

Universidad Politécnica Salesiana



Visión por Computador

Cálculo de los Momentos de Zernike: Un tutorial introductorio

autor: vlarobbyk

Nivel: Experto / Académico

Enfoque: Fundamento Matemático (LaTeX) + Implementación (PyTorch)

En este tutorial se explica cómo calcular de forma base los *Momentos de Zernike*, considerando una explicación matemática de cada paso a partir de una matriz que representa una imagen real.

1. Introducción

Los Momentos de Zernike son descriptores de imagen ortogonales. Su propiedad más importante es que sus magnitudes son invariantes a la rotación, lo que los hace ideales para reconocer patrones sin importar su orientación. Para calcular el momento A_{nm} de orden n y repetición m para una imagen $I(x, y)$, proyectamos la imagen sobre las funciones base de Zernike $V_{nm}(\rho, \theta)$:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y I(x, y) \cdot [V_{nm}(\rho, \theta)]^*$$

Donde:

- ρ : Radio normalizado ($0 \leq \rho \leq 1$).
- θ : Ángulo en coordenadas polares.
- $R_{nm}(\rho)$: Polinomio Radial.

La función base $* V_{nm}^*$ se define como:

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) \cdot e^{jm\theta}$$

Para este ejemplo, trabajaremos con el caso donde $n = 2, m = 0$.

- Al ser $m = 0$, la parte imaginaria $e^{j0} = 1$, por lo que trabajaremos solo con números reales.
- El cálculo del momento se simplifica a:

$$A_{2,0} = \frac{3}{\pi} \sum_x \sum_y I(x, y) \cdot R_{2,0}(\rho)$$

- El polinomio radial se simplifica a: $R_{2,0}(\rho) = 2\rho^2 - 1$ por las siguientes razones:

La simplificación ocurre porque estamos en un caso especial donde la repetición es cero ($m = 0$). Si desglosamos la fórmula general parte por parte para ver por qué desaparecen los términos complejos, partiendo de que la fórmula general es:

$$A_{nm} = \underbrace{\frac{n+1}{\pi}}_{\text{Factor de Escala}} \sum \sum I(x, y) \cdot \underbrace{[V_{nm}(\rho, \theta)]^*}_{\text{Base Conjugada}}$$

El Factor de Escala ($\frac{n+1}{\pi}$)

Sustituimos el valor del orden $n = 2$:

$$\frac{2+1}{\pi} = \frac{3}{\pi}$$

(Este es el origen del 3).

La Base Conjugada ($[V_{nm}]^*$) La función base completa es:

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) \cdot e^{jm\theta}$$

El conjugado (marcado con el asterisco *) invierte el signo del exponente imaginario:

$$[V_{nm}(\rho, \theta)]^* = R_{nm}(\rho) \cdot e^{-jm\theta}$$

Ahora, sustituimos $m = 0$:

El término angular se convierte en: $e^{-j(0)\theta} = e^0$.

Cualquier número elevado a la 0 es 1. Por lo tanto, la parte angular compleja desaparece y queda solo el número real 1.

$$[V_{2,0}]^* = R_{2,0}(\rho) \cdot 1 = R_{2,0}(\rho)$$

Conclusión: al no haber rotación ($m = 0$), la fase angular no importa. El momento se vuelve una simple suma ponderada por el polinomio radial $R_{2,0}$.

El Problema

Calcularemos el Momento de Zernike $A_{2,0}$ para una matriz de imagen de 7x7 que representa una cruz.

Matriz de Entrada (I):

$$I = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Por favor, considerar lo siguiente:

- **x (Horizontal)**: Representa las Columnas. Se desplaza de izquierda a derecha.
- **y (Vertical)**: Representa las Filas. Se desplaza de arriba a abajo.

Normalización de Coordenadas:

Este es el paso crítico. Los polinomios de Zernike solo existen dentro de un círculo unitario ($\text{radio} \leq 1$). Sin embargo, nuestra imagen es una rejilla discreta de píxeles con índices enteros (filas 0 a 6, columnas 0 a 6).

Debemos mapear cada píxel (y, x) al espacio (ρ, θ) .

Paso A: Encontrar el Centro Geométrico

En una imagen de tamaño $H \times W$, el centro (c_y, c_x) se define como:

$$c_y = \frac{H - 1}{2}, \quad c_x = \frac{W - 1}{2}$$

Para nuestra imagen 7x7:

$$c_y = \frac{6}{2} = 3, \quad c_x = \frac{6}{2} = 3$$

El píxel central es (3, 3).

Paso B: Definir las Coordenadas Cartesianas Centradas

Restamos el centro a los índices originales para mover el origen (0,0) al centro de la imagen:

$$y' = y - c_y = y - 3$$

$$x' = x - c_x = x - 3$$

Paso C: Calcular el Factor de Escala (R_{max})

Para que **todos** los píxeles caigan dentro del círculo unitario ($\rho \leq 1$), el píxel más lejano del centro debe tener una distancia de 1 en el nuevo sistema. Los píxeles más lejanos son las esquinas (ej. índice 0,0).

1. Coordenadas centradas de la esquina (0,0):

$$y' = 0 - 3 = -3$$

$$x' = 0 - 3 = -3$$

2. Distancia Euclídea máxima:

$$R_{max} = \sqrt{(-3)^2 + (-3)^2} = \sqrt{9+9} = \sqrt{18} \approx 4.2426$$

Paso D: Fórmula Final del Radio Normalizado (ρ)

Para cualquier píxel, su radio normalizado es su distancia al centro dividida por la distancia máxima:

$$\rho = \frac{\sqrt{(x-3)^2 + (y-3)^2}}{\sqrt{18}}$$

Nota para el estudiante: Si al calcular ρ obtienes un valor mayor a 1, revisa tu R_{max} .

Cálculo Manual (Paso a Paso)

A continuación se presentan algunos ejemplos de cálculo de las coordenadas normalizadas para los siguientes puntos:

(a) $c_1(x = 2, y = 0)$

(b) $c_2(x = 4, y = 2)$

(c) $c_3(x = 0, y = 0)$

Para el primer caso $c_1(x = 2, y = 0)$, aplicamos la fórmula obtenida a partir del centro, es decir ρ y sustituimos los valores correspondientes:

$$\rho = \frac{\sqrt{(2-3)^2 + (0-3)^2}}{\sqrt{18}}$$

$$\rho = \frac{\sqrt{1+9}}{\sqrt{18}}$$

$$\rho = \frac{\sqrt{10}}{\sqrt{18}}$$

$$\rho = 0.7453$$

Para el segundo caso $c_2(x = 4, y = 2)$, nuevamente ρ y sustituimos los valores correspondientes:

$$\rho = \frac{\sqrt{(4-3)^2 + (2-3)^2}}{\sqrt{18}}$$

$$\rho = \frac{\sqrt{1+1}}{\sqrt{18}}$$

$$\rho = \frac{\sqrt{2}}{\sqrt{18}}$$

$$\rho = 0.333$$

$$\rho_{matriz} = \begin{bmatrix} 1.00 & 0.85 & 0.75 & 0.71 & 0.75 & 0.85 & 1.00 \\ 0.85 & 0.67 & 0.53 & 0.47 & 0.53 & 0.67 & 0.85 \\ 0.75 & 0.53 & 0.33 & 0.24 & 0.33 & 0.53 & 0.75 \\ 0.71 & 0.47 & 0.24 & 0.00 & 0.24 & 0.47 & 0.71 \\ 0.75 & 0.53 & 0.33 & 0.24 & 0.33 & 0.53 & 0.75 \\ 0.85 & 0.67 & 0.53 & 0.47 & 0.53 & 0.67 & 0.85 \\ 1.00 & 0.85 & 0.75 & 0.71 & 0.75 & 0.85 & 1.00 \end{bmatrix}$$

Cálculo del Polinomio $R_{2,0}$

Aplicamos la fórmula general de Zernike:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{(n-|m|)}{2}} (-1)^s \frac{(n-s)!}{s! (\frac{n+|m|}{2} - s)! (\frac{n-|m|}{2} - s)!} \rho^{n-2s}$$

Para $n = 2, m = 0$, la sumatoria va de $s = 0$ a 1 :

$s = 0$:

$$(-1)^0 \frac{2!}{0! \cdot 1! \cdot 1!} \rho^2 = 1 \cdot \frac{2}{1} \rho^2 = 2\rho^2$$

$s = 1$:

$$(-1)^1 \frac{1!}{1! \cdot 0! \cdot 0!} \rho^0 = -1 \cdot 1 \cdot 1 = -1$$

Resultado: $R_{2,0}(\rho) = 2\rho^2 - 1$.

Como se aprecia dejamos expresada la fórmula en términos del radio (ρ).

Cálculo Numérico de la matriz ($n = 2, m = 0$)

A continuación se presentan algunos ejemplos de cálculo del polinomio para los valores distintos de cero (cruz con valores 1):

(a) Coordenada original $c_1(x = 3, y = 0)$:

1. Ubicación del Píxel

Coordenada original $c_1(x = 3, y = 0)$ (No olvidar que el centro de la imagen es: $(c_x = 3, c_y = 3)$).

2. Coordenadas Normalizadas (x', y')

Restamos el centro para saber dónde está respecto al origen:

$$x' = 3 - 3 = 0$$

$$y' = 0 - 3 = -3$$

3. Cálculo del Radio Normalizado (ρ)

Distancia al cuadrado del píxel (d^2):

$$d^2 = (x')^2 + (y')^2 = 0^2 + (-3)^2 = 9$$

Radio Máximo al cuadrado (R_{max}^2): (Distancia a la esquina)

$$R_{max}^2 = 18$$

Radio Normalizado al cuadrado (ρ^2):

$$\rho^2 = \frac{d^2}{R_{max}^2} = \frac{9}{18} = \mathbf{0.5}$$

Es decir, este píxel está situado exactamente a la mitad de la "energía" radial máxima (en términos cuadráticos).

4. Evaluación en el Polinomio

Recordemos la fórmula del polinomio para $n = 2, m = 0$:

$$R_{2,0}(\rho) = 2\rho^2 - 1$$

Sustituimos el valor de $\rho^2 = 0.5$:

$$R_{2,0} = 2(0.5) - 1$$

$$R_{2,0} = 1 - 1$$

$$R_{2,0} = \mathbf{0}$$

Actividad: Calcular el valor de la matriz para los siguientes casos:

(b) $c_2(x = 4, y = 2)$

(c) $c_3(x = 0, y = 0)$

Siguiendo esta lógica, con estas fórmulas se puede calcular la matriz del polinomio. Por favor, tomar en cuenta que luego de calcular la matriz del polinomio, se obtiene la matriz del momento (multiplicando por los valores de intensidad originales de la imagen):

La Matriz del Polinomio ($R_{2,0}$)

Estos son los valores del "filtro" de Zernike evaluados en cada celda de 7x7.

$$R_{2,0} = \begin{bmatrix} 1.00 & 0.67 & 0.33 & \mathbf{0.00} & 0.33 & 0.67 & 1.00 \\ 0.67 & 0.33 & 0.00 & \mathbf{-0.56} & 0.00 & 0.33 & 0.67 \\ 0.33 & 0.00 & -0.56 & \mathbf{-0.89} & -0.56 & 0.00 & 0.33 \\ \mathbf{0.00} & \mathbf{-0.56} & \mathbf{-0.89} & \mathbf{-1.00} & \mathbf{-0.89} & \mathbf{-0.56} & \mathbf{0.00} \\ 0.33 & 0.00 & -0.56 & \mathbf{-0.89} & -0.56 & 0.00 & 0.33 \\ 0.67 & 0.33 & 0.00 & \mathbf{-0.56} & 0.00 & 0.33 & 0.67 \\ 1.00 & 0.67 & 0.33 & \mathbf{0.00} & 0.33 & 0.67 & 1.00 \end{bmatrix}$$

Observación Clave: Si se observar los bordes de la fila y columna central (los ceros en

negrita), se puede notar que el polinomio vale exactamente 0 en los extremos de la cruz. Esto significa que esos píxeles no aportan nada al momento $A_{2,0}$.

**La Matriz de Contribución Final* ($I \cdot R_{2,0}$)

Aquí multiplicamos la matriz de arriba por la imagen (que es 0 en las esquinas y 1 en la cruz). Los ceros de la imagen "apagan" los valores del polinomio en las esquinas.

$$Matriz_Final = \begin{bmatrix} 0 & 0 & 0 & \mathbf{0.00} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mathbf{0.56} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mathbf{0.89} & 0 & 0 & 0 \\ \mathbf{0.00} & -\mathbf{0.56} & -\mathbf{0.89} & -\mathbf{1.00} & -\mathbf{0.89} & -\mathbf{0.56} & \mathbf{0.00} \\ 0 & 0 & 0 & -\mathbf{0.89} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\mathbf{0.56} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0.00} & 0 & 0 & 0 \end{bmatrix}$$

Ubicación	Índices (y,x)	Centrados (y',x')	$p^2=18y'^2+x'^2$	Polinomio $2p^2-1$	Contribución (I·Poly)
Centro	(3, 3)	(0, 0)	0/18=0	2(0)-1=-1	-1
Brazo Vertical	(2, 3)	(-1, 0)	1/18	2(1/18)-1=-8/9	-0.889
	(4, 3)	(1, 0)	1/18	2(1/18)-1=-8/9	-0.889
	(1, 3)	(-2, 0)	4/18=2/9	2(2/9)-1=-5/9	-0.556
	(5, 3)	(2, 0)	4/18=2/9	2(2/9)-1=-5/9	-0.556
	(0, 3)	(-3, 0)	9/18=0.5	2(0.5)-1=0	0
	(6, 3)	(3, 0)	9/18=0.5	2(0.5)-1=0	0
Brazo Horiz.	(3, 2)	(0, -1)	1/18	2(1/18)-1=-8/9	-0.889
	(3, 4)	(0, 1)	1/18	2(1/18)-1=-8/9	-0.889
	(3, 1)	(0, -2)	4/18	2(2/9)-1=-5/9	-0.556
	(3, 5)	(0, 2)	4/18	2(2/9)-1=-5/9	-0.556
	(3, 0)	(0, -3)	9/18	0	0
	(3, 6)	(0, 3)	9/18	0	0
SUMA TOTAL					-6.7778

Cálculo Final del Momento

La fórmula incluye un factor de normalización constante:

$$A_{nm} = \frac{n+1}{\pi} \times \text{Suma}$$

$$A_{2,0} = \frac{3}{\pi} \times (-6.7778)$$

$$A_{2,0} \approx 0.9549 \times (-6.7778) = \mathbf{-6.472}$$

Ejemplo de cálculo en un caso con números imaginarios

Para este caso, vamos a suponer que tenemos la siguiente matriz que igualmente representa una imagen de 7 filas y 7 columnas:

Matriz de Entrada (I):

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

1. Preparación de Fórmulas ($n = 4, m = 2$)

A. El Polinomio Radial $R_{4,2}(\rho)$ Si calculamos el polinomio radial obtendremos este resultado:

$$R_{4,2}(\rho) = 4\rho^4 - 3\rho^2$$

Nota: Como ejercicio, realice el cálculo para este caso.

B. Parte Angular Compleja La fórmula de Zernike usa el conjugado de la función base:

$$\text{Base}^* = e^{-jm\theta} = \cos(m\theta) - j \cdot \sin(m\theta)$$

Para $m = 2$:

$$\text{Angular} = \cos(2\theta) - j \cdot \sin(2\theta)$$

2. Cálculo Paso a Paso: Píxel *coordenada(2, 2)*

Como podemos ver en la matriz, en la coordenada $c_1(2, 2)$ existe un 1.

Paso A: Coordenadas y Radio

- Centro: (3, 3) (se obtiene igual que en el cálculo anterior, ya que la matriz tiene el mismo tamaño).
- Posición: $y = 2, x = 2$.
- Coordenada Normalizada o Centrada:
 - $y' = 2 - 3 = -1$
 - $x' = 2 - 3 = -1$
- Radio Cuadrado (ρ^2):
 - $d^2 = (-1)^2 + (-1)^2 = 2$
 - $R_{max}^2 = 18$
 - $\rho^2 = 2/18 = 0.111$
- Radio a la Cuarta (ρ^4):
 - $(2/18)^2 = 0.0123$

Paso B: Evaluar el Polinomio $R_{4,2}$ Sustituimos ρ^2 y ρ^4 :

$$R_{4,2} = 4 \left(\frac{1}{81} \right) - 3 \left(\frac{1}{9} \right)$$

Para restar, trabajamos como denominador con 81:

$$R_{4,2} = \frac{4}{81} - \frac{27}{81} = -\frac{23}{81} \approx -0.2839$$

(Es un número Real negativo).

Paso C: Evaluar la Parte Angular

En esta parte del cálculo veremos cómo aparece la parte imaginaria.

Necesitamos el ángulo θ del punto $(-1, -1)$ (coordenada normalizada).

En el círculo trigonométrico, $(-1, -1)$ está a 225° (o -135° o $\frac{-3\pi}{4}$ radianes).

La fórmula requiere el doble del ángulo (2θ) porque $m = 2$:

$$2\theta = 2 \times 225^\circ = 450^\circ$$

$$450^\circ - 360^\circ = 90^\circ$$

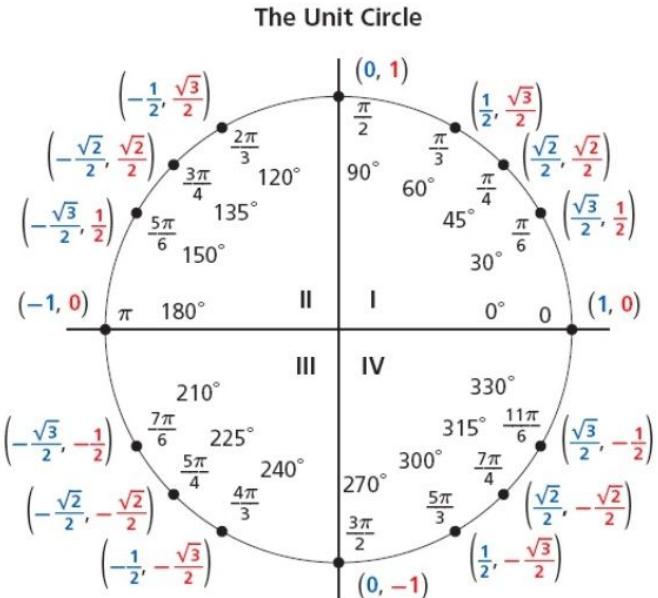
Para contar con la referencia de cómo se obtiene el ángulo a partir de la Coordenada, a continuación se presenta la imagen del círculo trigonométrico:

El Círculo Unitario

- Un **círculo unitario** es un círculo con un radio de 1 unidad.

$$\begin{aligned}\sin \theta &= \frac{y}{r} = \frac{y}{1} = y \\ \cos \theta &= \frac{x}{r} = \frac{x}{1} = x \\ \tan \theta &= \frac{y}{x}\end{aligned}$$

Así que las coordenadas de cualquier punto en el círculo puede ser escrita de la forma $(\cos \theta, \sin \theta)$



Fuente: montenegro735, FUNCIONES TRIGONOMÉTRICAS: MÉTODO DEL CÍRCULO UNITARIO.

URL: [Genially](#)

Ahora aplicamos Euler ($\cos -j \sin$):

- $\cos(90^\circ) = 0$
- $\sin(90^\circ) = 1$

$$\text{Angular} = 0 - j(1) = -j$$

Paso D: Fusión Final

Multiplicamos: (Pixel de la Imagen $c_1(-1, -1)$) \times (Polinomio) \times (Angular).

$$\text{Contribución} = 1 \times (-0.2839) \times (-j)$$

$$\text{Contribución} = +0.2839j$$

Resultado: Como se observa, la contribución de este píxel es puramente imaginaria positiva. No tiene parte real.

Reflexión final:

Como se ha podido apreciar, cuando $m \neq 0$ (como en el caso $n = 4, m = 2$), el Momento de Zernike resultante es un número complejo.

1. ¿Por qué es un número complejo?

Matemáticamente, la razón de ello, es el término angular de la fórmula:

$$e^{-jm\theta} = \cos(m\theta) - j \sin(m\theta)$$

- Cada píxel aporta una "flecha" en el plano complejo.
- Al hacer la suma total, se están sumando muchas flechas.
- A menos que la imagen sea perfectamente simétrica (donde las flechas hacia arriba cancelan a las de abajo), la suma final tendrá una componente Real y una componente Imaginaria.

Por ejemplo, si tenemos partes complejas en la **Matriz_Final**, el momento de Zernike tendría la siguiente forma:

$$A_{4,2} = \text{Parte Real} + j(\text{Parte Imaginaria})$$

2. ¿Para qué sirve que sea complejo? (El secreto de Zernike)

Este es el concepto más importante. El hecho de que sea un número complejo nos da dos datos vitales:

A. La Magnitud (Módulo) $|A_{nm}|$

Nos dice "**cuánto**" se parece la imagen a ese polinomio
sin importar si la imagen está girada .

- **Propiedad de Invarianza:** Si tomamos una imagen y la rotamos 90 grados, el número complejo cambia, pero su **magnitud** sigue siendo idéntica.
- Esta es la característica que usamos para **reconocimiento de patrones**.

B. La Fase (Ángulo) ϕ

Nos dice la **orientación** de la imagen.

- Si calculamos $A_{4,2}$ y obtenemos un ángulo de 45°, y luego rotamos la imagen, el ángulo del resultado cambiará acorde a esa rotación.

Matriz Comparativa: Tipos de Momentos de Zernike

Característica	Caso Radial / Simétrico	Caso Angular / Rotacional
Condición	$m = 0$	$m \neq 0$
Ejemplo	$A_{2,0}, A_{4,0}$	$A_{1,1}, A_{2,2}, A_{4,2}$
Tipo de Resultado	Número Real Puro	Número Complejo ($a + bj$)

Característica	Caso Radial / Simétrico	Caso Angular / Rotacional
Término Angular	$e^{j0} = 1$ (Se anula la parte imaginaria)	$e^{jm\theta} = \cos(m\theta) - j \sin(m\theta)$
Interpretación Física	Mide la distribución radial (anillos) de la intensidad desde el centro.	Mide la simetría rotacional y la orientación de la forma.

```
In [ ]: import numpy as np
import requests
import cv2
import torch

from torch.nn import functional as F
from io import BytesIO
from PIL import Image
from matplotlib import pyplot as pp
import matplotlib.patheffects as pe
from IPython.display import display, Math, HTML

image = torch.tensor()

# La siguiente función permite convertir un tensor de Pytorch o un arreglo en una m
def tensor_to_latex_manual(tensor, decimals=1):
    """
    Convierte un tensor/array a LaTeX
    evitando saltos de línea automáticos.
    """

    # 1. Convertir a Numpy asegurando que esté en CPU y sin gradientes
    if isinstance(tensor, torch.Tensor):
        tensor = tensor.detach().cpu().numpy()

    tensor = np.array(tensor)

    # 2. Validar dimensiones (si es vector 1D, convertir a 2D)
    if tensor.ndim == 1:
        tensor = tensor[None, :]

    rows, cols = tensor.shape
    latex_lines = []

    # 3. Construcción manual fila por fila
    for i in range(rows):
        # Formateamos cada número individualmente.
        # Esto garantiza que NO haya saltos de línea ocultos.
        row_values = [f"{x:.{decimals}f}" for x in tensor[i]]

        # Unimos con '&' y cerramos la fila con '\\'
        line = " & ".join(row_values) + r" \\"
        latex_lines.append(line)

    return r"\begin{bmatrix}" + "\n" + "\n".join(latex_lines) + "\n" + r"\end{bmatrix}
```

A continuación se visualiza la ROI que se usará para realizar el cálculo del HOG:

```
In [ ]: #latex_code = tensor_to_latex_manual(roi_ref)
#display(HTML('<h2>ROI de la Imagen para el cálculo (S):</h2> <br />'))
#display(Math(latex_code))
```