

Atividade 3: Exploração de imagens

Rodney Rick
RA 181.814
Disciplina MO444
Prof. Dr. Anderson Rocha

Resumo—O objetivo deste trabalho é estudar a representação de dados de imagens, transformá-los, rechonhecê-los e agrupar devido suas característica e avaliar os resultados finais obtidos. Também abordado discussão dos dados de saída e problemas enfrentados para as execuções.

I. INTRODUÇÃO

Nas seções seguintes o conjunto de dados para esta tarefa será apresentado, como foi extraídos valores, técnicas de agrupamentos e para melhores melhores explorações dos dados, usado uma técnica de redução de dimensionalidade.

Para o escopo, foi necessário a extração de alguns dados do conjunto de imagens a fim de gerar novos dados, como histograma de *RGB* e transformação de *HOG* (explicados posteriormente) a qual essa etapa representa a extração dos dados. Ou seja, transformar para obter mais recursos explorar.

Utilizou-se uma técnica de agrupamento (*clustering*) conhecida e muito usado como *k-Means*, e posterior reduções de dimensionalidade, usando *PCA* para re-executar o agrupamento e verificação de melhoria de resultado obtido.

Ao final do relatório (seção IV) faz-se o levantamento dos resultados e discussão das técnicas aplicadas conforme a resolução. Na seção de Apêndice A tem alguma informação extra dos dados.

Este trabalho visa responder questões como qual o número de grupos mais adequado de agrupamento a fim de disponibilizar e representar o conjunto total de imagens. Quais os pontos fortes e fracos das técnicas aplicadas e como interpretar dados com algoritmos não-supervisionados (no caso para agrupamentos). A aplicação de uma técnica de redução de dimensionalidade poderia melhorar o resultado final? Esses e outros questionamentos são transformado em algumas respostas para este trabalho.

Para esse trabalho foram considerados algumas leitura essenciais para este trabalho:

- Capítulo 11 e 12 do livro de Carvalho, A.;
- Capítulo 10 (seções referentes a *k-Means*) do livro de Duda et al.;
- Capítulo 9 (*k-Means*) e 12 (*PCA*) do livro de Bishop;
- e, notas de aula e artigos na internet.

II. O CONJUNTO DE IMAGENS E SUA EXPLORAÇÃO

O conjunto de dados é composto de uma base de imagens que contém 81.442 amostras de imagens. Cada amostra é composta por um objeto e um fundo preto. Ainda, cada contém um e somente um objeto, porém ao averiguar algumas



Figura 1: Quatro imagens contidas na coleção do mesmo objeto.

imagens, percebe-se que o mesmo objeto se repete, no entanto, com posicionamento diferente em cada amostra.

Para este trabalho, utilizou-se as bibliotecas *OpenCV* (*Open Source Computer Vision Library*) e *scikit-image* para a captura de atributos detectores e descritores presentes nas imagens da amostragem, cálculo de contraste e brilho para melhoria da imagem original a fim de obter mais características que possam aumentar na identificação das imagens.

As figuras 2 e 12 (esta presente no Apêndice A), demonstra o tratamento e melhoria da imagem original trabalhando em atributos como brilho e contraste. Ainda presente na figura 2, segue a extração de dados referente as componentes de cores *RGB*. No gráfico acima (com menos picos) representa a extração das componentes da imagem original e no gráfico abaixo, apresenta as componentes após a melhoria de brilho e contraste. É possível perceber que existe, pós-tratamento, mais atributos a serem aproveitados da imagem tratada que na imagem original.

No entanto, mesmo com o tratamento e a extração dos dados de cores, percebe-se que existe um valor muito extremo que deve ser removido antes das análises. Na imagem original, a cor preta é muito presente, sendo necessária sua remoção. Na imagem tratada, a mesma cor preta não é mais presente, mas suas similares (como escala de cinza mais próxima do preto). Torna-se agora necessário a remoção de duas cores, o fundo e o "chão" do objeto.

Dado o tratamento devido para ambas as imagens, alinha-se as componentes formando um único vetor de atributos. Como existe 256 cores para cada componente e ao mesclar os três atributos em um vetor, tem-se 768 posição de cores que descrevem uma imagem. Porém, para dar ênfase nos objetos, na imagem original, desconsidera-se um valor de máximo

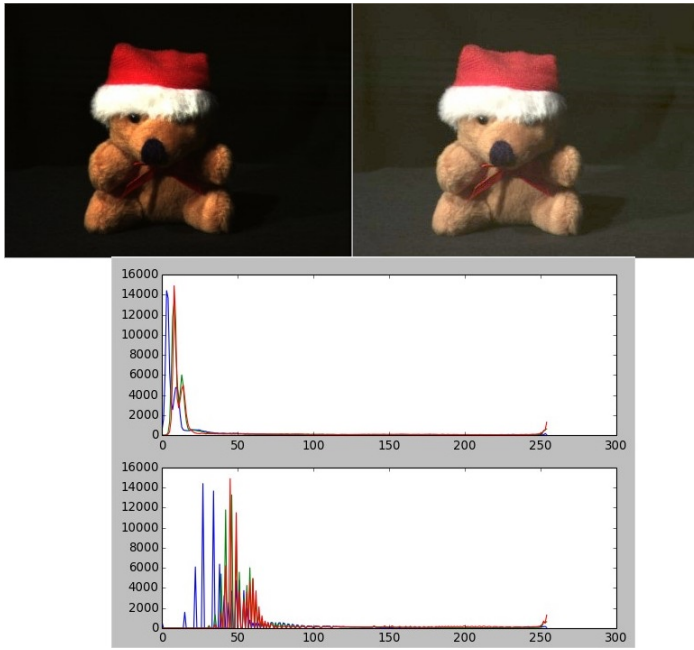


Figura 2: Imagem original (esq.), imagem com aumento de contraste e brilho (dir.) e sua decomposição das Componentes de RGB.

(seria a composição do preto) para cada componente e o vetor final tem 765 posições. O mesmo se realiza na imagem tratada, porém com valor de 762 posições no vetor de saída, sendo a remoção das duas cores mais presentes.

Ainda, para oferecer mais resultados e de formas diferentes para captura de imagens e de seus atributos, foi utilizado a técnica *Histogram of Oriented Gradients (HOG)* que trabalha com a detecção dos objetos disponíveis em uma imagem ou vídeo. A figura 3 exibe dois objetos originais e suas transformações com *HOG*. Uma transformação de imagem gerada pelo *HOG* tem os seguintes premissas:

- normalização global imagem;
- calcular e gerar a imagem gradiente em x e y;
- criação dos histogramas de gradiente;
- normalizando em todo blocos;
- achatamento das características em um vetor.

Para a base de treino foi considerado 40.000 imagens. Não foi possível usar um maior valor de exemplares na base de treinamento devido ao tempo de processamento. Após o tratamento das imagens e agora em posse de dados numérico para utilização das análises, pode-se trabalhar algumas técnicas conforme a seção III demonstra.

III. OS ALGORITMOS DE AGRUPAMENTOS

Para atender ao requisitos de agrupamento de imagens conforme proposto como trabalho, esta seção trabalha-se com as técnicas: (a) *k-Means* (subseção III-A) e (b) *PCA* (subseção III-B).

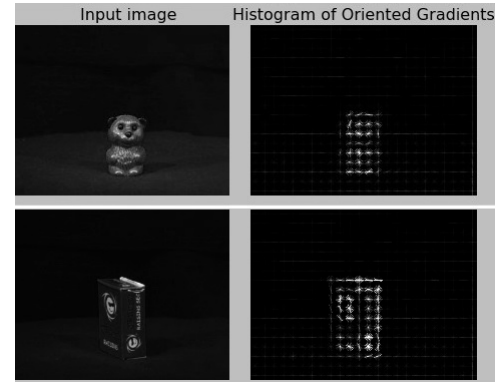


Figura 3: Imagem trabalhada usando *HOG*

A. *k-Means*

Criada para constatar e aglomerar objetos semelhantes, a técnica de *k-Means* é um método de quantização vetorial muito usado em mineração de dados. Com o propósito de particionar n observações em k clusters, na qual cada observação pertence ao um cluster, dado que o observado deve estar presente na média mais próxima de um determinado cluster.

Dado que é um técnica que reside no escopo das técnicas não-supervionadas (para alguns considerada também como semi-supervisionada), *k-Means* trabalha por padrão com a distância euclidiana, adotando para o ponto de seu cluster k o ponto médio. A equação 1 demonstra o cálculo do ponto médio do centro de um cluster, também conhecido como centróide. As variáveis c_i , k descrevem a centróide i do cluster k_i , n é o número de objetos presentes, x_j é o objeto em questão analisado e μ_i é o ponto médio nesta iteração.

$$c_i = \sum_{j=0}^k \sum_{x_j=0}^n \|x_j - \mu_i\|^2 \quad (1)$$

Os critérios de parada são até que as centróides não se movimentem mais ou até um número de iterações máxima seja atingido. Afinal, caso as centróides nunca converjam, torna-se necessário outro critério de parada.

Porém existem um ponto importante em questão, qual seria o k ideal para a composição do melhor do número de agrupamentos a fim de maximizar o número de objetos assistidos e ser o menor k possível?

Para esse tipo essa resposta existe o método de *Elbow* (figura 4), o qual descreve que dado um k inicial, deve-se iterar k (até um valor máximo desejado) cada solução comparando com o resultado anteriormente obtido. E, no momento em que os resultados r_i e r_{i+1} não apresentarem significativas melhoras, k_i é o valor para número de clusters mais adequado para este problema abordado. Ao qual variância entre os valores resultados das somas dos quadrados das distâncias de todos os centróides até os objetos presentes em seu agrupamento é o melhor para a composição da solução.

No entanto, esse método tem uma abertura. Quando existir um caso onde é quase imperceptível essa variância alta, como

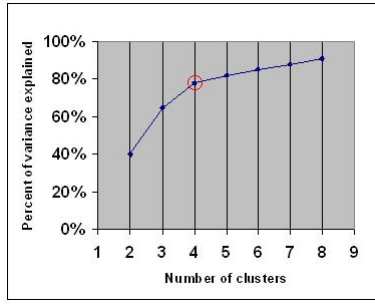


Figura 4: Critério de *Elbow* [4].

deveria proceder? Nesse trabalho, foi encontrado tal ponto. Se o valor de k é muito grande, o próximo resultado deve ser um valor adequado, e para este caso foi avaliado como existindo uma variância maior nos próximos 3 valores obtidos e adotando também o critério de tempo de execução. A figura 5, apresenta o valor de $k = 2880$ e um tempo de solução adequado com relação ao valor anterior e bem menor que os próximos 3 resultados ($k = [2890, 2900, 2910]$).

Tabela I: Tabela de tempo de execução de *k-Means*

k	tempo de execução
2800	4 min 32s
2850	5 min 12s
2880	5 min 38s
2880	6 min 47s
2900	7 min 17s
2910	8 min 25s

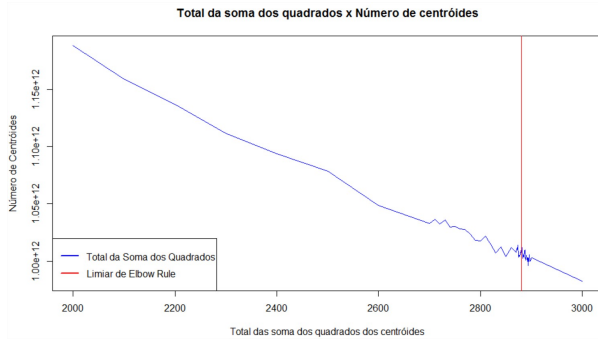


Figura 5: Soma total das distâncias dos objetos até as centróides c_i de cada *cluster*

B. PCA

A análise de *PCA* (*Principal component analysis*) é definida matematicamente como a transformação linear ortogonal, onde os dados são trabalhados em um sistema de variância. Dado tal transformação, para cada amostra presente, o valor da maior variância projeta-se ao longo da primeira coordenada (também conhecida chamado primeira componente), a segunda

maior variância fica ao longo da segunda coordenada, e assim por diante. A figura 6 demonstra a projeção dos dados, após também efetuado a rotação dos mesmo e trazendo para o ponto de cruzamento dos eixos X e Y.

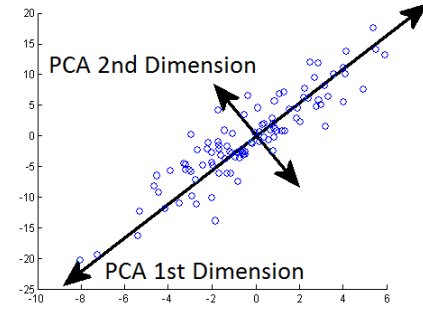


Figura 6: Decomposição das Componentes do PCA.

Demonstrado (simplificadamente) o conceito, os resultados obtidos da técnica *PCA* para os dados originais, localização das 2880 centróides com a técnica de *Elbow* ilustra-se na figura 7. Aplicando *PCA* agora para as imagens trabalhadas em brilho e contraste representam 2300 centróides, descoberto com *Elbow* (figura 8).

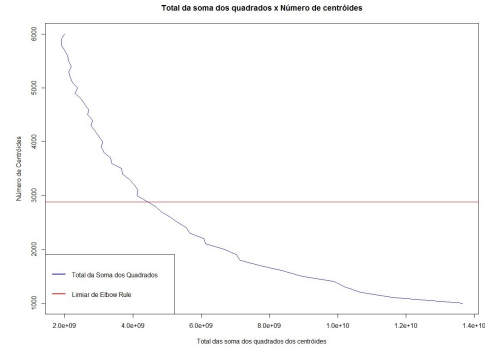


Figura 7: Aplicação de valores para k e limiar de *Elbow* para amostras originais.

Por curiosidade, escolhe dois atributos quaisquer do processamento do *PCA* e resolvi gerar o *plot* dos dados. E, ao verificar que existiam 7 grandes *clusters* distintos. Esse número específico (7) foi encontrado em um primeiro momento usando *Elbow* para gerar entre 2 a 50 *clusters* inicialmente para esse trabalho para todos os dados explorados. Fazendo alguns experimentos de dois em dois atributos, percebe-se que esses 7 grandes agrupamentos se sustentam e aparecem mais vezes. Na figura 9, exibe 2300 centróides representadas pela cor preta, os pontos coloridos representam os atributos das imagens escolhidas para análise e em vermelho os grandes *clusters* percebidos. Esses grandes *clusters* devem representar na coleção de imagens as principais características que as diferenciam em primeiro momento.

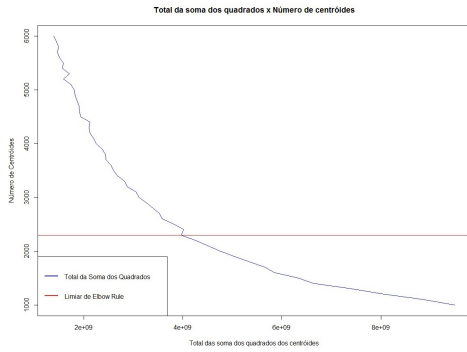


Figura 8: Aplicação de valores para k e limiar de *Elbow* para amostras com melhor contraste e brilho.

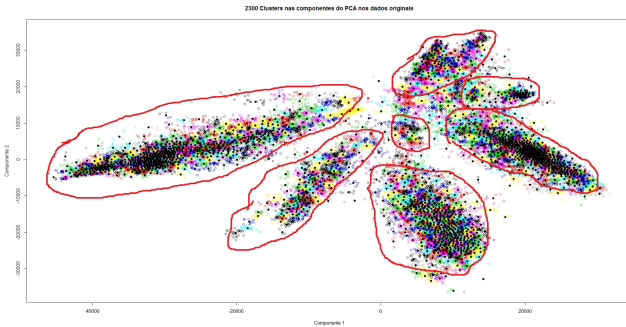


Figura 9: Dados das imagens transformadas usando contraste e brilho e depois vetorizados (pontos coloridos) convertidos usando *PCA* e posteriormente usando *k-Means*, onde cada *cluster* é agrupado pela centróide (pontos pretos) mais próxima.

Ainda trabalhando *PCA* para as amostras com contraste e brilho, atuou-se usando *HOG* (conforme explicado a técnica anteriormente). Porém os resultados analisando 10.000 amostras e com um valor para k acima de 4.500 para *clusters*, mais do que esse valor tornava-se inviável devido a demora de resultados obtidos no processamento dos *clusters*. Não foi possível obter um resultado adequado para um "melhor" valor de k usando o método de *Elbow*, também usado critério de tempo de execução para exibição nos resultados apresentados na figura 10.

Como para *HOG* o método de *Elbow*, com os critérios de tempo e variação entre as iterações não convergiram para um valor mais adequado e aplicados para esta tarefa não atenderam, foi escolhido um valor menor que $k = 2000$ (menor que os valores anteriores aplicados - $k = [2300, 2880]$).

A figura 11 demonstra o resultado de saída de um dos *clusters* gerado após o processamento.

IV. CONCLUSÕES

Para este trabalho, houve um tempo grande no aprendizado das técnicas de processamento de imagens (para não simplesmente sua execução) e um tempo de processamento para a extração muito alto, porém pode-se levantar alguns pontos importantes:

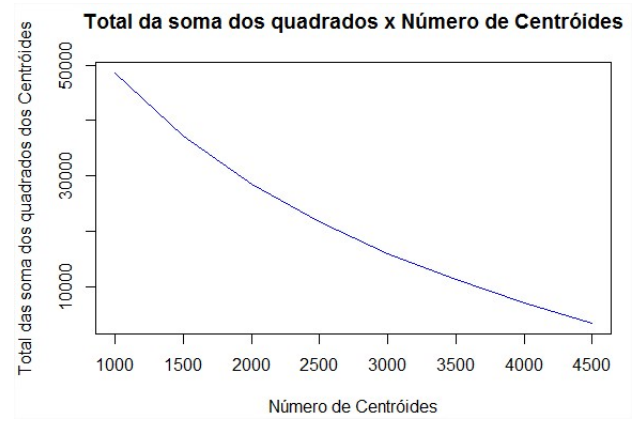


Figura 10: Dados das imagens transformadas usando contraste e brilho e depois vetorizados (pontos coloridos) convertidos usando *PCA* e posteriormente usando *k-Means*, onde cada *cluster* é agrupado pela centróide (pontos pretos) mais próxima.

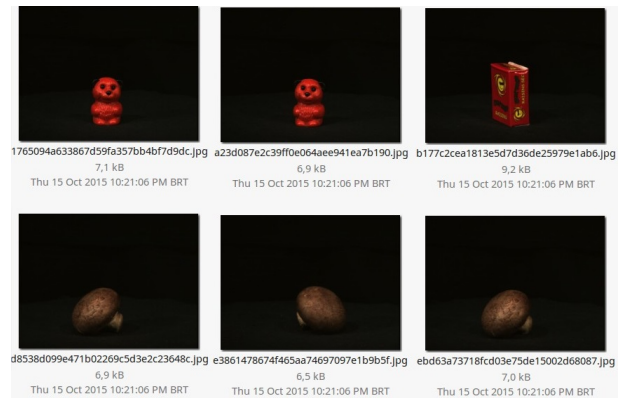


Figura 11: Imagens agrupadas no mesmo *cluster*, bem parecidas em forma e cor, mas representando objetos diferentes.

- Quais as melhores escolhas para extração de imagens e criação dos agrupamentos?
- Processar somente usando *RGB* é suficiente?
- Escolha como *BIC* (*Bayesian Information Criterion*) seria melhor? E *HOG*, atendeu a este escopo?
- PCA* realmente foi útil? Qual o maior ganho?

Exercitando um pouco os questionamentos acima tem-se alguns pontos a serem levantando:

item (a)

Grande parte do início do trabalho deve-se a falta de experiência com imagens e como explorar, quais os melhores caminhos a traçar com as leituras para melhores explorar os atributos. Sobre a criação dos agrupamentos (*clusterings*), algumas técnicas como *k-Means*, *X-Means* e *k-Medoids* foram melhor entendidas, mas que por experiências passadas, a escolha final foi *k-Means*.

item (b)

Sobre o trabalho com histograma de usando *RGB*

parecia inicialmente suficiente, porém depois de muito teste foi decidido aplicar novas ensaios usando *HOG* e resultou muito mais interessantes para os dados agrupados. Porém *HOG*, teve um ponto muito importante comparado com *RGB*. Enquanto este resultava em um vetor de tamanho 768 (256 para cada cor) o *HOG* resultou em um vetor de saída de tamanho 3456 posições onde só foi possível executar após o processamento de *PCA* para redução de dimensionalidade dos dados de entrada.

item (c)

A escolha por *HOG* pareceu mais interessante (e atendeu muito bem ao resultado final) devido aos vários artigos disponíveis para esse tratamento comparado ao *BIC* (*Bayesian Information Criterion*). No entanto a escolha, dependeu de muito processamento e tempo de execução o que onerou bem próximos trabalhos e/ou uso de outras técnicas. *BIC* teria sido o próximo passo.

item (d)

PCA realmente foi mais que crucial para a execução deste trabalho, principalmente pela execução. Devido a sua alta redução da dimensão dos dados vetorizados, foi possível trabalhar com *HOG*, o que anteriormente não se mostrou muito viável devido ao vetor de 3456 e a quantidade de amostras (10.000) e para *RGB* uma quantidade menor de atributos, por volta de 768, mas com 40.000 amostras.

REFERÊNCIAS

- [1] Carvalho, A., *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. LTC. [Online]. Available: <https://books.google.com.br/books?id=4DwelAEACAAJ>
- [2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [4] Wikipedia. (2015) Determining the number of clusters in a data set. [Online]. Available: https://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set#The_Elbow_Method

APÊNDICE

Na figura 12 representa o tratamento de imagem conforme tratado no início da seção II.

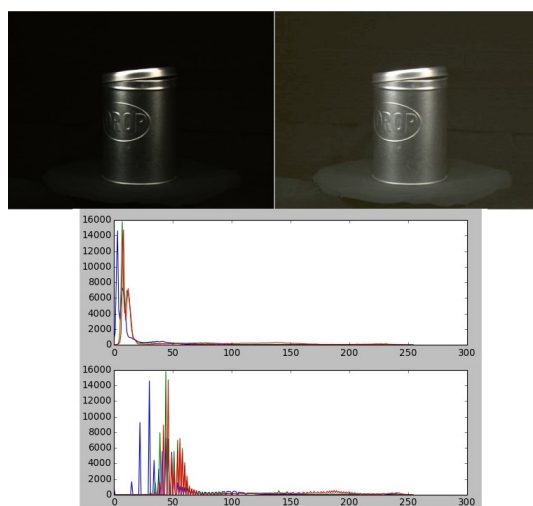


Figura 12: Imagem original (esq.), imagem com aumento de contraste e brilho (dir.) e sua decomposição das Componentes de RGB.