

# Sistema Reconhecedor de Comida

Daniela Costa<sup>1</sup>, Polyana Costa<sup>1</sup>

<sup>1</sup>Universidade Federal do Maranhão

## Introdução

Reconhecer comida está longe de ser uma tarefa simples, uma vez que é muito comum os itens alimentares apresentarem variações que vão desde a cor até a forma que estes podem ter. No entanto, identificar alimentos pode ser muito útil, por exemplo, em acompanhamentos de dietas, tratamento de obesidade e em outras aplicações que, assim como essas, geram impactos diretos na qualidade de vida. Sendo assim, o objetivo deste trabalho é propor um protótipo de sistema reconhecedor de comida, que através de informações como cor e descoberta de keypoints tenta classificar uma imagem de acordo com algumas categorias pré-definidas, tais como: carne, salada, sorvete e lasanha.

## Proposta

Nossa proposta consiste na criação de protótipos para o reconhecimento de alimentos. Foram desenvolvidos dois sistemas que aplicam diferentes técnicas para extração de características. O primeiro utiliza color histogram para a classificação com SVM, sendo voltado para dispositivos mobile. Já o segundo, executado somente em computador, aplica Scale Invariant Feature Transform (SIFT) combinado ao Bag of Word (BOW) nas imagens adquiridas. A figura 1 traz a interface da versão aplicativo. Na seção de resultados são apresentados os testes executados afim de verificarmos a acurácia de cada uma das técnicas propostas.

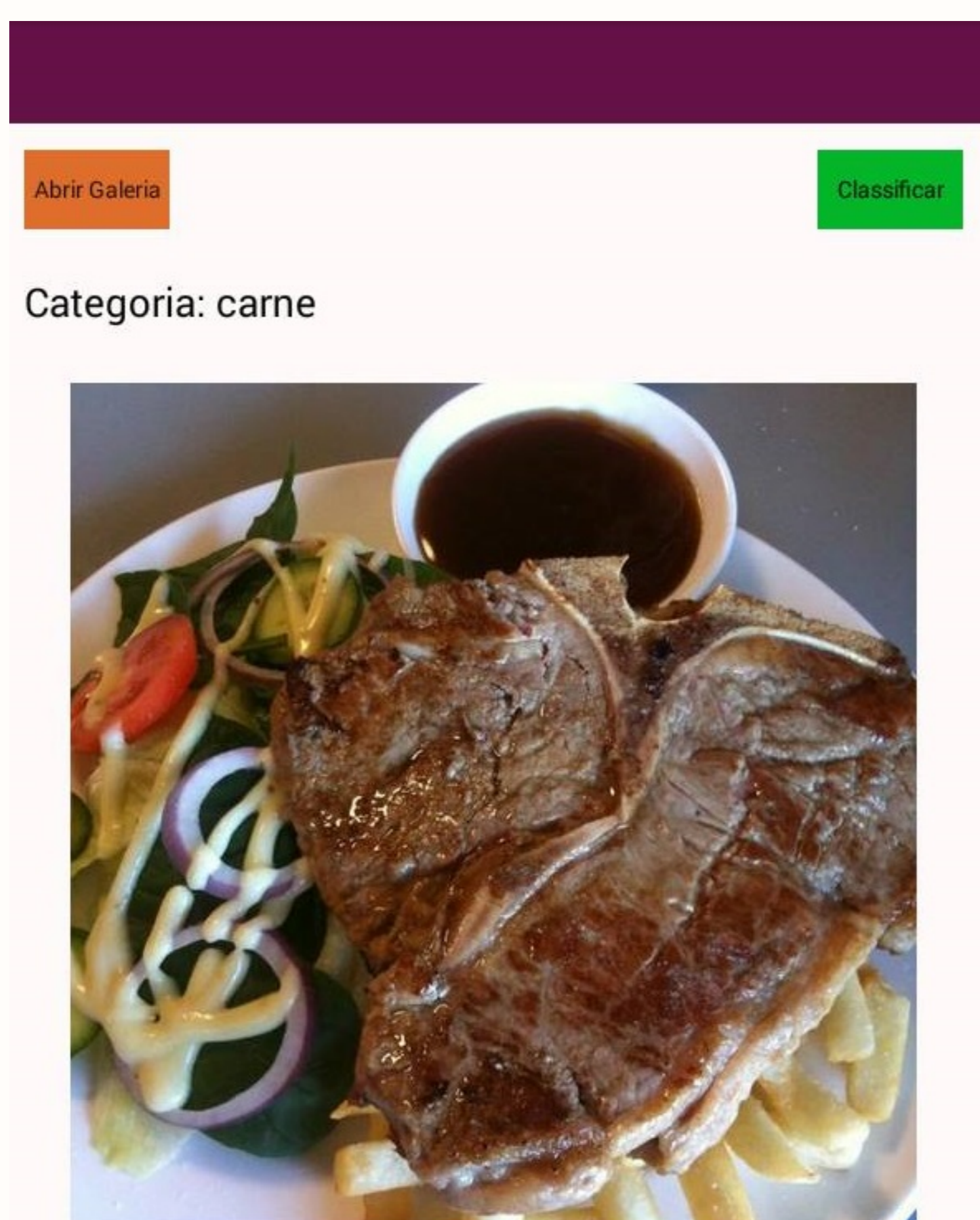


Figure 1: Versão aplicativo

## Metodologia

As imagens utilizadas tanto nas etapas de treinamento, quanto nas de teste são provenientes do The Food-101 Data Set [1], o qual possui mais de 100 categorias. Dentre as quais foram escolhidas as classes com imagens de carne, cupcake, lasanha, macarrão, sorvete e salada. O diagrama apresentado pela figura ?? elenca as etapas executadas pela versão desktop do sistema reconhecedor de alimento.

Inicialmente as imagens adquiridas pelo dataset[1] são convertidas para HSV. Logo após, são submetidas ao SIFT, o qual gera um conjunto de keypoints e descritores para a imagem. Tais descritores são acumulados em uma matriz de features, que depois é repassada para o BOW. Aplicar BOW foi necessário, pois os descritores gerados pelo SIFT possuíam tamanho muito variável, o que comprometia a adição destes na matriz de treinamento. Concluída esta etapa, inicia-se o treinamento com SVM. O tamanho da base de treinamento varia de teste para teste, mas o mínimo foi de 90 imagens.

Para a versão mobile não é usado SIFT, em seu lugar utiliza-se o histograma normalizado da imagem convertida para HSV, o qual servirá como entrada para a classificação. Como treinar em dispositivo móvel não é algo recomendável, o aplicativo precisa apenas carregar o arquivo de treinamento em xml (gerado pela execução em computador), e já estará hábil para identificar as categorias treinadas. Até o momento, as classes utilizadas para essa versão foram carne, salada e sorvete cujos resultados são discutidos na próxima seção.

## Resultados

### Acurácia de classificação em aplicativo

Como já dito anteriormente, as categorias treinadas para o reconhecimento na versão mobile foram carne, salada e sorvete. Para esse teste, o SVM foi treinado com 90 imagens (30 de cada classe) e testado com 30 imagens (10 de cada classe). A acurácia obtida é refletida no gráfico da figura 2, onde podemos observar que a classe carne obteve um índice de 80% de reconhecimento, sorvete com 60% e salada com 90%.

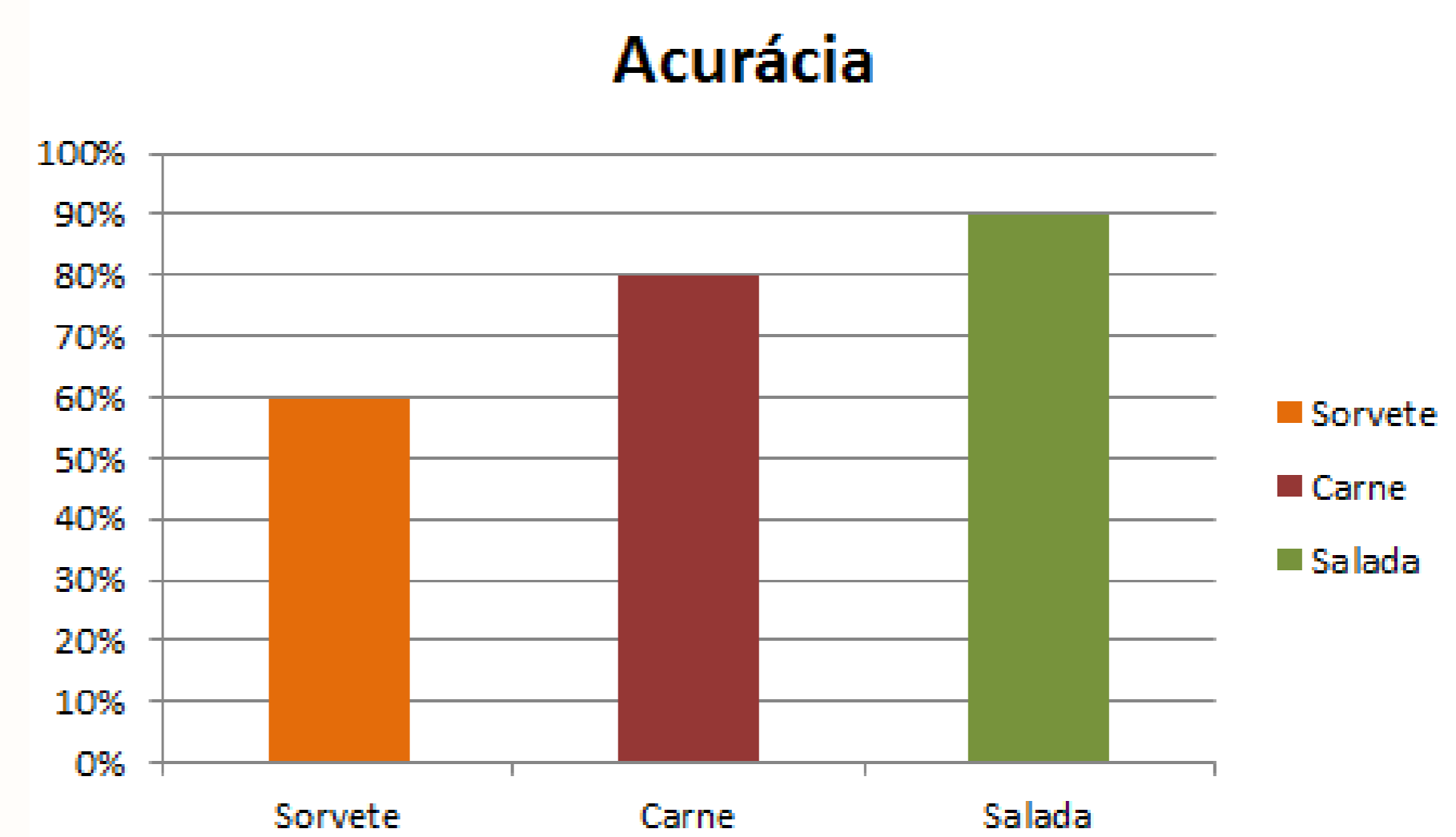


Figure 2: Acurácia com Color Histogram

### Acurácia de classificação em desktop

A versão desktop contou com testes utilizando SIFT e outro teste utilizando Color Histogram. Para o treinamento com color histogram, foram utilizadas 3 classes: Cupcake, Macarrão e Lasanha, cada classe com 50 imagens. Para realizar os testes, utilizou-se 15 imagens de cada classe e os resultados são mostrados na figura 3. Podemos ver que as taxas de reconhecimento para Cupcake e Lasanha mostraram-se satisfatórias, com 73% e 87% de acurácia, respectivamente. A classe macarrão não obteve muito sucesso, com acurácia de apenas 40%.

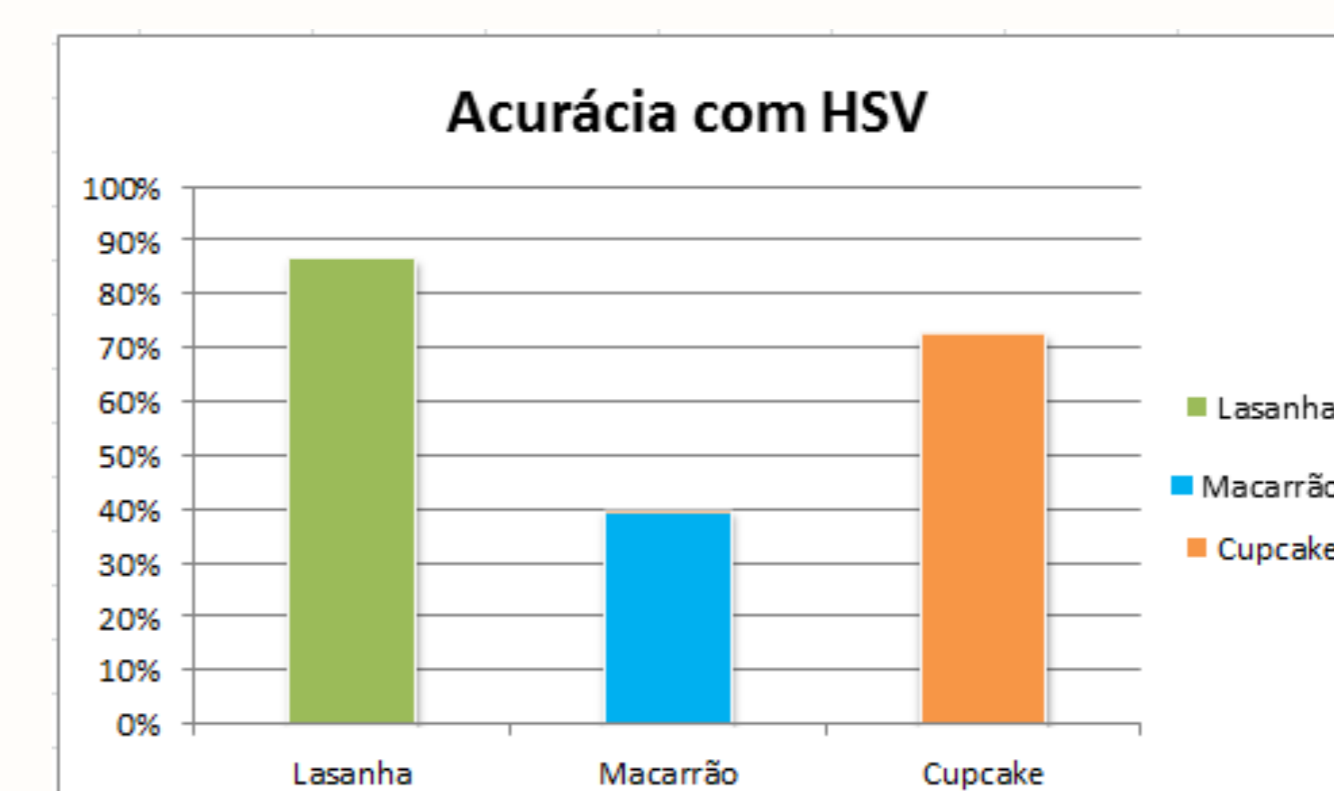


Figure 3: Resultado da classificação para: cupcake, macarrão e lasanha

Um dos testes feito para a versão desktop, considerou 4 classes: carne, salada, sorvete e lasanha. No treinamento foram utilizadas 120 imagens (30 de cada classe), enquanto que na de teste foram 40 imagens (10 de cada classe). A figura 4 mostra a taxa de reconhecimento para cada categoria. Nela podemos ver que com SIFT o reconhecimento foi mais eficiente. A classe salada continuou tendo o melhor índice. A taxa de identificação para sorvete superou a da carne (ficando 80%, enquanto que a carne obteve 70%). Já a classe lasanha não obteve sucesso como as outras, ficando com 40%.

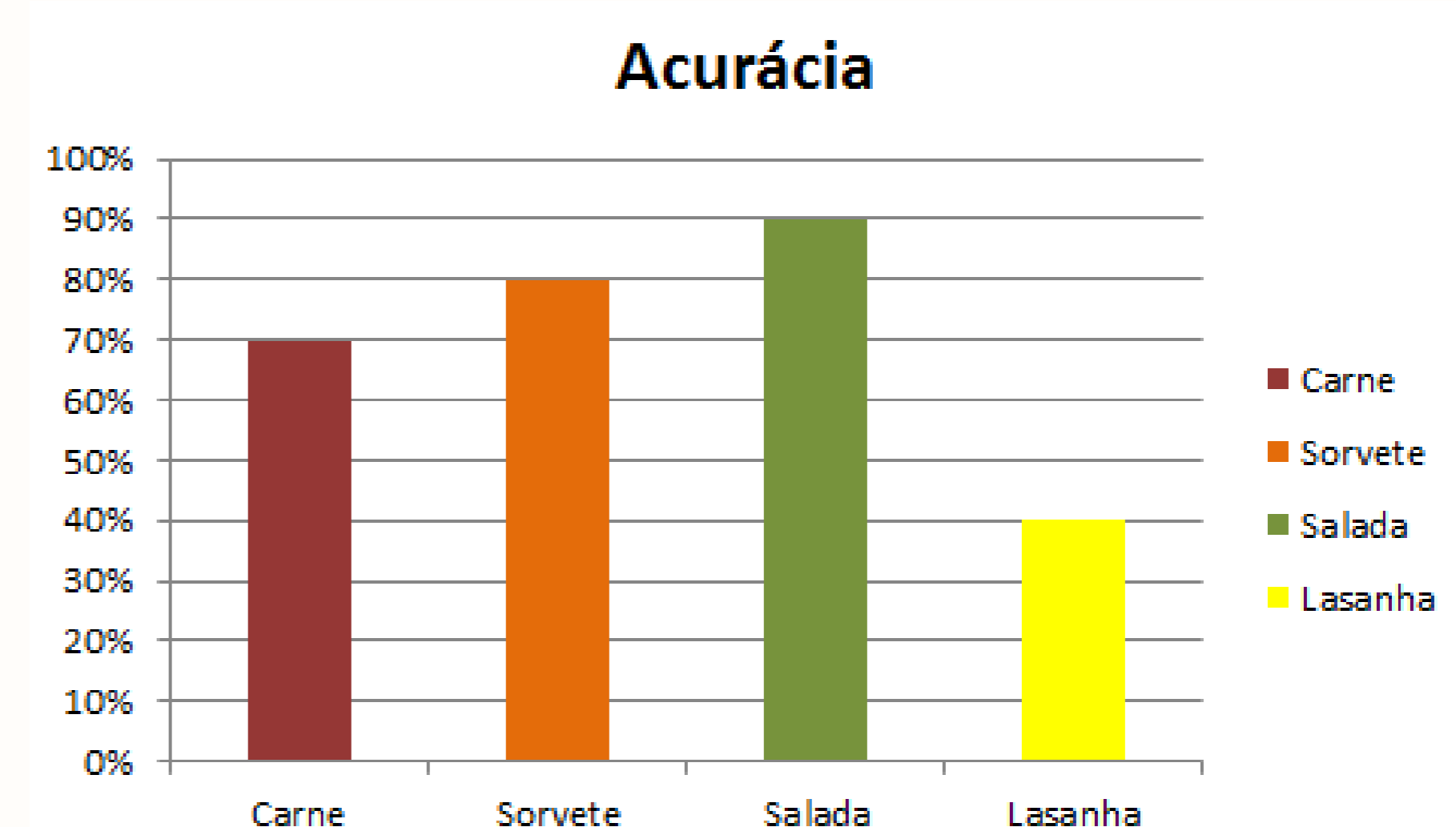


Figure 4: Resultado de classificação para carne, sorvete, salada e lasanha (utilizando sift)

## Conclusão

O objetivo deste trabalho foi apresentar o que seria o início de um sistema capaz de identificar alimentos. Basicamente, foram desenvolvidos e comparados dois sistemas que aplicam técnicas diferentes para o reconhecimento, sendo o primeiro voltado para dispositivos móveis e o segundo para computadores. As avaliações feitas em ambos mostram a robustez da técnica com SIFT em relação à que aplica color histogram. No entanto, a praticidade oferecida pelo aplicativo também deve ser levada em consideração, fazendo com que um possível trabalho futuro foque justamente nessa junção de praticidade e robustez na classificação de alimentos.

## References

[1] [http://www.vision.ee.ethz.ch/datasets\\_extra/food-101/](http://www.vision.ee.ethz.ch/datasets_extra/food-101/)