Bag-of-Words in Image Annotation

Tsai, Chih-Fong. "Bag-of-words representation in image annotation: A review." *ISRN Artificial Intelligence* 2012 (2012).

Lu Yu

iMorpheus.ai

Friday 24/11/2017

12:00PM (GMT+8)

Bag-of-Words (BoW)

- Describes image by histogram of “word” occurrences
- Represents feature on whole image level, compared with feature descriptor that captures local information
- No geometric information
- Brings up semantic gap between low level image feature and high level human recognition

Image representation procedure

BoW from images involves the following steps:

1. Detect regions/points of interest
2. Compute local descriptors over those regions/points
3. Quantize the descriptors into words to form the visual vocabulary
4. Find the occurrences in the image of each specific word in the vocabulary for constructing the BoW feature (or a histogram of word frequencies)

Image representation pipeline

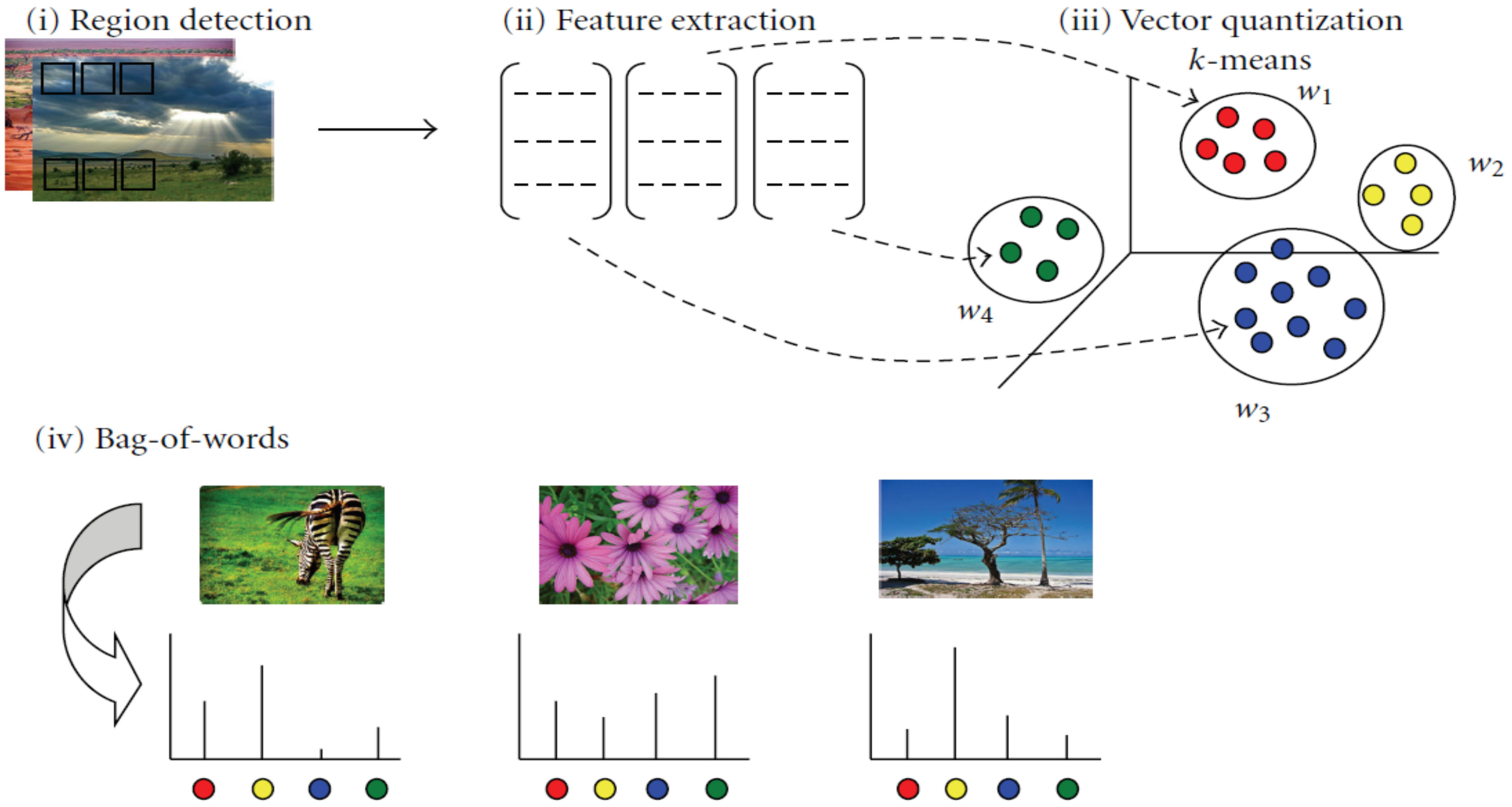
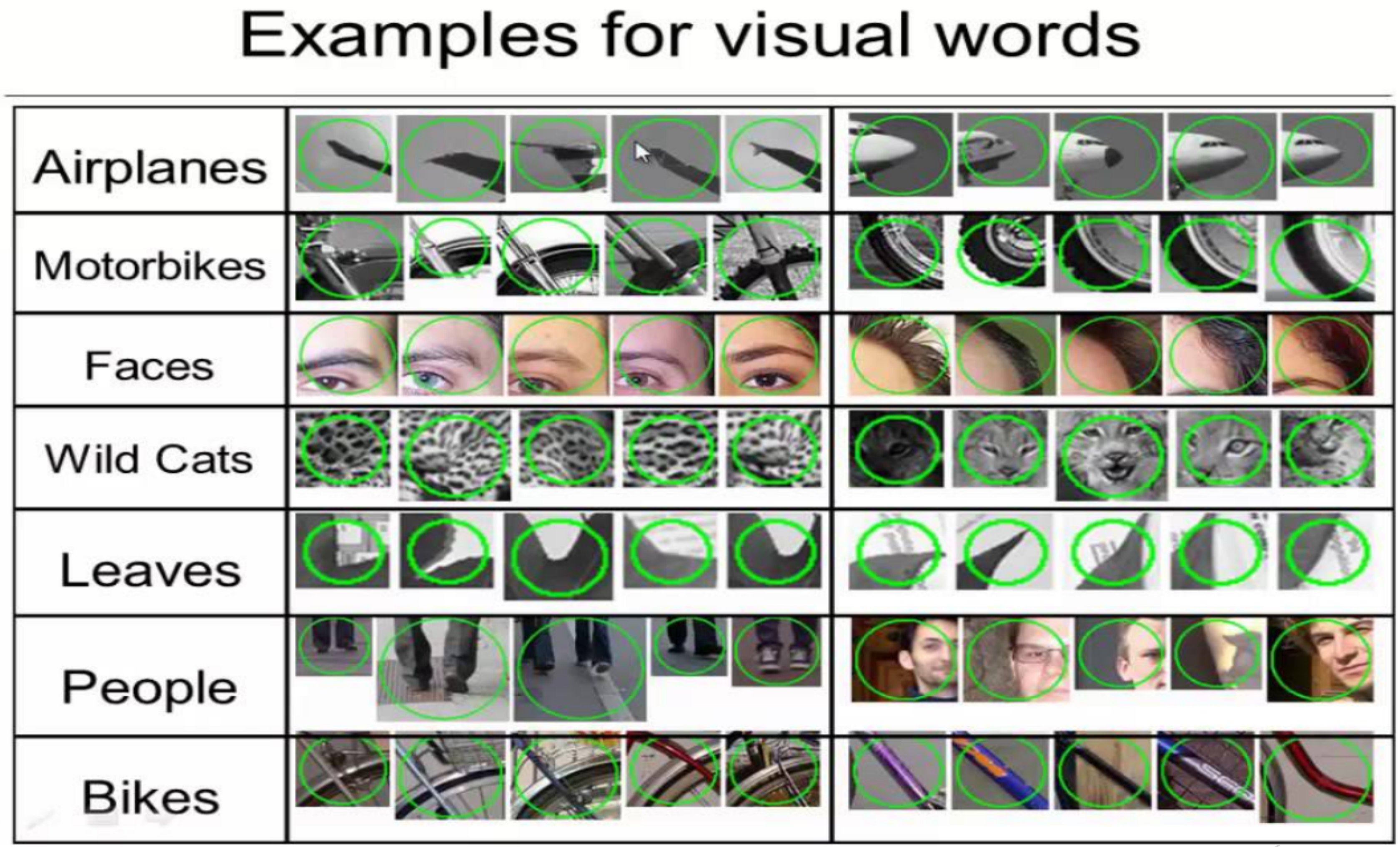


FIGURE 1: Four steps for constructing the bag-of-words for image representation.

Visual words



BoW applications

- Image classification
- Image query and retrieval
- Object detection
- Scene categorization
- New class detection

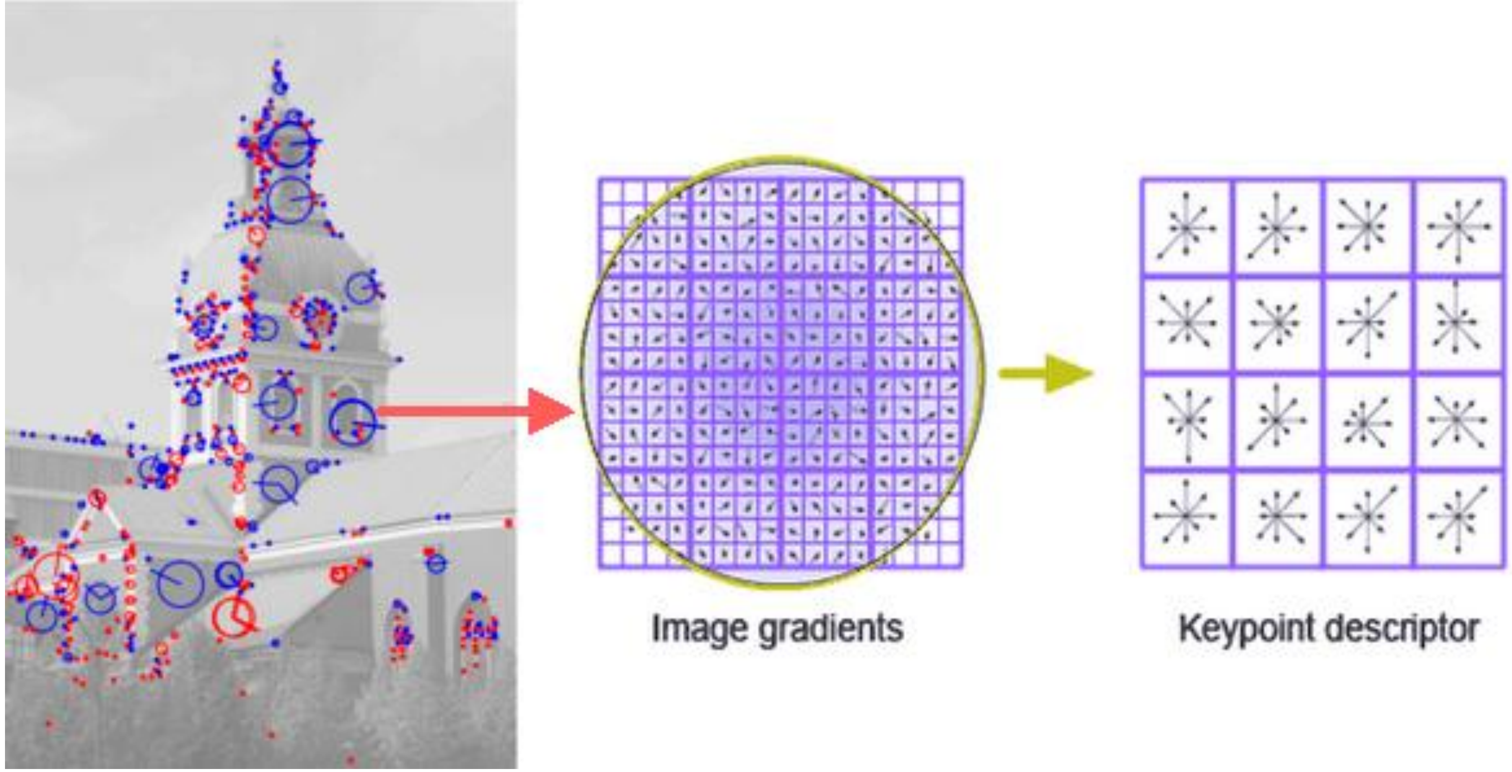
Region/point detection - DoG

- Difference of Gaussians
- Subtract one blurred version of original image from another, less blurred version of the original

$$\Gamma_{\sigma,K\sigma}(x,y) = I * \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} - I * \frac{1}{2\pi K^2\sigma^2} e^{-(x^2+y^2)/(2K^2\sigma^2)}$$

Feature extraction - SIFT

- Scale-invariant feature transform

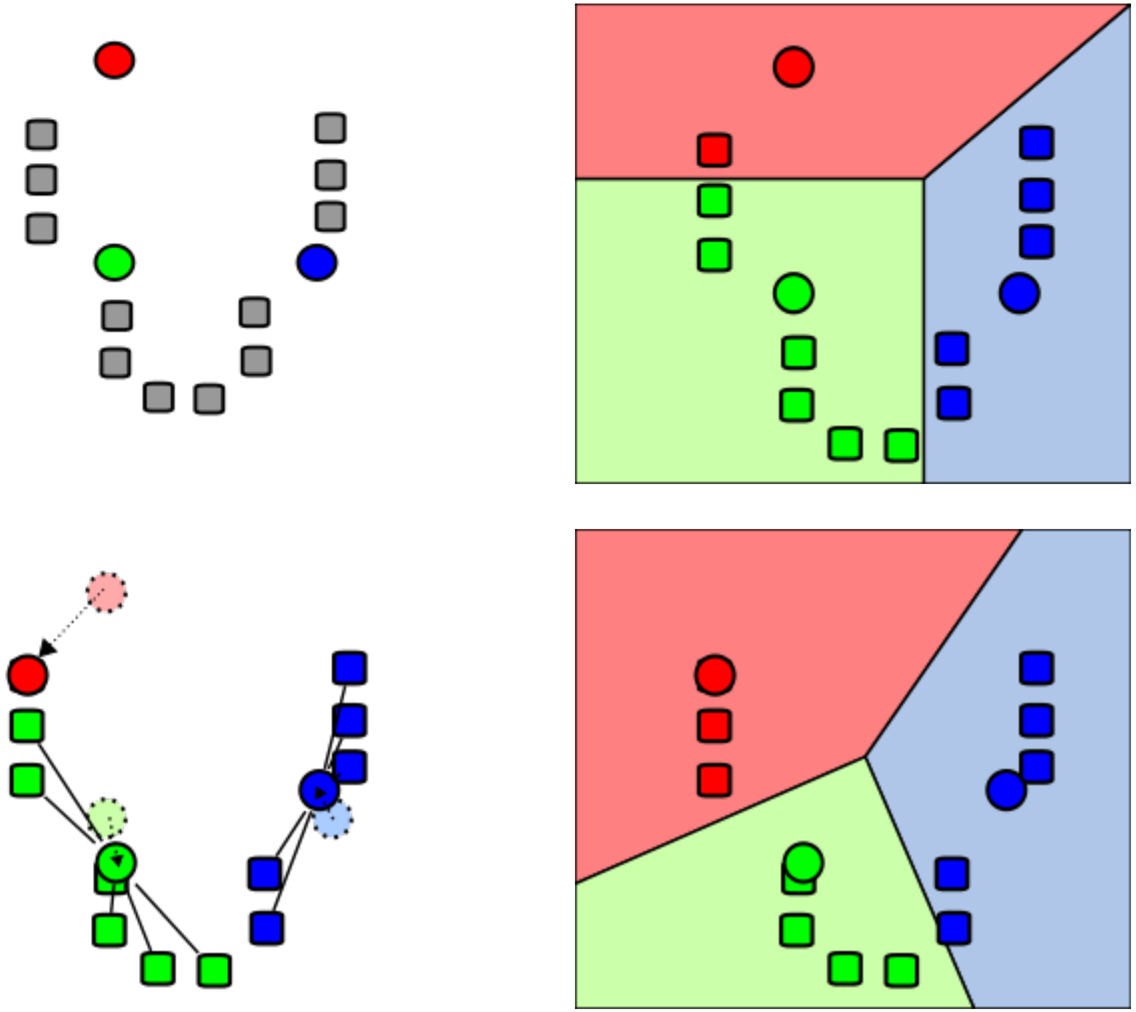


K-means clustering

Partitions n observations (points) into k clusters without prior classification rule

1. Choose k initial centers (can be random)
2. Calculate point distances to centers
3. Points closest to each center form cluster
4. Update cluster center by taking mean
5. Repeat steps 2 to 4 until convergence

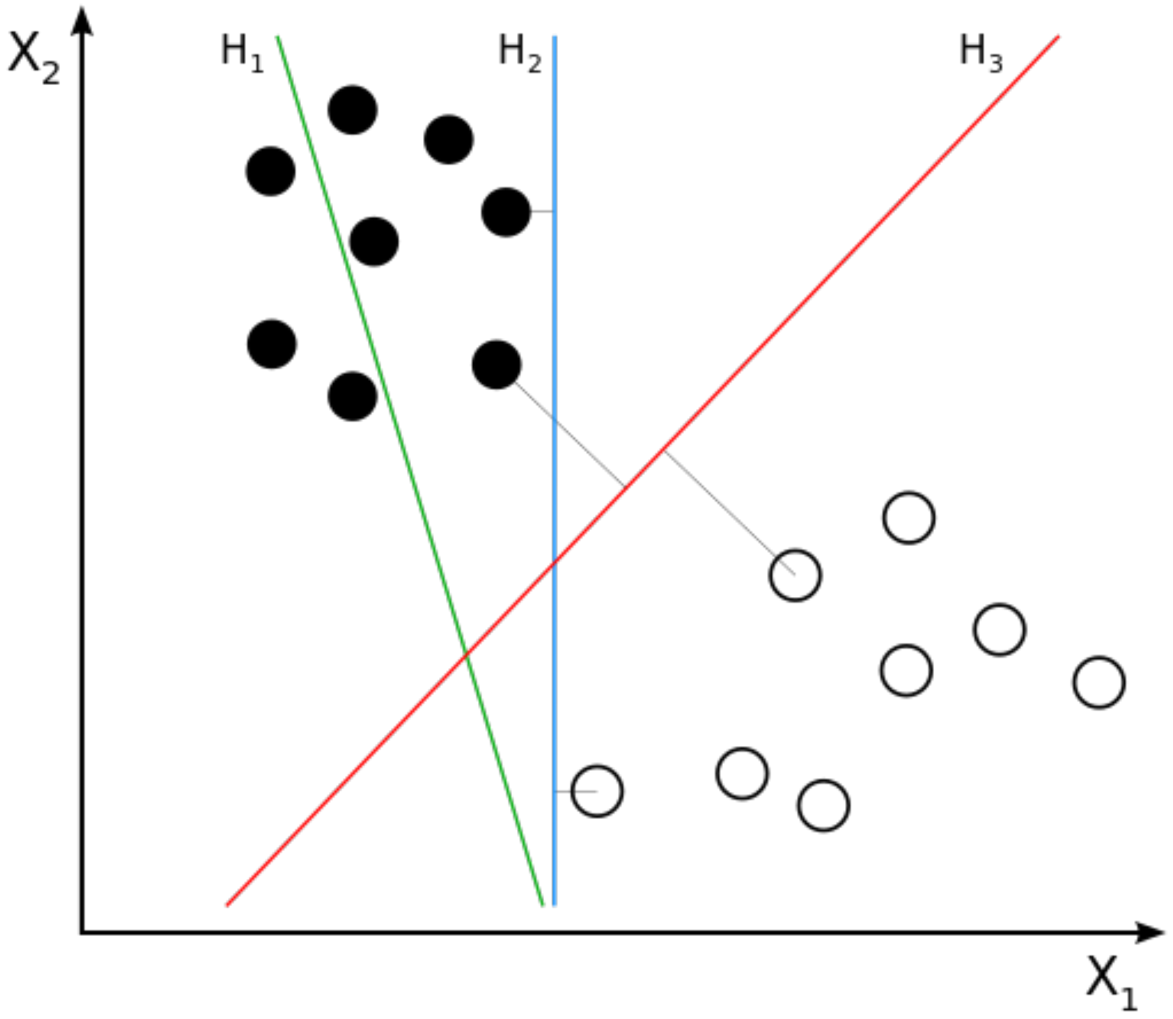
K-means clustering



Support vector machine (SVM)

- Supervised machine learning model
- Deterministic binary classifier
- Data: high dimensional vector
- Training set is labeled by two categories
- SVM model tries to find a hyperplane that separates two categories as far as possible
- New data (unlabeled) are categorized by the hyperplane (rule) already learned

Support vector machine (SVM)



Limitations of BoW

- Lack of spatial or geometric description
- Lack of relation among words

BoW in ORB SLAM

- K-means clustering into hierarchical tree: leaf nodes are words and other nodes are clustering centers
- For loop closure, ORB SLAM calculates and compares bag-of-words vector with keyframes in covisibility map.
- DBoW2: open source C++ library on GitHub.

iMorpheus Journal Club

Next Friday, 1/12/2017 12:00PM GMT+8:

Evaluation of Local Detectors and Descriptors for Fast Feature Matching

微信 **Wechat**

Website : **http://imorpheus.ai**
Email Address : **live@imorpheus.ai**