ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Xây dựng hệ thống truy vấn ảnh dựa vào văn bản ngoại cảnh

HV: Hồ Trần Nhật Thủy
GVHD: PGS,TS. Lý Quốc Ngọc

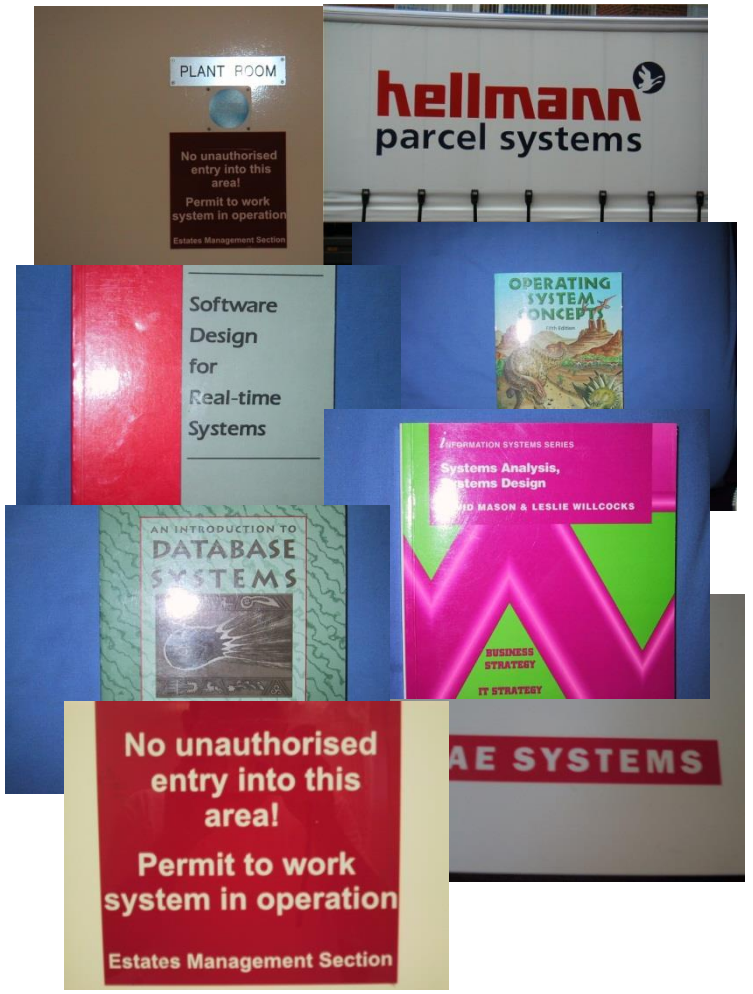
Nội dung

1. Phát biểu bài toán
2. Hướng tiếp cận
3. Kết quả thực nghiệm
4. Kết luận và hướng phát triển

Giới thiệu

- Văn bản trong ảnh là một đối tượng chứa nhiều thông tin ngữ nghĩa quan trọng
- Khai thác nội dung văn bản trong ảnh có nhiều ứng dụng
  - Các thiết bị hỗ trợ người khiếm thị
  - Hiểu nội dung ảnh đa ngôn ngữ
  - Hệ thống dẫn dường tự động
  - Lập chỉ mục ảnh
  - ...

Phát biểu bài toán



Phát biểu bài toán

- Cho trước tập dữ liệu ảnh chứa văn bản ngoại cảnh
  - Phát hiện, rút trích, nhận dạng văn bản ngoại cảnh trong ảnh
    - Đầu vào: ảnh chứa văn bản ngoại cảnh
    - Đầu ra: vị trí các vùng văn bản, chuỗi ký tự tương ứng
  - Cho phép người dùng thực hiện truy tìm các ảnh có chứa từ khóa mong muốn.
    - Đầu vào: Câu truy vấn dưới dạng từ khóa hoặc ảnh
    - Đầu ra: tập ảnh được sắp hạng theo độ tương đồng về nội dung văn bản so với ảnh truy vấn

Hướng tiếp cận

Phát hiện và rút trích văn bản

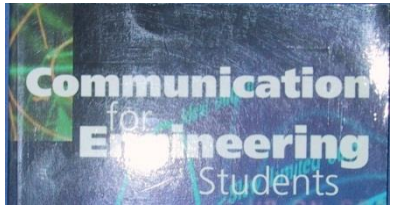
OCR

OCR post-correction

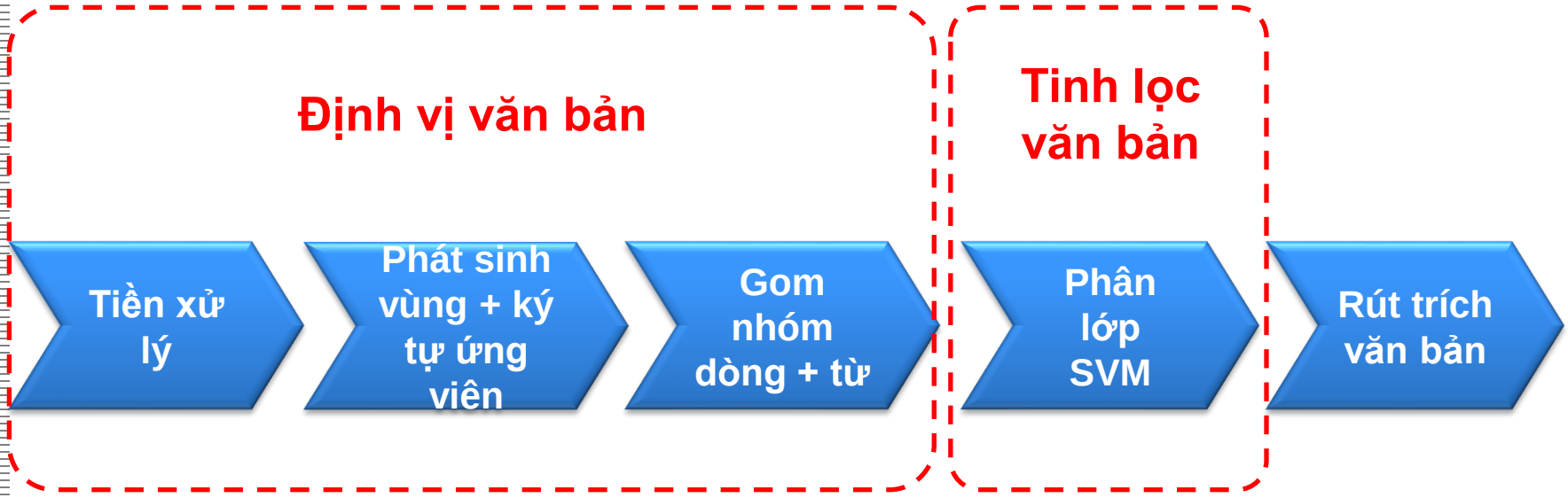
Tổ chức dữ liệu và truy vấn

Các thách thức

- Kích thước, kiểu chữ, màu sắc, vị trí, hướng khác nhau
- Nền phức tạp
- Sự chiếu sáng
- Ảnh mờ
- …



Mô hình phát hiện và rút trích văn bản



Tiền xử lý

- Phép tái tạo ảnh (*reconstruction)*
  - Mục đích: Rút trích các đối tượng liên kết có cường độ sáng hơn vùng nền xung quanh
  - Cách thực hiện:
    - Đầu vào: ảnh mức xám I, ảnh J (có 4 biên trùng ảnh I)
    - Bước dilation: Với mỗi pixel $p \in I$

$$K(p) \leftarrow \max\{J(q), q \in N_G(p) \cup \{p\}\}$$

$N_G(p)$: các pixel trong vùng lận cận G của p

    - Bước minimum: Với mỗi pixel $p \in I$

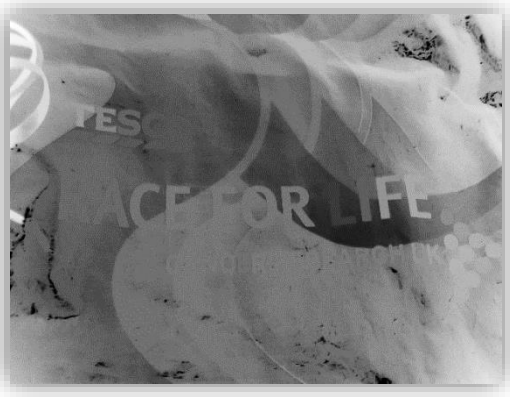
$$J(p) \leftarrow \min\{K(p), I(p)\}$$

Mô hình phát hiện và rút trích văn bản

Tiền xử lý → Phát sinh vùng + ký tự ứng viên → Gom nhóm dòng + từ → Phân lớp SVM → Rút trích văn bản

I

reconstruction

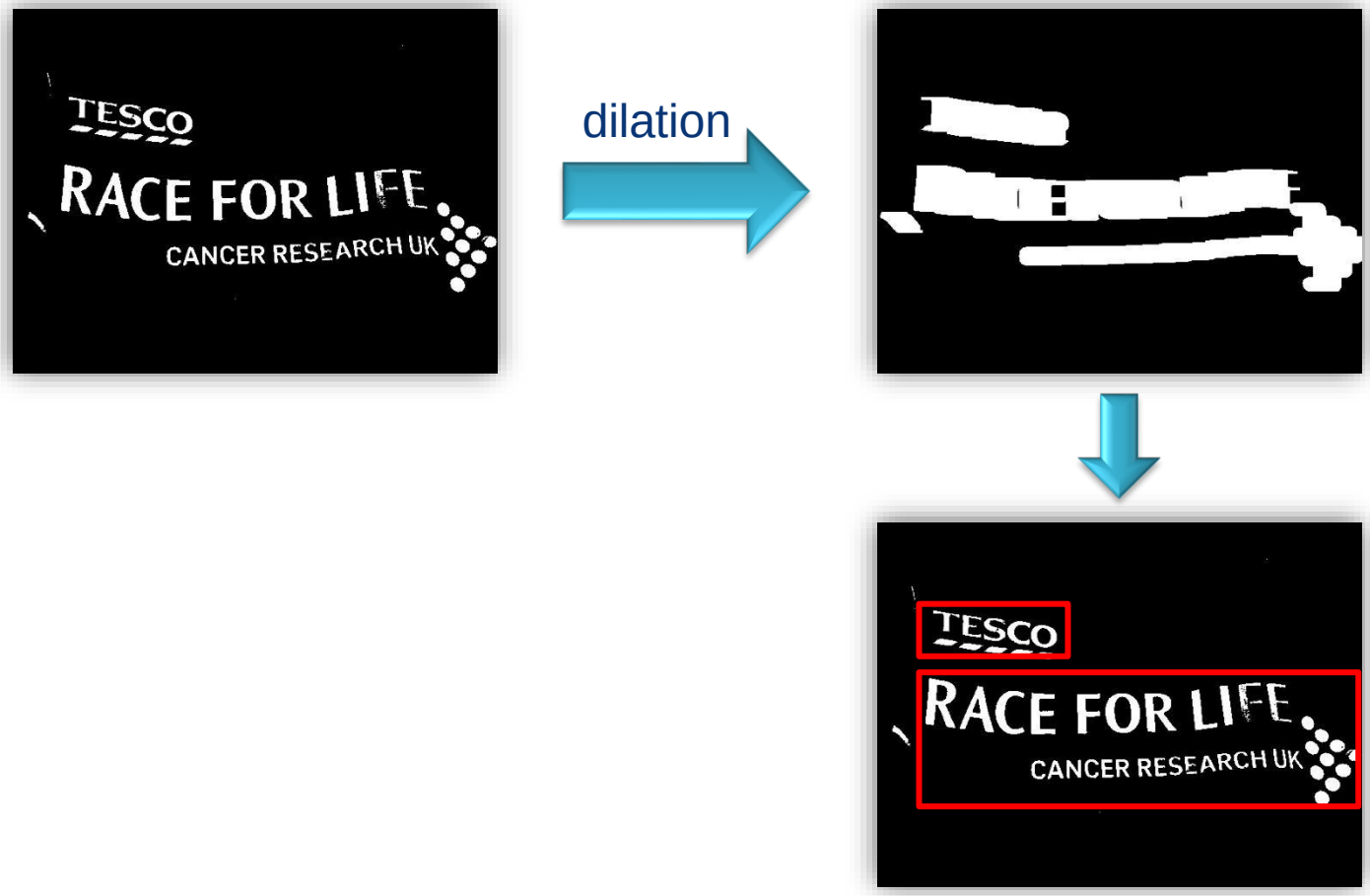


J

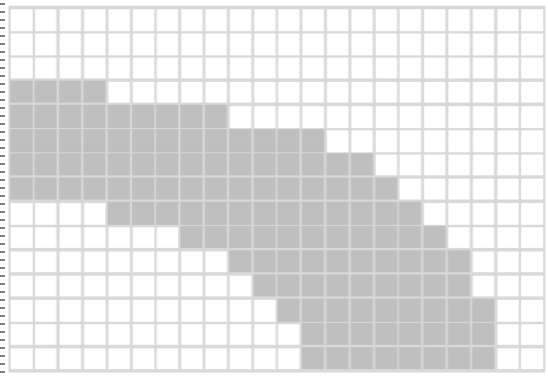
I-J

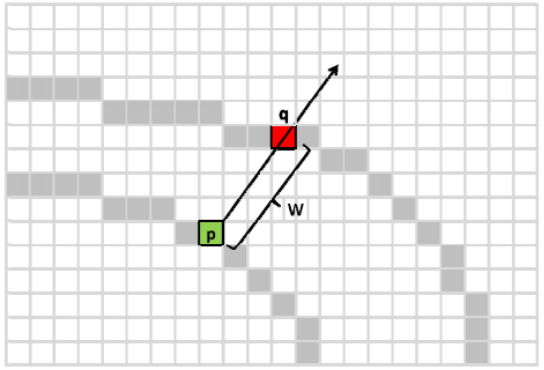
Mô hình phát hiện và rút trích văn bản

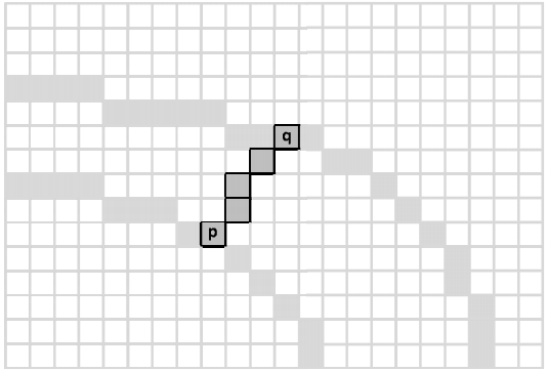
Tiền xử lý → **Phát sinh vùng + ký tự ứng viên** → Gom nhóm dòng + từ → Phân lớp SVM → Rút trích văn bản



Stroke Width Transform







	4	4	4	4					
	4	4	4	4					
	4	4	4	4					
	4	4	4	4					
	4	4	4	4					
	4	4	4	3					
	4	4	3	3	3	4	4	4	4	4
	4	3	3	3	4	4	4	4	4	4
		3	3	4	4	4	4	4	4	4
			4	4	4	4	4	4	4	4

B. Epshtein, E. Ofek, and Y. Wexler, Detecting text in natural scenes with stroke width transform, in Proc. IEEE Conf. CVPR, 2010.

Mô hình phát hiện và rút trích văn bản

Tiền xử lý → **Phát sinh vùng + ký tự ứng viên** → Gom nhóm dòng + từ → Phân lớp SVM → Rút trích văn bản

Stroke Width Transform →

Std/mean < 0.5 →

Mô hình phát hiện và rút trích văn bản

Tiền xử lý → Phát sinh vùng + ký tự ứng viên → **Gom nhóm dòng + từ** → Phân lớp SVM → Rút trích văn bản

Nhóm thành dòng

Tỉ lệ độ rộng nét
Tỉ lệ chiều cao
Khoảng cách giữa 2 ký tự

Mô hình phát hiện và rút trích văn bản

Tiền xử lý → Phát sinh vùng + ký tự ứng viên → **Gom nhóm dòng + từ** → Phân lớp SVM → Rút trích văn bản

Tách từ

$$T = Mean(D) + \beta \times Std(D) \quad (1)$$

D: khoảng cách giữa các ký tự trên một dòng

Mô hình phát hiện và rút trích văn bản

- Đầu vào: các từ ứng viên phát hiện được
- Mục tiêu: phân lớp các từ ứng viên

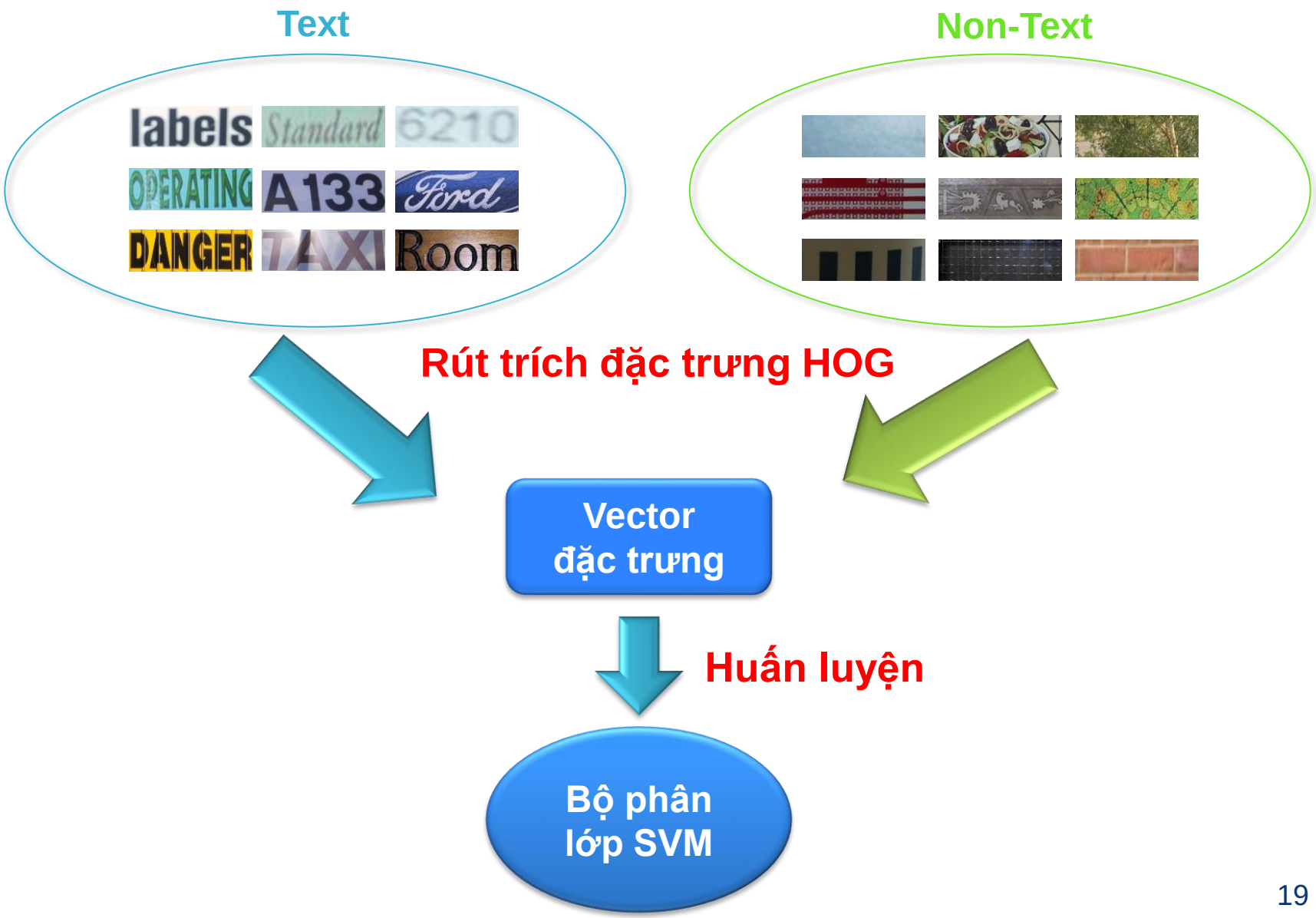
Histogram of Oriented Gradient (HOG)

- Được đề xuất bởi Navel Dalal và Bill Triggs năm 2005.
- Đặc trưng HOG được dùng nhiều trong lĩnh vực phát hiện và nhận dạng đối tượng
- Ý tưởng chính: đặc điểm, hình dáng của đối tượng có thể được biểu diễn khá tốt thông qua phân bố cường độ cục bộ của hướng cạnh

Support Vector Machines (SVM)

- Được đề xuất bởi Vapnik
- Đạt hiệu quả cao trong các bài toán phân lớp nhị phân
- Có khả năng huấn luyện một bộ phân lớp phi tuyến trong không gian đặc trưng có số chiều cao với số lượng các mẫu huấn luyện nhỏ.

Huấn luyện


Text
Non-Text
labels
Standard
6210
OPERATING
A133
Ford
DANGER
TAXI
Room
Rút trích đặc trưng HOG
Vector đặc trưng
Huấn luyện
Bộ phân lớp SVM

Mô hình phát hiện và rút trích văn bản

Tiền xử lý → Phát sinh vùng + ký tự ứng viên → Gom nhóm dòng + từ → Phân lớp SVM → Rút trích văn bản

Mô hình phát hiện và rút trích văn bản

- Trong mỗi từ đã phát hiện (R), tính mức xám trung bình (mean) và độ lệch chuẩn (std) của các điểm ảnh có SWT > 0

$$T_{R,1} = mean(R) - k_1 \times std(R) \qquad (2)$$

$$T_{R,2} = mean(R) + k_2 \times std(R) \qquad (3)$$

$$B_R(i,j) = \begin{cases} 0 & if\ T_{R,1} \le gray(i,j) \le T_{R,2} \\ 255 & other \end{cases}, \forall (i,j) \in R$$

Hướng tiếp cận

Phát hiện và rút trích văn bản

OCR

OCR post-correction

Tổ chức dữ liệu và truy vấn

Hiệu chỉnh kết quả OCR

- Khoảng cách Levenshtein: số lượng thao tác tối thiểu cần để chuyển chuỗi này thành chuỗi khác, với các thao tác thêm, xóa và thay thế ký tự.
  - VD: khoảng cách Levenshtein giữa hai chuỗi “kitten” và “sitting” là 3

- Độ khớp N-gram (TF):
  - r@pt0r $\rightarrow$ {##r, #r@, r@p, @pt, pt0, t0r, 0r#, r##}
  - raptor $\rightarrow$ {##r, #ra, rap, apt, pto, tor, or#, r##}

$$TF\left(r@pt0r, raptor\right) = \left|\{\#\#r, r\#\#\}\right| = 2$$

Hiệu chỉnh kết quả OCR

- Gọi w_{ocr} là kết quả OCR
- Với mỗi từ w thuộc từ điển D, tính khoảng cách Levenshtein $L(w_{ocr}, w)$
- Chọn các từ có $L(w_{ocr}, w)$ nhỏ nhất vào tập ứng viên CW
- Với mỗi $w=a_1a_2\ldots a_n \in$ CW, tính $TF(w_{ocr}, w)$ và $score(w)$
  - $$score(w) = \sqrt[n-2]{\prod_{i=1}^{n-2} P(a_{i+2}|a_i a_{i+1})} \qquad (4)$$
- Chọn w có tổng $(score(w) + TF(w_{ocr}, w))$ lớn nhất

Hướng tiếp cận

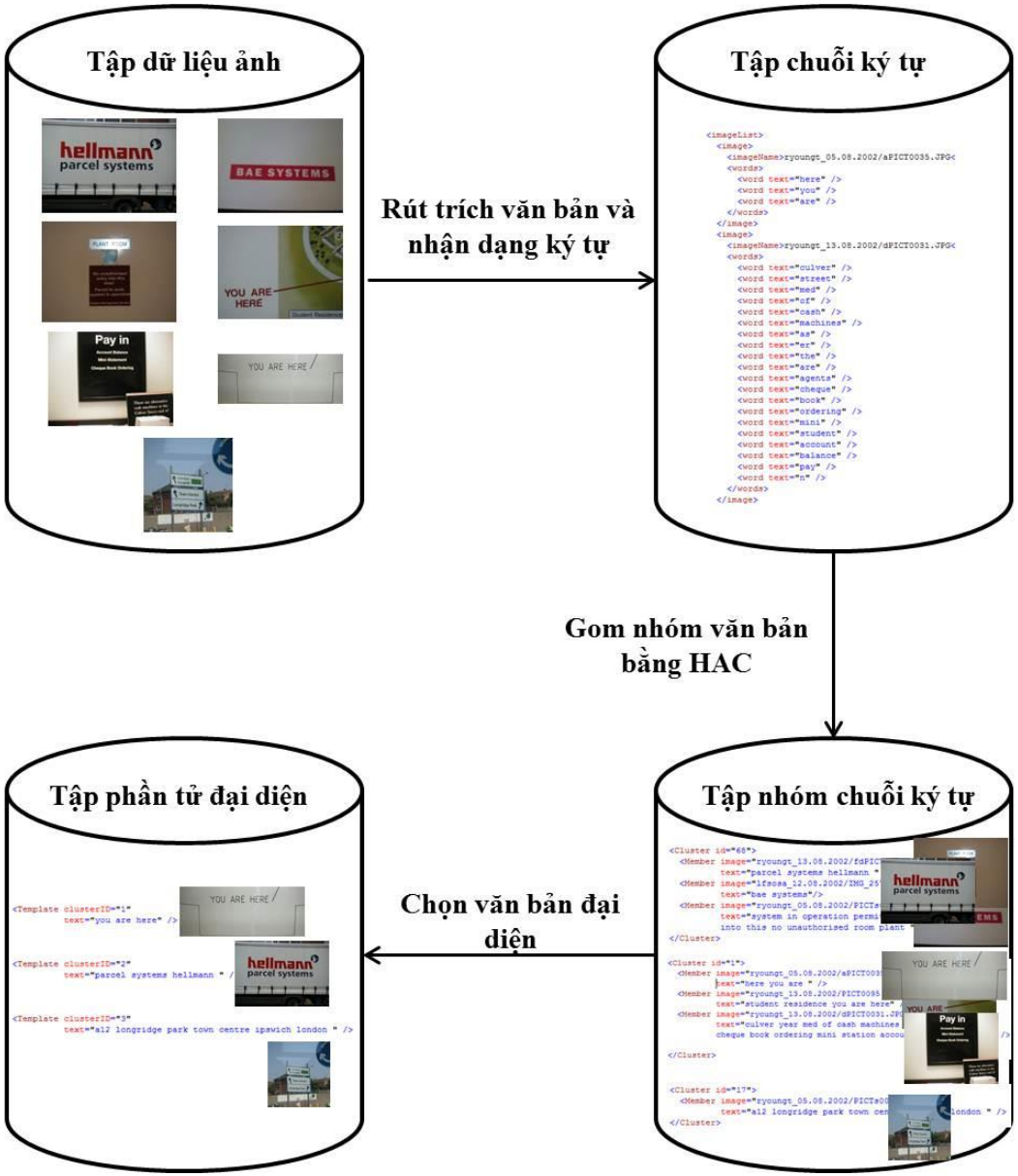
Phát hiện và rút trích văn bản

OCR

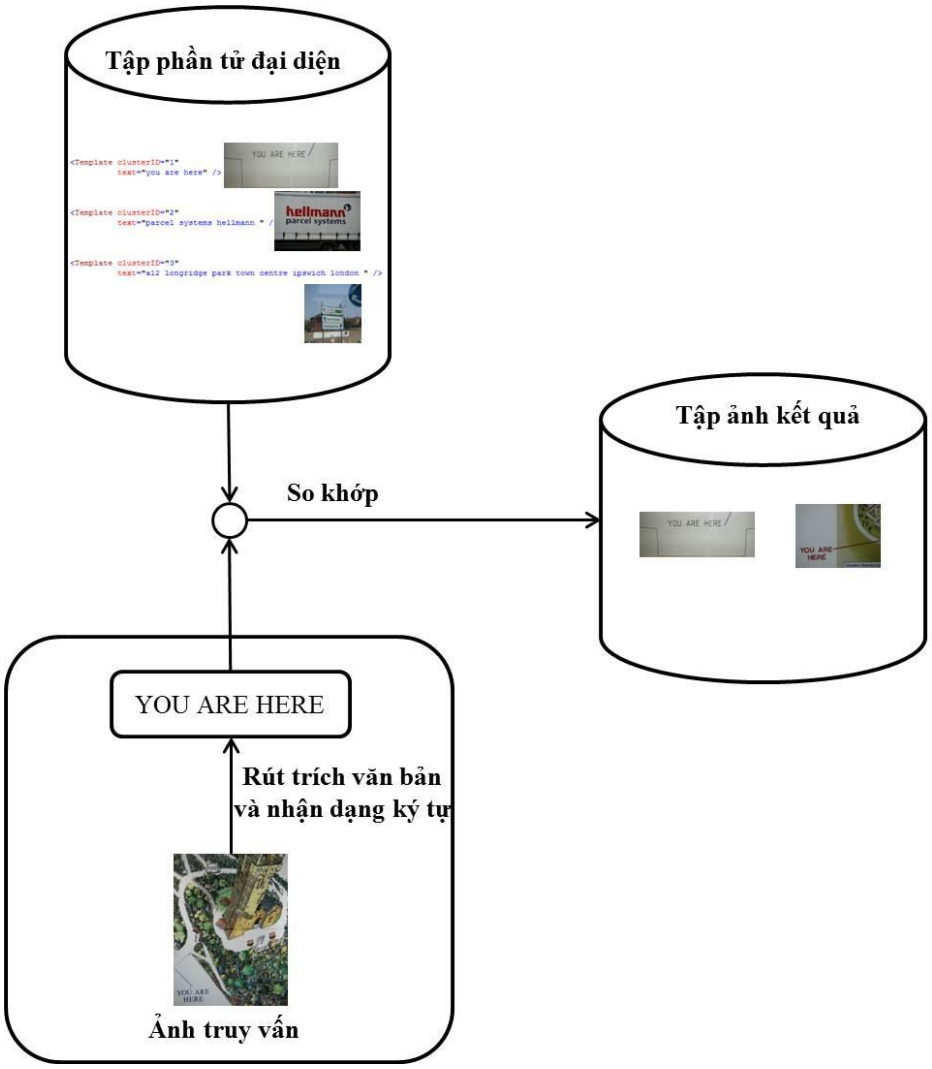
OCR post-correction

Tổ chức dữ liệu và truy vấn

Mô hình tổ chức dữ liệu



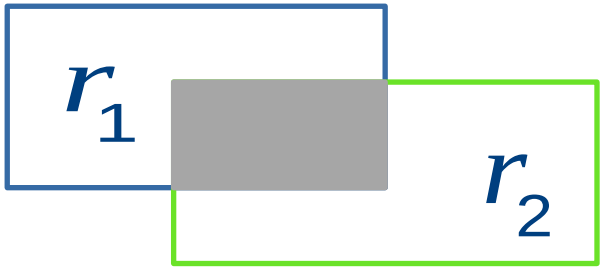
Mô hình truy vấn ảnh



- Cách thức truy vấn: bằng từ khóa hoặc bằng ảnh
- Độ đo dị biệt: khoảng cách Levenshtein
- Tìm kiếm chính xác: ảnh được chọn chứa ít nhất một từ thuộc chuỗi truy vấn $\xi = 0.0$
- Tìm gần đúng: sự khác biệt giữa hai chuỗi nhỏ hơn một ngưỡng ξ

Kết quả thực nghiệm

- Kết quả phát hiện văn bản
  - ICDAR 2003 dataset: 251 ảnh train, 249 ảnh test, kích thước từ 307×93 đến 1600×1200 pixels (có groundtruth)



$$m(r_1, r_2) = \frac{2a(r_1 \cap r_2)}{a(r_1) + a(r_2)}$$

$$P = \frac{\sum_{r_e \in E} m(r_e, T)}{|E|} \quad (5)$$

$$R = \frac{\sum_{r_t \in T} m(r_t, E)}{|T|} \quad (6)$$

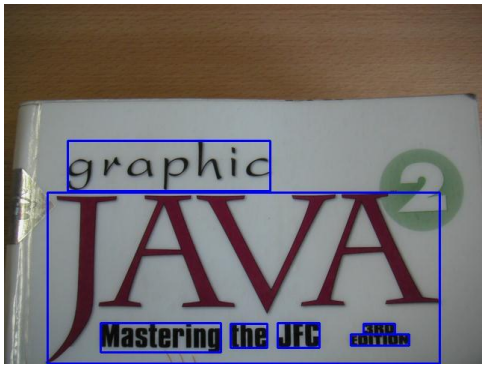
$$f = \frac{1}{\alpha / P + (1-\alpha) / R}$$

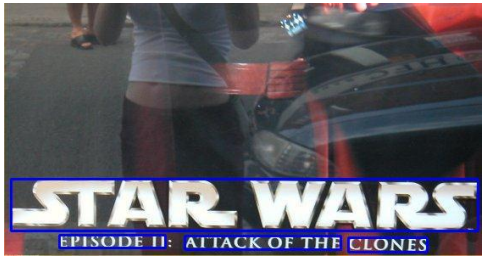
Kết quả phát hiện văn bản

- Kết quả trên tập thử nghiệm

Phương pháp	Precision	Recall	f
Kim's method [2011]	0.83	0.62	0.71
Phương pháp đề xuất	**0.78**	**0.62**	**0.69**
Epshtein [2010]	0.73	0.60	0.66
Yi [2011]	0.67	0.58	0.62
TH-TextLoc [2011]	0.67	0.58	0.62
Hinnerk Becker [2005]	0.62	0.67	0.62
Neumann [2011]	0.69	0.53	0.60
Alex Chen [2005]	0.60	0.60	0.58

Kết quả phát hiện văn bản





Kết quả truy vấn ảnh

- Kết quả truy vấn ảnh với $\xi = 0.0$

$$P = \frac{\text{Số ảnh tìm được đúng với độ dị biệt } \xi}{\text{Số ảnh tìm được với độ dị biệt } \xi}$$

$$R = \frac{\text{Số ảnh tìm được đúng với độ dị biệt } \xi}{\text{Số ảnh đúng thực có với độ dị biệt } \xi}$$

Cách thức truy vấn	Số lần truy vấn	Độ chính xác	Độ phủ
Bằng từ khóa	50	98.36%	74.41%
Bằng ảnh	25	91.20%	60.43%

Kết luận

- Xây dựng mô hình phát hiện và rút trích văn bản góp phần vượt qua một số thách thức: nền nhiễu loạn, không biết trước kiểu chữ, kích thước, màu sắc, bố cục, vị trí của văn bản.

- Đề xuất phương pháp hiệu chỉnh kết quả OCR giải quyết một phần khó khăn khi nhận dạng văn bản ngoại cảnh

- Đề xuất và xây dựng hệ thống truy vấn ảnh dựa vào văn bản ngoại cảnh. Đây là mô hình truy vấn ảnh mới.

Hướng phát triển

- Tìm kiếm và kết hợp với các đặc trưng khác để phân biệt văn bản và vùng nền tốt hơn.
- Phát triển các phương pháp nhận dạng văn bản
- Ứng dụng trên các thiết bị di động.
- Phát triển hệ thống để có thể xử lý trên chữ Việt

Tài liệu tham khảo

1. N. Dalal, Finding People in Images and Videos, 2006.
2. B. Epshtein, E. Ofek, and Y. Wexler, Detecting text in natural scenes with stroke width transform, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2963-2970, 2010.
3. S. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, ICDAR 2003 robust reading competitions, In Proceedings of ICDAR, pp. 682 – 687, 2003.
4. A. Shahab, F. Shafait, and A. Dengel, ICDAR 2011 robust reading competition challenge 2: Reading text in scene images In ICDAR 2011, pp. 1491–1496, 2011.
5. P. Soille, Morphological Image Analysis: Principles and Applications, Springer, 2003, pp. 182–198 .
6. V. N. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.

Bài báo

- Thuy Ho, Ngoc Ly, A scene text-based image retrieval system, IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2012. (Under review)

Cám ơn Thầy Cô và các bạn!