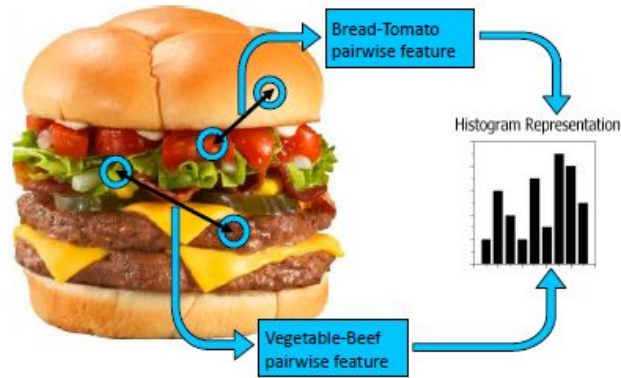


Food Recognition Using Statistics of Pairwise Local Feature

**Autores: Shulin (Lynn) Yang - University of Washington
Mei Chen - Intel Labs Pittsburgh
Dean Pomerleau - Robotics Institute, Carnegie Mellon
Rahul Sukthankar - Robotics Institute, Carnegie Mellon
2010, IEEE**

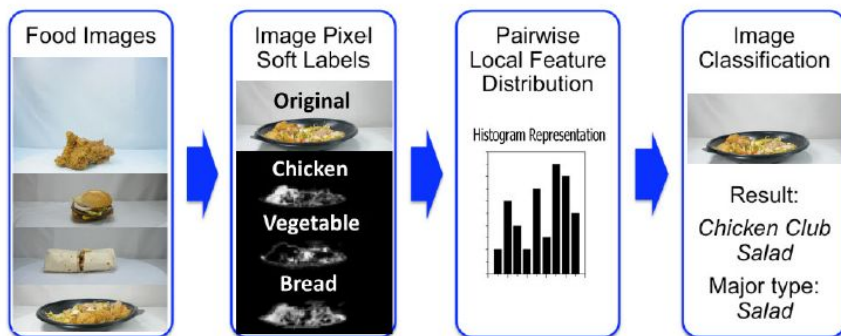
Introdução

- Comidas podem apresentar deformações e variações em sua forma, além de conter muitos ingredientes, nem todos visíveis;
- Métodos tradicionais de reconhecimento requerem a detecção de características significativas, como bordas, contornos e keypoints (difíceis de se extrair em imagens de comida);
- Os autores propõem caracterizar comidas por seus ingredientes e as relações espaciais entre eles;



Objetivos

1. O método desenvolvido usa características de pares de pixels para fazer estatísticas e formar um histograma multi-dimensional e usa SVM para classificar a imagem;
2. Um Soft Label é atribuído a cada pixel da imagem usando STF;
3. Um histograma multi-dimensional é construído a partir disso;
4. Finalmente, o histograma é tratado como um vetor de características num classificador, para reconhecer o item alimentício;



Referencial Teórico

STF

- Sematic texton forest é um método para categorização e segmentação de imagens que gera soft labels pra um pixel de acordo com características locais básicas (cor, por exemplo);
- O STF é adquirido por meio de conjuntos de árvores de decisão treinadas;
- O autor usou STF porque esse método de classificação usa conhecimento humano para rotular as imagens na parte do treino;
- É extremamente rápido tanto para treinar quanto para testar.

Mais informações : <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4587503>

Referencial Teórico

SVM

- Conjunto de métodos do aprendizado supervisionado que analisam os dados e reconhecem padrões e classifica os dados;
- O SVM padrão toma como entrada um conjunto de dados e prediz, para cada entrada, qual de duas possíveis classes a entrada faz parte;
- Usado no aprendizado de máquina.

Mais informações : <http://nca.ufma.br/~geraldo/vc/12.SVM.pdf> http://www.cs.columbia.edu/~kathy/cs4701/documents/jason_svm_tutorial.pdf

www.seer.ufrgs.br/rita/article/download/rita_v14_n2_p43-67/3543

Metodologia

— — —

Soft labeling

- Em vez de rotular cada pixel da imagem como membro de uma só categoria, cada pixel recebe um vetor de probabilidades correspondendo às chances daquele pixel ser um dado ingrediente;
- Todos os pixels da imagem são classificados em algumas categorias. Os autores usam nove, como podemos ver na figura:

Metodologia



(a)



(b)

Metodologia

— — —

Global ingredient representation

- Trata-se de um histograma unidimensional com 8 bins representando a frequência de cada ingrediente, com exceção do background em uma dada imagem de comida. É a forma mais simples de usar os soft labels gerados pelo STF;
- Esse método não captura as relações espaciais entre os ingredientes, o que é de extrema importância pra distingui-los.

Metodologia - Extração de Características Espaciais

- Nessa etapa, tenta-se capturar a relação espacial entre os ingredientes. Foram definidas quatro características para os pares de pixels:

Pairwise Feature	Feature Description	Expression	Fig.
Distance	The distance between two pixels P_1 and P_2	$D(P_1, P_2)$	(b)
Orientation	The orientation of the line between two pixels P_1 and P_2	$O(P_1, P_2)$	(c)
Midpoint category	The type of the pixel in the middle of two pixels P_1 and P_2	$M(P_1, P_2)$	(d)
Between-pair category	The type of all pixels between two pixels P_1 and P_2	$B(P_1, P_2)$	(e)

Figura --: Características individuais

$$D(P_1, P_2) = \log[|P_1, P_2| + 1]$$

$$[0^\circ, 360^\circ)$$

$$\bar{B}(P_1, P_2) = \{T_i | i = 1, 2, \dots, t\}$$

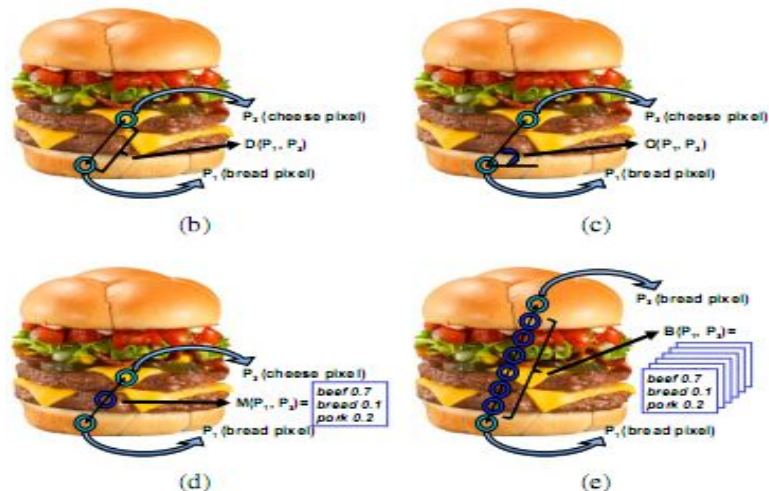


Figura --: Ilustração das quatro local features

Metodologia - Extração de Características Espaciais

- Além das características individuais, foram definidas duas compostas:

Joint Pairwise Feature	Feature Description	Expression
Distance and Orientation	The conjunction of distance and orientation feature between pixels P_1 and P_2	$DO(P_1, P_2)$
Orientation and Midpoint category	The conjunction of orientation and midpoint category feature between pixels P_1 and P_2	$OM(P_1, P_2)$

Figura --: Ilustração das quatro local features

Histograma de Representação da Distribuição de Características de Pares (PFD)

- É o histograma que irá representar a relação espacial entre os ingredientes usando as características vistas anteriormente.
- Como calcular essas características para todos os pixels da imagem pode ser computacionalmente caro, são escolhidos, de forma randômica, N ($N=1000$) pixels da porção que não faz do background, o que corresponde à $(1000!/2!*998!)$ pares de pixels pela combinação simples;
- As características locais ($D(P_i,P_j)$, $O(P_i,P_j)$, $M(P_i,P_j)$ e $B(P_i,P_j)$), depois de calculadas para os N pixels, são acumuladas no histograma de distribuição.
- As duas primeiras dimensões têm 8 bins, representando as 8 categorias de pixel. As dimensões da distância e orientação têm 12 bins cada uma; já as dimensões para as categorias do midpoint e between-pair têm 8 bins.
- Para obter a invariância quanto à escala e rotação, o histograma de distribuição é normalizado. Utilizam-se os valores de moda da distância e orientação.

Histograma de Representação da Distribuição de Características (PFD)

Histogram Type	Histogram Description	Size
Distance	label × label × distance	8×8×12
Orientation	label × label × orientation	8×8×12
Midpoint category	label × label × midpoint	8×8×8
Between-pair category	label × label × between-pair	8×8×8
Distance and Orientation	label × label × distance × orientation	8×8×12×12
Orientation and Midpoint category	label × label × orientation × midpoint	8×8×12×8

Figura --: Representação do Histograma de Distribuição (PFD)

Classificação

— — —

- Na etapa de classificação, cada imagem é representada como um histograma multidimensional. Esse diagrama pode ser do tipo **PDF** (baseado na distribuição das características em pares) ou **GIR** (representa a distribuição de ingredientes na imagem, não capturando a relação espacial entre os ingredientes).
- Para classificação foi utilizado Support Vector Machine (SVM) com kernel x^2 , dado pela função 1. Já o valor aproximado de x é definido pela função 2..

$$(1) k_{\chi^2}(x, y) = e^{-d_{\chi^2}(x, y)}$$

$$(2) d_{\chi^2}(x, y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$$

- onde x e y representam os histogramas de distribuição de características (PDFs) de duas imagens, e x_i e y_i são dois bins dos histogramas.
- Com esta matriz do kernel precomputada, o SVM é aplicado para classificar as imagens em várias categorias de alimentos. A implementação do SVM foi dada pela biblioteca LIBSVM.

Experimentos

- Como Dataset foi utilizado o **Pittsburgh Food Image Dataset (PFID)**. Esta base reúne uma coleção de imagens e vídeos de fast-food de 13 cadeias de restaurantes. Os experimentos focaram num conjunto de 61 categorias de itens alimentares específicos (por exemplo, Big Mac McDonald) com fundo mascarado. Cada categoria de alimento contém três diferentes instâncias do alimento (comprado em dias diferentes diferentes ramos da cadeia de restaurante), e seis imagens a partir de seis pontos de vista (60 graus de separação) de cada instância de alimento.
- **1º Passo: Treinar STF** - 16 imagens de comida manualmente segmentadas.
- **2º Passo: Aplicação do STF** - em 1098 imagens para geração dos soft labels (vetores de probabilidades de tamanho 9);
- **3º Passo: Criação e Normalização do Histograma de Distribuição (PFDs)** - representação invariante à rotação e escala.
- **4º Passo: Classificação com SVM.**
- **5º Passo: Comparação** - a acurácia obtida foi comparada com os resultados apresentados por duas abordagens conhecidas: Color Histogram e Bag of SIFT.

Resultados

- Acurácia da Classificação

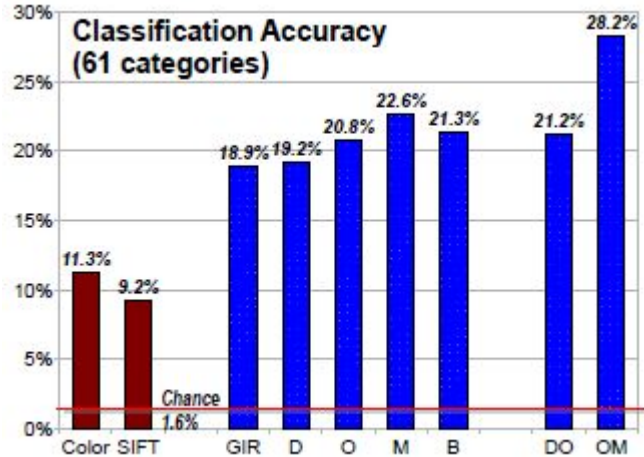


Figura --: Acurácia para as 61 categorias

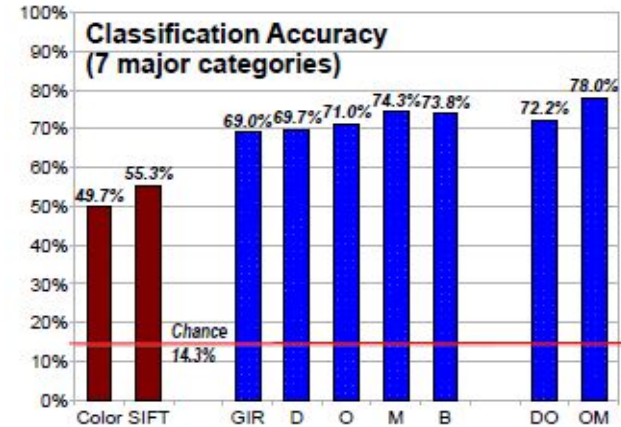


Figura --: Acurácia para as 7 categorias - sanduíches, saladas, frango, pães / pastelarias, donuts, bagels, e tacos

Pontos Fracos

- Dificuldade em reconhecer certos tipos de comida

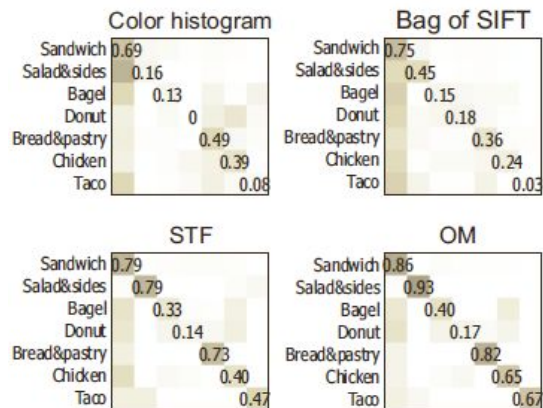


Figura --: Matrizes Comparativas



Figura --: (esquerda) Arby's Roast Turkey and Swiss; (direita) Arby's Roast Turkey Ranch Sandwich

Conclusão

— — —

Embora o Reconhecimento de comida seja uma área nova, esta vem apresentando um crescimento notável. Como muitas técnicas padrão para o reconhecimento de objetos se mostram inadequadas para este problema, este artigo propôs uma nova abordagem baseada na exploração das características espaciais de alimentos em combinação com métodos estatísticos para a rotulagem dos pixels da imagem, permitindo o desenvolvimento de sistemas práticos para reconhecimento de alimentos.

Os experimentos foram feitos utilizando o Pittsburgh Food Image Dataset (PFID), e demonstraram que o método proposto supera significativamente os métodos tradicionais como o pacote de características baseado em SIFT e o histograma de cor.