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 scene khó phát hiện hơn so với văn bản caption, 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