Bag of Visual Words

Apr. 7, 2019

Sunho Kim

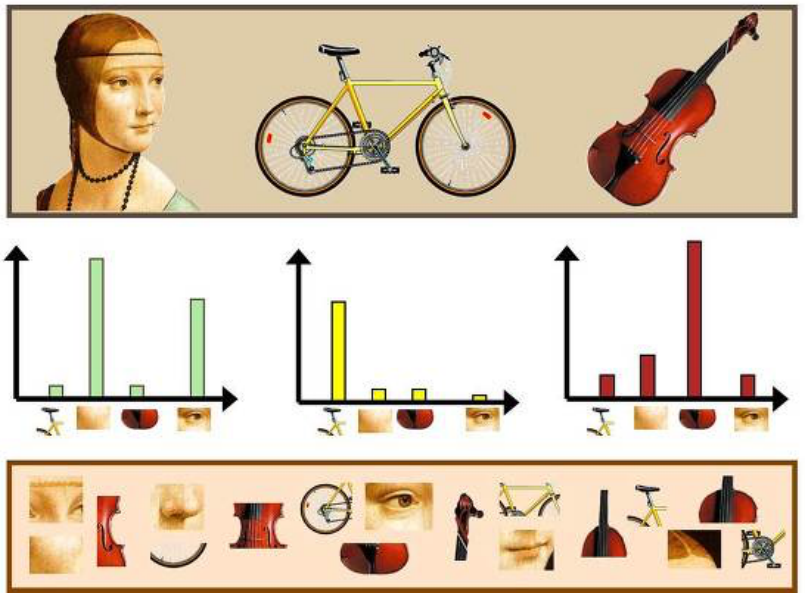
Bag of Visual Words

- Limitation of previous methods
  - 단순 특징점 매칭을 사용한다면?
    - 많은 시간 소요
    - 조명 환경이 다를 경우, 디스크립터를 이용한 비교가 불안정하다.

- Bag of Visual Words
  - 영상의 특징을 '단어'로써 설명하자?!
  - Process
    1. 영상 내에서 '단어'의 개념을 결정하여 하나의 '사전'을 구성한다.
    2. 단어(히스토그램)을 통해 영상 전체를 설명한다. 이 설명은 벡터 형태로 변환된다.
    3. 서로 다른 영상 간 유사성을 검사한다.

Bag of Visual Words

- Basic concepts
  - Process detail
    1. Feature extraction: SIFT, ORB, …
    2. Clustering: 추출된 특징점들에 대해 군집화 수행 후, Cluster center를 찾아 이를 codeword로 정의한다. (k-means clustering)
    3. Codebook generation: 전 단계에서 정의된 codeword들로 구성됨. Codebook 내 codeword 갯수는 clustering 과정에서 몇 개의 cluster로 수행할지에 따라 조절될 수 있다.
    4. Image representation: 각각의 영상들을 앞서 생성한 codeword들의 히스토그램으로 표현한다.
    5. Learning and recognition: 히스토그램을 기반으로 학습 및 인식 과정 수행



https://darkpgmr.tistory.com/125

Visual codebook

- K-means clustering
  - 총 N개의 데이터 $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$(각 데이터는 D개 차원으로 구성됨)를 K개의 클러스터로 나누자
  - D차원 벡터 $\boldsymbol{\mu}_k$ (k=1, 2, …, K)로 각 집단의 중심 표현
  - 목표: 각각의 데이터 포인트로부터 가장 가까운 $\boldsymbol{\mu}_k$까지 거리의 제곱합들이 최소가 되도록!
    - 적절한 $\boldsymbol{\mu}_k$를 찾고, 각 데이터 포인터들을 해당하는 각 집단에 할당할 것
  - Distortion measure function

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

    - $r_{nk} \in \{0,1\}$: 어떤 n 번째 샘플 $\mathbf{x}_N$이 k 번째 클러스터에 속하는 경우 $r_{nk} = 1$ 이고 아닌 경우 0이 된다.

http://norman3.github.io/prml/docs/chapter09/1.html
PRML 9장 1절

Visual codebook

- K-means clustering

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \mu_k\|^2$$

  - J를 최소화하는 $\boldsymbol{\mu}_k$와 r_{nk}를 구해야 한다.
    - 임의의 $\boldsymbol{\mu}_k$를 설정하고, 이 값이 고정된 상태에서 J를 최소화하는 r_{nk} 구하기 (Expectation)
      - J는 r_{nk}에 대해 선형 함수이며 $\mathbf{x}_k$는 모두 서로 독립적이다.
      - 즉, 최적화 과정에서는 다른 샘플과의 연관성을 고려할 필요 없이 각각에 대해 최적의 값을 찾으면 된다.
      - 각 클러스터 중심과 샘플의 거리를 측정해서 가장 가까운 클러스터를 선택하면 된다.

$$r_{nk} = \begin{cases} 1 & if\ k = argmin_j \|\mathbf{x}_n - \mu_j\|^2 \\ 0 & otherwise \end{cases}$$

    - 위 단계에서 얻은 r_{nk}를 고정하고, J를 최소화하는 $\boldsymbol{\mu}_k$ 구하기 (Maximization)
      - J는 $\boldsymbol{\mu}_k$에 대해서는 Quadratic. → 미분을 통해 최소가 되는 지점을 구할 수 있다!

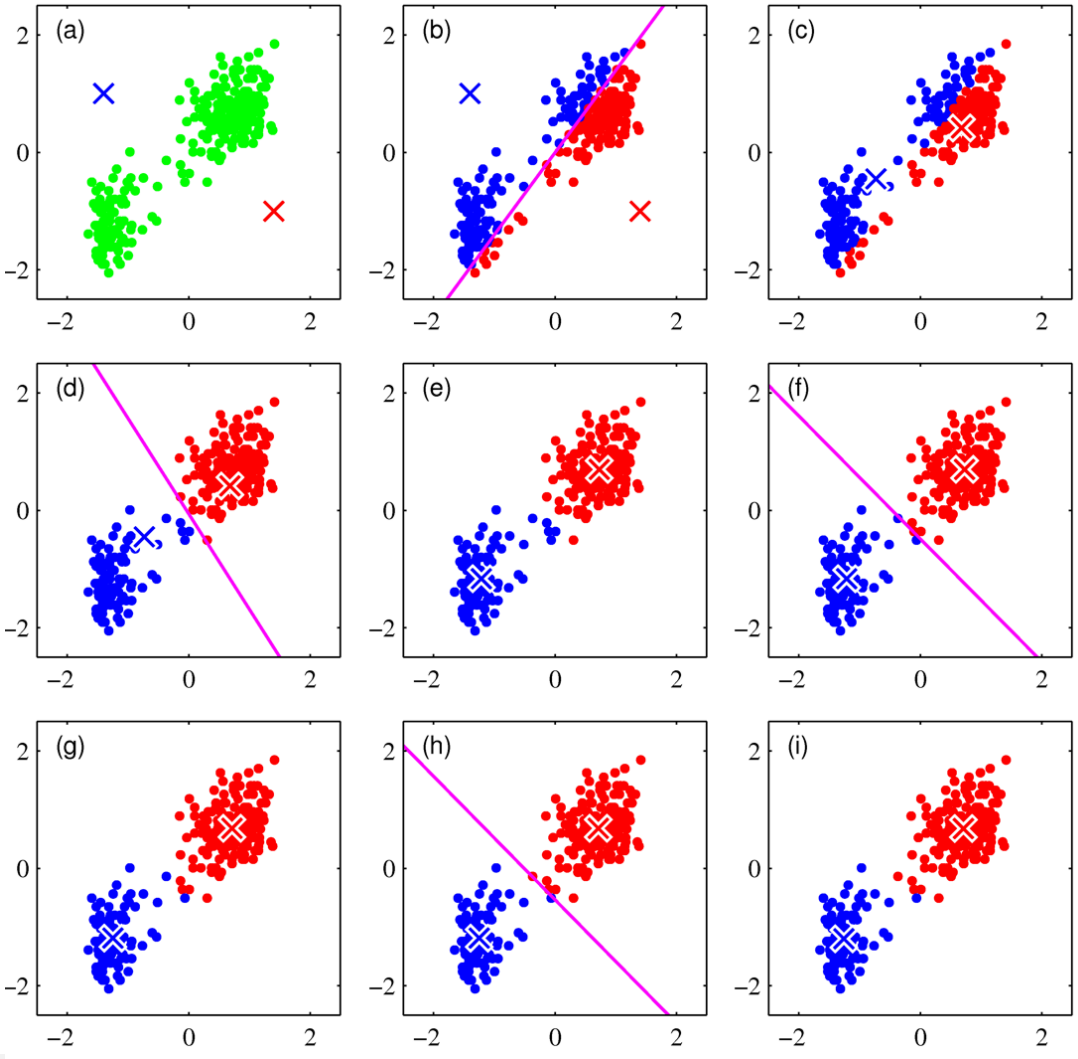
$$2\sum_{n=1}^{N} r_{nk}(\mathbf{x}_n - \mu_k) = 0 \qquad \mu_k = \frac{\sum_n r_{nk}\mathbf{x}_n}{\sum_n r_{nk}}$$

    - 두 값이 적당한 범위로 수렴하거나 반복 횟수가 최대에 도달할 때 까지 이 과정을 반복한다.
  - Expectation-Maximization(EM) 알고리즘

http://norman3.github.io/prml/docs/chapter09/1.html
PRML 9장 1절

Visual codebook

- K-means clustering



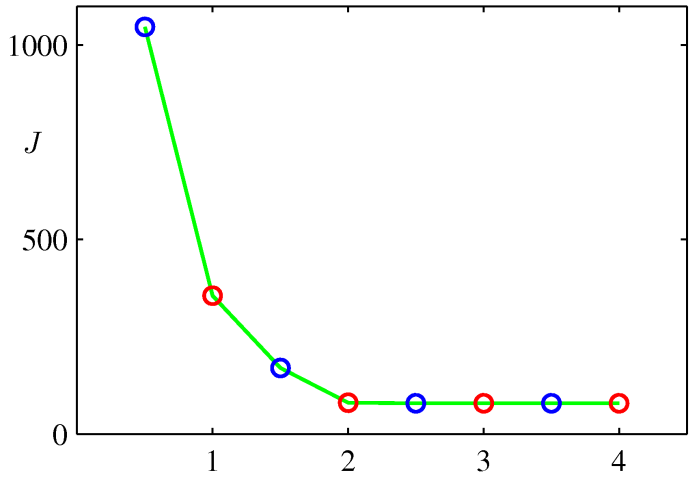
http://norman3.github.io/prml/docs/chapter09/1.html
PRML 9장 1절

Visual codebook

- K-means clustering
  - EM 과정을 반복함에 따라 그래프와 같이 J 값이 줄어든다.
  - 초기 $\boldsymbol{\mu}_k$설정을 어떻게 하느냐에 따라 성능 차이가 발생한다.
    - 샘플 내 한 점을 초기 $\boldsymbol{\mu}_k$로 설정하는 방법도 사용되곤 한다.
  - 학습 속도
    - Expectation 단계에서 평균 값과 모든 점들 간 비교 연산이 있어 많은 계산량 소모
    - 데이터를 트리 구조로 저장하여 이를 개선 (KD-Tree, Chow-Liu Tree)
  - 온라인 방식의 알고리즘 유도 방법
    - J를 $\boldsymbol{\mu}_k$에 대해 미분하며 회귀 함수를 얻고, 이에 대한 제곱근 값을 계산한다. (Robinson-Monro algorithm)

$$\mu_k^{new} = \mu_k^{old} + \eta_n(\mathbf{x}_n - \mu_k^{old})$$

    - η_k: learning rate.



http://norman3.github.io/prml/docs/chapter09/1.html
PRML 9장 1절

Similarity measurement

- TF-IDF
  - Term Frequency: 특정 단어가 문서 내에서 얼마나 자주 등장하는가
  - Document Frequency: 특정 단어가 다른 문서에서도 얼마나 자주 등장하는가
  - Inverse Document Frequency: 반대로 다른 문서에서는 잘 등장하지 않는 단어!
  - TF-IDF: 특정 단어가 해당 문서 내에서는 자주 등장하는데, 다른 문서에서는 자주 등장하지 않는다!

$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

https://www.bloter.net/archives/264262
https://nesoy.github.io/articles/2017-11/tf-idf

DBoW

Bags of Binary Words for Fast Place Recognition in Image Sequences

Dorian Gálvez-López, Juan D. Tardós

Robotics, Perception and Real Time Group
Departamento de Informática e Ingeniería de Sistemas
Instituto de Investigación en Ingeniería de Aragón
Universidad de Zaragoza, Spain

Gálvez-López, Dorian, and Juan D. Tardos. "Bags of binary words for fast place recognition in image sequences." *IEEE Transactions on Robotics* 28.5 (2012): 1188-1197.

Additional issues…

- 사전의 크기: 클 수록 비교 대상이 많아져 보다 정확도 높은 결과가 나올 수 있으나…
- 유사성 점수 처리
  - 절대적인 점수에만 의지하는 것이 도움이 되지 않을 수 있다.
  - 다른 사무실인데 유사한 스타일로 인테리어가 되어 있거나…
  - 선험적 유사도 활용, 유사도에 대해 절대적인 임계치를 부여하지 않는다.
- 키프레임
  - 너무 인접한 프레임들을 모두 키프레임으로 선정하는 것은 좋지 않아 (유사도)
  - 픽셀 간 평균 시차 or 트래킹이 유지되는 특징점의 갯수를 기반으로 키프레임 선정
- 루프 탐지 후 검증 필요 (기하학적)
- 머신 러닝 활용
  - SIFT, ORB 등이 아닌 러닝 기반의 특징 활용?

Qin, Tong, Peiliang Li, and Shaojie Shen. "Vins-mono: A robust and versatile monocular visual-inertial state estimator." *IEEE Transactions on Robotics* 34.4 (2018): 1004-1020.

Thank you