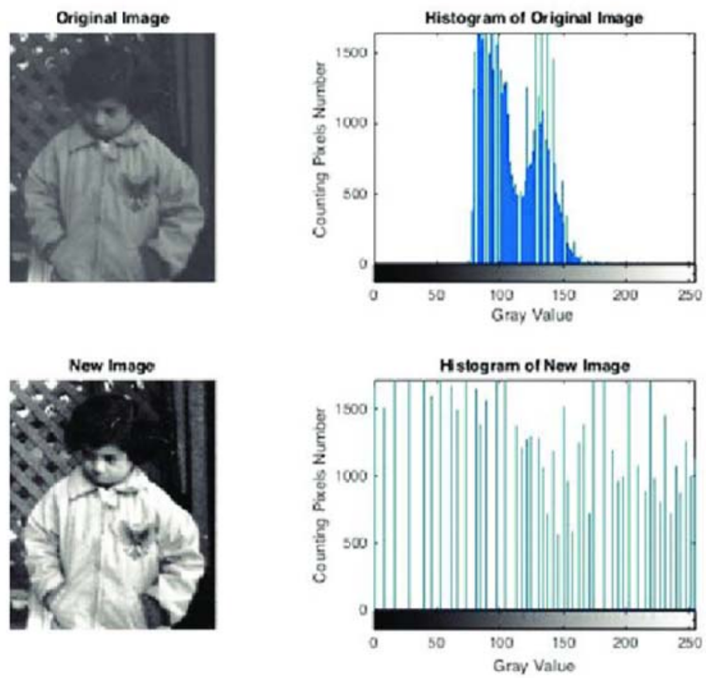
Feature Extraction

Features/Descriptors

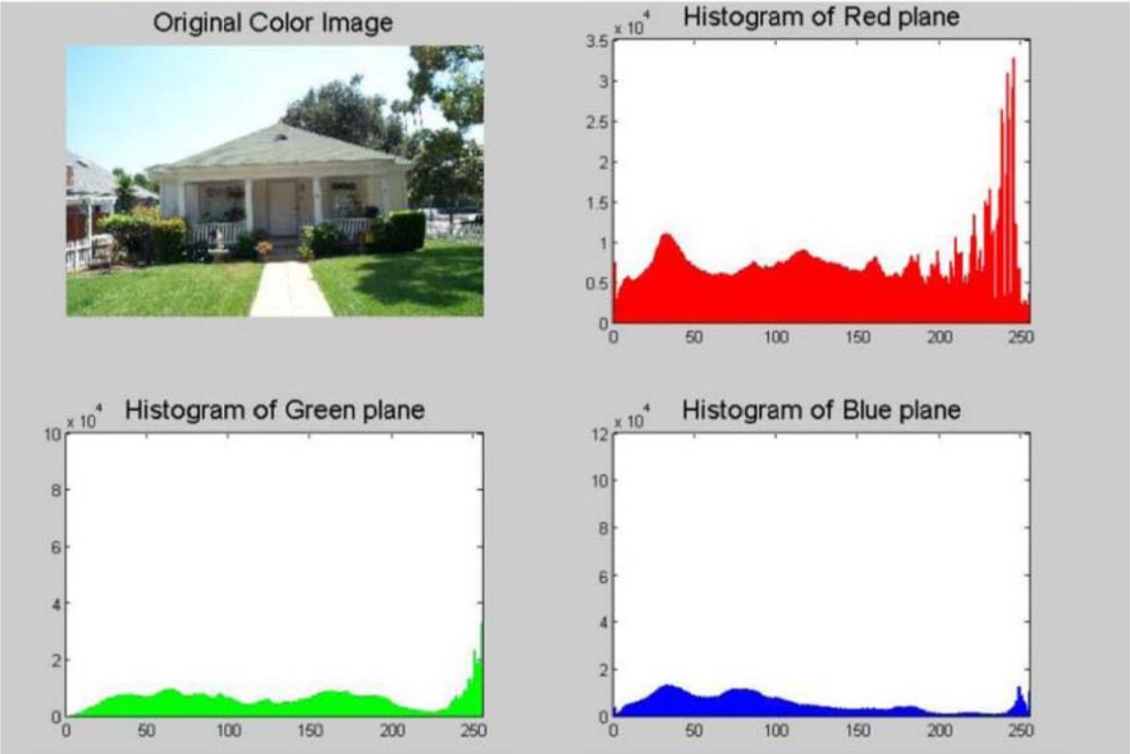
Features

- Shape features
- Region Features (Intensity, colour, texture)

Intensity Features

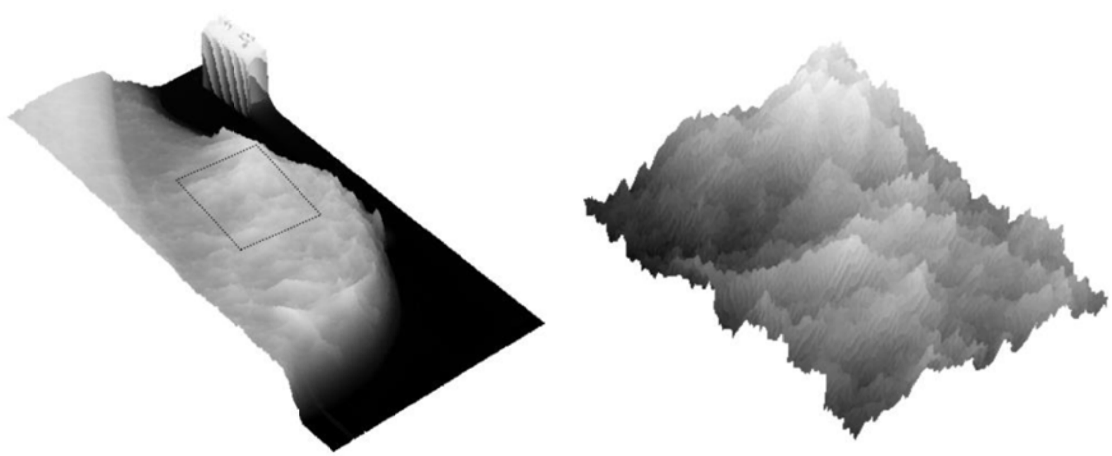


Colour Features



Texture Features

- **Texture** is a repeating pattern of local variations in image intensity:
  - Texture cannot be defined for a point.



Texture Features

For example, an image has a 50% black and 50% white distribution of pixels.

Three different images with the same intensity distribution, but with different textures.

Grey Level Co-occurrence

- One of the ways to extract texture information
- A **gray level co-occurrence matrix (GLCM)** contains information about the positions of pixels having similar gray level values.
- A co-occurrence matrix is a two-dimensional array, P, in which both the rows and the columns represent a set of possible image values.
- GLCM calculates how often a pixel with gray level (grayscale intensity or Tone) value ***i*** occurs either horizontally, vertically, or diagonally to adjacent pixels with the value ***j***

Grey Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix (GLCM):

- The GLCM gives a joint distribution, $p(i,j)$ of gray level pairs having intensities (i, j) within an image.
- Let $q(i,j)$ is the element of $GLCM$ of a given image f of size $M \times N$ containing the number of gray levels G ranging from 0 to $G-1$. Then q can be defined as the matrix element and given by,

$$q(i,j) = \sum_{x=1}^{M} \sum_{y=1}^{N} \begin{cases} 1, & \text{if } f(x,y) = i \text{ and } f(x+\Delta x, y+\Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

- Each entry of the GLCM is dependent on two parameters D (distance of separation between two neighboring resolution cells) and θ (direction of neighboring pixel w.r.t reference pixel).
- The value of D is set according the number of resolution level of 2D-DWT and θ can be $0^\circ, 45^\circ, 90^\circ$ and 135°.

Grey Level Co-occurrence Matrix: Directionality

- Each element of the normalized gray-level co-occurrence matrix ($NGLCM$) is defined as,

$$p(i,j) = q(i,j) / \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} q(i,j) \tag{2}$$

where n represents the size of $NGLCM$.

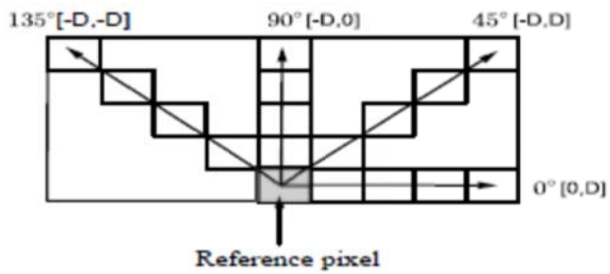


Figure 8: Directionality used in the gray-level co-occurrence matrix.

GLCM Computation (D=1, θ=0º)

I

1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

GLCM

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

GLCM Computation

Process of computation of GLCM and NGLCM:

	0	1	2	3
0	0	2	1	2
1	2	3	2	0
2	0	1	3	2
3	2	0	2	1

(a)

	0	1	2	3
0	0	1	4	0
1	1	0	3	1
2	4	3	0	3
3	0	1	3	0

(b)

	0	1	2	3
0	0	1	5	0
1	1	0	2	1
2	5	2	0	3
3	0	1	3	0

(c)

	0	1	2	3
0	0	0	0	3
1	0	0	2	1
2	0	2	6	0
3	3	1	0	0

(d)

	0	1	2	3
0	2	1	0	1
1	1	0	2	1
2	0	2	4	0
3	1	1	0	2

(e)

Figure 9: Computation of co-occurrence matrices. (a) Intensity values of input image with 4 gray levels. Different co-occurrence matrices (*GLCM*) for set distance $D = 1$ at four different directions such as (b) horizontal ($\theta = 0^\circ$), (c) vertical ($\theta = 90^\circ$), (d) right diagonal ($\theta = 45^\circ$), (e) left diagonal ($\theta = 135^\circ$).

Normalized GLCM

	0	1	2	3
0	0	0.0417	0.1667	0
1	0.0417	0	0.1250	0.0417
2	0.1667	0.1250	0	0.1250
3	0	0.0417	0.1250	0

(a)

	0	1	2	3
0	0	0.0417	0.2083	0
1	0.0417	0	0.0833	0.0417
2	0.2083	0.0833	0	0.1250
3	0	0.0417	0.1250	0

(b)

	0	1	2	3
0	0	0	0	0.1667
1	0	0	0.1111	0.0556
2	0	0.1111	0.3333	0
3	0.1667	0.0556	0	0

(c)

	0	1	2	3
0	0.1111	0.0556	0	0.0556
1	0.0556	0	0.1111	0.0556
2	0	0.1111	0.2222	0
3	0.0556	0.0556	0	0.1111

(d)

Figure 10: Normalized co-occurrence matrices ($NGLCM$) of corresponding co-occurrence matrices ($GLCM$) in Figure 9 at directions (a) $\theta = 0°$, (b) $\theta = 90°$, (c) $\theta = 45°$ and (d) $\theta = 135°$.

Numeric Features of GLCM

- Gray level co-occurrence matrices capture properties of a texture but they are not directly useful for further analysis, such as the comparison of two textures
- Numeric features are computed from the co-occurrence matrix that can be used to represent the texture more compactly.

Numeric Features of GLCM

Label	Feature	Computation
f1	Energy	$\sum_{i=1}^{G}\sum_{j=1}^{G}\{p(i,j)\}^2$
f2	Correlation	$\frac{\sum_{i=1}^{G}\sum_{j=1}^{G}(i,j)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
f3	Entropy	$-\sum_{i=1}^{G}\sum_{j=1}^{G}p(i,j)\log(p(i,j))$
f4	Sum variance	$\sum_{i=2}^{2G}(i-sum\ entropy)^2 p_{x+y}(i)$
f5	Sum average	$\sum_{i=2}^{2G} i p_{x+y}(i)$

Numeric Features of GLCM: Notations

$p(i,j)$	$\frac{x(i,j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} x(i,j)}$
$p_x(i)$	$\sum_{j=1}^{N} p(i,j)$
$p_y(j)$	$\sum_{i=1}^{N} p(i,j)$
μ_x	$\sum_{i=1}^{N} i \cdot p_x(i)$
μ_y	$\sum_{j=1}^{N} j \cdot p_y(j)$
σ_x^2	$\sum_{i=1}^{N} (i-\mu_x)^2 \cdot p_x(i)$
σ_y^2	$\sum_{j=1}^{N} (j-\mu_y)^2 \cdot p_y(j)$
$p_{x+y}(k)$	$\sum_{i=1}^{N}\sum_{j=1}^{N}{}_{i+j=k}\; p(i,j)$
$p_{x-y}(k)$	$\sum_{i=1}^{N}\sum_{j=1}^{N}{}_{\lvert i-j\rvert=k}\; p(i,j)$

Numeric Features of GLCM: Notations

μ_{x+y}	$\sum_{k=2}^{2N} k \cdot p_{x+y}(k)$
μ_{x-y}	$\sum_{k=0}^{N-1} k \cdot p_{x-y}(k)$
HX	$-\sum_{i=1}^{N} p_x(i) \cdot \log p_x(i)$
HY	$-\sum_{i=1}^{N} p_y(i) \cdot \log p_y(i)$
HXY	$-\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) \cdot \log p(i,j)$
$HXY\,1$	$-\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j) \cdot \log\left[p_x(i) \cdot p_y(j)\right]$
$HXY\,2$	$-\sum_{i=1}^{N}\sum_{j=1}^{N} p_x(i) \cdot p_y(j) \cdot \log\left[p_x(i) \cdot p_y(j)\right]$
$Q(i,j)$	$\sum_{k=1}^{N} \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$

Texture Features computed from GLCM

Feature	Original expression
Autocorrelation [31]	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i \cdot j)p(i,j)$
Cluster prominence [4]	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i+j-2\mu)^3 p(i,j)$
Cluster shade [4]	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i+j-2\mu)^4 p(i,j)$
Contrast [4]	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 p(i,j)$
Correlation [4]	$\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{i-\mu_x}{\sigma_x}\right)\left(\frac{j-\mu_y}{\sigma_y}\right)p(i,j)$
Difference entropy [4]	$-\sum_{k=0}^{N-1}p_{x-y}(k)\log p_{x-y}(k)$
Difference variance [4]	$\sum_{k=0}^{N-1}(k-\mu_{x-y})^2 p_{x-y}(k)$
Dissimilarity [31]	$\sum_{i=1}^{N}\sum_{j=1}^{N}\lvert i-j\rvert \cdot p(i,j)$
Energy [4]	$\sum_{i=1}^{N}\sum_{j=1}^{N}p(i,j)^2$
Entropy [4]	$-\sum_{i=1}^{N}\sum_{j=1}^{N}p(i,j)\log p(i,j)$

Texture Features computed from GLCM

Homogeneity [31]	$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{p(i,j)}{1+(i-j)^2}$
Information measure of correlation 1 [4]	$\frac{HXY-HXY1}{\max(HX,HY)}$
Information measure of correlation 2 [4]	$\sqrt{1-\exp[-2(HXY2-HXY)]}$
Inverse difference [38]	$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{p(i,j)}{1+\lvert i-j\rvert}$
Maximum probability [31]	$\max_{i,j} p(i,j)$
Sum average [4], μ_{x+y}	$\sum_{k=2}^{2N} kp_{x+y}(k)$
Sum entropy [4]	$-\sum_{k=2}^{2N} p_{x+y}(k)\log p_{x+y}(k)$
Sum of squares [4]	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu)^2 p(i,j)$
Sum variance [4]	$\sum_{k=2}^{2N}(k-\mu_{x+y})^2 p_{x+y}(k)$
Maximal Correlation Coefficient [4]	$\sqrt{\lambda_2(Q(i,j))}$

Haralick Texture Operator

- Haralick et al. suggested a set of 14 textural features which can be extracted from the co-occurrence matrix, and which contain information about image textural characteristics such as homogeneity, linearity, and contrast.

Haralick, R.M., K. Shanmugam, and I. Dinstein, "Textural features for image classification". IEEE Transactions on Systems, Man and Cybernetics: pp. 610-621. 1973. (26264 citations)

References

- Carlson, G.E. and W.J. Ebel. "Co-occurrence matrix modification for small region texture measurement and comparison". in IGARSS'88- Remote Sensing: Moving Towards the 21st Century, pp.519-520, IEEE, Edinburgh, Scotland. 1988.
- Argenti, F., L. Alparone, and G. Benelli, "Fast algorithms for texture analysis using co-occurrence matrices". IEE Proceedings, Part F: Radar and SIgnal Processing, 137(6): pp. 443-448. 1990.

Popular Feature Extraction Techniques

Image Transformation Based Feature Extraction Techniques

- Wavelet Transform
- Gabor transform

Other popular FE Techniques

LBP

SIFT

SURF

HOG