

Multimedia oral examination - Image Features

1.Why use image feature? Images are represented as different shapes of matrix in computer. But they cannot be directly compared by computer, we need to extract features that can reflect similarity as the medium to represent image. So we can use image features to measure the similarity between images.

2.CBIR(content based image retrieval) (1)What is CBIR? For a given image as a query,CBIR search for images that share the same or similar content with the query,rather than the metadata such as keywords, tags, or descriptions. "content" might refer to colors, shapes, textures, or any other information that can be derived from the image itself.

(2)Explain the framework of CBIR? First we extract features from the query image,then calculate the similarity/distance between the query image feature and image features in a image candidate which is pre-computed, and return the most similar Top-N images back to the user.

3.Describe the characteristics that image features should satisfy. There are some invariance properties of image features,including scale,rotation,flip,translation invariance.This is because they do not affect human's vision system, so image features also need to be robust to these transformations.

4.Global feature:what?pros&cons? Global feature are some statistical variables defined on the whole image,such as color histogram,color moments,HOG(histogram of oriented gradients) and VLAD.They are easy to compute,but not very distinctive.For example we can randomly change the coordinates of each pixel to build a quite different image,but this image share the same color distribution with the original one.

(1)color histogram it is a statistics on color distribution,that is the the number of pixels in each color range and concatenate them if there are different channels,such as RGB and HSV and Lab which are more distinctive.By color histogram,images are transformed from matrix to vectors.Flip and rotation invariant.

(2)color moment Color moments measures color distribution in an image in the same way as computing moments of a probability distribution.Color moments are scaling and rotation invariant, but they cannot handle occlusion successfully. **(3)HOG(histogram of oriented gradients)** HOG counts occurrences of gradient orientation on each dense grid of image to build a histogram

5.Why corners instead of edges? Corner points are more robust to transformation.For example,points along the edges and in the flat region will merge to each other after scaling and rotation,but corners will survive.

6.How to detect corners? We can take an observing window and move around,if change happens strongly in x and y directions,it is the corner point.if change happens along one direction,the window is on edge;otherwise the window is in a flat region.

To formalize the probing procedure above, we define an Energy function to measure changes in a window with width u and length v by subtracting pixel value on (x,y) and $(x+u,y+v)$.

Harris point detector detect corners by two steps. First calculate Harris function value for each pixel. In the second step, take the point that attains local maximum through Non-maximum Suppression, which means its Harris function value is higher than its 8 neighboring point.

7. local point and scale invariance (1) why scale invariance? We have detected corner points, but we cannot just compare pixel values of them, which is apparently not distinctive. We need to compare a local region, so we should determine the radius. If we fix the value of radius, it will not cover the same local structure as long as they do not share the same scale. So we need to find a characteristic scale for each local point. **(2) how to calculate characteristic scale?** The overall procedure is, first build Harris and LoG (or DoG) octaves with a series of σ s and select points that attain local maxima both with Harris function in X-Y plane and LoG function in scale space. There are some scale space functions like LoG, DoG and Hessian, characteristic scale is the one that makes corner point attain local maximum.

(3) Discuss about LoG, DoG, Harris and Hessian? Harris function is not a good scale space function because it is sensitive to noise, resulting in no significant peak along different scale. Laplacian-of-Gaussian is a good option but a little time consuming, DoG is a good approximation of LoG and more efficient to calculate. Hessian is also widely used. SURF (Speed-Up Robust Feature) is a fast approximation of Hessian by speed-up the convolution operation using integral image.

8. local point and rotation invariance given a local point and its scale, we can build a histogram of gradient for this local region, the peak of the histogram is the dominant orientation, we can rotate this local patch to the dominant orientation to achieve rotation invariance.

9. local point and affine invariance Affine transformation is caused by different viewpoints, involves rotation, scaling and skew. We fix this through transforming an ellipse region back to a circle region by normalizing it until two eigen values are approximately equal.

10. local point and flip invariance We can normalize patch to a fixed pose to make images with arbitrary flip comparable. We fix this by estimating the dominant angular moment, then flip image along x axis and rotate it.

11. SIFT SIFT is the most successful image local feature. First, we normalize the patch to $2d \times 2d$, then calculate gradient field. Partition gradient into 4×4 grid and compute gradient histogram on each block, then concatenate 16 histograms as the final feature. Normally it is an 128-dimensional vector.

There are two tricks: First, in order to normalize, we find a corresponding pixel in original patch for each one in target patch. If we reverse this operation, it will result in overlapping or holes in target patches, which will increase the noise. Second, when computing gradient histogram on each pixel, weight will be shared by two blocks, the portion depends on the distance between pixel and block center.

12. Applications of Image Local Feature Image stitching, image retrieval and visual object detection **Image stitching:** Extract SIFT feature and perform pair-wise match between images, then stitch matching pairs.

13.Global features v.s. Local features Global and Local features process images in different granularity,local features are finer.Both of them have their pros and cons. Global features are easy to compute,but not very distinctive.For example we can randomly change the coordinates of each pixel to build a quite different image,but this image share the same color distribution with the original one. Local features,for example SIFT features, are more distinctive but in high dimension,and only matched locally which could be good and bad.And Gradient based feature cannot distinguish different textures,so SIFT is sensitive to texture.