Mutual Information

Author: Willams. M. Wells III

Yen-Yun Yu

2011, Mar. 29

Outline

1. Image Registration
2. Mutual Information
    - Introduction
    - Mathematic Background
3. Result

Image Registration (1/3)

To register two images means to align them so that features overlap.

1. Rigid Body Registration: Translation and Rotation.
2. Affine Registration: Translation, Scale, Rotation and Shear.
3. Non-Rigid Body Registration: Imaging Warping and deformation

- Comparison of one subject to another
- Monitoring of changes in an individual

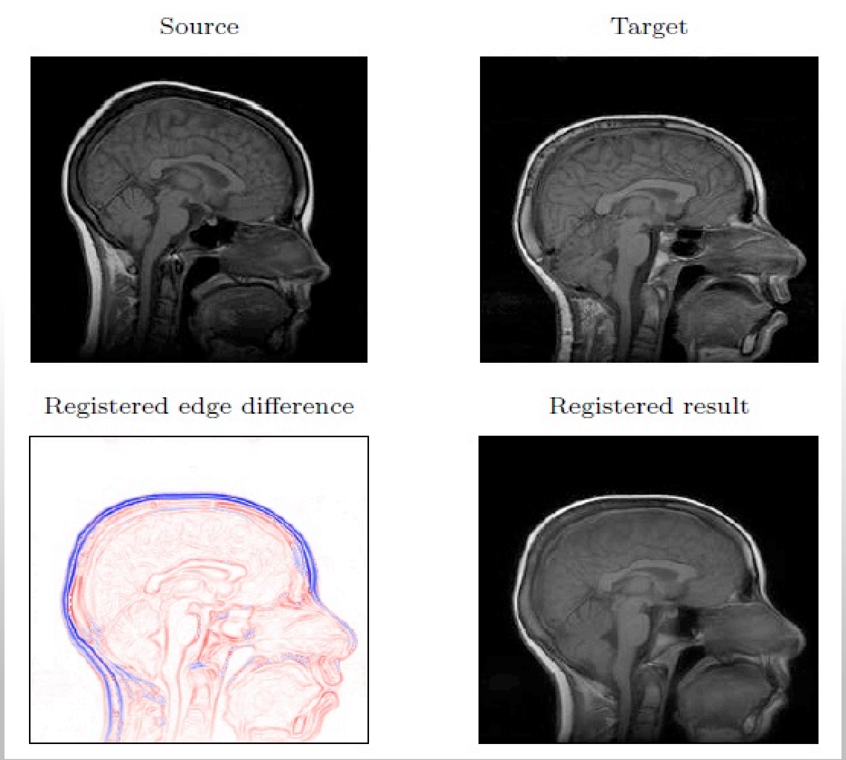


Image Registration (2/3)

- Sum of Squared Intensity Difference
- Cross-correlation
- Mutual Information

- Intensity Base: Their basic principle is to search the one that maximizes a criterion measuring the the intensity similarity of corresponding voxels.

Image Registration(3/3)

- Sum of Squared Intensity Difference
- Cross-correlation
- Mutual Information
  - Histogram based
  - Robust against the image degradation
  - No segmentation required

Outline

1. Image Registration
2. Mutual Information
    - Introduction
    - Mathematic Background
3. Result

Entropy and Mutual Information(1/4)

Entropy and Joint Entropy for a random variable x

$$h(x) = -\int p(x) \ln p(x) dx$$

$$h(x, y) = -\int\int p(x, y) \ln p(x, y)\, dx dy$$

Mutual Information

$$I(x, y) = h(x) + h(y) - h(x, y)$$

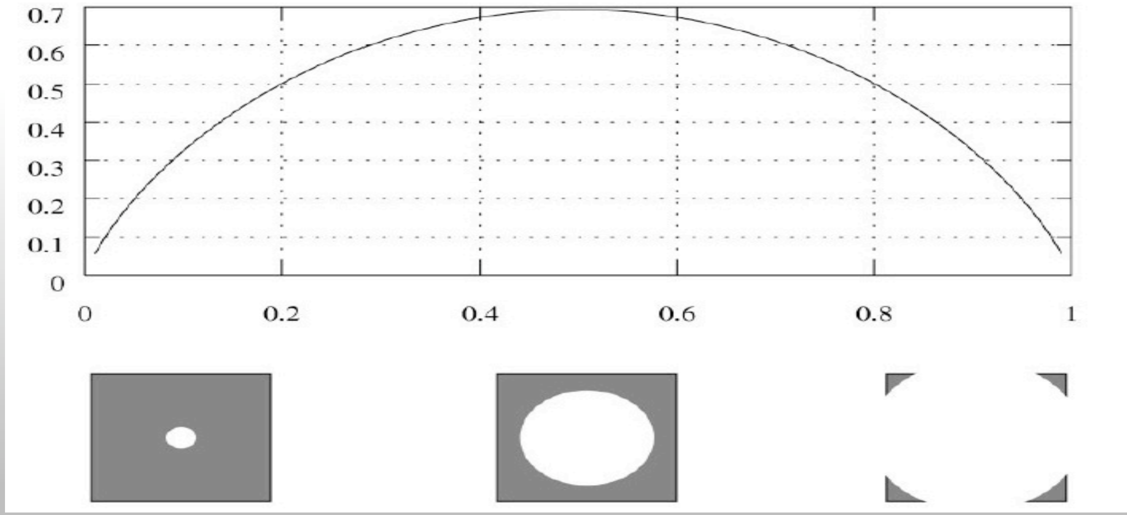
Advantage in using mutual information over joint entropy is it includes the individual input’s entropy

Binary Entropy Example(2/4)

The concept of uncertainty can provide a useful basis for information theory. Some value are rare and some are not. Base on information theory, we can form the observed probability distribution for predicting what value a pixel/voxel has. If all probability are equal, the prediction is hard; If only one probability is higher than the others, the prediction is easy.

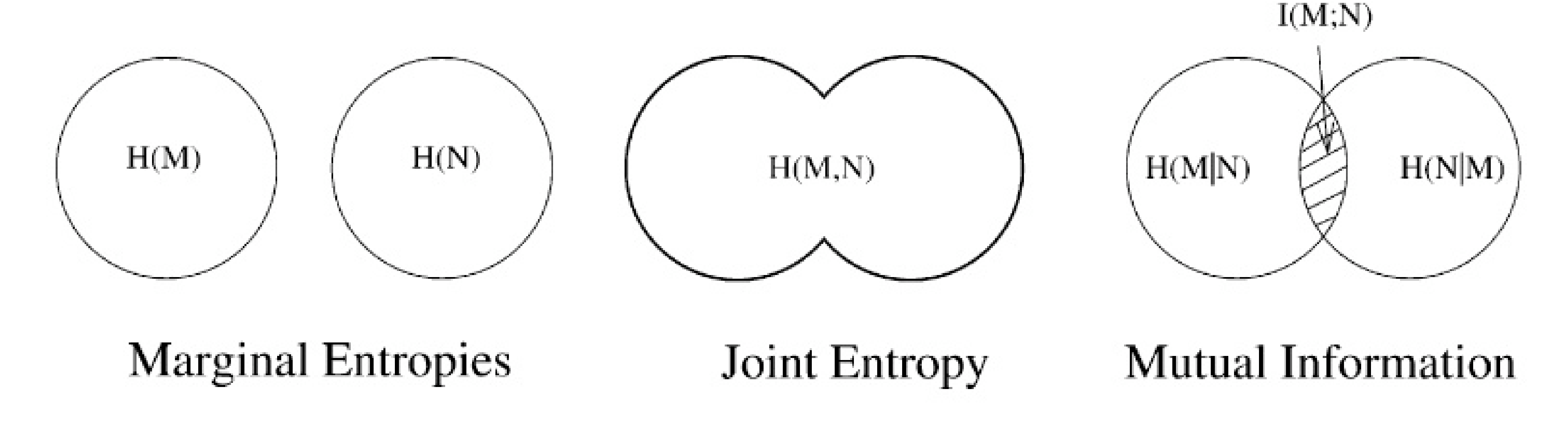
$$
\begin{aligned}
P_X(0) &= p \\
P_X(1) &= 1 - p \\
H(X) &= -(p \ \log_2(p) + (1-p) \ \log_2(1-p)) = h(p)
\end{aligned}
$$

The binary entropy is 0 for p = 0 and p = 1 and has a maximum for p=0.5.



Mutual Information(3/4)

1. By minimizing the joint entropy of two images, we can get the transformation matrix and which part should be overlapped.
2. We can maximize the I(M,N) for overlap invariance.



Max. Mutual Information(4/4)

We seek an estimate of the transformation that registers the reference volume u and test volume v by maximizing their mutual information.

$$\hat{T} = \arg\max_{T} I(u(x),\ v(T(x))) \quad . \qquad (1)$$

Given that T is a transformation from the coordinate frame of the reference volume to the test volume, v(T (x)) is the test volume voxel associated with reference volume voxel u(x).

Outline

1. Image Registration
2. Mutual Information
    - Introduction
    - Mathematic Background
3. Result

Parzen Window(1/5)

Our first step in estimating entropy from a sample is to approximate the underlying probability density p(z) by a superposition of functions centered on the elements of a sample A drawn from z:

$$p(z) \approx P^*(z) \equiv \frac{1}{N_A} \sum_{z_j \in A} R(z - z_j) \qquad (3)$$

Window function is a Gaussian density function.

$$G_\psi(z) \equiv (2\pi)^{\frac{-n}{2}} |\psi|^{\frac{-1}{2}} \exp(-\frac{1}{2} z^{\mathrm{T}} \psi^{-1} z) \ ,$$

where ψ is the (co-)variance of the Gaussian.

Estimating Entropy(2/5)

Unfortunately, evaluating the entropy integral is difficult if not impossible.

$$h(z) \approx -E_z[\ln P^*(z)] = -\int_{-\infty}^{\infty} \ln P^*(z)dz$$

This integral can however be approximated as a sample mean:

$$h(z) \approx -\frac{1}{N_B} \sum_{z_i \in B} \ln P^*(z_i) \quad , \qquad (4)$$

where N_B is the size of a second sample B.

We may now write an approximation for the entropy of a random variable z as follows:

$$h(z) \approx h^*(z) \equiv \frac{-1}{N_B} \sum_{z_i \in B} \ln \frac{1}{N_A} \sum_{z_j \in A} G_\psi(z_i - z_j) \quad . \quad (5)$$

Derivative of Entropy(3/5)

Next we examine the entropy of v(T (x)), which is a function of the transformation T . In order to find a maxima of entropy or mutual information, we may ascend the gradient with respect to the transformation T .

$$\frac{d}{dT}h^*(v(T(x))) = \frac{1}{N_B}\sum_{x_i \in B}\sum_{x_j \in A} W_v(v_i, v_j)(v_i - v_j)^{\mathrm{T}}\psi^{-1}\frac{d}{dT}(v_i - v_j) \quad (6)$$

using the following definitions:

$$v_i \equiv v(T(x_i)) \;,\; v_j \equiv v(T(x_j)) \;,\; v_k \equiv v(T(x_k)) \;,$$

weighting factor

$$W_v(v_i, v_j) \equiv \frac{G_{\psi_v}(v_i - v_j)}{\sum_{x_k \in A} G_{\psi_v}(v_i - v_k)} \;.$$

Derivative of Mutual Information(4/5)

In order to seek a maximum of the mutual information, we will calculate an approximation to its derivative

$$\frac{d}{dT}I(T) \approx \frac{d}{dT}h^*(u(x)) + \frac{d}{dT}h^*(v(T(x))) - \frac{d}{dT}h^*(u(x), v(T(x)))$$

1. The reference volume is not a function of the transformation. As a result its derivative is zero.
2. The entropy of the test volume is dependent on the variance of the window functions,
3. The joint entropy of two random variables, can be evaluated by $w = [u(x), v(T(x))]^{\mathrm{T}}$ --> h(w)

Max. Mutual Information(5/5)

$$\widehat{\frac{dI}{dT}} = \frac{1}{N_B} \sum_{x_i \in B} \sum_{x_j \in A} (v_i - v_j)^{\mathrm{T}} \quad [W_v(v_i, v_j)\psi_v^{-1} - W_w(w_i, w_j)\psi_{vv}^{-1}] \frac{d}{dT}(v_i - v_j) \ .$$

The weighting factors are defined as

$$W_v(v_i, v_j) \equiv \frac{G_{\psi_v}(v_i - v_j)}{\sum_{x_k \in A} G_{\psi_v}(v_i - v_k)} \quad , \text{ and } \quad W_w(w_i, w_j) \equiv \frac{G_{\psi_w}(w_i - w_j)}{\sum_{x_k \in A} G_{\psi_w}(w_i - w_k)}$$

using the following notation (and similarly for indices j and k),

$$u_i \equiv u(x_i) \ , \ v_i \equiv v(T(x_i)) \ , \text{ and } w_i \equiv [u_i, v_i]^{\mathrm{T}} \ .$$

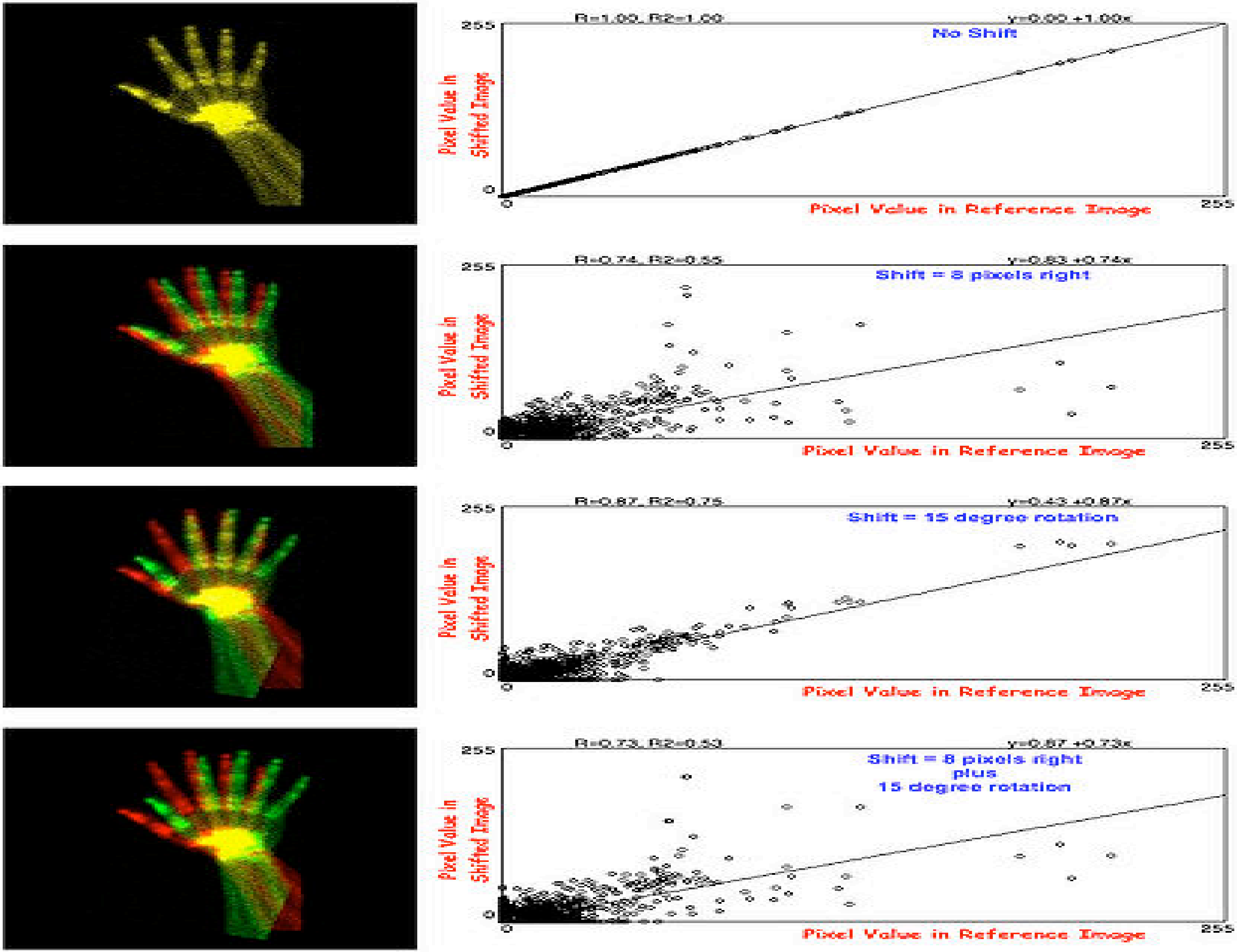
For Example, We can seek a maximum of mutual information by using a stochastic analog of gradient descent.

Repeat:

$$\begin{aligned} A &\leftarrow \{\text{sample of size } N_A \text{ drawn from } x\} \\ B &\leftarrow \{\text{sample of size } N_B \text{ drawn from } x\} \\ T &\leftarrow T + \lambda \widehat{\frac{dI}{dT}} \end{aligned}$$

Outline

1. Image Registration
2. Mutual Information
    - Introduction
    - Mathematic Background
3. Result



R=1.00, R2=1.00
y=0.00 +1.00x
No Shift
R=0.74, R2=0.55
y=0.83 +0.74x
Shift = 8 pixels right
R=0.87, R2=0.75
y=0.43 +0.87x
Shift = 15 degree rotation
R=0.73, R2=0.53
y=0.87 +0.73x
Shift = 8 pixels right
plus
15 degree rotation
Pixel Value in Shifted Image
Pixel Value in Reference Image
0
255

Mutual Information base tracking

Question