

ĐỒ ÁN MÁY HỌC

INTERESTING OF IMAGES



Khoa Công nghệ thông tin
Đại học Khoa học tự nhiên TP HCM

MỤC LỤC

Các nội dung chính	1
Thông tin nhóm	2
Phân công	2
Nội dung	3
1. Unusualness	3
2. Aethesis	4
3. General Preferences.....	4
4. Tổng hợp	5

Các nội dung chính

- Unusualness
- Aethesis
- General Preferences
- Tổng hợp, kết quả thực nghiệm

Thông tin nhóm

MSSV	Họ Tên	Email
1312621	Lê Đông Triều	Trieule95@gmail.com
1312657	Huỳnh Văn Tuấn	Tunhuh95@gmail.com
1312681	Trần Tuấn Tú	Trantuantu020995@gmail.com

Phân công

Tên	Công việc
Lê Đông Triều	Đọc hiểu toàn bộ nội dung bài báo để viết báo cáo. Tính các đặc trưng trên tiêu chí tính thẩm mỹ của một ảnh.
Huỳnh Văn Tuấn	Thực hiện cài đặt để tính các đặc trưng liên quan đến tính bất thường của một bức ảnh. Huấn luyện các đặc trưng để tính độ hấp dẫn của các ảnh.
Trần Tuấn Tú	Thực hiện cài đặt để tính mối tương quan giữa hạng mục cảnh (scenes catagories) và độ hấp dẫn của một bức ảnh.

Nội dung

Tính toán máy học cho việc dự đoán độ thú vị

Ở mục này chúng tôi đề xuất những đặc trưng quan trọng là: tính bất thường, tính thẩm mỹ và những sở thích chung. Sau đó chúng tôi sử dụng chúng để dự đoán độ thú vị của những bức ảnh. Bằng cách đó chúng tôi xây dựng nhờ vào tập các nét đặc trưng chính và dùng chúng trước khi làm việc với các bức ảnh và mở rộng với các đặc trưng phụ của một bức ảnh.

Một cách chính thức, cho một bức ảnh I chúng ta tìm một điểm số s cho độ thú vị của bức ảnh. Hướng đi để đạt được việc này gồm 2 bước: (i) khám phá những đặc trưng khác nhau để lọc ra những dấu hiệu cho độ thú vị và (ii) kết hợp những đặc trưng riêng lẻ.

1. Unusualness

Theo một số nguyên cứu cho thấy rằng, sự bất thường là một dấu hiệu quan trọng để con người cảm thấy ảnh này hấp dẫn hơn ảnh kia.

Để có được các đặc trưng trong sự bất thường ta sẽ sử dụng hai phương pháp để tính một là tính các ngoại lệ trên toàn cục của ảnh (global outlier), và các bộ phận cấu thành nên ảnh đó (Composition of parts).

Global outlier:

Sử dụng thuật toán LOF (Local outlier factor) ước lượng mật độ của một điểm đến k điểm gần nhất của nó trên các global image descriptors như raw RGB $s_{pixel}^{unusual}$, GIST $s_{GIST}^{unusual}$, Spatial Pyramid on SIFT Histogram $s_{pyr}^{unusual}$.

Composition of parts:

Chúng ta sử dụng các superpixel bằng thuật toán SLIC, mỗi superpixel sẽ được xem như một node của đồ thị. Đồ thị có cạnh nối khi và chỉ khi hai node là hai superpixel kề nhau.

Sử dụng hàm năng lượng sau:

$$E(L) = \sum_{i \in S} D_i(l_i) + \lambda \sum_{\{i,j\} \in N} V(l_i, l_j).$$

Trong đó, S là tập các superpixel, N là tập các superpixel láng giềng.

Hàm chi phí $D_i(l_i)$ khoảng cách Eclidean trong không gian mô tả (descriptor space) của superpixel i đến superpixel láng giềng gần nhất.

Hàm chi phí $V(l_i, l_j)$ là 0 nếu $l_i = l_j$, 1 nếu ngược lại.

Theo một số thực nghiệm thì chọn $\lambda = 0.02$ là cho hiệu suất tốt nhất.

Đồng thời để có chi phí nhỏ nhất thì ta sử dụng thuật toán GraphCut. Từ các giá trị nhận tối ưu của một ảnh ta sẽ tính $s_{compose}^{unusual} = E(L) / |S|$

2. Aethesis

Để dự đoán tính thẩm mỹ, chúng tôi tập trung vào hình ảnh khi nhìn cảm thấy dễ chịu.

Độ sắc sỡ. Chúng tôi đo lường độ sắc sỡ dựa trên đề xuất của Datte và Wang, với Earth Mover distance (trong không gian màu LUV) của biểu đồ màu của một bức ảnh H_I đến biểu đồ màu thông thường H_{uni} . Một biểu đồ màu thông thường là sắc sỡ nhất có thể, vì vậy nó nhỏ hơn khoảng cách của nhiều bức ảnh sắc sỡ, $s_{colorful}^{aesth} := -EMD(H_I, H_{uni})$.

Tính kích thích. Machadijk và Hanbury rút trích điểm số cảm xúc từ những raw pixels. Những đặc trưng này dựa trên kinh nghiệm. Tính kích thích là một yếu tố cho sự thú vị, chúng tôi sử dụng chỉ số kích thích như trong “Affective image classification using features inspired by psychology and art theory”. Điểm số kích thích được tính bằng cách lấy trung bình tất cả các pixel trong ảnh $s_{arousal}^{aesth} := \sum_p -0.31 \text{ brightness}(p) + 0.6 \text{ saturation}(p)$.

Độ phức tạp. Để bắt được độ phức tạp của một bức ảnh, chúng tôi so sánh kích thước của bức ảnh sau khi nén JPEG với kích thước gốc, $s_{complex}^{aesth} := \frac{\text{bytes}(\text{compress}(I))}{\text{bytes}(I)}$. Chúng tôi dùng JPEG như một dạng nén mất dữ liệu, cái mà nén một hình ảnh nhưng mắt người không thể nhận biết được đã mất một số dữ liệu so với ảnh gốc, nếu điểm số nén cao thì $s_{complex}^{aesth}$ thấp, nên ảnh đó có ít thông tin quan trọng.

Độ tương phản. Chúng tôi tính toán khoảng cách nhỏ nhất của 98% khối của một biểu đồ thang độ xám để thu được $s_{contrast}^{aesth}$.

Sự sắp xếp cạnh. Chúng tôi tính toán phạm vi bao của ảnh chứa 98% cấu cạnh ở mỗi kích thước. Phạm vi bao càng nhỏ thì càng ít sự lộn xộn và nền đều hơn. $s_{edges}^{aesth} := 1 - w_x w_y$

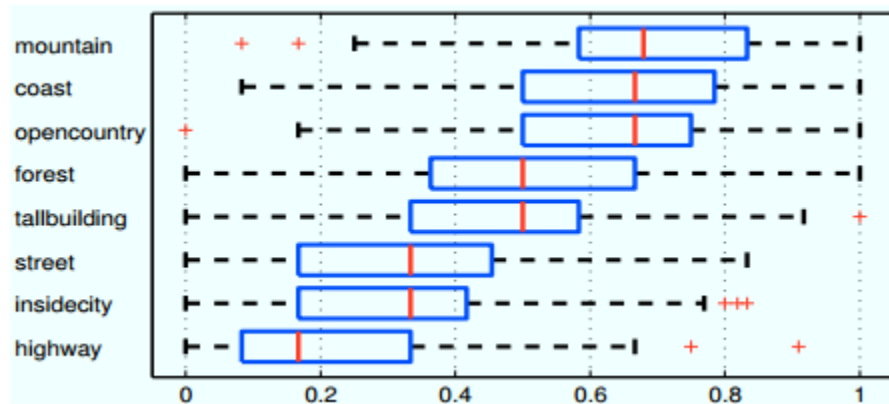
3. General Preferences

Các hình ảnh thuộc các cảnh khác nhau có mối liên quan trực tiếp đến độ ấn tượng của một bức ảnh.

Để phân lớp được các hạng mục cảnh (scene catagories), chúng ta có thể rút trích đặc trưng của bức ảnh. Ở đây, ta rút trích đặc trưng GIST, RGB pixel, SIFT và Color histogram

sau đó train chúng trên Support Vector Regressor (v-SVR) sử dụng RBF kernel. Tối ưu tham số với $\nu = 0.5$, $\gamma = \{2^{-15}, 2^{-13}, \dots, 2^1, 2^3, 2^5\}$ và $C = \{2^{-5}, 2^{-3}, \dots, 2^7, 2^9\}$

Điểm số interesting của mỗi bức ảnh dựa và hạng mục cảnh được xây dựng bằng cách tính toán số lượt xem của mỗi bức ảnh. Mỗi bức ảnh sẽ được xem qua bởi 11.9 chuyên gia. Số lượt xem của ảnh sẽ được thống kê và xấp xỉ tỉ lệ thuận với điểm số thú vị của một bức ảnh. Sau khi thống kê, mối tương quan giữa hạng mục cảnh và độ thú vị của một bức ảnh như sau:



Những cảnh tự nhiên sẽ gây ấn tượng hơn những cảnh nhân tạo, những ngọn núi sẽ có điểm thú vị cao, ngược lại những con đường sẽ có điểm thú vị thấp.

4. Tổng hợp

Sử dụng linear model trên hàm độ lỗi là bình phương độ lỗi nhỏ nhất. Để thực hiện huấn luyện trên một tập ảnh trong tập train, để tìm ra các trọng số thích hợp cho việc tính độ hấp dẫn của những ảnh khác. Để thực hiện một cách hiệu quả, ta thực hiện làm sạch dữ liệu để có hiệu suất tốt nhất, đồng thời các đặc trưng cũng phải được chuẩn hóa, ta dùng hàm sigmoid để các giá trị phải trong đoạn từ $[0 - 1]$.

Những gì đã làm được

Tính được độ hấp dẫn dựa trên độ bất thường của các đặc trưng như pixel, GIST, Spatial Pyramid on SIFT histogram.

Thực hiện cài đặt rút trích các đặc trưng GIST, SIFT và train chúng trên SVM để phân loại hạng mục cảnh của ảnh (Scenes categories) từ đó trả về kết quả về tính liên quan giữa độ hấp dẫn với hạng mục cảnh của ảnh.

Ở đây, rút trích đặc trưng GIST, SIFT sau đó train chúng trên Support Vector Regressor (v-SVR). Đối với đặc trưng GIST, sử dụng RBF kernels với $\sigma = 0.6$, đối với đặc trưng SIFT, sử dụng histintersection kernels, trong hai trường hợp đều áp dụng $\lambda = 0.2$.

Dữ liệu được train gồm 600 ảnh JPEG từ 15 hạng mục cảnh khác nhau ở kích thước 256x256 (Mỗi hạng mục có 40 ảnh).

Độ chính xác khoảng 80% trên việc phân loại cảnh.

Những gì chưa làm được:

Chưa cài đặt được đặc trưng compose của ảnh.

Chưa cài đặt được các đặc trưng về tính thẩm mỹ.

Chưa huấn luyện được các đặc trưng raw RGB, color histogram để tính độ tương quan giữa độ hấp dẫn và các hạng mục cảnh.

Kết quả thực nghiệm của nhóm:

Tuy chỉ huấn luyện được trên một số đặc trưng để tính độ hấp dẫn nhưng kết quả chạy cũng không có gì bất ổn lắm. Độ chính xác tương đối và dao động không nhiều giữa các lần thử.

