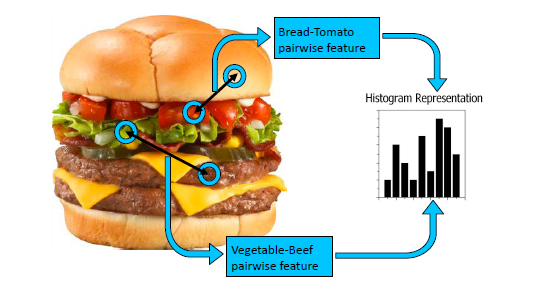
Nhận dạng thực phẩm sử dung thống kê theo cặp của đặc trưng cục bộ

Nhận dạng thực phẩm là khó khăn vì các đối tượng biến dạng có sự khác biệt lớn về ngoại hình. Chúng tôi cho rằng chìa khóa để nhận biết thực phẩm đó là mối liên hệ giữa các thành phần với nhau(Như là thịt và bành mỳ trong bánh sandwich). Chúng tôi đề xuất một đại diện mới cho các hạng mục thực phẩm là tính toán các cặp thống kê giữa các đặc trưng cục bộ. Chúng tôi tích lũy những số liệu thống kê trong một biểu đồ đa chiều, cái mà sử dụng như một vector đặc trưng cho phân loại. Thực nghiệm của chúng tôi cho thấy rằng đại diện đề xuất này chính xác đáng kể hơn khi xác định thực phẩm so với các phương pháp đã có trước đó.

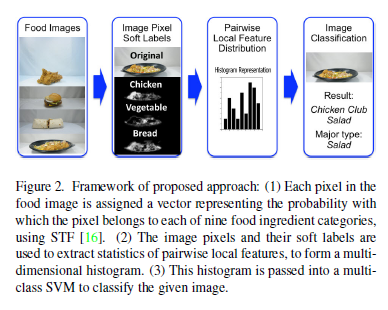
1. Giới thiệu

Nhận dạng thực phẩm tự động đang nổi lên như một đề tài nghiên cứu quan trọng trong việc nhận dạng đối tượng bời vì nhu cầu công cụ đánh giá đối tượng tốt hơn để chống béo phì. Mục tiêu của hệ thống là cho phép mọi người hiểu rõ hơn về nội dùng thành phần dinh dưỡng của các chế độ ăn uống mà họ lựa chọn. Và cung cấp cho các chuyên gia y tế lượng tiêu thụ thức ăn của các bệnh nhân của họ. Người dùng không cần báo cáo 1 cách chính xác lượng thực phẩm mà họ sử dụng.

Thật không may, nhận dạng đối tượng cơ bản tiếp cận dựa trên tập hợp số liệu thống kê của mô tả đặc trưng cục bộ hiện đang kém về việc này bởi vì thực phẩm thường biến dạng và biến thể.Thậm chí trong trường hợp đối với các mặt hàng tương đối chuẩn từ danh sách các thực phẩm ăn nhanh. Nghiên cứu của chúng tôi là có động lực bởi vì việc quan sát một mặt hàng thực phẩm chủ yếu có thể được đặc trưng bới các thành phần của nó và mối quan hệ không gian tương ứng giữa chúng. Ví dụ bánh mỳ thường là 1 lớp thịt và bao quanh bởi 2 phía là lát bánh mỳ,trong khi món salad bao gồm các loại rau có bố cục sắp xếp không đều nhau. Giả thuyết của chúng tôi đưa ra là việc phát hiện bất cứ thành phần nào đưa ra cũng có thể là không đáng tin cậy,ta vẫn có thể lấy thông tin đầy đủ từ việc tập hợp các số liệu thống kê và các loại nguyên liệu theo cặp và sự sắp xếp theo không gian giữa chúng để xác định sự đáng tin cậy mặt hàng thực phẩm.



Mặc dù giả thuyết này mang tính trực quan, Áp dụng nó vào việc nhận dạng thực phẩm là khó khăn vì nhiều lý do.Đầu tiên, thậm chí một thực phẩm có chứa nhiều thành phần nhìn thấy riêng biệt ,nhưng không phải tất cả đều có thể nhìn thấy được trong 1 hình ảnh đưa ra do bị che lấp hoặc bị biến đổi do quá trình chế biến. Thứ 2, mặc dù độ chính xác hoàn hảo là không yêu cầu ở cấp độ thành phần(nguyên liệu), các phương pháp đề xuất không yêu cầu phân đoạn cấp pixel để tạo ra phân phối thông qua các thành phần từ các miếng ghép(bản vá) nhỏ cục bộ.Cuối cùng, có thể có sự biến đổi đặc biệt trong nội bộ về kích thước quan sát và mối quan hệ không gian giữa các thành phần, đòi hỏi chúng ta phải xây dựng các đại diện có thể đối phó, giải quyết sự xuất hiện các biến đổi này 1 cách có nguyên tắc.



Tiếp cận của chúng tôi được minh họa trong hình 2 và có thể tóm tắt như sau,Đầu tiên chúng tôi gán nhãn (phân phối thông qua thành phần ) cho mỗi pixel ảnh bằng cách sử dụng STF(sematic texon forest) . Ví dụ, sự phân bố tương ứng với 1 điểm ảnh trong mảnh màu đỏ có thể có khả năng cao cho "Cà chua" và thấp thấp "Phomat",trong khi các điểm ảnh khác lại có giá trị cao cho "Bánh" và "Phomat" và thấp cho "Cà chua".Tiếp theo chúng ta xây dựng 1 biểu đồ nhiều chiều các đặc trưng nơi mà mỗi bin tương ứng với 1 cặp nhãn thành phần và mối quan hệ hình học rời rạc giữa hai điểm ảnh.Ví dụ, theo quan sát thì cặp "bánh" và "cà chua" nhãn là 20-30pixel ngoài 1 góc 40-50 độ. Kể từ khi chúng tôi sử dụng nhãn mềm, mỗi quan sát từ 1 cặp pixel góp phần góp phần xác suất vào mỗi bin. Như vậy, Việc tổng hợp các biểu đồ từ nhiều mẫu của các cặp điểm ảnh trong ảnh phục vụ như là đại diện của phân bố theo không gian của các thành phần cho hình ảnh đó. Cuối cùng, chúng tôi xử lý biểu đồ này như là vector đặc trưng để nhận dạng mỗi thực phẩm. Đại diện của chúng tôi là không những củ thể cho thực phẩm mà còn phù hợp để nhận biết các đối tượng hoặc những cảnh nhận có thành phần trực quan sắp xếp theo cấu trúc không gian dự đoán được.

1. Công việc liên quan

Hiện có ít công việc liên quan đến nhận dạng thực phẩm. Shroff đề xuất một hệ thống có khả năng đeo được để nhận dạng thực phẩm theo dõi lượng calorie. Bolle đã phát triển một hệ thống sản xuất tự động ID gọi là VeggieVision để giúp quá trình thanh toán sản phẩm. Yang giới thiệu bộ dữ liệu PFID cho việc nhận dạng thực phẩm cùng với các tiêu chuẩn bằng cách sử dụng hai thuật toán: color histogram và BoS đặc trưng. Russo theo dõi việc sản xuất thức ăn nhanh bằng video. Wu and Yang phân tích video ăn uống để nhận dạng mặt hàng thực phẩm ướng tính lượng calorie tiêu thụ. Nhận dạng đối tượng đã là lĩnh vực nghiên cứu phổ biến trong thị giác máy tính nhiều năm qua. Chúng bao gồm các phương pháp tiếp cận dựa trên đặc trưng cục bộ như mô tả SIFT hoặc các đặc trưng toàn cục như color histograms hoặc đặc trưng GIST, hình ảnh thường được biểu diễn như là "túi" chứa các đặc trưng mà không có thong tin rõ ràng. Cách tiếp cận khác, ví dụ như "Mô hình chòm sao" và nhiều mở rộng của nó, đối tượng mô hình như là bộ sưu tập các bộ phận với một sắp xếp không gian dự đoán được. Điều đặc biệt quan tâm trong nhận dạng thực phẩm đó là phương pháp cái mà giải quyết vấn đề biến dạng đối tượng một cách rõ ràng. Trong ngữ cảnh hình dạng, lấy n pixel từ đường nét của 1 hình dạng , có được n-1 vector bằng cách kết nối một vector với những cái khác, và sử dụng những pixel này như là những mô tả về hình dạng tại các điểm ảnh. Leordeanu đề xuất nhận dạng sử dụng tương tác theo cặp cảu các tính năng đơn giản.

Cách tiếp cận tốt giải quyết nhiều vấn đề trong nhận dạng đối tượng. Nhưng có 2 vấn đề làm cho họ khó áp dụng trong nhận dạng thực phẩm. Thứ nhất, tất cả các phương pháp tiếp cận đều yêu cầu xác định các điểm đặc trưng có ý nghĩa như cạnh, đường viền, điểm chính , mốc. Nhưng những đặc trưng này thường không có sẵn trong ảnh của thực phẩm. Thứ 2, khai thác hình dạng giống nhau là khó vì thực phẩm thường có hình dạng khá vô định.

Để vượt qua nhưng thử thách như vậy, cách tiếp cận của chúng tôi không dựa trên các phát hiện của các tính năng như cạnh hay điểm chính . Thay vào đó, chúng tôi sử dụng thống kê đặc trưng cục bộ được xác định qua cặp lựa chọn ngẫu nhiên của các điểm ảnh.

Sự thống kê đặc trưng cục bộ theo cặp hiệu quả ghi lại đặc điểm hình dạng quan trọng và mối quan hệ trong không gian giữa các thành phần thực phẩm, tạo điều kiện cho sự nhận dạng chính xác hơn.

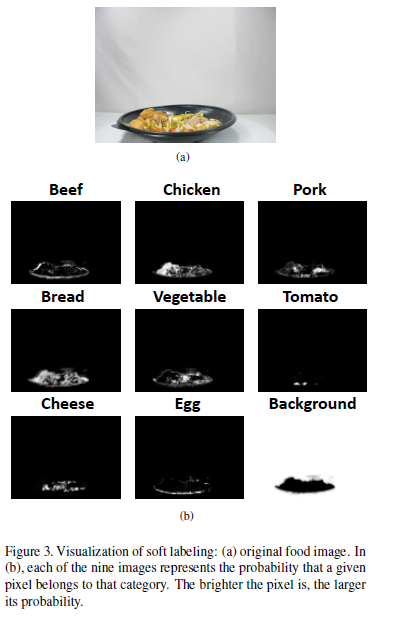
1. Phân bố theo cặp các đặc trưng cục bộ(PFD)

Trong phân này, chúng tôi đưa ra định nghĩa về đại diện hình ảnh là thống kê theo cặp các đặc trưng cục bộ. Sử dụng đại diện này , mỗi ảnh có thể được đại diện bởi một biểu đồ đa chiều

3.1 Nhãn mềm của Pixel

Trước khi có được thống kê của các đặc trưng cục bộ, chúng tôi phân loại tất cả các tất cả các điểm ảnh thành nhiều loại dựa trên sự xuất hiện của các mảnh ghép xung quanh điểm ảnh. Đặc trưng cục bộ sẽ được thu thập 1 cách riêng biệt dựa trên các loại pixel khác nhau.Chúng tôi sử dụng 9 loại handpicked đại diện cho 8 thành phần trong thức ăn nhanh: beef, chicken, pork, bread, vegetable, tomato/tomato sauce, cheese/butter, egg/other,cộng thêm 1 loại cho nền.

Thay vì ghi nhãn mỗi điểm ảnh là một thành viên của một thể loại duy nhất, chúng tôi sử dụng nhãn mềm để tạo ra một vector của các xác suất tương ứng với khả năng rằng các điểm ảnh được thuộc về mỗi trong chín loại. Ghi nhãn mềm hoãn cam kết về danh tính của một điểm ảnh, cho phép cho sự không chắc chắn về các nhãn đúng cho một điểm ảnh, và cho sự phân biệt giữa các loại thành phần mờ.

Nhiều phương pháp có thể được sử dụng để phân loại pixel. Chúng tôi đã chọn để sử dụng Semantic Texton Forest (STF). STF là một phương pháp để phân loại hình ảnh và phân đoạn mà tạo ra nhãn mềm cho một điểm ảnh dựa trên đặc điểm ở mức độ thấp cục bộ (như màu sắc của điểm ảnh lân cận). Điều này đạt được bằng cách sử dụng cụm công của cây quyết định rằng mỗi được đào tạo trên một phần nhỏ của hình ảnh phân đoạn-thủ công. Một điểm ảnh được phân loại bằng cách giảm dần mỗi cây từ gốc đến lá dựa trên tính năng cấp thấp tính xung quanh các điểm ảnh được đưa ra. Lá chứa phân lớp xác suất (trên thành phần) là thuật toán của chúng tôi sử dụng để tính toán các nhãn mềm mại của nó. Chúng tôi đã chọn để sử dụng STF bởi vì nó kết hợp kiến thức của con người trong các hình thức đào tạo nhãn hình ảnh bằng tay, và bởi vì nó đã được chứng minh hiệu quả để ghi nhãn mềm trong lĩnh vực đầy thách thức khác như nhiệm vụ PASCAL Visual Object Classes (VOC).

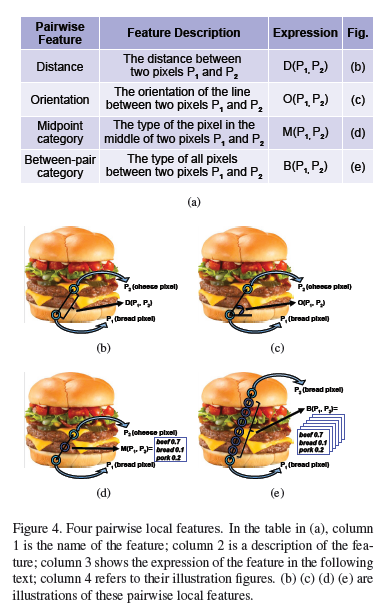
Với STF, mỗi điểm ảnh được gán một vector với xác suất K, trong đó K là số lượng các loại pixel. K = 9 trong trường hợp của chúng tôi.

3.2 Đại diện thành phần toàn cục(GIR)

Cách đơn giản nhất để sử dụng các nhãn điểm ảnh mềm mại được tạo bởi STF để đại diện cho một hình ảnh thực phẩm là để tạo ra một biểu đồ 1 chiều với 8 bin đại diện thường xuyên làm thế nào mỗi 8 loại thành phần thực phẩm xuất hiện trong hình ảnh. Để tạo một biểu đồ tổng thể đại diện cho sự phân bố của các thành phần cho hình ảnh, chúng tôi tổng hợp các nhãn mềm cho 8 thành phần thực phẩm cho tất cả các điểm ảnh không nền trong một hình ảnh. Sau đó, chúng tôi chuẩn hóa các biểu đồ bằng số điểm ảnh không nền trong ảnh. Đại diện thành phần toàn cục này (Gir) là trực quan và dễ dàng để tính toán, nhưng nó không nắm bắt được mối quan hệ không gian giữa các thành phần rất quan trọng để phân biệt một thực phẩm với cái khác. Nó sẽ phục vụ như là một chuẩn mực hữu ích để so sánh với, cặp đặc trưng cục bộ phức tạp hơn mô tả dưới đây.

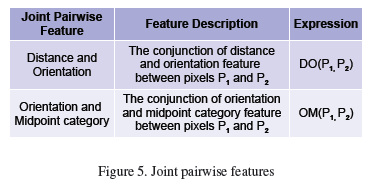
3.3 Cặp đặc trưng

Chúng tôi cố gắng để nắm bắt các mối quan hệ không gian giữa các điểm ảnh của các thành phần thực phẩm khác nhau sử dụng các tính năng cặp địa phương.

Khoảng cách cặp phản ánh khoảng cách giữa hai điểm ảnh trong một hình ảnh. Tương tự như các đặc mô tả trong phương pháp Shape Context, các đặc khoảng cách cặp được định nghĩa là lớn hơn khi hai điểm ảnh trong cặp gần nhau. Chúng tôi sử dụng công thức Sturges ', k = log2 [n + 1] và sử dụng các bản ghi của các khoảng cách giữa các cặp điểm ảnh P1 và P2.

D(P1,P2) = log[|P1, P2| + 1]

Định hướng cặp, DO (P1, P2), được định nghĩa là góc của đường thẳng nối các cặp điểm ảnh. Nó dao động trong [0; 360], như là một đặc trưng chung của khoảng cách và định hướng, DO (P1, P2), một đặc trưng chung của các định hướng và điểm giữa, OM (P1; P2), như thể hiện trong hình. 5.



3.4 Biểu đồ đại diện cho phân phối cặp đặc trưng

Tính toán tập hợp đầy đủ các đặc trưng cặp trong một hình ảnh có thể tốn kém chi phí tính toán, cho một hình ảnh với M pixel, chúng tôi sẽ cần phải xem xét M/2 cặp. nó đủ để xem xét M/2 chứ không phải là M\*M cặp vì mối quan hệ của chúng tôi là cặp đối xứng; chúng tôi chuyển đổi trực tiếp định hướng đến một hình thức đối xứng bằng trung bình tính theo cả hai hướng. Tuy nhiên, M/2 vẫn có thể gây tốn lớn cho hình ảnh có độ phân giải cao.

Như vậy, để tính toán hiệu quả, chúng tôi lấy mẫu ngẫu nhiên N (N = 1000) pixels từ phần không chứa nền của hình ảnh và ước tính số liệu thống kê cặp sử dụng các N / 2 cặp pixel. Chúng tôi sử dụng một tập hợp P đại diện cho các điểm ảnh N chúng tôi chọn ngẫu nhiên từ một hình ảnh: P = (P1, P2, .... Pn). Các nhãn mềm của N điểm ảnh được đại diện là S = {Sik | j = 1,2,3, ... n; k = 1,2,3,4,5,6,7,8}. Chúng tôi tính toán cặp đặc trưng cục bộ (D (Pi; Pj), O (Pi; Pj), M (Pi; Pj) và B (Pi; Pj)) cho tất cả các pixel, và tích lũy giá trị của họ vào một phân phối. Sự phân bố được cân bằng các nhãn mềm của hai điểm ảnh. Cụ thể, trọng lượng này là sản phẩm của các xác suất mà một điểm ảnh được gán cho nhãn thành phần tương ứng với bin nhận được.

Chúng tôi sử dụng một biểu đồ đa chiều để đại diện cho sự phân bố cáccặp đặc trưng cục bộ, như chi tiết trong hình. 6. Hai chiều đầu tiên của biểu đồ là nhãn thành phần. Các chiều khác là những trọng số cặp đặc trưng tính toán cho các cặp điểm ảnh.Chiều đầu tiên và thứ hai của biểu đồ đa chiều trên có 8 thùng, đại diện cho tám loại pixel (trừ background). Các chiều khác có hoặc 12 thùng cho khoảng cách cặp hoặc định hướng, hoặc 8 thùng cho loại điểm giữa hoặc giữa đôi.

3.5 Chuẩn hóa Histogram

Chúng tôi chuẩn hóa sự phân bố của một số cặp đặc trưng nhất định để đạt được sự bất biến của phép biến đổi như xoay và tỷ lệ. Chúng tôi tính toán các chế độ về khoảng cách cặp hoặc định hướng, và thay đổi kích thước tương ứng của biểu đồ đa chiều để nó tập trung vào chế độ. Kể từ khi chúng tôi áp dụng một logarit để tính năng khoảng cách, sự thay đổi quy mô của một hình ảnh trở nên tương ứng để chuyển dịch biểu đồ của nó trong chiều khoảng cách, dẫn đến bất biến tỷ lệ. Tương tự như vậy, tập trung các biểu đồ định hướng vào kết quả giá trị của nó trong chế độ bất biến phép quay. Vì chúng ta chỉ đo được khoảng cách tương đối giữa các điểm ảnh, đại diện của chúng tôi cũng là bất biến dịch

3.6 Phân loại với phân phối đặc trưng cục bộ

Sử dụng PFD hoặc đại diện thành phần toàn cục, mỗi hình ảnh được biểu diễn như là một biểu đồ đa chiều. Chúng tôi sử dụng Support Vector Machine (SVM) để phân loại bằng nhân X2. Xấp xỉ đối xứng X2 được định nghĩa như sau : 

x và y là đại diện biểu đồ PFD của 2 ảnh. xi,yi là đại diện cho 2 bin của biểu đồ. Nhân X2 được định nghĩa như sau: 

Với tính toán hạt nhân ma trận trước, chúng tôi phân loại thực phẩm bằng SVM.

1. Phương pháp thực nghiệm

Trong phần này, chúng tôi lại rút gọn dữ liệu của chúng tôi, cách tiếp cận cơ bản và chi tiết việc thực hiện các phân khúc cấp pixel sử dụng cái mà chúng tôi tạo ra thành phần nhãn mềm của chúng tôi. Kết quả thí nghiệm được trình bày trong phần 5.