Bag of Visual Words

Alan Peixinho

Bag of Visual Words

Bag of Visual Words

- Descritor baseado em **características locais**.
- Capaz de adaptar-se a diversos problema de categorizacão.
- Robusto a oclusões parciais.
- Pode ser invariante a alterações de escala e rotação.

Bag of Words

Surgiu como uma adaptação do modelo de Bag of Words (classificação de documentos de texto) para problemas de classificação e recuperação de images por conteúdo.

Histograma de Palavras

In brightest day
In blackest night
No evil shall escape my sight
Let those who worship evil's might
Beware my power, Green Lantern's Light

Dicionário

good	0
evil	2
love	0
day	1
night	1
...	
rain	0

Bag of Words

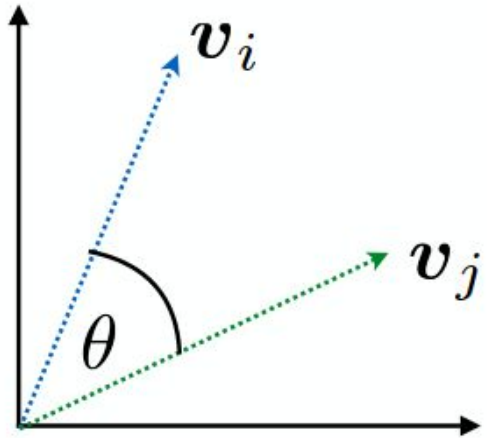
O descritor do documento *d* é representado pela probabilidade de ocorrência de cada palavra que compõe o dicionário.

Sendo *h* o histograma de palavras, e c o número de palavras presentes no dicionário, temos:

$$bow(d) = \frac{h(d)}{c(d)}$$

Similaridade de Documentos

A similaridade entre dois documentos pode então ser quantificada, como a similaridade de cossenos entre os descritores.



$$s(v_i, v_j) = \frac{v_i \cdot v_j}{\|v_i\| \, \|v_j\|}$$

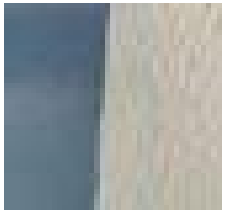
Palavras Visuais

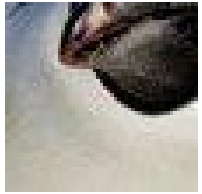
Nos problemas de imagem, teremos o que chamamos de **palavras visuais**.

As palavras visuais são pequenas porções da imagem, capazes de descrever de forma satisfatória a imagem como um todo.

Palavras Visuais

É possível descrever o objeto contido na imagem, utilizando apenas palavras visuais?





Palavras Visuais

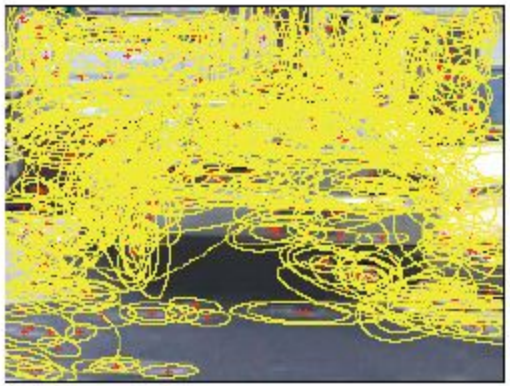
Pipeline

1. Amostragem de Regiões
2. Extração de Características
3. Aprendizado do Dicionário Visual
4. Histograma de Palavras Visuais.

Amostragem de Regiões

Quais regiões da imagem devem ser consideradas palavras visuais?

Amostragem de Regiões



Sparse, at interest points

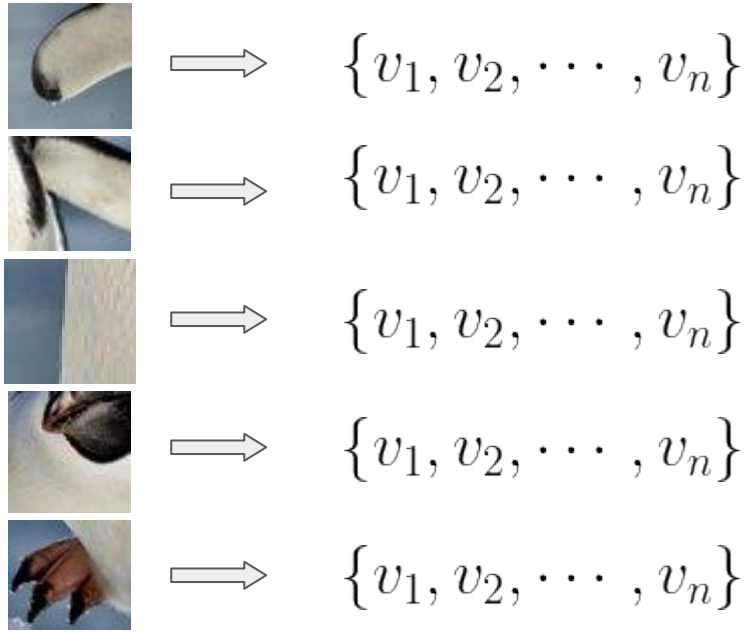
Dense, uniformly

Randomly

Extração de Características

Amostradas as regiões em todas as imagens de aprendizado, extraímos seus respectivos vetores de características.

Um dos descritores mais utilizados é o **SIFT**.



Aprendizado do Dicionário Visual

Assim como na categorização de textos, precisamos definir um dicionário com as palavras visuais a serem consideradas.

Mas como definimos quais palavras visuais utilizar?

Aprendizado do Dicionário Visual

A abordagem mais comum é utilizar técnicas de agrupamento (clustering) sobre todas as palavras visuais.

Essas técnicas de agrupamento buscam encontrar grupos de dados similares, dentro de um conjunto maior.

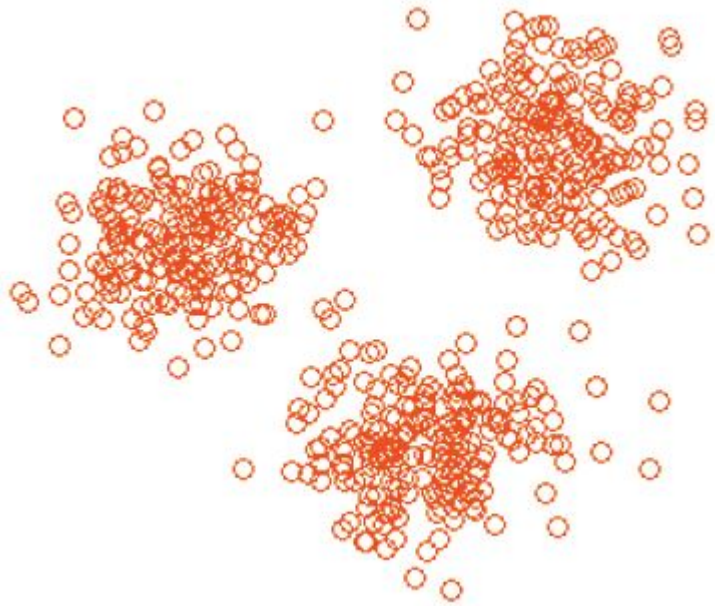
A técnica de agrupamento mais simples da literatura é o **k-means**.

K-Means

Definido o número k de grupos

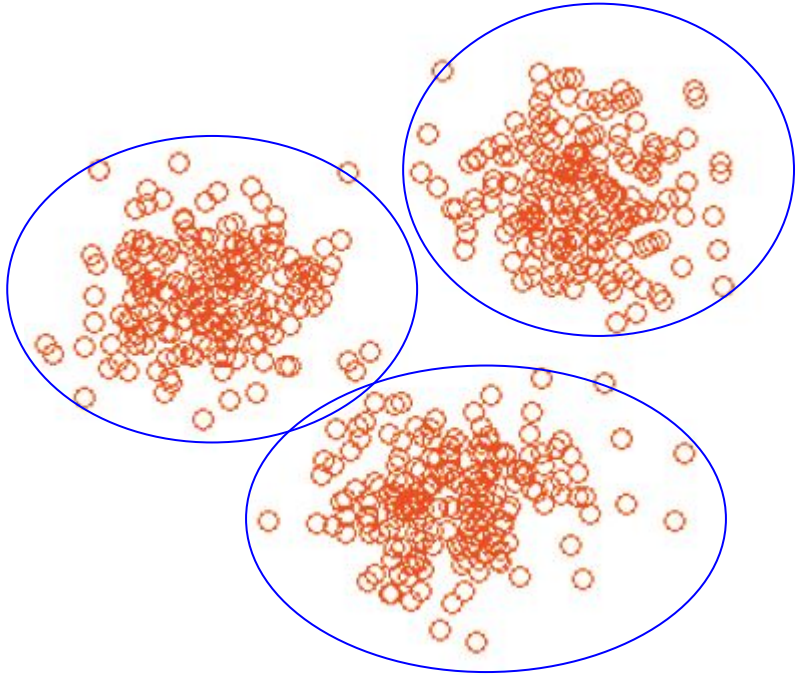
1. Selecione um número k de centróides aleatóriamente.
2. Para cada amostra, associe-a ao centróide mais próximo.
3. Calcule a média dos elementos de cada centróide.
4. Repita as etapas 2 e 3 até que não ocorram mudancas significativas.

KMeans



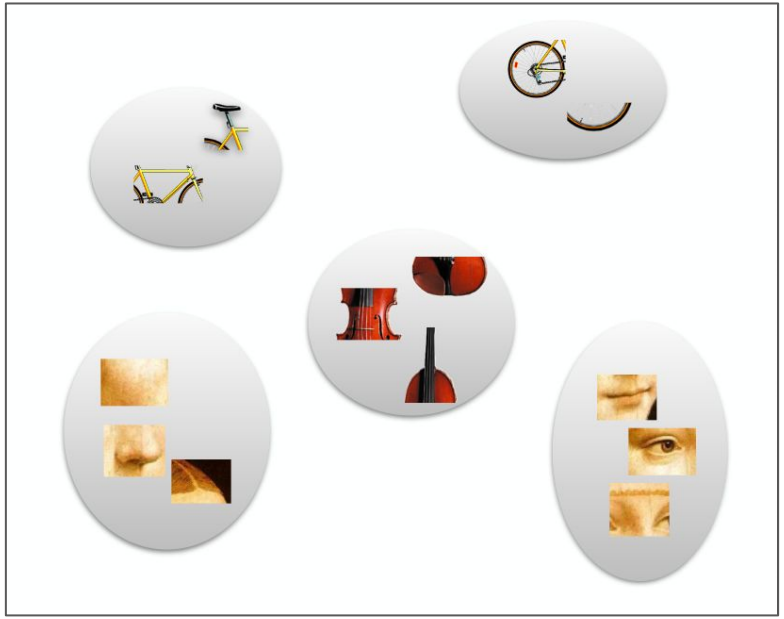
KMeans

K=3



Aprendizado do Dicionário Visual

Os centróides definidos pelo agrupamento serão as palavras visuais que comporão o nosso dicionário visual.



Histograma de Palavras Visuais

Definido o dicionário visual, para cada imagem devemos “contar” a frequência das palavras visuais.

Temos duas abordagens de realizar esta contagem:

- ➔ Hard Assignment
- ➔ Soft Assignment

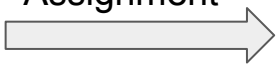
Hard Assignment

Para cada região amostrada da imagem associamos a palavra visual mais próxima.

Contamos a frequência de cada uma das palavras visuais, criando um histograma de palavras visuais.

Hard Assignment

Assignment



Assignment Vector

A	0
B	1
C	0
...	
N	0

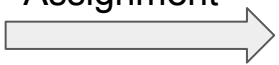
Soft Assignment

Para cada região amostrada da imagem calculamos a probabilidade de pertença para cada palavra visual.

Somamos a probabilidade de pertença de cada uma das palavras visuais ao dicionário, criando um histograma de palavras visuais.

Soft Assignment

Assignment



Assignment Vector

A	0.001
B	0.250
C	0.01
...	
N	0.05

Histograma de Palavras Visuais

Sendo a_i o assignment vector da i-ésima palavra visual de uma imagem.

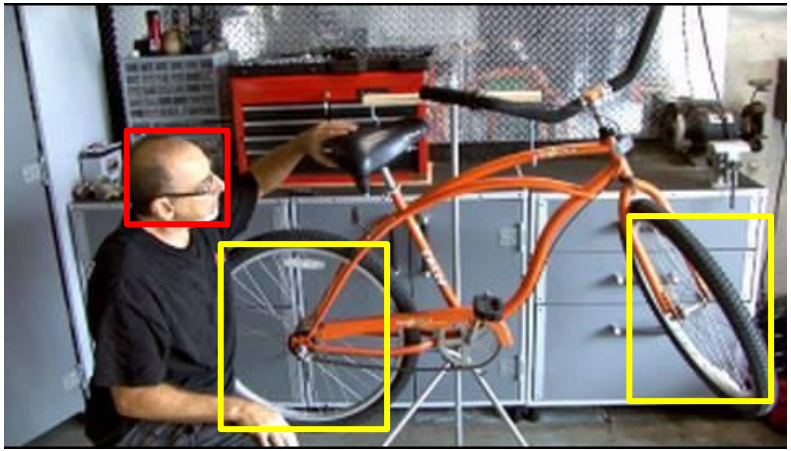
O descritor da imagem é o histograma h, definido por:

$$h = \sum_{i=0}^{n} \frac{a_i}{n}$$

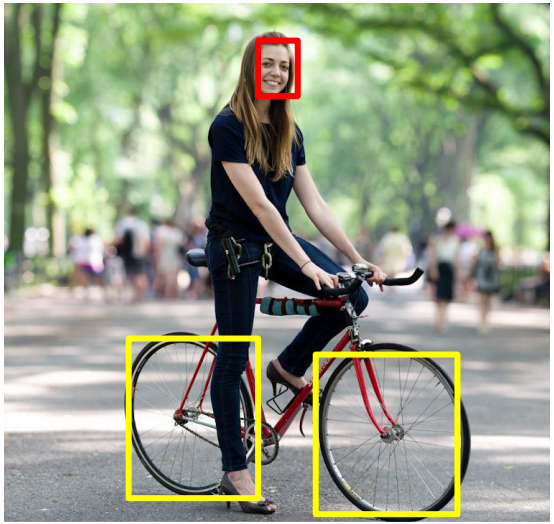
Informacão Espacial

Como pode se imaginar, o BOW possui grande desvantagem em problemas onde a localizacão espacial apresenta fator relevante.

Mechanic



Athlete



Informacão Espacial

Uma das abordagens mais comuns para se lidar com o problema são as piramides espaciais.

Onde o descritor BOW é computado para janelas da imagem em diferentes resolucões e concatenado.

Informacão Espacial