Statistics for Data Science

Tutorial 2

Abijith J. Kamath

December 27, 2023

Digital Images

- Image are 2D arrays of dimensions $M \times N$
- Each *pixel* (i, j) takes values in $\{0, 1, \dots, 255\}$
- Image histogram gives the number of pixels that take a particular intensity
  $h[k]$ is the count of the pixels that take the intensity level k
- Normalised histogram $p[k] = \frac{1}{MN} h[k]$

Which has 'Better' Contrast?

(a) Image 1

(b) Image 2

Histogram Equalisation

Lemma

Let X be a continuous random variable with an invertible cumulative distribution function F_X, and let $Y = F_X(X)$. Then, $Y = \mathcal{U}[0,1]$.

Proof.

Since $Y = F_X(X)$, $0 \leq Y \leq 1$.

$$\begin{aligned} \forall y \in [0,1],\ F_Y(y) &= P[Y \leq y], \\ &= P[F_X(X) \leq y], \\ &= P[X \leq F_X^{-1}(y)], \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

Therefore, $Y = \mathcal{U}[0,1]$. $\square$

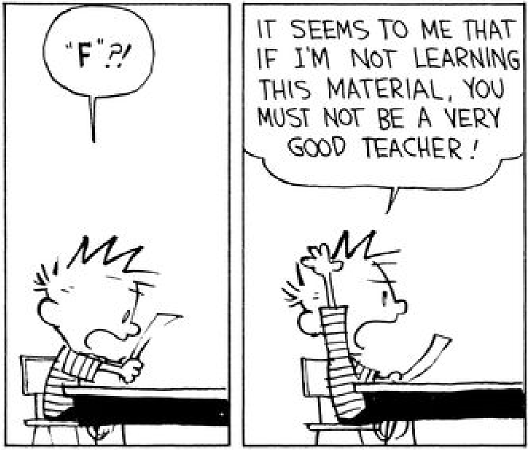
Discrete Implementation of Histogram Equalisation

- Given image $I[i,j]$
- Compute the normalized histogram $p[k]$
  Note: $p[k]$ is the relative number of pixels with intensity k
- Histogram equalisation

$$\boxed{J[i,j] = \sum_{k=0}^{I[i,j]} p[k]}$$

- Warning: Not all distribution functions are invertible; discrete implementations introduce errors; output histogram may not be the uniform distribution.

Fin.


"F"?!
IT SEEMS TO ME THAT IF I'M NOT LEARNING THIS MATERIAL, YOU MUST NOT BE A VERY GOOD TEACHER!