Multimedia oral examination - Feature Matching and Aggregation

1.Near-duplicate Video/Image Retrieval

(1)Whis is near-duplicate Image and Video?

Near-duplicate means images or videos which are identified as “essentially the same” by humans.They are close to each other, but different in details,for example photometric variations or the same scene captured by different cameras.

(2)why is it challenging?

It is a challenging problem because in Large-scale, it is very time consuming and memory consuming.For example,the computation costs of cross-matching a long video are counted in CPU years.

3.What are the Point-to-point matching schemes?

There are three types of matching schemes,including M2O, O2O and OOS. M2O refers to many-to-one match, which means that more than one point can be matched to the same point. O2O refers to a one-to-one match, which means that only the most closest pair of many-to-one matching will be retained. OOS refers to one-to-one and symmetric,which means that only matching pairs that are the closest for both sides will be retained.

One-to-one is more robust than many-to-one,One-to-one and symmetric performs the best.

4.Geometric verification : Least Square Approach

(1)What is Least Square Approach?

Least Square Approach can perform geometric verification on matching points.

The idea of Least Square Approach is that matching points between two images should satisfy an affine transformation with the same parameters(4+2=6).

So we can use a set of matching pairs to estimate parameters,and the estimated affine model can be used to filter false matches.

(2)Pros and Cons?

Advantage: It can perform geometric verification and filter false matches,so it is useful if we require precise estimation.

Disadvantage: It requires large number of correct matches and it is very very slow. And estimating parameters is not a trivial operation.

5.Bag-of visual words (BoW)

(1)What is BoW?

BoW refers to Bag-of visual words. It is based on a feature vocabulary trained by K-means,so each local feature can be quantized into a centroid. Given all local features of an image,the image can be represented by term frequency TF/IDF of visual words.

(2)How to do Online Retrieval with BoW?

Given all local features of an image and pre-established vocabulary,the image can be represented using term frequency (TF) that each word,which is a vector share the same length with the vocabulary.

The resulting vector can be used to mesaure distance between two images.

(3)Pros and Cons

Advantages: If the vocabulary is large enough, the resulting vector is very sparse.So we can use inverted file to organize images, which makes matching becomes very efficient compared to point-to-point matching.

Disadvantages: Many details get lost because of quantization. For example,feature points mapped to the same cluster are with zero distances to each other,so it may introduce many mis-match and false match.

False matches can be reduced by Hamming embedding and Weak Geometric Constraint

6.Hamming Embedding

(1)What is Hamming Embedding ?

In BoVW,feature points mapped to the same cluster are with zero distances,Hamming Embedding helps to estimate the intra-distance within a cluster and reduce noisy matches.A simple idea is to prune matches hold Hamming distance larger than a threshold.

(2)how to do offline training and online Embedding?

In offline training procedure,for each cluster,we project a large set of SIFT features belongs to this cluster into a low dimensional space and compute the median value of each element. Therefore ,for a given SIFT feature,we do the same projection and compare projected element with the precomputed median value to generate a binary embedding vector. The binary vector is attached to each quantized feature and used to verify the visual word match.A simple idea is to prune matches hold Hamming distance larger than a threshold.

(3)pros and cons?

It helps to reduce false matches but will introduce extra memory cost.

7.Reciprocal Geometric Verification(RGV,相互几何验证)

(1)What is RGV?

For a pair of pictures being compared, the scale and rotation angle between them can be estimated by two independent approaches and they should be close to the same for near-duplicate.So the difference between the estimated parameters can be used to re-weight distance between BoWs

The two estimation methods one uses the straight line between two matching points, one uses the dominant orientation of the local patch.

(2)pros and cons?

Reciprocal Geometric Verification can use ignorable cost to enlarge the gap between near-duplicate and non near-duplicate.

7.Vector of Locally Aggregated Descriptor(VLAD,局部特征聚合描述符)

(1)What is VLAD ? How to do ?

In BoVW， each vector is represented by the nearest neighbor in vocabulary which make many details get lost.VLAD aggregate the subtraction of feature and the word which is a more detailed description.

(2)pros and cons?

It helps to reduce false matches and memory complexity is low,but will introduce too many noises.

8.Summary

For BoVW, BoVW+HE, VLAD and VLAD+PQ measured by mean Average Precision,VLAD performs pretty well with much lower memory complexity;BoVW+HE performs the best