INTRODUCTION TO AUDIOVISUAL PROCESSING

Exercises v1.0 (Spring 2019-2020)

Signal Processing Group (Ferran Marqués, Francesc Rey)

This document includes a set of problems covering the INTRODUCTION TO AUDIOVISUAL PROCESSING course of UPC. It combines four kinds of exercises:

- **Preliminary exercises**, which are examples of the background knowledge that students are supposed to have from previous courses,
- **Basic exercises**, which are mainly to get used to the basic concepts of each topic and to get some practice on them,
- **Exam exercises**, which are based on those exercises that have been proposed in previous exams and further develop the basic concepts, and
- **Further research exercises**, which are advanced exercises proposing open questions to be studied as a research topic.

The goal of this document is to help students assimilate the course contents. It also represents what may be expected from them at the mid-term control and the final exam.

This document may include mistakes, typos, errors or inaccuracies.
Please report them, preferably by email, to ferran.marques@upc.edu or francesc.rey@upc.edu

Table of Contents:

Topic 1: Statistical signal modelling

Preliminary Exercises

1.1. Demonstrate that $\sigma_X^2 = var(X) = E\{[X - E\{X\}]^2\} = E\{X^2\} - E^2\{X\}$.

1.2. Given the mean and variance of a random variable X, compute the mean and variance of $Y = kX$.

1.3. Given the pdf of a random variable X ($f_X(x)$), compute the pdf of the sum of X and a constant value a: $Z = X + a$. Given the mean and variance of X, compute the mean and variance of Z.

1.4. Compute the mean and the variance of the sum of a set of independent random variables: $Z = \sum_{i=1}^{N} X_i$

1.5. Compute the mean and the variance of the weighted sum of a set of independent random variables: $Z = \sum_{i=1}^{N} w_i X_i$

1.6. Compute the mean and variance of the difference between two independent random variables: $Z = X - Y$

1.7. Given two random variables (X, Y), demonstrate that $c_{X,Y} = r_{X,Y} - m_X m_Y$

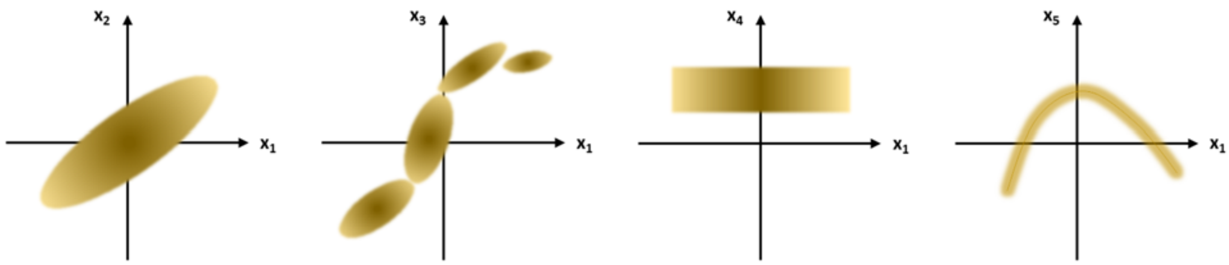
Basic Exercises

1.8. As first task at your new job in the multimedia company you are working for, you have been asked for reviewing an old piece of code that is a part of the company's basic coder. The soft is not correctly documented so it is complicated to understand the algorithm and the reasons for the various choices and steps in it. The first thing that surprises you is the fact that there are four different variables (x2, x3, x4, and x5) that are linearly predicted based on an initial one (x1).

You suspect that this may come from successive patches that have been included in the code, so you want to analyze if these predictions are sound.

AT THIS POINT, YOU MAY PROPOSE YOUR OWN ANALYSIS STRATEGY

To do so, you collect sets of samples of all these variables at a large number of time instants. The following plots show the joint probability density function (pdf) of each of these predicted variables (x_2, x_3, x_4, and x_5) and x_1.



Case 1 **Case 2** **Case 3** **Case 4**

a) **Discuss** in which of these four cases you would apply linear prediction to exploit the correlation between the variables. If there is any case on which you would not apply it, discuss which technique you would use.

b) In order to quantify the usefulness of linear prediction in these cases, as well as in future cases, you decide to implement a routine that computes a measure of the linear dependency between two variables. Discuss the main features of the covariance, the correlation and the correlation coefficient and propose one measure for that task.

c) Given the previous study, which are the following steps you will perform? Which will the content of your report to your team leader be?

1.9. The dynamic range of the difference between two images i1 and i2 [-1,1] is higher than the dynamic ranges of the original images [0,1].

a) Define by an equation the mapping $s = f(r)$ that allows us to visualize the difference image with the same dynamic range as the original images such that positive (negative) values are displayed as light (dark) gray.

b) Which difference, $s = f(i_2 - i_1)$ or $s = f(i_1 - i_2)$, is displayed as the third image of the following figure? (Justify your response)

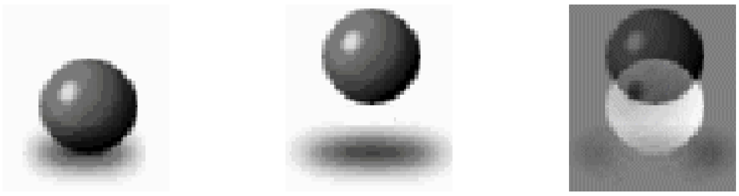


Image i_1 *Image i_2* *Difference*

1.10. Assuming that your input (r) and output (s) are discrete variables, discuss the implementation of a Range Transform using a Look-Up-Table. Write a pseudo-code of the operation.

1.11. Compute the Mean Square Error achieved when quantifying a variable X, that presents a uniform probability density function ($\mathrm{pdf}_X(x) = K$, $x \subset [0, L\Delta]$), using a uniform quantizer with L levels. Assume that the representative for each quantization interval is the central value of the interval.

1.12. Show that a signal that is quantized with a uniform, scalar quantizer increases its quality (SNR) in 6 dBs with every additional bit used in the quantizer. Assumptions:

- The signal $x[n]$ is always within the interval $[-A_x, A_x]$,
- A *mid-rise* uniform quantizer of B bits is used,
- The signal $x[n]$ is uniformly distributed within the quantization step D, and
- The signal power is approximated by $s_x^2 = kA_x^2$ where k is a constant value that depends on the kind of signal.

Hint: $\log_{10}[2] \approx 0{,}3$

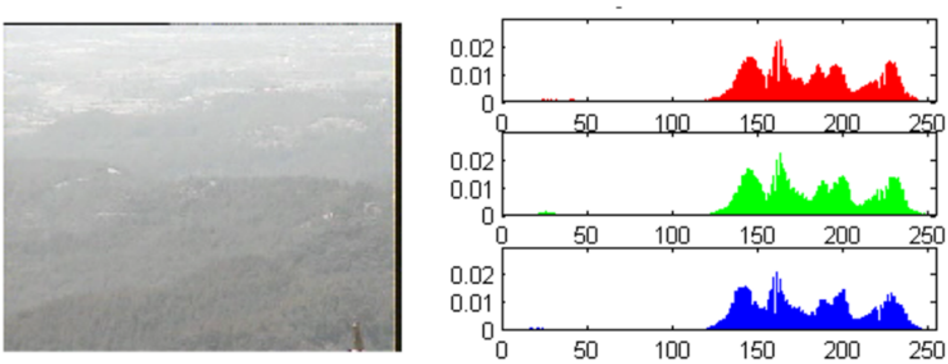
1.13. To transform the grey level values of an image whose range is [0, 255], the following range transform $s = r^2$ is proposed with normalized values ($0 \leq r \leq 1$). Therefore, the levels of the input image must be divided by 255 before applying the transform. After transforming, the image is de-normalized to have an output image with natural values (including zero) in the range [0, 255].

a) Which are the zones of grey level expansion and compression?
b) Can the transform be inverted?
c) If this transform is used as first step before applying a 4-level mid-rise uniform quantizer, which are the intervals of the input image values that will be quantized to this four output levels?
d) How could you build a non-uniform quantizer using this transform?

1.14. The following 4x4 image, initially quantized with 64 levels, is to be quantized with only 2 levels. Obtain the image after optimal quantization (with two levels) in the sense of minimum mean quadratic error. Which is the quadratic error achieved? Compare the result with the quadratic error that is obtained using a 2-level uniform quantization. Note that the resulting levels have to be integer values.

10	10	20	20
10	10	20	20
10	10	10	60
60	10	60	60

1.15. In forest fire monitoring, sometimes it is necessary to apply a range transform to palliate the weather conditions. In the following figure we present a typical (blurred) image and the normalized histogram of its RGB components.

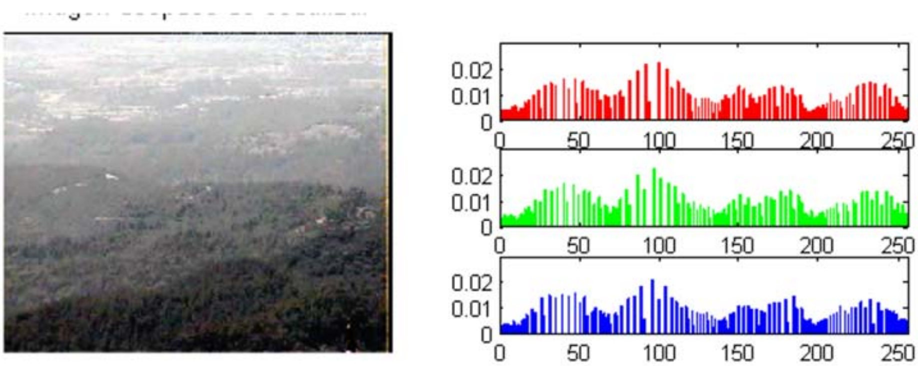


a) Propose a range transform (no equalization) to improve the visualization of the image.
b) Explain the necessary steps to equalize the image.

For the R component, the histogram has values equal to zero (0) for $r \leq 124$). In addition, the following 5 values of the normalized histogram are given:

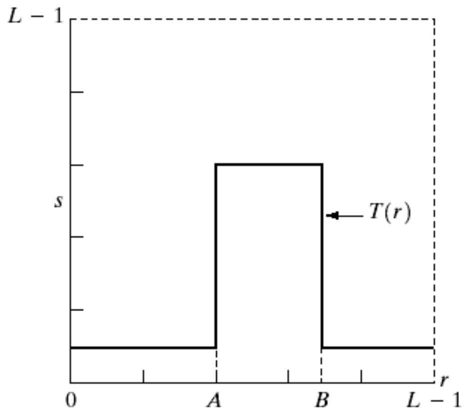
p(125)	p(126)	p(127)	p(128)	p(129)
0,001	**0,001**	**0,0015**	**0**	**0,002**

c) Comment to which levels these five input levels will be mapped after equalization
d) Discuss which transform (question (a) or (b)) has been applied to obtain the results presented in the following figure.

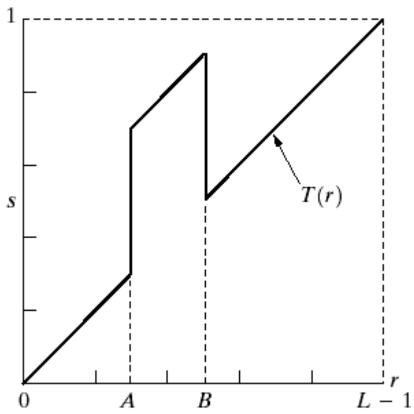


Exam Exercises

1.16. Gray level "slicing" is a special family of mapping transformations which consist in highlighting a specific range of gray levels in an image. Most are variations of the two basic approaches shown below:



Mapping #1

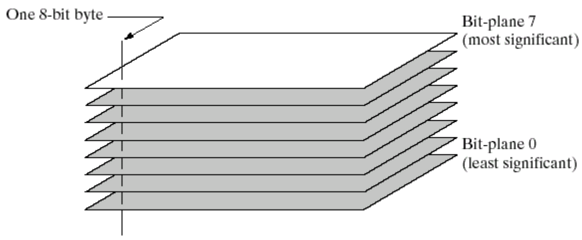


Mapping #2

Please answer the following questions as briefly as possible. Short (but complete) answers will be better marked:

1. What is the highlighted range of gray level values in Mapping #1 and Mapping #2?
2. Could you explain the basic difference between the images resulting from Mapping #1 and Mapping #2?
3. Could you think of a simple application (e.g. in X-ray or satellite imagery) where such mappings could be useful?

Instead of highlighting gray-level ranges, we might be interested in highlighting the contribution to the total appearance by specific bits. Suppose that each pixel in an image is represented by 8 bits. Imagine that the image is composed of eight 1-bit planes, ranging from bit-plane 0 for the least significant bit to bit-plane 7 for the most significant bit. In terms of 8-bit bytes, plane 0 contains all the lowest order bits in the bytes of the image pixels and plane 7 contains all the highest order bits, as shown in the figure below:



Separating a digital image into its bit-planes (bit-plane slicing) is useful for analyzing the relative importance played by each bit of the image.

4. What kind of image information is contained in the higher order bits (especially the top four)? What is left for the lower bit-planes?
5. The images below correspond to bit-planes 3 and 7 of the original image on the left. Could you tell which image corresponds to which bit-plane and why?

Original image

bit-plane X

bit-plane Y

6. Propose a set of gray-level slicing transformations $s = T_i(r)$, $0 \leq i < 8$ capable of producing all the individual bit-planes (i) of an 8-bit monochrome image.
7. If the four lower bit-planes are set to zero and the image is reconstructed back only using the information remaining in the four higher bit-planes, what will the reconstructed image look like? What is this operation equivalent to?

1.17. The pixel values r of a gray level discrete image $\boldsymbol{A}$ are in range $100 \leq r \leq 163$. The image histogram $h_A[r]$ is constant and can be expressed as:

$$h_A[r] = \begin{cases} 1 & 100 \leq r \leq 163 \\ 0 & other\ r \end{cases} \quad (1)$$

The gray level image $\boldsymbol{B}$ is the same size as image $\boldsymbol{A}$, and its histogram $h_B[r]$ has a unique non-zero value:

$$h_B[r] = \begin{cases} K & r = 64 \\ 0 & other\ r \end{cases} \quad (2)$$

a) What is the size NxN of the images, assuming they are square? What is the value of K?

b) Specify the histogram of the difference image $\boldsymbol{A-B}$, $h_{\boldsymbol{A-B}}[r]$.

c) Can the difference image be obtained through a mapping applied on $\boldsymbol{A}$? If yes, define precisely such mapping $s=f(r)$.

d) Define the representation levels of the 2-bit quantizer optimal, in Mean Squared Error (MSE), for image $\boldsymbol{A}$.

e) Justify that the following images $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$i_A[m,n] = 8m + n + 100 \quad \forall [m,n] \quad (3)$$

$$i_B[m,n] = 64 \quad \forall [m,n] \quad (4)$$

where $[m,n]$, $0 \leq m,n < N$, represent the image coordinates, have histograms corresponding to equations (1) and (2).

In the sequel, we will assume that the images $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined by equations (3) and (4).

g) A pseudo-color mapping is applied on image $\boldsymbol{B}$ to create image $\boldsymbol{C}$. Compute the pixel values of image $\boldsymbol{C}$ assuming that the mapping is defined as:

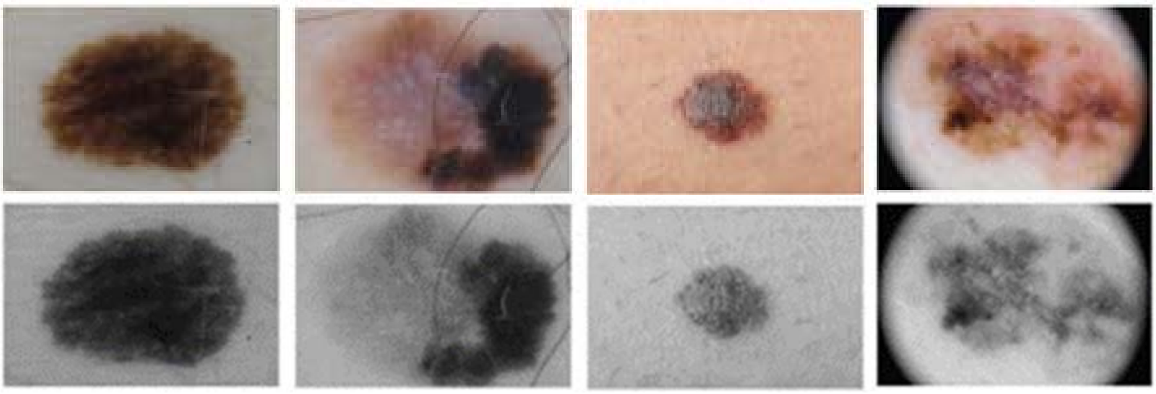
$$LUT[r] = \left(R[r], G[r], B[r]\right),\ with \begin{cases} R[r] = h_A[r] \\ G[r] = h_B[r] \\ B[r] = h_{A-B}[r] \end{cases} \quad for\ 0 \leq r \leq 255 \quad (5)$$

1.18. A la teva empresa d'anàlisi de dades avui tens una reunió amb una representació de la Unitat de Dermatologia d'un dels hospitals universitaris de la teva ciutat. Per part de l'hospital, venen a la reunió la metge Cap de la Unitat de Dermatologia i el responsable TIC de la unitat.

Aquesta unitat té un gran volum de dades clíniques dels milers de casos de melanoma (càncer de pell) que han tractat en els darrers anys, així com de la seva evolució. El que volen és poder crear una base de dades útil, aglutinant tota aquesta informació, que els permeti treballar amb aquestes dades i col·laborar amb la teva empresa per tal de fer models predictius de la gravetat i evolució de les lesions de pell que hagin de tractar en el futur.

Per tal de mostrar-vos el tipus de dades, us ensenyen amb una *tablet* la fitxa típica d'un malalt en la que es veu un gran volum de dades alfanumèriques i algunes imatges de les lesions, tant en color com en nivell de grisos. En aquesta aplicació observes que, al costat de les imatges, hi ha un botó que permet activar dues possibilitats: "Stretched" i "Equalized".

Li preguntes al responsable TIC sobre el significat d'aquest botó. Ell et comenta que no és expert en imatge, però que ha observat que moltes de les dermatoscòpies que prenen no utilitzen tot el marge dinàmic de la imatge. Així, amb aquest botó s'està fent una lleugera transformació a la imatge per a que el metge la pugui visualitzar millor: l'opció "Stretched" expandeix el marge dinàmic de manera lineal i la funció "Equalized" equalitza l'histograma. Et remarca que es tracta només d'una eina de visualització, i que les dades en memòria sempre es guarden amb els seus valors originals.



Exemples d'imatges en color i en nivells de grisos de la base de dades.

La Cap de la Unitat, veient el teu interès, et pregunta què vol dir exactament aquests canvis en les imatges.

a) Explica-li a la metge què pretenen aquestes dues transformacions i la diferència entre les dues.

El responsable TIC et diu que, de fet, no està 100% segur de la implementació que ha fet de l'opció "Equalized", ja que no acaba d'entendre alguns resultats que obté.

b) Escriu amb pseudo-codi la funció "Equalized" la qual, a partir d'una imatge en nivell de grisos, et retorna una altra imatge amb el seu histograma equalitzat.

El responsable TIC, després d'analitzar breument el teu pseudo-codi, diu que sí, que això és el que ell ha implementat. Aleshores et pregunta pels resultats que obté d'aquesta funció i que no acaba d'entendre. En concret et pregunta:

c) En imatges en nivell de grisos, i especialment quan la imatge té un fons negre (veure la quarta imatge en els exemples), el resultat de la funció "Equalized" no utilitza tampoc tot el marge dinàmic de la imatge de sortida. En concret ha vist que, sovint, els valors més propers a zero no són utilitzats. Raona si això és correcte o no. Proposa un exemple simple d'histograma que ajudi a il·lustrar el teu raonament.

d) En imatges en color, el resultat d'aplicar la funció "Equalized" a cadascuna de les components de la imatge per separat (R, G, B) i després fusionar aquesta informació en una sola imatge dóna resultats molt estranys i ha decidit no incloure aquesta possibilitat a l'aplicació. Raona per què succeeix això.

Finalment, la cap de la unitat et demana quina creus que ha de ser la manera de tractar les imatges abans de visualitzar-les per als metges: deixar-les amb el rang original, ampliar el rang mitjançant l'operador "Stretched" o ampliar-lo mitjançant l'operador "Equalized".

e) Raona-li a la metge quina opció creus que és la més adient per a la visualització per part dels metges. Centra el teu raonament en el cas d'imatges de nivells de grisos.

1.19. Se dispone de una imagen digital x[m,n] de tamaño 256x256 cuantificada con un número de bits suficiente como para considerar que el nivel de gris toma valores continuos menores o iguales que A. Con el objetivo de almacenar cada píxel con menos bits, se hace un "re-cuantificación" de la imagen mediante un cuantificador de 3 bits y margen dinámico [0,A].

Se considera en primer lugar un <u>cuantificador uniforme</u>.

a) Indica de forma gráfica los intervalos de cuantificación y los niveles de reconstrucción x_q a la salida del cuantificador, teniendo en cuenta que el margen dinámico del cuantificador va de [0,A].

b) Calcula el tamaño en bits de la imagen cuantificada.

c) Calcula la potencia del error de cuantificación en función de A. Explica las hipótesis utilizadas para hacer el cálculo.

A continuación se considera un cuantificador de Max Lloyd (cuantificador de mínimo error cuadrático medio) con el mismo margen dinámico [0,A]. Sus niveles de reconstrucción a la salida del cuantificador, para la imagen dada, son x_q={0.0954·A, 0.1458·A, 0.1949·A, 0.2415·A, 0.2770·A, 0.3090·A, 0.3509·A, 0.4025·A}. La potencia del error de cuantificación en este caso es $1.5A^2 \cdot 10^{-4}$

d) A la vista de los niveles de reconstrucción escogidos, ¿qué se puede deducir de la función de densidad de probabilidad de los valores de la imagen x[m,n]?

e) ¿Cuál será el tamaño en bits de la imagen cuantificada?

A continuación representaremos en una pantalla tanto la imagen original x[m,n], como las imágenes cuantificadas $x_{q,uniforme}$[m,n] y x_{q,Max_Lloyd}[m,n] (ver imágenes de la hoja adjunta). Para la representación se utilizan 256 niveles de gris repartidos uniformemente entre 0 y 255 (el nivel 255 representa el valor A). Para x_{q,Max_Lloyd}[m,n] el histograma normalizado, p[r], que mide el uso de estos 256 niveles es:

r	0 ... 23	**24**	...	**37**	...	**50**	...	**62**	...	**71**	...	**79**	...	**89**	...	**103**	... 255
$p[r]$	0	**0.155**	0	**0.087**	0	**0.147**	0	**0.153**	0	**0.142**	0	**0.146**	0	**0.085**	0	**0.085**	0

f) Calcula la transformación s=T(r) que permite ecualizar el histograma de x_{q,Max_Lloyd}[m,n]. Si has de aproximar el nuevo nivel por un número entero, redondea al entero más cercano.

g) Finalmente, indica de forma justificada a qué imagen x[m,n], $x_{q,uniforme}$[m,n], $x_{q,MaxLloyd}$[m,n] y T(x_{q,Max_Lloyd}[m,n]), corresponde cada una de las imágenes de la hoja adjunta. ¿A cuál de estas imágenes corresponden cada uno de los dos histogramas representados?

Imagen 1

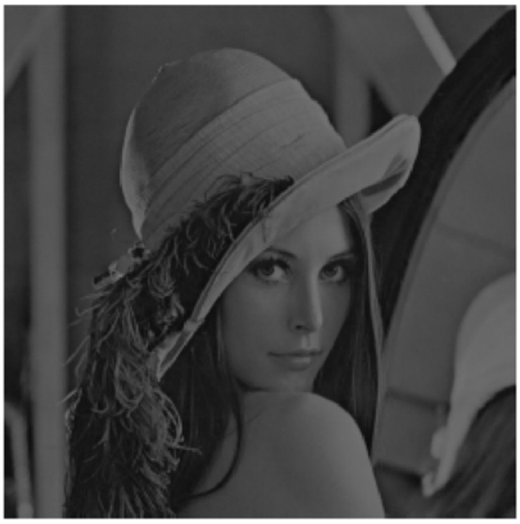
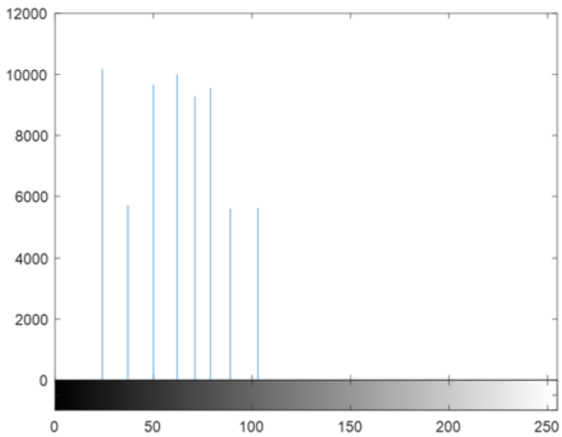
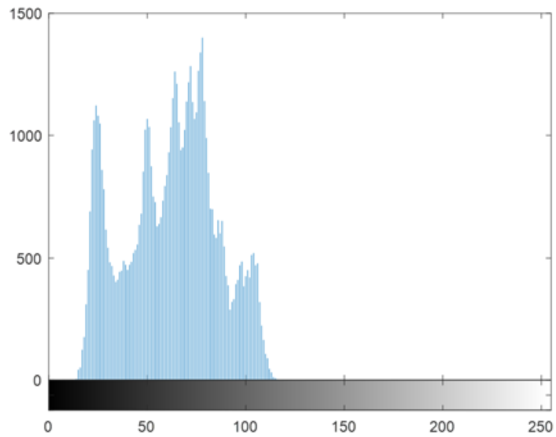


Imagen 2

<table>
<tr><td></td><td></td></tr>
<tr><td>Image 3</td><td>Imagen 4</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Histograma a)</td><td>Histograma b)</td></tr>
</table>