1. Uniformity Measures

It is often important to determine if a set of numbers is uniform or if two sets of numbers are close together or far apart. In image processing one may examine regions to see if a region is uniform. This is the same as determining if the pixel values in the region are uniform. One may also examine two adjacent regions and desire to determine if the gray-levels of the regions are close (the regions are candidates for merging) or if the gray-levels of the regions are far apart (the regions are not candidates for merging). There are many different measures of uniformity [Haralick and Shapiro, 1985]. that have been utilized. Some are examined in the following sections.

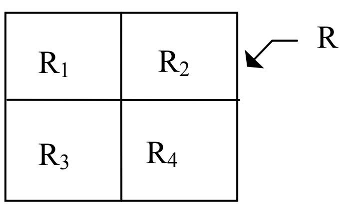
A uniformity function attempts to measure the similarity between adjacent regions so that the best representation of objects is obtained. It can be desirable to use different features (measures) if they give different properties of the region. The measures should not be correlated. If one has different measures to use in the uniformity function, then they can be combined into one uniformity function. Let $U_1,U_2,U_3,\ldots,U_N$ be uniformity functions. Then $U=F(w_1U_1,w_2U_2,w_3U_3,\ldots,w_NU_N)$ is a uniformity function with weights w_i that can be adjusted to reflect the importance of each individual uniformity measure. The weights make the system adaptive. They could be selected by the user or adjusted automatically.

1.1 Measures of Uniformity

It is often important to measure the distance between two feature vectors. Two common measures are the sum of the squared differences (SSD) and the sum of the absolute differences (SAD). The SSD is good when the noise is Gaussian while the SAD is justified when the noise is Exponential [Sebe, Lew, Huijsmans, 2000].

Most uniformity measures use variance in the equations. The mean of a region is defined to be $\mu=\frac{1}{N}\sum\{g(p)\mid p\,\varepsilon\,R\}$ and the variance is defined to be $\sigma^2=\frac{1}{N}\sum\{(g(p)-\mu)^2\mid p\,\varepsilon\,R\}$ where N is the number of points in region R. A more uniform region has a low variance.

Let's consider some examples of uniformity measures. Consider the case of region R where we consider to R into the sub regions R_1, R_2, R_3, and R_4. Let μ be the mean of R and μ_i be the mean of R_i. One simple test could be to consider the region R_i uniform if $|\mu-\mu_i|<\varepsilon$. If all the regions were uniform then R would not be split into the sub regions. Another simple test would be to consider the pixels in a region as at vector **u**. If one has two such vectors from regions R1 and R2 then a measure of the distance between the two regions is $d(\mathbf{u}_1,\mathbf{u}_2)=\frac{<\mathbf{u}_1,\mathbf{u}_2>}{\|\mathbf{u}_1\|\|\mathbf{u}_1\|}$ [Duda and Hart, 1973, pp. 216]. Using matrices with the vectors in a column matrix we have that $d(\mathbf{u}_1,\mathbf{u}_2)=\frac{\mathbf{u}_1^t * \mathbf{u}_2}{\|\mathbf{u}_1\|\|\mathbf{u}_1\|}$. This measure reflects the cosine of the angle between the two vectors.



Another measure of uniformity is $U_1(R)=\frac{\sum_{i=1}^{4} N_i\sigma_i^2}{N}$ where N_i =#(R_i) is the number of pixels in region R_i and N is the number of pixels in region R. The variance σ_i^2 is a measure of the uniformity of each sub region R_i. A larger variance indicates a region is less uniform. If the σ_i^2 are small, then this indicates that the split region will be more uniform. If the σ_i^2 are small, then $U_1(R)$ is small indicating that the region R may be split into sub regions.

Another measure is $U_2(R)=\frac{\sum_{i=1}^{4}|\mu_i-\mu|}{\sum_{i=1}^{4}\sigma_i^2}$. If $U_2(R)$ is large then R is not uniform and should be split into the four sub regions. $U_2(R)$ is large when $\sum_{i=1}^{4}|\mu_i-\mu)|$ is large and $\sum_{i=1}^{4}\sigma_i^2$ is small.

This says the sub regions differ from R but are themselves uniform. The split regions will be more uniform that the large region.

Consider now an example calculation of these uniformity measures shown in the following figure.. For this example, $\mu=\frac{48}{16}=3, \mu_1=2, \mu_2=3, \mu_3=2, \mu_4=5$.

R1		R2	
2	2	4	4
2	2	2	2
1	1	5	5
3	3	5	5
R3		R4	

Figure 1. Four Regions

Therefore, $|\mu-\mu_i|<2$ for the mean comparison test. In addition $\sigma_1^2=\frac{\sum_{(x,y)\in R_1}(g(p)\text{-}2)^2}{4}=0$ $\sigma_2^2=\frac{\sum_p (g(p)-3)^2}{4}=\frac{1+1+1+1}{4}=1$, $\sigma_3^2=\frac{\sum_{(x,y)\in R_1}(g(p)\text{-}2)^2}{4}=1$

$\sigma_4^2=\frac{\sum_{(x,y)\in R_1}(g(p)\text{-}5)^2}{4}=0$. Therefore, $U_1(R)=\frac{\sum_{i=1}^4 n_i\sigma_i^2}{n}=\frac{0+4+4+0}{16}=\frac{1}{2}$.

For the calculation of $U_2(R)$ we get $U_2(R)=\frac{\sum_{i=1}^{4}|\mu_i\text{-}\mu|}{\sum_{i=1}^{4}\sigma_i^2}=\frac{4}{2}=2$.

An additional measure of uniformity for regions is $U_3=1-\sum_i \frac{w_i\sigma_i^2}{\sigma_{max}^2}$ [Ng and Lee, 1996].

Here $\sigma_{max}^2=\frac{(g\max-g\min)^2}{2}$ where gmax and gmin are the max and minimum gray-level values in the regions and w_j is a weight associated with region R_j. The quantity σ_{max}^2 is a normalizing factor. Two regions have good separation if the between class variance is large and the within class variances are small. This indicates that the regions should not be merged. Therefore, if U_3 is large the regions should not be merged and a small U_3 indicates the regions should be merged. For the previous example we get $\sigma^2_{max} = \frac{(g\max - g\min)^2}{2} = \frac{(5-1)^2}{2} = 8$ and $U_3 = 1 - \sum_{R_j} \frac{w_j \sigma_j^2}{\sigma^2_{max}} = 1 - \frac{(0+0+1+1)}{8} = 1 - \frac{1}{4} = .75$. The w_j are each 1 in this example.

1.2 Additional Measures of Uniformity.

The Fisher distance function is another measure of region uniformity. Consider two regions R_1 and R_2. The equation $U_4(R_{12}) = \frac{(N_1 + N_2)(\mu_1 - \mu_2)^2}{N_1\sigma_1^2 + N_2\sigma_2^2}$ is the Fisher distance [Zhu and Yuille, 1996]. If the value is low, the regions should be merged. The term N is the number of points in both regions while σ^2 is the variance of the combined regions. If σ_i^2=0 for every i. then this means the regions are uniform and should not be merged. If the numerator is also zero, one would consider merging them. For the above example, we get $U_4(R_{12}) = \frac{(N_1 + N_2)(\mu_1 - \mu_2)^2}{N_1\sigma_1^2 + N_2\sigma_2^2} = \frac{8*1^2}{4(0+1)} = 2$ for regions one and two. For regions one and three we get $U_4(R_{13}) = 32$. The numerator is a measure of the separation of the means and the denominator is a measure of the within class variance.

The fisher measure of class separation is given by $J = \frac{|\mu_a - \mu_b|}{\sigma_a^2 + \sigma_b^2}$ [Schalkoff, 1992, pp. 91]. If the number is small the regions should be separated.

If one has multivariate data, then another form of the equations for the fisher distance are given below [Duda and Hart, 1973 , pp 117].

$J = \frac{|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b|}{s_a^2 + s_b^2}$, where

$\boldsymbol{\mu}_a$ is the mean vector of the samples from class A or R_1 for our examples ,

$\boldsymbol{\mu}_b$ is the mean vector of the samples from class B or R_2 for our examples.

$s_a^2 = \sum_{\mathbf{u}_i \in A} (\mathbf{u}_i - \boldsymbol{\mu}_a)^2$, and $s_b^2 = \sum_{\mathbf{u}_i \in B} (\mathbf{u}_i - \boldsymbol{\mu}_b)^2$.

Another measure of uniformity is derived from statistical comparisons [Jain, Kasturi, and Schunck, 1995, pp. 99]. Consider two adjacent regions R_1 and R_2. Let σ_1 be the standard deviation of region one, σ_2 be the standard deviation of region two and σ be the standard deviation of the combined regions. Let N_i be the number of pixels in region R_i and N be the number of pixels in region R. The measure of uniformity is $U_5 = \frac{\sigma^N}{\sigma_1^{N_1} + \sigma_2^{N_2}}$. If U_5 is small then merge the two regions otherwise keep them separate. Again using regions one and two in the above example we get $\mu = 2.5$ and $\sigma^2 = .75$ for the two combined regions. Therefore, $U_5 = \frac{\sigma^N}{\sigma_1^{N_1} + \sigma_2^{N_2}} = \frac{.86^8}{0^4 + 1^4} = .32$.

The Mahalanobis distance function may also be used as a uniformity measure. The measure is linearly related to the log of the probability that a measurement vector is drawn from a given distribution [Duda and Hart, 1973 , pp. 24 ; Cootes and Taylor, 2001]. There is an assumption that the distributions are Gaussian [Duda and Hart, 1973 , pp. 24]. The Mahalanobis distance is $\mathrm{mahd}(\mathbf{u}_s) = (\mathbf{u}_s - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{u}_s - \boldsymbol{\mu})$ where $\mathbf{u}_s$ m_s is a sample measurement vector, $\boldsymbol{\mu}$ is the mean vector, and Σ is the covariance matrix of samples from a given class. To use the distance function one would compare the distance of the point to different regions and add the point to the region where it has the smallest distance. The covariance matrix is singular when the sample vectors are linearly dependent otherwise the matrix has an inverse. From the previous example if we let z = 5 then the mahd distance from z to region 2 is 4 while the mahd distance from z to region 3 is 9.

The Bhattacharyya distance measures the distance between two Gaussian multivariate distributions. It is given in the following equation. The general form is $\text{bhath} = \ln \int_{-\infty}^{\infty} (f * g)^{\frac{1}{2}}$ [Landgrebe, 2002]. The terms f and g are the class conditional density function. The equation expressed in terms of the first two moments is given in the following equation.

$$\text{bhatd} = \frac{1}{8}(\boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\mu}_{\mathbf{b}})^{t}\left[\frac{\Sigma_{a} + \Sigma_{b}}{2}\right]^{-1}(\boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\mu}_{\mathbf{b}}) + \frac{1}{2}\ln\left[\frac{\left|\frac{\Sigma_{a} + \Sigma_{b}}{2}\right|}{\sqrt{|\Sigma_{a}||\Sigma_{b}|}}\right].$$

The first term gives the distance in terms of the means and the second term gives the distance in terms of the difference between class covariance matrices. The optimal Bayes classifier error is related to this distance. For regions one and three of the previous example the bhatd distance is 1.

The Kullback relative information measure is given by $K = \sum_{i=1}^{N_u} u(i)\log\left(\frac{u(i)}{v(i)}\right)$ where $u=(u(1), u(2), \ldots, u(N_u))$ and $v=(v(1), v(2), \ldots, v(N_u))$ $v = \{v_i\}$ are two discrete probability density vectors [Sebe, Lew, Huijsmans, 2000]. One can use this measure with pixels from an image by arranging the pixels as a vector and normalizing the sum to one. In continuous form the equation for the Kullback-Leibler distance is $D(p_1 \| p_2) = \int p_1(k)\log\left(\frac{p_1(k)}{p_2(k)}\right)dk$ [Johnson and Sinanovic, 2001] where p_1 and p_2 are two probability functions. The Kullback distance is not symmetric. One symmetric measure is the J divergence given by $J(p_1,p_2) = \frac{D(p_1 \| p_2) + D(p_2 \| p_1)}{2}$ [Johnson and Sinanovic, 2001; Ifarraguerri and Chang, 2000]. In our example the distance between regions 1 and 2 is $J(p_1,p_2) = \frac{D(p_1 \| p_2) + D(p_2 \| p_1)}{2} = .0578$ while the distance between regions 1 and 3 is . $J(p_1,p_3) = \frac{D(p_1 \| p_3) + D(p_3 \| p_1)}{2} = .1373$.

Another symmetric measure is given by $\frac{1}{R(p_1,p_2)} = \frac{1}{D(p_1 \| p_2)} + \frac{1}{D(p_2 \| p_1)}$ [Johnson and Sinanovic, 2001]. formed as a harmonic mean calculation. It is called the resistive measure due to its analogy to the formula for computing the equivalent resistance of parallel resistors. The resistive distance is thought to be the better symmetric distance because it more accurately reflects the average error probability of an optimal classifier and can be easily computed [Johnson and Sinanovic, 2001].

Another measure of distance is the Chernoff distance $C(p_1,p_2) = \max\{-\log u(t) \mid 0 \le t \le 1\}$ where $u(t) = \int [p_1(k)]^{1-t}[p_2(t)]^t dk$ [Johnson and Sinanovic, 2001]. The Bhattacharyya distance is $B(p_1,p_2) = -\log(u(1/2))$.

1.3 Feature Size Reduction

One may well define many different features to characterize an object. It is not desirable to keep a large number of features in the final characterization of objects. Therefore, one should reduce the feature size. If one has a training set and a pattern classifier then one can perform searches and select the subset of features that gives the best classification results [Ganster, Pinz, Rohrer, Wilding, Binder, and Kittler, 2001]