

CMP 197 - Fundamentos de Visão Computacional

Relatório do trabalho 2

Alleff Dymytry Pereira de Deus (00331157), Letícia dos Santos (00275604)

O objetivo desse trabalho é criar um sistema de *Content-based image retrieval* (CBIR) usando informação de forma como medida de similaridade, sendo que CBIR consiste em fornecer uma imagem de consulta e recuperar imagens semelhantes em uma base de dados. Foi desenvolvido um sistema que recebe como entrada uma imagem de busca e um número inteiro positivo N. Então, são retornadas as N imagens da base de dados mais semelhantes à imagem de busca, em ordem decrescente de similaridade. Idealmente, a imagem de busca deve ser a primeira imagem a ser recuperada, se estiver presente na base de dados. Todas as imagens e tabelas estão em Anexo A e Anexo B, respectivamente. Os códigos das abordagens e implementações podem ser encontrados neste [repositório](#).

Também foi implementada uma rotina que aplica uma variação aleatória de escala e rotação.

O relatório está dividido em seções. Inicialmente, são apresentadas as operações aplicadas sobre as imagens em alto nível nas Seções 1 e 2. Após, o algoritmo implementado é descrito em detalhes na Seção 3. Testes são apresentados e discutidos na Seção 4.

1. Operações sobre única imagem

- *Threshold*: utilizado para eliminar ruídos e uniformizar as imagens.
- Descritor hog: fornece um *histogram of oriented gradients* (HOG) que permite caracterizar a imagem. Uma característica interessante é que as bordas ficam evidenciadas. Isso contribui para a solução do problema de encontrar imagens semelhantes, pois as imagens possuem somente a borda externa do objeto que representam.
- Rotação: adiciona um ângulo em relação a imagem original, este efeito é aplicado para que seja possível encontrar imagens dentro do banco que possuam alguma orientação diferente em relação a imagem de entrada.
- Rotação e escala aleatórias: pré-requisito do trabalho. Adiciona mais dificuldade para achar imagens semelhantes. É importante destacar que a implementação utilizada evita o corte do objeto no limite da imagem ao realizar a rotação.
- Detecção de bordas com *Canny*: detecta a borda contínua que existe em cada imagem do banco. Também elimina ruídos.

2. Técnicas para comparação entre imagens

- Distância euclidiana:
 - `euclidean_1D(self,imgA,imgB)`: distância euclidiana calculada sobre o valor de dois vetores.
 - `euclidean_2D(self,imgA,imgB)`: distância euclidiana calculada sobre o valor dos pixels de duas imagens.
- Erro médio quadrático:
 - `mse_1d(self,imgA,imgB)`: erro médio quadrático aplicado em dois vetores.
 - `mse_2d(self,imgA,imgB)`: erro médio quadrático aplicado aos valores de pixels de duas imagens.

- Distância cosseno
 - `distance.cosine(fd_q,fd_db)`: distância do cosseno aplicado em dois vetores. Sendo que a distância entre os dois vetores é o produto escalar entre esses vetores.

3. Algoritmo implementado

O algoritmo implementado pode ser dividido em 3 etapas distintas entre si, sendo elas: (3.1) Ajustes dos descritores; (3.2) Comparação entre as imagens; e (3.3) Avaliação das métricas.

3.1. Ajuste dos descritores

Nesta etapa do algoritmo, o objetivo é ajustar a imagem de entrada com os descritores escolhidos para fazer a comparação com as demais imagens do sistema. As operações de transformações, utilizadas como descritores das imagens, são aplicadas em todas as imagens presentes no banco de imagens.

3.2. Comparação entre as imagens

Após a etapa de ajuste dos descritores, é necessário realizar a comparação entre as imagens utilizando os descritores e guardar estes valores obtidos. Assim, são aplicadas as técnicas propostas neste trabalho, que foram apresentadas na Seção 2, entre a imagem de entrada e de cada imagem dentro do banco de imagens. O resultado de cada combinação de métrica e descritor para cada imagem é anexado ao seu respectivo vetor. As tabelas do Anexo B mostram as combinações utilizadas.

Então, os vetores são ordenados e são selecionadas as N melhores imagens de cada vetor. Logo, cada combinação de métrica e descritor possui um vetor de tamanho N e esses vetores são entrada da próxima etapa.

3.3. Avaliação das métricas

É obtida a precisão de cada combinação de métrica e descritor, que é calculada segundo a Equação 1 utilizando os elementos do vetor resultante da etapa anterior. Para obter a "Quantidade de imagens relevantes da categoria" na Equação 1, é utilizado o nome de cada categoria como peça chave para avaliar se a imagem que foi retornada é da mesma categoria da imagem de busca, ou seja, se é relevante. Após a precisão ser calculada, é selecionada o vetor que obteve maior precisão para ser o resultado final.

$$P = \frac{\text{Quantidade de imagens relevantes da categoria}}{N \text{ imagens retornadas}} \quad (1)$$

Por exemplo, a consulta foi feita com a imagem fork01 e N=4. A métrica erro médio quadrático com o descritor HOG retornou fork01, bird03, camel04 e car05, resultando em precisão de 0,25. Já a métrica distância euclidiana com descritor Canny retornou fork01, fork03, fork10 e fork19, precisão de 1. O restante das combinações obtiveram desempenhos piores. Nesse contexto, a métrica distância euclidiana utilizando Canny possui a maior precisão, logo as imagens obtidas por essa combinação na etapa anterior são o resultado final do algoritmo.

4. Testes e discussão

Com o algoritmo implementado, foram realizados diversos testes com imagens de entrada diferentes. O banco de imagens utilizado possui 18 categorias de figuras diferentes, no qual todas possuem o fundo branco e o objeto contido na imagem em preto. Contudo, cada categoria possui características diferentes, sendo elas: a forma, o tamanho e a rotação.

As imagens que foram utilizadas como consulta pertencem a 6 categorias diferentes e podem ser vistas na Figura 1. Foram realizados testes com o valor de $N = 6$ como padrão. Além disso, existem duas variações para obter as N imagens: comparação da imagem de entrada com o banco de imagens; e comparação utilizando ambas as imagens rotacionadas e escaladas de forma aleatória.

Dentre as 6 categorias utilizadas, a categoria *fork* foi utilizada como exemplo para avaliar o funcionamento do algoritmo implementado. Os resultados obtidos para essa categoria podem ser vistos na Tabela 1 e as N imagens mais semelhantes podem ser vistos na Figura 2(a) (sem modificações nas imagens) e na Figura 2(b) (com rotação e escala modificada).

Analisando os resultados da categoria *ray*, pode-se notar que dentre as N imagens retornadas a imagem de entrada não foi devolvida como um dos resultados, este resultado pode ser notado na Figura 6(a) e 6(b). Isto ocorreu devido a escolha de utilizar uma medida de precisão, no qual a imagem de entrada não estava presente no conjunto que obteve a maior precisão. Assim, é possível discutir acerca da escolha de devolver sempre o conjunto que possua a imagem de entrada ou o conjunto que realmente apresente a maior precisão.

Pode-se obter a análise de que a rotação inserida nas imagens, gera um fator de dificuldade para o encontro de imagens semelhantes. Utilizando os resultados, notou-se que metade dos testes não obtiveram uma precisão igual ou acima de 50% quando inserido a rotação e a escala nas imagens. Já os testes utilizando as imagens originais, foi possível obter todas as categorias que foram escolhidas com o valor de precisão igual ou acima de 50%.

Anexos

A. Imagens

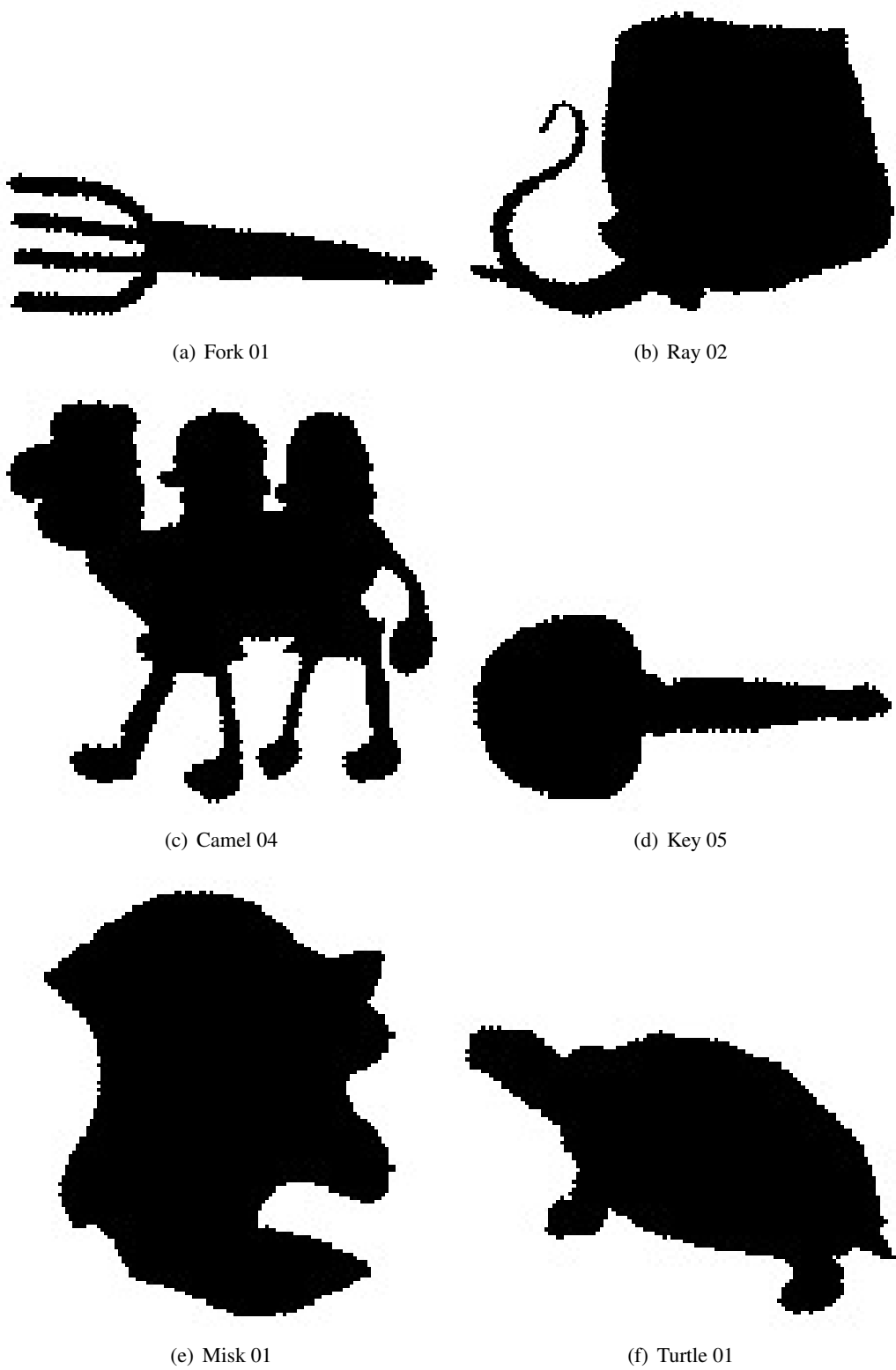
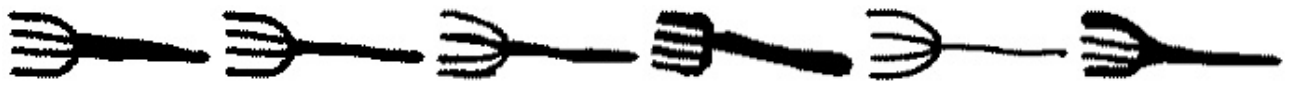
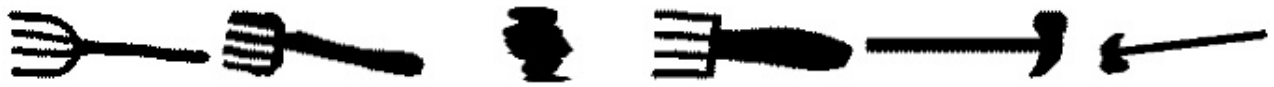


Figura 1: Imagens Originais



(a) CBIR



(b) CBIR(Escalado,Rotacionado)

Figura 2: Resultado fork01



(a) CBIR



(b) CBIR(Escalado,Rotacionado)

Figura 3: Resultado camel04



(a) CBIR



(b) CBIR(Escalado,Rotacionado)

Figura 4: Resultado key05



(a) CBIR



(b) CBIR(Escalado,Rotacionado)

Figura 5: Resultado Misk01



(a) CBIR



(b) CBIR(Escalado,Rotacionado)

Figura 6: Resultado ray02



(a) CBIR



(b) CBIR(Escalado,Rotacionado)

Figura 7: Resultado turtle01

B. Tabelas

Descritor	Métrica	P
HOG	Euclidiana	1.0
HOG	MSE	1.0
HOG	Cosseno	1.0
Rot 45°	Euclidiana	0.166
Rot 90°	Euclidiana	0.0
Rot 135°	Euclidiana	0.5
Rot 180°	Euclidiana	0.166
Rot 225°	Euclidiana	0.166
Rot 270°	Euclidiana	0.0
Rot 315°	Euclidiana	0.5
Canny	Euclidiana	0.333

Tabela 1: Valor de P para os N resultados da imagem *fork01*

Descritor	Métrica	P
HOG	Euclidiana	0.333
HOG	MSE	0.333
HOG	Cosseno	0.5
Rot 45°	Euclidiana	0.333
Rot 90°	Euclidiana	0.0
Rot 135°	Euclidiana	0.0
Rot 180°	Euclidiana	0.0
Rot 225°	Euclidiana	0.0
Rot 270°	Euclidiana	0.166
Rot 315°	Euclidiana	0.0
Canny	Euclidiana	0.0

Tabela 2: Valor de P para os N resultados da imagem *fork01* aplicando rotação e escala randômica

Tabela 3: Valor de P para os N resultados da imagem *fork01*

Descritor	Métrica	P
HOG	Euclidiana	0.5
HOG	MSE	0.5
HOG	Cosseno	0.833
Rot 45	Euclidiana	0.0
Rot 90	Euclidiana	0.0
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	0.0
Rot 225	Euclidiana	0.0
Rot 270	Euclidiana	0.0
Rot 315	Euclidiana	0.0
Canny	Euclidiana	0.166

Tabela 4: Valor de P para os N resultados da imagem *camel04*

Descritor	Métrica	P
HOG	Euclidiana	0.0
HOG	MSE	0.0
HOG	Cosseno	0.166
Rot 45	Euclidiana	0.0
Rot 90	Euclidiana	0.166
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	0.0
Rot 225	Euclidiana	0.166
Rot 270	Euclidiana	0.0
Rot 315	Euclidiana	0.0
Canny	Euclidiana	0.0

Tabela 5: Valor de P para os N resultados da imagem *camel04* aplicando rotação e escala randômica

Descritor	Métrica	P
HOG	Euclidiana	1.0
HOG	MSE	1.0
HOG	Cosseno	1.0
Rot 45	Euclidiana	0.0
Rot 90	Euclidiana	0.0
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	1.0
Rot 225	Euclidiana	0.0
Rot 270	Euclidiana	0.0
Rot 315	Euclidiana	0.0
Canny	Euclidiana	1.0

Tabela 6: Valor de P para os N resultados da imagem *Misk01*

Descritor	Métrica	P
HOG	Euclidiana	0.166
HOG	MSE	0.166
HOG	Cosseno	0.166
Rot 45	Euclidiana	0.166
Rot 90	Euclidiana	0.5
Rot 135	Euclidiana	0.166
Rot 180	Euclidiana	0.5
Rot 225	Euclidiana	0.166
Rot 270	Euclidiana	0.333
Rot 315	Euclidiana	0.333
Canny	Euclidiana	0.0

Tabela 7: Valor de P para os N resultados da imagem *Misk01* aplicando rotação e escala randômica

Descritor	Métrica	P
HOG	Euclidiana	0.333
HOG	MSE	0.333
HOG	Cosseno	0.333
Rot 45	Euclidiana	0.833
Rot 90	Euclidiana	0.0
Rot 135	Euclidiana	0.833
Rot 180	Euclidiana	0.0
Rot 225	Euclidiana	0.0
Rot 270	Euclidiana	0.0
Rot 315	Euclidiana	0.333
Canny	Euclidiana	0.166

Tabela 8: Valor de P para os N resultados da imagem *ray02*

Descritor	Métrica	P
HOG	Euclidiana	0.0
HOG	MSE	0.0
HOG	Cosseno	0.0
Rot 45	Euclidiana	0.666
Rot 90	Euclidiana	0.166
Rot 135	Euclidiana	0.166
Rot 180	Euclidiana	0.333
Rot 225	Euclidiana	0.166
Rot 270	Euclidiana	0.166
Rot 315	Euclidiana	0.0
Canny	Euclidiana	0.0

Tabela 9: Valor de P para os N resultados da imagem *ray02* aplicando rotação e escala randômica

Descritor	Métrica	P
HOG	Euclidiana	0.5
HOG	MSE	0.5
HOG	Cosseno	0.5
Rot 45	Euclidiana	0.166
Rot 90	Euclidiana	0.666
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	0.0
Rot 225	Euclidiana	0.166
Rot 270	Euclidiana	0.0
Rot 315	Euclidiana	0.0
Canny	Euclidiana	0.166

Tabela 10: Valor de P para os N resultados da imagem *turtle01*

Descritor	Métrica	P
HOG	Euclidiana	0.0
HOG	MSE	0.0
HOG	Cosseno	0.333
Rot 45	Euclidiana	0.166
Rot 90	Euclidiana	0.166
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	0.0
Rot 225	Euclidiana	0.0
Rot 270	Euclidiana	0.333
Rot 315	Euclidiana	0.0
Canny	Euclidiana	0.0

Tabela 11: Valor de P para os N resultados da imagem *turtle01* aplicando rotação e escala randômica

Descritor	Métrica	P
HOG	Euclidiana	1.0
HOG	MSE	1.0
HOG	Cosseno	1.0
Rot 45	Euclidiana	0.0
Rot 90	Euclidiana	0.0
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	0.0
Rot 225	Euclidiana	0.0
Rot 270	Euclidiana	0.0
Rot 315	Euclidiana	0.0
Canny	Euclidiana	0.666

Tabela 12: Valor de P para os N resultados da imagem *key05*

Descritor	Métrica	P
HOG	Euclidiana	0.166
HOG	MSE	0.166
HOG	Cosseno	0.166
Rot 45	Euclidiana	0.166
Rot 90	Euclidiana	0.333
Rot 135	Euclidiana	0.0
Rot 180	Euclidiana	0.666
Rot 225	Euclidiana	0.0
Rot 270	Euclidiana	0.166
Rot 315	Euclidiana	0.333
Canny	Euclidiana	0.0

Tabela 13: Valor de P para os N resultados da imagem *key05* aplicando rotação e escala randômica