Assignment 1. Advanced Computer Vision

Petrariu Florin-Iustinian

18 March 2023

1. Exercise 1

1.1 What is Computer Vision?

Computer vision is an interdisciplinary scientific field that enables computers and systems to derive meaningful information from digital images. Being also a field of artificial intelligence (AI), it seeks to understand and automate tasks that the human visual system can do.

1.1 Why is it important?

Computer Vision tasks include methods like acquiring, processing, analyzing and understanding digital images and all these tasks can help many industries and also areas of our life. Thus, computer vision is involved in numerous real-world applications like Medical imaging, Autonomous vehicles and Robotics. Overall, it has the potential to improve productivity and safety.

1.2 How Computer Vision relates to Computer Graphics?

Although Computer Graphics deals with the generation, rendering, and use of 3D representations, along with Computer vision share a common goal of understanding and processing visual information.

1.3 Give the most important 3 applications of computer vision in your opinion and argument your selection and ranking.

In my opinion, the most important applications of computer vision are in medicine. Medical imaging has the potential to improve health and save a life. There are many implications in the healthcare sector and some of them include the ability to recognize patterns and make diagnoses, systems that can help detect brain tumors, surgical guidance and more. Autonomous vehicles are on rank 2 because can make our roads safer and also can increase accessibility for people who are unable to drive. Some of the important applications include: recognizing pedestrians, and road signs, analyzing their surroundings in real-time and recognizing also other vehicles to safely navigate the road. On the last place is manufacturing (quality control) because automated quality control minimizes safety risks, and increases production efficiency. Camera-based systems can collect real-time data and leverage computer vision and machine learning algorithms to analyze it and this helps in identifying defects in the production line more efficiently.

2. Exercise 2

2.1 What is light reflection?

Light reflection is the phenomenon where light waves bounce off a surface (like a mirror) and change direction. The amount of light that is reflected depends on the properties of the surface and also the angle at which light reflects off a surface is determined by the angle at which it strikes the surface as well as the angle of the surface itself. In course, we have three types of reflection: *diffuse*, *specular* and *mixed*.

2.2 Can the angle of refraction be different than the angle of incidence? When and why? Justify your answers.

Yes, the angle of refraction can be different than the angle of incidence. This can occur at the point where light travels from one medium to another of different densities (different refractive index). It is called the phenomenon of refraction which causes the direction of light waves to change with a different refractive index. For example, the angle of refraction is smaller than the angle of incidence when a light ray travels from a rarer medium to a denser medium. As mentioned in the course, the relationship between the angle of incidence and the angle of refraction is given by *Snell's law*.

3. Exercise 3

3.1 Describe the rainbow and the sunrise. Which phenomena, types of light interaction with matter, are involved in the creation of the rainbow and sunrise effects? Be as thorough as possible in your response and justify.

A Rainbow is a phenomenon that is the result of the refraction and reflection of light. Both refraction and reflection are phenomena that involve a change in a wave's direction. More than that, the radius of a rainbow is determined by the water droplets's refractive index. A refractive index is a measure of how much a ray of light refracts as it passes from one medium to another. Also, the rainbow is the result of the dispersion of light in water droplets in the atmosphere. A rainbow shows up as a spectrum of light: familiar colors that include green, red, orange and violet and when sunlight hits a rain droplet, some of the light is refracted and separated into its component colors due to the difference in the refractive index of each color. This separation is known as dispersion which causes the colors of the rainbow to appear in a specific order: *ROYGBIV*. Thus, the spectrum is separated, producing a rainbow. This is because the electromagnetic spectrum is made of light with many different wavelengths, and each is reflected at a different angle. For example, red has the longest wavelength of visible light, about 650 nanometers and usually appears on the outer part of a rainbow's arch.

A sunrise, on the other hand, is caused by the scattering of sunlight in the Earth's atmosphere. Scattering of light is the phenomenon in which light rays deviate from their original path upon striking an obstacle like dust, gas molecules, or water vapors. In the case of sunrise, the scattering of light in the Earth's atmosphere results in the colorful hues of the sky. Another type of light interaction in sunrise is refraction: when the sun is near the horizon and light travels a longer diagonal path to reach our eyes which causes its lighting to be refracted or bent.

In the end, the interaction of light with matter is crucial for both phenomena. The creation of both involves the types of light interactions with matter, such as refraction, reflection, dispersion, scattering, and illumination.

4. Exercise 4

4.1 What is depth of field?

Depth of field (DoF) refers to the area in front of and behind the subject that is in focus, while the rest of the image appears blurred or out of focus. Therefore, is the distance between the nearest and furthest elements in a scene that appear to be acceptably sharp in an image. It is influenced by several factors, including the distance between the camera and the subject, the focal length of the lens and the aperture of the lens.

4.2 If the focal length is assumed fixed, can aperture be adjusted to help in handling objects at far and near distance, respectively?

Aperture refers to the opening of a lens's diaphragm through which light passes. It is calibrated in f/stops. If the focal length is assumed fixed we can adjust departure to help in handling objects at far and near distances, respectively.

The lower the f/stop → the larger the opening in the lens → the less depth of field → the blurrier in the background. This can be useful when the main subject needs to be in focus but the background can be blurred : (e.g.: in portrait photography).

The higher the f/stop → the smaller the opening in the lens → the greater the depth of field → the sharper the background. This can be useful when the entire scene needs to be in focus: (e.g.: landscape photography)

4.3 When a small depth of field is desirable?

As mentioned in the subsection above, a small depth of field is desirable in portrait photography where the subject is in focus and the background is blurred. In this way, we can move the attention to the subject due to the blur surroundings.

5. Exercise 5

5.1 The task is fabric classification into a couple of categories. The fabric image is captured with a camera facing a planar fabric. The fabric image shows a patch from the fabric and nothing else. It is known that fabrics can come in all colors and shapes, regardless of their category. What color space and which channel(s) should be used?

There are numerous color spaces with different properties in literature. For the task at hand, we can use the LAB color space (also known as *CIELAB* or *CIE* L^*a^*b. The L^*a^*b space consists of a luminosity L, chromaticity layer a indicating where the color falls along the red-green axis, and chromaticity layer b indicating where the color falls along the blue-yellow axis. We use this color space because it separates the luminance (brightness) component from the chromatic (color) components, making it a good choice for color-based classification tasks. For the channels, a and b described above would be the most useful as they capture the chromatic information of the fabric. The luminance may also be useful, but maybe not enough informative for distinguishing between fabric categories, as it captures the brightness information.

5.2 What features and descriptors would be the best for this task? Are these scale invariant? Justify your answers.

In addition to color features, texture features can also be useful for fabric classification. We can use some simple analysis of texture like range (difference between maximum and minimum intensity values in a neighborhood) or variance (sum of the squares of the differences between the intensity of the central pixel and its neighbors). Although the statistical measures described so far are easy to calculate, they do not provide any information about the repeating nature of texture. In this way, a gray level co-occurrence matrix (*GLCM*) contains information about the positions of pixels having similar gray levels and is a commonly used method for texture analysis in image processing. As presented in the course, the *GLCM* can be used to extract various statistical features that describe the texture of an image, such as contrast, correlation, energy, homogeneity and entropy and these can be used as inputs to a machine learning model. This method is not scaled invariant meaning that is sensitive to changes in image scale: different GLCM matrices for different image scales.

Furthermore, we can also use Gabor filters to capture the local structure and orientation information of an image, which can be useful for distinguishing between different fabric categories. Also, these filters can be applied to the fabric image at different scales and the output response can then be used as a feature vector. Is important to mention that Gabor filters are not scale-invariant, which means that they may not be able to detect the same texture features at different scales

6. Exercise 6

6.1 A gray image with pixel grey-level values is represented on 8 bits [0..255]. All the pixel values belong in fact to the range [33..58]. What is the minimum number of bits that can be used for representing the values in this image without loss of information/details?

If the pixel values belong to the range [33..58] this means that we need to represent $58 - 33 + 1 = 26$ values. Therefore, the minimum number of bits that can be used for representing the values in this image without loss of information/details is 5 bits ($2^5 = 32$), which is the minimum number of bits to represent 26 possible values.

6.2 Given a budget of 3 bits devise a quantization procedure such that to minimize the quantization error. In what situation the uniform quantization is the best solution? Justify your answers.

Quantization error is the difference between the analog signal and the closest available digital value signal after the analog signal has been sampled and rounded to a discrete digital value. If we only have a budget of 3 bits, this means that we have 8 possible quantization levels, each represented by a 3-bit value. In order to minimize the quantization error we can divide the range of possible analog signals value into $2^3 = 8$ equally sized intervals. To assign a quantization level we can take the midpoint value of each interval. By choosing the interval boundaries such that the range of possible analog signal values is evenly divided among the intervals we try to minimize the quantization error, on average.

A uniform quantization is the best solution when we know more information about the signal that is being quantized, such as if the signal has a uniform distribution over its range. If this is the case, then the method described above should produce good results and will capture the signal's characteristics well, thus minimizing the error. If the signal has a non-uniform distribution, then a non-uniform quantization procedure may be introduced to minimize the quantization error because can allocate more quantization levels where more detail is required.

7. Exercise 7

7.1 Given a sequence 1, 3, 4, 7, 100, 3, 6, 2, 3 what are the results obtained by applying a mean filter, a median filter, a maximum filter, and a minimum filter, respectively. For each filter the support window is of length 3.

We could also use a padding of 0, but as it is not specified We will not pad the sequence. Results after applying each filter:

- mean filter: {2.66, 4.66, 37, 36.66, 36.33, 3.66, 3.66},
- median filter: {3, 4, 7, 7, 6, 3, 3 }
- maximum filter: {4, 7, 100, 100, 100, 6, 6 }
- minimum filter: {1, 3, 4, 3, 3, 2, 2}

7.2 If it is known that this sequence of values comes from a distribution with mean 3 and standard deviation 3, what is the "outlier" from the sequence?

If we know the mean and the standard deviation of the distribution then if a value is a certain number of standard deviations away from the mean, that data point is identified as an outlier. This specific number is called the threshold. We will set the threshold to value 3. In this case, our interval will be (-6,12) and if a value is outside this interval then is an outlier. From the initial sequence, we can see that the outlier is the value 100. The new sequence, without this value, is: $\{1, 3, 4, 7, 3, 6, 2, 3\}$. Furthermore, we can take advantage of some intuitions created in the previous response: we can observe that the median filter eliminates the big value, but the median filter is affected by it and we know that the median filter is effective at removing outliers. More than that, the max filter also could be useful for detecting peaks or abrupt edges as we see in the previous response that the value 100 is selected multiple times. Besides this, based on the exercise below we can also calculate a Z-score for each value and a probability to have a value bigger than 99 is very unlikely, so we can consider this value as an outlier.

7.3 If this is a normal distribution and we can sample an infinite amount of values, what is the probability to have values above 25? Justify your answers.

We can use the $Z - score$ to find the probability to have values above 25. We will note this probability as $P(Z > 25)$ and use the formula for $Z = \left(\frac{x-\mu}{\sigma}\right)$.

$$P(Z > 25) = P\left(\frac{Z-3}{3} > \frac{25-3}{3}\right) \tag{1}$$

$$= P(Z > 7.33) = 1 - \phi(7.33) = 1 - 0.999999999999947 \simeq 0 \tag{2}$$

We can see that the probability of having values above 25 is very close to 0, this means that it is extremely unlikely to observe values above 25 in this distribution.

8. Exercise 8

8.1 For what types of noise median filtering is clearly preferable over Gaussian filtering? Justify your answer.

In order to create some intuition (a more empirical approach) I made the implementations for the two filters applied to images with noise (type of noises described in the exercise). After this, the **MSE** results were calculated and the one with the lowest value is considered more efficient.

The *MSE* can be calculated by the formula:

$$MSE = \frac{1}{M \cdot N} \cdot \sum_{x=1}^{m} n \sum_{y=1}^{m} [g(x,y) - f(x,y)]^2$$

where $g(x,y)$ is the filtered image and $f(x,y)$ is the original image.

- **Uniform noise, i.e. distributed by uniform distribution**: *MSE* for Median filter is 62.5903 and for the Gaussian filter is 60.3314. Uniform distribution adds some noise to each pixel value, uniformly and independently and this means that we don't have many extreme values. In such cases, a Gaussian filter that takes into account the local spatial information of the image could obtain better results.

- **Gaussian noise, i.e noise distributed by independent normal distribution**: *MSE* for Median filter is 43.2221 and for Gaussian filter is 41.2553. Gaussian noise is drawn from a Gaussian (normal) distribution with a certain standard deviation and the Gaussian filter is better to reduce the noise because is drawn from the same distribution. In this way, the Gaussian filter is a good choice for images with Gaussian noise.

- **Exponential noise model**: *MSE* value of $118,672$ for Median noise and 147.428 for Gaussian noise. In this case, the median filter is better because exponential noise is impulsive and introduces high-intensity spikes or low-intensity dips in the image. Since the median filter is a non-linear filter, it can effectively remove impulsive noise without blurring the edges and details in the image.

- **Salt and pepper noise**: For the median filter, we have a value of *MSE* equal to 14.0184 while for the Gaussian filter, we have a value equal to 85.5401. This is because median filters are particularly effective at removing outliers (extreme values) because they take the median value of a series of numbers for which the extreme values usually appear at the ends or at the beginning. For this kind of noise, salt and pepper, outliers are often present due to a bit of error in transmission and the median filter is effective at removing this impulsive noise.

- **Rayleigh noise**: *MSE* value for median filter is 195.367 and for Gaussian filter is 197.122. Rayleigh noise is drawn from a Rayleigh distribution. In this case, the two filters obtain similar results, however median filter seems to be better. This is because the gaussian filter relies on the assumption that the noise in the image is normally distributed, which is not the case for Rayleigh. The Rayleigh noise can introduce outliers which are better removed by the median filter.