Nhiều phương pháp cho việc rút trích văn bản từ ảnh và video. Tuy nhiên không có phương pháp rút trích văn bản tổng quát bởi

Văn bản được nhúng

trong shaded or textured background

trong ảnh có độ tương phản thấp hoặc phức tạp

trong ảnh có sự thay đổi về kích thước font, loại font, màu sắc, hướng, và sự canh lề

Phân loại ảnh chứa văn bản

+ ảnh document: ảnh này chủ yếu là ảnh xám

+ ảnh document nhiều màu: ảnh này chứa nhiếu màu sắc khác nhau

Phân loại văn bản trong ảnh

+ văn bản caption(caption text or superimposed text, or artificial text): được đặt lên ảnh một cách nhân tạo

+ văn bản scene (scene text or graphics text): chúng tồn tại một cách tự nhiên trong ảnh

Văn bản nhân tạo khó phát hiện hơn so với văn bản caption, giá trị nội dung văn bản và ít có nhiều nghiên cứu về nó,

Việc phân loại những kỹ thuật khác nhau là tương đối khó. Do đó, ta cần định nghĩa một thuật ngữ, và tóm tắt những đặc trưng của văn bản có thể được sử dụng cho những thuật toán TIE (text information extraction)

Sau đây liệt kê một số thuộc tính được sử dụng trong các bài báo

1. Hình học:
   1. Kích thước
   2. Sự canh lề: Với văn bản caption, những ký tự trong xuất hiện thành từng nhóm và thường nằm theo chiều ngang, đôi lúc, có thể theo chiều đứng; trong khi đó văn bản scene thì không.
   3. Khoảng cách giữa các ký tự: Khoảng cách giữa những ký tự trong văn bản thì gần như giống nhau.
2. Màu sắc: Những ký tự trong văn bản thường có màu sắc giống nhau hoặc na ná nhau. Thuộc tính này cho phép ta có thể dùng phương pháp dựa trên những thành phần kết nối cho việc phát hiện văn bản. Hầu hết, các nghiên cứu vẫn tập trung trên việc tìm kiếm những chuỗi văn bản có cùng một màu. Tuy nhiên, những ảnh video, ảnh những tài liệu có màu sắc phức tạp thì những chuỗi văn bản có thể có hơn 2 màu khác nhau, để gây hiệu ứng, đặc biệt có thể có hơn 2 màu trong cùng một từ
3. Chuyển động: Văn bản trong video có thể xuất hiện trong vài frame liên tiếp, nó có thể chuyển động hoặc không chuyển động. Thuộc tính được dùng để theo vết và cải tiến văn bản. Văn bản caption thường di chuyển theo một hướng cố định: hoặc ngang, hoặc dọc. Văn bản scene thường di chuyển tùy ý phụ thuộc vào sự di chuyên của đối tượng hoặc camera.
4. Cạnh: hầu hết văn bản scene và caption được thiết kế để dễ đọc. Do đó, nó cạnh của văn bản so với background thường rõ, và không bị nhòe.
5. Nén: Nhiều hình ảnh kỹ thuật số được ghi lại, truyền và xử lý dưới dạng nén. Do đó, để hệ thống TIE nhanh, thì văn bản có thể được rút trích mà không cần giải nén.
6. Stroke\*: là hiệu ứng, sao cho những ký tự đen trong văn bản được bao quanh với viền màu trắng. Điều này giúp tăng tính dễ đọc của văn bản trên background có màu sắc không đồng nhất.



Rút trích văn bản là gì

Hệ thống TIE có

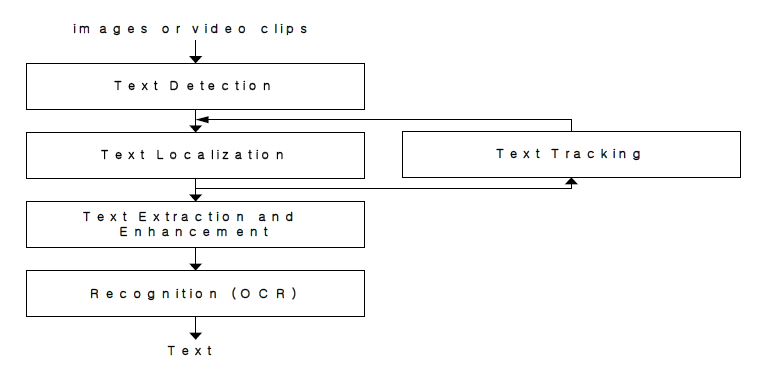
Input: là một ảnh tĩnh hoặc một chuỗi ảnh (video)

+ ảnh có thể là ảnh grayscale hoặc ảnh màu; bị nén hoặc không bị nén

+ văn bản trong ảnh di chuyển hoặc không di chuyển

Vấn đề TIE có thể được thành những vấn đề con sau:

1. Phát hiện văn bản(dectection)
2. Định vị văn bản (localization)
3. Theo vết văn bản (tracking)
4. Rút trích và nâng cao chất lượng văn bản (extraction and enhancement)
5. Nhận dạng văn bản (recognition)



Hình 5 – Kiến trúc của hệ thống TIE

Phát hiện văn bản là xác định có hay không sự hiện diện của văn bản trong ảnh/frame cho trước.

Định vị văn bản là xác định vị trí của văn bản trong ảnh và sinh ra bounding box bao xung quanh văn bản.

Theo vết văn bản: mục địch giảm thời gian xử lý cho việc định vị văn bản và duy trì ràng buộc vị trí của nó so với các frame liên tục.

Rút trích văn bản là quá trình tách văn bản khỏi background bằng việc phân đoạn những thành phân văn bản từ background. Kết quả là ta thu được ảnh nhị phân, và ta cần nâng cao chất lượng của ảnh nhị phân này trước khi đưa nó vào OCR bởi vì những vùng văn bản thường có độ phân giải thấp và bị nhiễu.

Nhận dạng văn bản: ta sử dụng những công nghệ OCR để biến đổi những văn bản trong ảnh nhị phân thành văn bản thuần túy (plain text).

1. Phát hiện văn bản:

Trong giai đoạn, chúng ta chưa có thông tin biết trước: có hay không có sự hiện diện của văn bản trong ảnh cho trước, nhiệm vụ của giai đoạn này là xác định điều đó.

**Kim** [1995]sử dụng phương pháp phát hiện sự thay đổi scene (scene-change detection change) để chọn ra những frame từ video như là ứng cử viên chứa văn bản. Phương pháp này không được mô tả chi tiết nhưng đại loại anh ấy sử dụng những giá trị ngưỡng thấp để phát hiện thay đổi scene bởi vì vùng văn bản chỉ chiếm một phần tương đối nhỏ trong ảnh. Quá trình này được thực hiện lại sau khoảng thời gian 2 giây 🡪nhạy trong việc phát hiện thay đổi scene.

**Smith và Kanade** [1998] định nghĩa sự thay đổi scene dựa trên sự khác nhau giữa hai frame liên tục 🡪 thu được độ chính xác tới 90% trong phát hiện thay đổi scene.

**Gargi et al.** phát hiện văn bản bằng việc giả sử rằng số lượng những block intracode trong P-, B- frames trong video bị nén MPEG tăng ***khi văn bản xuất hiện***

**Lim et al.** đưa ra giả sử đơn giản hơn: văn bản thường có cường độ (intensity) cao hơn background. Tác giá đếm số lượng pixel mà có cường độ lớn một ngưỡng được cho trước và những pixel này có sự khác biệt màu sắc lớn so với những láng giềng của chúng. Những frame mà có số lượng lớn những pixel như thế được xem là frame chứa văn bản. 🡪 đơn giản, nhanh. Tuy nhiên, có vấn đề với những văn bản bị đảo màu.

* Giai đoạn phát hiện văn bản có liên quan gần tới giai đoạn định vị văn bản và theo vết văn bản.

1. Định vị văn bản

Các phương pháp định vị văn bản được chia làm hai loại: dựa trên vùng và dựa trên vân. Một số phương pháp kết hợp cả hai loại

* 1. Những phương pháp dựa trên vùng

Những phương pháp này sử dụng thuộc tính màu sắc hoặc grayscale trong vùng văn bản hoặc sự khác nhau giữa chúng so với thuộc tính background tương ứng. Chúng được chia thành 2 nhóm tiếp cận: dựa trên trành phần kết nối (connected component) và dựa trên cạnh (edge-based). Cả hai theo hướng tiếp cận bottom-up bằng việc

* + - * xác định những cấu trúc (thành phần) con như CCs và edge
      * sau đó merge những cấu trúc con này để tạo những cấu trúc (thành phần) lớn hơn
      * xác định những khung viền bao quanh văn bản.

**Chú ý:** Một số kết hợp cả hai.

1. Những phương pháp dựa trên vùng

Ohya et al. đưa ra phương pháp 4 giai đoạn: 1. nhị phân hóa dựa trên ngưỡng cục bộ; 2. Phát hiện những thành phần ký tự ứng viên bằng việc dùng sự khác nhau về mức xám; 3. Nhận dạng dựa vào độ tương đồng giữa ký tự ứng viên và ký tự mẫu trong cơ sở dữ liệu; 4. Thao tác relaxation để cập nhật độ tương đồng.

* Phương pháp này có giới hạn là chỉ áp dụng cho văn bản thẳng đứng và không kết nối.

**Locating text in complex color images**

<http://portal.acm.org/citation.cfm?id=844379.844664>

Video bản tin: thời sự/chương trình ca nhạc, kinh tế-tài chính/chứng khoán.

Việc rút trích văn bản trong

Giai đoạn 1: video bản tin thời sự giúp ta tóm tắt các bản tin/sự kiện chính trong video; video chương trình ca nhạc giúp ta biết được những ca sĩ nào đã tham gia chương trình này, và họ đã biểu diễn những ca khúc nào, và ai và người sáng tác những ca khúc đó.

Giai đoạn 2: video bản tin kinh tế-tài chính giúp ta biết tỷ giá hối đoái, giá vàng, tình hình cổ phiếu, chiều hướng của các nhà đầu tư.

Dữ liệu thử nghiệm: Lấy các video từ trên mạng.