

KHỚP MẪU TEMPLATE MATCHING

Báo cáo môn học: Lý thuyết nhận dạng

Giảng viên: TS. Nguyễn Hữu Tiến

Sinh viên: Nguyễn Minh Quân, KSTN Toán Tin K61

Tháng 6 năm 2020

Nội dung

- ① Tổng quan về khớp mẫu
- ② Độ đo dựa trên kỹ thuật tìm kiếm đường đi tối ưu
- ③ Độ đo dựa trên độ tương quan
- ④ Sử dụng thư viện OpenCV khớp mẫu

Where's Waldo?



Bài toán

Trong lĩnh vực xử lý ảnh (computer vision),

xác định một hình ảnh mẫu X (template/sample) có trong hình ảnh nguồn Y (source image) hay không, nếu có thì khoanh vùng khu vực đó.

Trong lĩnh vực xử lý ngôn ngữ (natural language processing),

tìm ra tất cả các lần xuất hiện của mẫu X trong văn bản Y.

Trong lĩnh vực xử lý âm thanh (speech recognition)

nhận dạng giọng nói.

Template matching

template/ pattern là một bản sao, sự giống nhau, sự tương đồng, một ví dụ đại diện, ...



matching để so sánh về sự tương đồng, kiểm tra sự giống nhau.

Độ đo dựa trên kỹ thuật tìm kiếm đường đi tối ưu (Measure based on optimal searching techniques)

Các chuỗi (vector mẫu) có độ dài khác nhau

$$r(i), i = 1, 2, \dots, I \text{ và } t(j), j = 1, 2, \dots, J$$

Biểu diễn bằng lưới hai chiều, với I điểm theo chiều ngang và J điểm theo chiều dọc. Mỗi điểm (i, j) trên lưới là khoảng cách giữa $r(i)$ và $t(j)$.

Đường đi từ nút đầu (i_0, j_0) đến nút cuối (i_f, j_f) là một tập hợp các nút

$$(i_0, j_0), (i_1, j_1), \dots, (i_f, j_f)$$

.

Độ đo dựa trên kỹ thuật tìm kiếm đường đi tối ưu

Hàm giá trị D được tính theo công thức

$$D = \sum_{k=0}^{K-1} d(i_k, j_k)$$

$d(i_k, j_k)$ là khoảng cách giữa hai phần tử tương ứng của hai chuỗi, K là số node trên đường đi.

Khoảng cách (distance) giữa hai chuỗi được định nghĩa là giá trị nhỏ nhất của D trên tất cả các đường dẫn có thể.

Nguyên tắc tối ưu Bellman (Bellman's optimality principle)

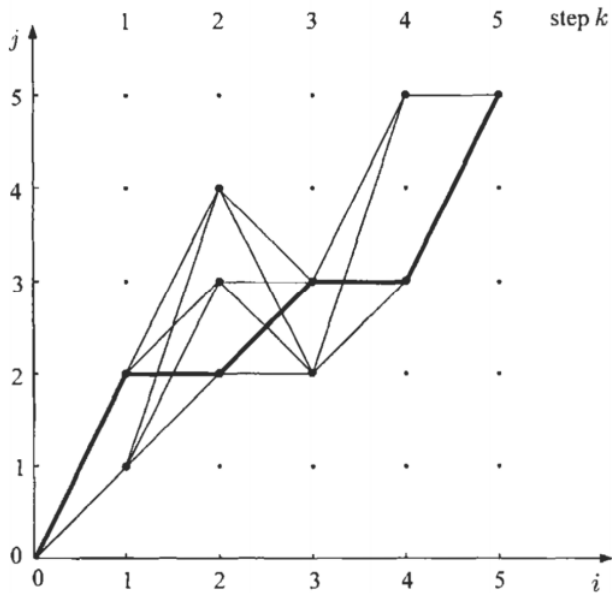
Thuật toán nhanh: Nguyên tắc Bellman

$(i_0, j_0) \xrightarrow{otp} (i_f, j_f)$ đường đi tối ưu

$$(i_0, j_0) \xrightarrow[(i,j)]{otp} (i_f, j_f) = (i_0, j_0) \xrightarrow{otp} (i, j) \oplus (i, j) \xrightarrow{otp} (i_f, j_f)$$

$$D_{min}(i_k, j_k) = \min_{i_{k-1}, j_{k-1}} [D_{min}(i_{k-1}, j_{k-1}) + d(i_k, j_k | i_{k-1}, j_{k-1})]$$

Nguyên tắc tối ưu Bellman



Khoảng cách chỉnh sửa (The Edit distance)

The Edit distance giữa hai chuỗi A và B , kí hiệu là $D(A, B)$, được định nghĩa là tổng nhỏ nhất của số thay đổi C , chèn I và xóa R cần thiết để thay đổi mẫu A thành mẫu B

$$D(A, B) = \min_j [C(j) + I(j) + R(j)]$$

trong đó j chạy qua tất cả các biến có thể để chuyển đổi A thành B .

Ứng dụng: chỉnh sửa tự động và truy xuất văn bản, tin sinh học,...

Khoảng cách Euclide (Euclidean distance)

Xét l là ảnh nguồn (ảnh xám) và g là ảnh mẫu (ảnh xám) có kích thước $m \times n$.

$$d(l, g, r, c) = \sqrt{\sum_{i=1}^n \sum_{j=1}^m (l(r+i, c+j) - g(i, j))^2}$$

trong đó (r, c) là tọa độ góc cao bên trái của mẫu g trên hình l .

Độ cong vênh thời gian trong nhận dạng giọng nói (Dynamic time warping in speech recognition)

Giả thiết: Nhận dạng từ rời rạc hoặc bị cô lập

Trong phân tích chuỗi thời gian, sự biến dạng thời gian động (DTW) là thuật toán

- ▶ Đo độ tương tự giữa hai chuỗi thời gian
- ▶ Có thể thay đổi tốc độ

Độ đo dựa trên độ tương quan (Measures based on correlations)

Đối tượng nghiên cứu là ảnh xám (ảnh đen trắng).

Xét ma trận ảnh mẫu với kích thước $M \times N$

$$r(i, j), i = 0, 1, \dots, M - 1, j = 0, 1, \dots, N - 1$$

và ma trận ảnh nguồn kích thước $I \times J$

$$t(i, j), i = 0, \dots, I - 1, j = 0, \dots, J - 1, M \leq I, N \leq J$$

Mục tiêu: Xây dựng một phương pháp phát hiện hình ảnh $M \times N$ trong $t(i, j)$ khớp với $r(i, j)$.

Độ đo dựa trên độ tương quan

$$D(m, n) = \sum_{i=m}^{m+M-1} \sum_{j=n}^{n+N-1} |t(i, j) - r(i - m, j - n)|^2$$

Tìm kiếm (m, n) sao cho $D(m, n)$ nhỏ nhất.

Khó khăn: Khối lượng tính toán lớn, khó xác định ngưỡng và nhạy cảm với tín hiệu nhiễu.

Hệ số tương quan chéo (cross-correlation coefficient)

$$D(m, n) = \sum_i \sum_j |t(i, j)|^2 + \sum_i \sum_j |r(i, j)|^2 - 2c(m, n)$$

trong đó

$$c(m, n) = \sum_i \sum_j t(i, j)r(i - m, j - n)$$

Hệ số tương quan chéo

$$c_N(m, n) = \frac{c(m, n)}{\sqrt{\sum_i \sum_j |t(i, j)|^2 \sum_i \sum_j |r(i, j)|^2}}$$

Độ tương quan (Correlations)

$$cor = \frac{\sum_{i=0}^{N-1} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{N-1} (x_i - \bar{x})^2 \sum_{i=0}^{N-1} (y_i - \bar{y})^2}}$$

trong đó,

x, y : là hình ảnh mẫu xám, hình ảnh nguồn

\bar{x}, \bar{y} : mức xám trung bình trong ảnh mẫu, trong ảnh nguồn

$N = m \times n$ kích thước ảnh mẫu.

Xem xét và cải tiến tính toán

(Computational consideration and improvement)

Tính toán trực tiếp

Số phép cộng và phép nhân $(2p + 1)^2 MN$

Cải thiện

- ▶ Tìm kiếm logarit hai chiều (Two-dimensional logarithmic search)
- ▶ Tìm kiếm phân cấp (Hierarchical search)
- ▶ Phương pháp tuần tự (Sequential methods)

Khớp mẫu biến dạng

(Deformable template matching)

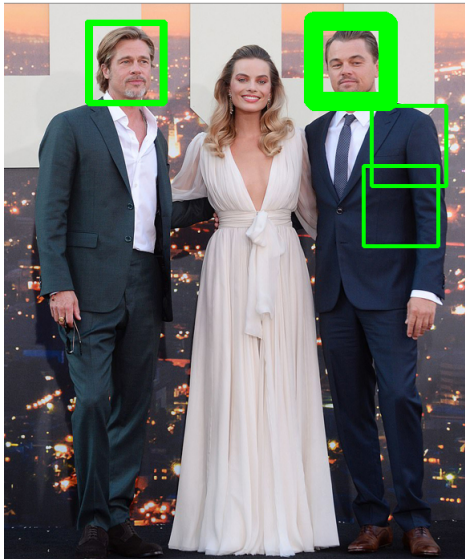
Template matching OpenCV python



Template matching OpenCV python



Template matching OpenCV python



Tài liệu tham khảo

Chapter 8: Template matching, Pattern Recognition, Sergios Theodoridis, paper 321-349.



THANK
YOU
FOR
WATCHING
MY SLIDESHOW