

Algoritmo Baseado nas k -Medias para Estimação Automática do Número de Palavras Visuais

Camila L. Silva¹, Jonathan A. Silva¹

¹Laboratório de Robótica de Ponta Porã (LaRPP)
Universidade Federal de Mato Grosso do Sul (UFMS)
Ponta Porã, MS – Brazil

{camila.leite002, jonathannos}@gmail.com

***Abstract.** This work presents a k -Means based algorithm to automatically estimate the number of clusters to compose a bag of visual words. As a complementary contribution, we purpose consider the spacial information of the clusters to describe images. As expected results, we believe that the spacial information of the clusters can be relevant to describe images, as well as find the natural number of visual words*

1. Introdução

A identificação de quais objetos estão presentes em uma imagem constitui tema ativo na literatura, e, um dos motivos para tal é a sua vasta área de atuação, como no auxílio ao reconhecimento de pessoas, e itens, em câmeras de segurança, e sintetizadores de vídeo para voz [Yang et al. 2007]. Contudo, para que seja possível o reconhecimento, é importante extrair características que descrevam os objetos na imagem, para então um algoritmo de classificação induzir um modelo.

Neste trabalho será abordado o método Bag of Visual Words [Pedrosa et al. 2015], que, basicamente, consiste na definição de um dicionário de vocabulários visuais, desenvolvidos a partir de pontos de interesses extraídos da imagem – tipicamente via SIFT [Lowe 1999] ou SURF [Bay et al. 2008]. Em geral, o algoritmo das k -Médias é utilizado para determinar os vocabulários – palavras – visuais, porém, uma das limitações do algoritmo consiste na determinação, a priori, do número ideal de grupos – k – a ser usado, já que tal informação pode não estar clara em determinadas bases de imagens, como as que possuem um grande volume de dados, ou as com característica volúveis.

Objetivando automatizar a estimação de k na etapa de determinação de um vocabulário visual, é proposta a aplicação do algoritmo OMR k [Naldi et al. 2011] para a estimação automática do vocabulário de palavras visuais, já que, dentre os métodos de agrupamento de dados utilizados atualmente na literatura, ainda não há levantamentos relevantes sobre a utilização do conceito abordado por [Naldi et al. 2011] em técnicas que envolvem manipulação de dados em imagens, como o BoVW.

Simultaneamente, propomos a inserção, nos histogramas das imagens, de características relacionadas à estrutura dos grupos, para complementar a caracterização de uma imagem, considerando que a abordagem comumente utilizada na literatura relaciona apenas o histograma de palavras visuais. Além disso, as informações espaciais desses grupos podem ser relevantes na caracterização dos mesmos, uma vez que, com a abordagem de histogramas, a informação espacial é ignorada.

2. Solução Proposta

Para realizar o levantamento de dados para indução de um classificador, foi selecionado o método Bag of Visual Words. Tal técnica consiste na criação de um dicionário de palavras-visuais, por meio da extração de características nas imagens. Em seguida, forma-se um histograma, contendo a informação relacionada com a frequência de ocorrência de cada palavra visual na imagem. A partir de um conjunto de treinamento, contendo as informações desses histogramas e suas respectivas informações de classe, para cada imagem é possível aplicar um algoritmo de classificação [Pedrosa et al. 2015].

Dentre todas as etapas do método, as que serão exploradas pelo trabalho são: a etapa de agrupamento de características, para formação de palavras visuais, e complementação das características do histograma, com informações das estruturas dos grupos, no caso, as distâncias entre grupos e intra grupos.

O processo de agrupamento envolve a identificação de um conjunto de grupos que descrevam um conjunto de dados, objetivando-se maximizar a homogeneidade entre os objetos de um mesmo grupo, e, a heterogeneidade entre objetos de grupos distintos. Um dos principais objetivos de tal tratamento dos dados, é possibilitar a transformação das informações em conhecimento [Martins et al. 2016].

Uma das dificuldades na tarefa de agrupamento, envolve a determinação prévia do número de grupos, como necessário no algoritmo de k -Médias. Assim, propomos a utilização do algoritmo OMR k , que realiza múltiplas execuções do k -médias, variando gradativamente o número k de grupos, até que o critério de parada seja alcançado. Posteriormente, como proposto por [Naldi et al. 2011], o critério de parada é definido como o agrupamento que possui melhor valor, de acordo com o critério da Silhueta Simplificada [Campello and Hruschka 2006].

Além da estimativa do valor de k , propomos ainda, na etapa de elaboração do histograma, a utilização das informações espaciais das palavras-visuais. Ou seja, além de usar o número de vezes que determinado vocabulário-visual surge no histograma, como característica na imagem, empregar, também, as silhuetas e informações referentes aos grupos aos quais os vocabulários pertencem.

3. Considerações Finais

As propostas não possuem foco apenas em melhorias na extração das características das imagens, como é frequentemente visto na literatura. Busca-se avanço no processo de formação das k palavras-visuais, através da estimação automática do número k , pois, segundo a abordagem clássica, tal número deve ser pré-determinado, condição que pode não ser óbvia em prática.

Juntamente, temos a inclusão de novas informações ao histograma, com dados que dizem respeito aos grupos formadores das palavras-visuais de uma imagem. Tais inclusões têm potencial de ser importantes na construção de um classificador, já que as informações espaciais dos grupos, analisadas pelas suas silhuetas, podem fornecer informações relevantes para caracterização de uma imagem, como as distâncias entre os grupos e de seus elementos entre si.

Por possuir caráter iniciativo, e de ainda não estar munido de resultados definitivos, motiva-se o uso do OMR k em detrimento de outras técnicas pelos bons resultados

apresentados em pesquisas que o utilizam, como em [de Andrade Silva et al. 2017] no agrupamento em dados dinâmicos, caráter positivo que é buscado na pesquisa atual.

Da mesma forma, a utilização de características globais dos grupos na descrição das imagens pode ser justificada pelas conclusões obtidas em [Yang et al. 2007], que mostram o efeito otimista deste tipo de abordagem ao utilizar, além de palavras visuais, o conceito de frases visuais.

References

- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.
- Campello, R. J. and Hruschka, E. R. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157(21):2858–2875.
- de Andrade Silva, J., Hruschka, E. R., and Gama, J. (2017). An evolutionary algorithm for clustering data streams with a variable number of clusters. *Expert Systems with Applications*, 67:228–238.
- Lev, G., Sadeh, G., Klein, B., and Wolf, L. (2016). Rnn fisher vectors for action recognition and image annotation. In *European Conference on Computer Vision*, pages 833–850. Springer.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.
- Martins, E. S., Ribeiro, M., Lisboa-Filho, J., Reijo, F., Freddo, A., and Reis, L. P. (2016). Clustering of spatial data for knowledge extraction. In *Information Systems and Technologies (CISTI), 2016 11th Iberian Conference on*, pages 1–6. AISTI.
- Naldi, M. C., Campello, R. J., Hruschka, E. R., and Carvalho, A. (2011). Efficiency issues of evolutionary k-means. *Applied Soft Computing*, 11(2):1938–1952.
- Pedrosa, G. V., Traina, A. J. M., et al. (2015). Making an image worth a thousand visual words. In *Workshop de Visão Computacional, XI*. Universidade de São Paulo–USP.
- Yang, J., Jiang, Y.-G., Hauptmann, A. G., and Ngo, C.-W. (2007). Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 197–206. ACM.