**Interview Challenge:**

**Computer Vision & Machine Learning for Image Retrieval Systems**

**Image Matching**

The core functionality of Image Retrieval is based on the ability to identify matching images.

1) Assume I and J are images of the same painting in different resolutions, different aspect ratios,

with and without compression artifacts, etc. What is your method of choice for confirming the

two images show the same content? And why did you choose that particular approach?

*Ideal case*

*Assuming the painting has the same scale in both images (the same part of the painting spans the same image proportion), and there is no aspect ratio distortion, translations, nor perspective distortion, i.e. images are centered plane projections; the following processing steps could be applied:*

* *Crop both images in both dimensions to the smaller dimension. This solves the issue of aspect ratio difference.*
* *Resample both images to a resolution coarser than that of the image with coarsest one. This reduces the computational cost of subsequent steps and allows comparison with classic metrics. Possibly, a resolution as low as 256x256 is good enough to distinguish classical painting, but finer resolution could be necessary to distinguish imitations. If one of the images has a resolution low enough, could be better to down-sample only the image with the finest resolution.*
* *If one of the images could be in grey-scale, convert both images to grey-scale.*
* *If the images are to be compared with distance metrics, e.g. as mean square error, at least one of the images should be intensity rescaled in order to match the intensity scale of the other image. If the images are to be compared with correlation metrics, e.g. dot product, intensity re-scaling is not necessary.*

*If centering is not guaranteed, the central part of one image (e.g. 1/3 by 1/3) could be used as template for template matching in the other image.*

*Once the metric is correctly working, a threshold should be defined using Signal Detection Theory on a labeled dataset, not necessarily large. The threshold could be selected by minimizing a cost function dependent on the cost of False Positives and of False Negatives. If these costs are not provided, the threshold value that maximizes informedness (Youden's J statistic) could be used.*

*Non-ideal case (e.g. photographs of the painting)*

*If scale is different or perspective distortion or likely to be in at least one of the images, a more sophisticated approach will be necessary. For example, a Key Point Matching strategy based on Scale Invariant Feature Transform (SIFT) could be exploited, which is already implemented in opencv and its patent expired in the year 2020.*