, tạo ở đầu ra một tín hiệu tỉ lệ với ảnh ở đầu vào về độ sáng. Nếu số hóa tín hiệu này, ta sẽ có ảnh số.

Các thanh bán dẫn được hợp thành bởi các phần tử ảnh silic rời rạc, gọi là photosites có điện áp ra tỉ lệ với cường độ ánh sáng tới. Các thanh bán dẫn đó bố trí trong 1 hoặc hai hàng và ta có hai loại: cảm biến dòng (line scan sensors) và cảm biến diện (area sensors). Cảm biến dòng bao gồm một hàng photosites và tạo ảnh hai chiều do chuyển động giữa cảnh và detector. Ví dụ cảm biến dòng được sử dụng nhiều trong các bộ quét ảnh phẳng. Còn cảm biến theo diện bao gồm một ma trận các photosites và vì vậy có khả năng cho hình ảnh như đèn vidicon. Ưu điểm nổi bật của cảm biến bằng các thanh bán dẫn là nó chịu đựng rung với tốc độ cao (1/10.000s). Nó thích hợp cho các ứng dụng mà trong đó cần làm "đông cứng" lại các chuyển động.

Công nghệ dùng trong các bộ cảm biến bán dẫn dựa trên cơ sở linh kiện phóng nạp ghép CCD (charge - coupled devices). Hình 1.4 là cảm biến CCD quét theo dòng, bao gồm một hàng photosites, 2 cổng chuyển dịch dùng để cát nhip nội dung các điểm ảnh thành các thanh ghi dịch, và cổng đầu ra dùng để cát nhip nội dung của các thanh ghi dịch vào bộ khuếch đại. Đầu ra bộ khuếch đại, tín hiệu điện áp tỉ lệ với nội dung của hàng photosites.

Các thanh CCD theo diện mặt tương tự như các sensors theo dòng, các photosites được xếp với kích thước một ma trận và liên kết cổng ghi - dịch, tách các cột photosites (hình 1.4b).

Các nội dung của các photosites (đánh số chẵn) lần lượt tạo cổng thành các thanh ghi dịch theo chiều đứng và sau đó thành các thanh ghi dịch theo chiều ngang. Nội dung tín hiệu từ thanh ghi được đưa vào một bộ khuếch đại có đầu ra là một dòng video. Lập lại thuật toán này cho các dòng đánh số lẻ, sẽ hoàn chỉnh màn hình thứ 2 của một ảnh truyền hình quét xen kẽ. Độ phân giải của các sensors theo diện là từ 32 x 32 điểm ảnh (loại sensors thấp) đến 256 x 256 điểm ảnh đối với sensor có độ phân giải trung bình. Linh kiện sensors cho độ phân giải cao hơn sẽ đạt 640 x 480 điểm ảnh, còn độ phân giải tương đối cao là 1280 x 1024 điểm ảnh. Độ phân giải đặc biệt cao sẽ



Hình 1.4. Cảm biến quét dòng bằng CCD (a) và cảm biến mặt bằng CCD (b).

cho 2048 x 2048 điểm ảnh (CCD chip). Để có ảnh số, tín hiệu đầu ra của camera được đưa vào bộ digitizer (bộ số hóa).

1.4.2. Lưu trữ (storage)

Một hình ảnh 8 bit với kích thước 1024 x 1024 pixels cần bộ nhớ 1 triệu bytes. Bộ nhớ số cho các ứng dụng xử lý ảnh nằm trong 3 lĩnh vực :

1. lưu trữ ngắn, dùng trong khi xử lý ;
2. lưu trữ tức thời, dùng cho việc gọi lại nhanh,
3. lưu trữ văn bản, đặc trưng sự kiện ít xảy ra.

Lưu trữ được đo bằng bytes (8 bit), kbytes (1000 bytes), Mbytes (1000.000 bytes), Gbytes (1000.000.000 bytes), Tbytes (tera bytes - 10^{12} bytes).

Một phương pháp dùng bộ nhớ ngắn (short term storage) là bộ nhớ máy tính (computer memory). Loại khác là các boards đặc trưng, gọi là frame buffers (bộ nhớ đệm ảnh), nó có thể nhớ một hoặc nhiều ảnh và xảy ra nhanh, thường bằng tốc độ video (30 hoặc 25 ảnh / giây). Phương pháp sau cho phép xem "zoom" ảnh (cận cảnh và toàn cảnh) liên tục, như cuộn (scroll) - dịch theo chiều đứng, kéo dẫn - dịch theo chiều ngang. Số lượng bộ nhớ trong một card nhớ đệm bị giới hạn bởi kích cỡ vật lý của card và mật độ nhớ của chips nhớ. Trên một card nhớ đệm đơn, thường có thể nhớ 32 Mbytes.

Bộ nhớ tức thời (on - line storage) nói chung có dạng đĩa từ (magnetic disks), ví dụ một đĩa từ thường hàng trăm Mbytes. Loại bộ nhớ từ - quang MO (magneto - optic storage) sử dụng một laser và vật liệu đặc biệt để nhớ được 1 Gbyte trên đĩa 5 1/4". Đặc trưng chính của bộ nhớ tức thời là nhận số liệu nhanh. Cho nên băng từ ít được dùng cho bộ nhớ tức thời trong xử lý ảnh. Jukeboxes chứa 30 - 100 đĩa quang có độ phân giải cao, các ứng dụng lưu trữ tức thời đòi hỏi khả năng đọc - ghi. Các jukeboxes quang làm việc cùng một nguyên tắc như jukebox nhạc, có nghĩa là một hệ thống cơ được dùng để đưa (retrieve) các đĩa quang vào hộp đĩa quang (drives).

Cuối cùng, lưu trữ văn bản (archival storage) được đặc trưng bởi các yêu cầu lưu trữ lớn hàng loạt, nhưng ít xảy ra. Các băng từ và đĩa quang thường được dùng cho lưu trữ văn bản. Băng từ mật độ cao (6400 bytes/inch) có thể lưu trữ một ảnh 1 Mbyte trong 13 ft (13.0,3 m = 3,9 m) băng. Vấn đề chính với băng từ là tuổi thọ ngắn (~ 7 năm) và cần phải có kiểm tra. Công nghệ đĩa quang ghi nhận một lần - đọc nhiều lần WORM (write - once - read - many), hiện nay có thể lưu trữ cỡ 16 Gbyte trên đĩa 5 1/4". Không giống như MO, đĩa WORM có hệ số hình dạng lớn hơn, có khả năng lưu trữ gần 6 Gbytes trên đĩa 12", và trên 10 Mbytes trên đĩa 14". Mặc dù vậy chúng không thể xoá được. Đĩa WORM có tuổi thọ 30 năm mà không cần bảo quản đặc biệt. Nếu lưu trữ trong một jukebox, các đĩa WORM có thể phục vụ như là các linh kiện lưu trữ tức thời mà các hoạt động đọc (read - only) chiếm ưu thế. Bộ nhớ WORM 1 Tbyte có thể đặt trong một jukebox với thể tích chiếm ít hơn 150 ft³. Dung lượng này chuyển dịch 1 triệu hình ảnh 8 bit với kích thước 1024 x 1024 pixel.