Chapter 14
Image Similarity Measures

Abstract. The subject of this chapter is image similarity measures. These measure provide a quantitative measure of the degree of match between two images, or image patches, A and B. Image similarity measures play an important role in many image fusion algorithms and applications including retrieval, classification, change detection, quality evaluation and registration. For the sake of concreteness we shall concentrate on intensity based similarity measures.

14.1 Introduction

Comparing two input images, or image patches, is a fundamental operation in many image fusion algorithms [21, 25, 26]. A meaningful image similarity measure [1] has two components: (1) A transformation T. This extracts the characteristics of an input image and represents it as multi-dimensional feature vector. (2) A distance measure D. This quantifies the similarity between the two images, where D is defined in the multi-dimensional feature space.

Mathematically, we represent a similarity measure between two images A and B as

$$S(A,B) = D(T(A),T(B)) \ .$$

The following example illustrates the use of a similarity measure in a content-based image retrieval (CBIR) system.

Example 14.1. A CBIR System [7, 13]. A CBIR system aims to recover images from an image repository or database, according to the user's interest. In the CBIR system each image in the database is represented as a multi-dimensional

[1] We use the term "similarity measure" as a general term which includes both similarity measures (which reach their maximum value when $A = B$) and dissimilarity, or distance, measures (which reach their minimum value when $A = B$). Apart from mutual information MI all the measures discussed in this chapter are dissimilarity measures.

feature vector which is extracted from a series of low-level descriptors, such as a color histogram, a co-occurrence matrix, morphological features, wavelet-based descriptors or Zernike moments. The subjective similarity between two pictures is quantified in terms of a distance measure which is defined on the corresponding multi-dimensional feature space. Common distance measures are: the Minkowski distance, the Manhattan distance, the Euclidean distance and the Hausdorff distance.

A similarity measure $S(A,B)$, or a distance $D(T(A),T(B))$, is metric when it obeys the following:

1. $S(A,B) \geq 0$ or $D(T(A),T(B)) \geq 0$,
2. $S(A,B) = 0$ or $D(T(A),T(B)) = 0$ if, and only if, $A = B$,
3. $S(A,B) = S(B,A)$ or $D(T(A),T(B)) = D(T(B),T(A))$,
4. $S(A,C) \leq S(A,B) + S(B,C)$ or
   $D(T(A),T(C)) \leq D(T(A),T(B)) + D(T(B),T(C))$.

Many studies on image similarity [21] suggest that practical and psychologically valid measures of similarity often obey the first three conditions but do not obey the fourth condition (known as the triangle inequality) and are therefore non-metric.

In designing the similarity measure we choose the transformation T according to what image characteristics are important to the user. The following example illustrates these concerns for a stereo matching algorithm.

Example 14.2. Stereo matching Algorithm [11]. In a stereo matching, or disparity, algorithm we compare two images A and B which are two views of the same scene taken from slightly different viewing angles. The image similarity measures $S(A,B)$ should therefore be insensitive to changes due to specular reflections, occlusions, depth discontinuities and projective distortions [5, 11]. At the same time, $S(A,B)$ should be sensitive to any other changes in A and B.

It is should be clear that there is no universal similarity measure which can be used in all applications. In selecting a suitable similarity measure we find it useful to broadly divide them into two groups:

Global Measure. These measures return a single similarity value which describes the overall similarity of the two input images. The global measures may be further divided into measures which require the input images to be spatially registered and those which do not require the input images to be spatially registered.

Local Measures. These measures return a similarity image or map which describes the local similarity of the two input images. By definition the local similarity measures require the input images to be spatially registered.

Although useful, the above division of the similarity measures into two classes should not be regarded as absolute. In many cases, we may convert a global similarity measure into a local similarity measure and vice versa. The following example illustrates a case of the former.

Example 14.3. Global to Local Similarity Measures. The mean square error (mse) is a simple global similarity measure. Given two $M \times N$ spatially registered input images A and B, the global mse measure is defined as follows:

$$mse_G = \sum_{m=1}^{M} \sum_{n=1}^{N} (A(m,n) - B(m,n))^2 / MN \, .$$

Clearly we may apply the mse measure to individual pixels. In this case we obtain a local mse map $mse_L(m,n)$ which is defined as follows:

$$mse_L(m,n) = (A(m,n) - B(m,n))^2 \, .$$

More generally, we may calculate the mse over a local window. Let $W(m,n)$ define a $L \times L$ window centered at (m,n), where we assume L is an odd number (Fig. 14.1). Then we may designate the gray-levels of the pixels in A which lie in $W(m,n)$ as $\widetilde{A}(p,q|m,n), p,q \in \{1,2,\ldots,L\}$, where

$$\widetilde{A}(p,q|m,n) = A(m+p-1-\lfloor L/2 \rfloor, n+q-1-\lfloor L/2 \rfloor) \, .$$

Similarly, $\widetilde{B}(p,q|m,n), p,q \in \{1,2,\ldots,L\}$ designates the gray-levels of the pixels in B which lie in $W(m,n)$. In this case, we define a local windowed mse map, $mse_W(m,n)$, as follows:

$$\begin{aligned} mse_W(m,n) &= \sum_{p=1}^{L} \sum_{q=1}^{N} \left(\widetilde{A}(p,q|m,n) - \widetilde{B}(p,q|m,n)\right)^2 / L^2 \, , \\ &= \sum_{p=m-\lfloor L/2 \rfloor}^{m+\lfloor L/2 \rfloor} \sum_{q=n-\lfloor L/2 \rfloor}^{n+\lfloor L/2 \rfloor} (A(p,q) - B(p,q))^2 / L^2 \, . \end{aligned}$$

Example 14.4. Local to Global Similarity Measures. Given two spatially registered $M \times N$ binary images A and B, a local similarity algorithm returns a local similarity measure $S_L(m,n)$ for each pixel $(m,n), m \in \{1,2,\ldots,M\}, n \in \{1,2,\ldots,N\}$. We may obtain a global similarity measure by aggregating the $S_L(m,n)$ values, e. g. by finding the global maximum of the $S_L(m,n)$ values:

$$S_G = \max_{(m,n)} (S_L(m,n)) \, .$$

A(1,1)	A(1,2)	A(1,3)	A(1,4)	A(1,5)
A(2,1)	A(2,2)	A(2,3)	A(2,4)	A(2,5)
A(3,1)	A(3,2)	A(3,3)	A(3,4)	A(3,5)
A(4,1)	A(4,2)	A(4,3)	A(4,4)	A(4,5)
A(5,1)	A(5,2)	A(5,3)	A(5,4)	A(5,5)

Fig. 14.1 Shows an input image with gray-levels $A(m,n), m,n \in \{1,2,\ldots,5\}$. Centered at $(3,4)$ we have a 3×3 window $W(3,4)$. The gray-levels in $W(3,4)$ are $\widetilde{A}(3,3) = \{A(2,3),A(2,4),A(2,5),A(3,3),A(3,4),A(3,5),A(4,3),A(4,4),A(4,5)\}$.

We start our discussion with the global similarity measures which do not require image registration.

14.2 Global Similarity Measures without Spatial Alignment

In this section we consider global similarity measures which do not require spatial alignment. These are similarity measures which compare the probability distributions or gray-level histograms of the two images. In general these similarity measures are robust against changes in illumination. However, because they do not require spatial alignment, their discrimination power is low.

14.2.1 Probabilistic Similarity Measures

The probabilistic similarity measures are global measures which do not require the input images to be spatially registered. By converting the input images to probability distributions, they are robust against changes in illumination and are widely used when the images have been captured under widely varying illumination and viewing conditions or by different sensor types.

Let A and B denote the two input images. We convert the pixel gray-levels $a \in A$ and $b \in B$ to a common gray scale x (see Chapt. 6). Let $p(x)$ and $q(x)$ denote the probability of a transformed gray-level x appearing in A and B, then several commonly used probabilistic similarity measures are:

Chernoff

$$S_C = -\log \int_x p^{\alpha}(x) q^{1-\alpha}(x) dx \,, \qquad 0 < \alpha < 1 \,.$$

Bhattacharyya

$$S_B = -\log \int_x \sqrt{p(x)q(x)} dx \,.$$

Jeffrey's-Matusita

$$S_{JM} = \sqrt{\int_x \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx} \,.$$

Kullback-Leibler

$$S_{KL} = \int_x p(x) \log \frac{p(x)}{q(x)} dx \,.$$

If the images A and B are spatially registered then we may use the sliding window procedure to generate local probabilistic similarity maps. Let $W(m,n)$ define a local $L \times L$ window centered at (m,n). If $\widetilde{A}$ and $\widetilde{B}$ designate the gray-levels of the pixels in A and B which lie in $W(m,n)$ (cf. Ex. 14.3) and $\widetilde{p}(x)$ and $\widetilde{q}(x)$ designate the corresponding transformed local (window) probability densities, then the local probabilistic similarity maps are

$$\begin{aligned}
\widetilde{S}_C(m,n) &= \int_x \widetilde{p}^{\alpha}(x) \widetilde{q}^{1-\alpha}(x) dx \,, \\
\widetilde{S}_B(m,n) &= \int_x \sqrt{\widetilde{p}(x)\widetilde{q}(x)} dx \,, \\
\widetilde{S}_{JM}(m,n) &= \sqrt{\int_x \left(\sqrt{\widetilde{p}(x)} - \sqrt{\widetilde{q}(x)}\right)^2 dx} \,, \\
\widetilde{S}_{KL}(m,n) &= \int \widetilde{p}(x) \log \frac{\widetilde{p}(x)}{\widetilde{q}(x)} dx \,,
\end{aligned}$$

Note The typical window size used in the local probability similarity measures is 20×20. This is needed to ensure we have sufficient pixels to accurately calculate the local probability densities $\widetilde{p}(x)$ and $\widetilde{q}(x)$.

Example 14.5. Color Image Segmentation [17]. The goal of image segmentation is to decompose the input image into a set of meaningful or spatially coherent regions sharing similar attributes. The algorithm is often a crucial step in many video and computer vision applications such as object localization or recognition. A simple image segmentation is the K-means cluster algorithm in which we divide the pixels into K clusters. Given a input image I in a given color space, we may characterize each pixel (x,y) in I by its local histogram

$\widetilde{\mathbf{H}}(x,y)$:

$$\widetilde{\mathbf{H}}(x,y) = (\widetilde{H}_1, \widetilde{H}_2, \ldots, \widetilde{H}_M)^T ,$$

where $\widetilde{H}_m$ is the number of pixels in the local window $W(x,y)$ whose color values fall in the mth bin. We then apply the K-means algorithm as follows. Initially we define K cluster centers by randomly selecting K histograms. Let $C_1, C_2, \ldots, C_K$ denote the K cluster centers or histograms. Each pixel (x,y) is associated with a given cluster

$$\delta_k(x,y) = \begin{cases} 1 \text{ if } (x,y) \text{ is associated with } C_k , \\ 0 \text{ otherwise} . \end{cases}$$

Then the K-means algorithm attempts to find the set of cluster centers $C_k, k \in \{1,2,\ldots,K\}$, such that the overall error

$$E = \sum_{(x,y)} \sum_{k=1}^{K} \delta_k(x,y) D\left(\widetilde{H}(x,y), C_k\right) ,$$

is a minimum, where $D(\widetilde{H}(x,y), C_k)$ is an appropriate distance (similarity) measure between $\widetilde{H}(x,y)$ and C_k. Mignotte [17] recommends using the Bhattacharyya distance.

14.2.2 χ^2 Distance Measure

If we represent the transformed distributions $p(x)$ and $q(x)$ as discrete distributions p_k and q_k, then we may use the χ^2 distance as a dissimilarity measure.

Let A and B denote two input images with gray-levels $a \in A$ and $b \in B$. We convert the gray-levels $a \in A$ and $b \in B$ to a discrete common scale x by defining K pairs of corresponding bins $[a'_k, a''_k)$ and $[b'_k, b''_k)$. Then the χ^2 distance between gray-level distributions of A and B is

$$\chi^2 = \sum_{k=1}^{K} \frac{(m_k - n_k)^2}{m_k + n_k} , \tag{14.1}$$

where m_k is the number of gray-levels $a \in A$ which fall in the interval $[a'_k, a''_k)$ and n_k is the number of gray-levels $b \in B$ which fall in the interval $[b'_k, b''_k)$.

Example 14.6. Face Recognition With Local Binary Patterns [1]. Ref. [1] describes an efficient image representation based on the local binary pattern (LBP) texture features (see Sect. 3.4). Given a training set of K facial images $A^{(k)}, k \in \{1,2,\ldots,K\}$, we divide each image A_k into R regions. *Note*: We assume the training images are spatially aligned.

For each pixel (m,n) in $A^{(k)}$ we extract its local binary pattern $LBP(m,n)$ which is a label $l, l \in \{1,2,\ldots,L\}$. Then for each region $r, r \in \{1,2,\ldots,R\}$, we construct a histogram (vector)

$$\mathbf{H}_r^{(k)} = (H_r^{(k)}(1), H_r^{(k)}(2), \ldots, H_r^{(k)}(L))^T ,$$

where $H_r^{(k)}(l)$ is the number of pixels in the rth region of A_k which have a LBP label equal to l. Each training image $A^{(k)}$ is thus represented by R histograms $\mathbf{H}_r^{(k)}, r \in \{1,2,\ldots,R\}$. Given a test image B we spatially align it to the training images and carry out the above process. Let $\mathbf{h}_r, r \in \{1,2,\ldots,R\}$, denote the corresponding histograms (vectors). Then we identify the test image B as belonging to the k^*th individual if

$$k^* = \arg\min_k \left(\sum_{r=1}^{R} \chi^2(\mathbf{h}_r, \mathbf{H}_r^{(k)}) \right) ,$$

where

$$\chi^2(\mathbf{h}_r, \mathbf{H}_r^{(k)}) = \sum_{l=1}^{L} \frac{(h_r(l) - H_r^{(k)}(l))^2}{h_r(l) - H_r^{(k)}(l)} .$$

The method showed high performance on difficult face recognition experiments.

The original χ^2 distance as defined in (14.1) defined between histograms $H_A = (m_1, m_2, \ldots, m_K)$ and $H_B = (n_1, n_2, \ldots, n_K)$ and not between the discrete probability distributions p_k and q_k. However, (14.1) may be easily converted to a probability distance measure by replacing m_k and n_k by $p_k = m_k/M$ and $q_k = n_k/N$, where $M = \sum_{k=1}^{K} m_k$ and $N = \sum_{k=1}^{K} n_k$.

Example 14.7. Probability Binning [19]. In probability binning we use variable width bins such that each bin contains the same relative number of observations of A. If $p_k = m_k/M$ and $n_k = q_k/N$ denote, respectively, the relative number of observations of A and B in the kth bin, where $M = \sum_{k=1}^{K} m_k$ and $N = \sum_{k=1}^{N} n_k$, then $m_1 = m_2 = \ldots = m_K$ and the probability binning χ^2 test is

$$\chi^2_{PB} = \sum_{k=1}^{K} \frac{|p_k - q_k|^2}{p_k + q_k} .$$

Given χ^2_{PB} we can define a normalized scale for it as follows. Let

$$T(\chi^2_{PB}) = \max\left(0, \frac{\chi^2_{PB} - \mu}{\sigma}\right) ,$$

then $T(\chi^2_{PB})$ represents the difference between the probability distributions of A and B as the number of standard deviations above μ, where $\mu = K/\min(M,N)$ is the minimum difference between F and G for which a confident decision of histogram difference can be made and $\sigma = \sqrt{K}/\min(M,N)$ is an appropriate standard deviation for χ^2_{PB}. Recently, Baggerly [3] has proposed a more accurate scale as follows:

$$T(\chi^2_{PB}) = \frac{\frac{2MN}{M+N}\chi^2_{PB} - (K-1)}{\sqrt{2(K-1)}} .$$

As in (14.1) we may define a local χ^2 measure:

$$\widetilde{\chi}^2(x,y) = \sum_{k=1}^{K} \frac{(\widetilde{m}_k(x,y) - \widetilde{n}_k(x,y))^2}{\widetilde{m}_k(x,y) + \widetilde{n}_k(x,y)} ,$$

where $\widetilde{m}_k(x,y), \widetilde{n}_k(x,y)$ denote, respectively, the number of pixels in $\widetilde{A}, \widetilde{B}$ which have a gray-level which falls in the kth histogram bin.

14.2.3 Cross-Bin Distance Measures

The global similarity measures considered until now (S_C, S_B, S_{JM}, S_{KL} and χ^2), all suppose the gray-levels a and b are measured on a common gray-scale. These similarity measures are therefore sensitive to any errors involved in defining the common gray-scale.

A discrete similarity measure which is less sensitive to any errors involved in defining a common gray-scale is the Earth Mover's distance (EMD) [20, 23]: Let m_k and n_k be the number of pixels in A, B which fall, respectively, in the kth histogram bin. Then the Earth Mover's distance between $\mathbf{m} = (m_1, m_2, \ldots, m_K)^T$ and $\mathbf{n} = (n_1, n_2, \ldots, n_K)^T$, is defined as

$$d_{EMD}(\mathbf{m},\mathbf{n}) = \min_{\alpha_{h,k}} \sum_{h=1}^{K} \sum_{k=1}^{K} c(h,k) ,$$

subject to

$$\alpha_{ij} \geq 0 ,$$
$$\sum_k \alpha_{hk} = m_h ,$$
$$\sum_h \alpha_{hk} = n_k ,$$

where $c(h,k)$ is an appropriate cost function. The earth mover's distance may be understood as an optimization technique which finds the minimum transportation cost. In this case, $c(h,k)$ is the cost of moving a unit mass from the hth bin to the kth bin and α_{hk} is the number of mass units carried from h to k.

Example 14.8. Mallow's Distance [12]. If we use normalized distributions $\mathbf{p} = (p_1, p_2, \ldots, p_K)^T$, where $p_k = m_k/M$ and $\mathbf{q} = (q_1, q_2, \ldots, q_K)^T$, where $Q_k = n_k/N$, then the EMD becomes the Mallow's distance [12]. If the histograms are one-dimensional and we use the following cost function $c(i,j) = |i-j|/K$, then

$$d_{mallow}(\mathbf{p},\mathbf{q}) = d_{EMD}(\mathbf{p},\mathbf{q}) = \frac{1}{K}\sum_{k=1}^{K} |P_k - Q_k| \; ,$$

where

$$P_k = \sum_{h=1}^{k} p_k \qquad \text{and} \qquad Q_k = \sum_{h=1}^{k} q_h \; .$$

The circular EMD [18] is a variant of the EMD which is used when one of the variables is circular in nature e.g. an angle.

Example 14.9. Circular Earth Mover's Distance [18]. If $\mathbf{p}$ and $\mathbf{q}$ are one-dimensional and $c(h,k) = |h-k|/K$, then the corresponding circular EMD is:

$$d_{CEMD} = \min_{h \in \{1,2,\ldots,K\}} \left(\frac{1}{K}\sum_{k=1}^{K} |\widetilde{P}_{hk} - \widetilde{Q}_{hk}| \right) ,$$

where

$$\widetilde{P}_{hk} = \begin{cases} \sum_{i=h}^{k} p_i & \text{if } k \geq h \; , \\ \sum_{i=h}^{K} p_i + \sum_{i=1}^{k} p_i & \text{if } k < h \; , \end{cases}$$

$$\widetilde{Q}_{hk} = \begin{cases} \sum_{i=h}^{k} q_i & \text{if } k \geq h \; , \\ \sum_{i=h}^{K} q_i + \sum_{i=1}^{k} q_i & \text{if } k < h \; , \end{cases}$$

An important consideration in the χ^2 and other histogram distance measures is the optimal selection of the histogram bins. This is an important issue: If the bin width is too narrow then the histogram is very noisy while if the bin width is too wide then the histogram is too smooth. In both cases, the discrimination power of the distance measure will be adversely affected. Recently [6] have described a simple semi-empirical formula for estimating the optimal number of bins in a regular histogram. We assume the pixel gray-levels are defined in the interval $[0,1]$. If there are N pixels, then the optimal number of bins is k^*:

$$k^* = \arg\max_k \left(L(k) - R(k) \right) ,$$

where

$$L(k) = \sum_{l=1}^{k} H(l) \log_2 \left(\frac{kH(l)}{N} \right) ,$$
$$R(k) = k - 1 + (\log_2 k)^{2.5} .$$

We now consider global similarity measures which require spatial alignment of the two input images A and B.

14.3 Global Similarity Measures with Spatial Alignment

In this section we consider the family of global similarity measures which require spatial alignment of the two input images. These similarity measures tend to return values change monotonically with increasing spatial misalignment. For this reason, these similarity measures are often used for spatial alignment algorithms (see Chapt. 4). We start with the mean square error (mse) and the mean absolute error (mae) which are probably the simplest measures [2].

14.3.1 Mean Square Error and Mean Absolute Error

The mean square error (mse) and the mean absolute error (mae) are defined as follows:

$$mse = \sum_k (a_k - b_k)^2 / K ,$$
$$mae = \sum_k |a_k - b_k| / K ,$$

where a_k and b_k are, respectively, the gray-levels of the kth pixel in A and B.

The mse and mae should be used when the input images have been captured with the same sensor under similar conditions, i. e. the photometric transformation between corresponding pixel gray-levels should be close to the identity transformation. Both measures are sensitive to outliers although the mae is less sensitive (more robust). In this case, we may robustify the mse and the mae by replacing the summations in the above equations by an α-trimmed summation:

[2] The mse and mae increase with increasing misalignment. The correlation coefficient and mutual information decrease with increasing misalignment.

$$mse_\alpha = \frac{1}{K-2\alpha}\sum_{k=\alpha+1}^{K-\alpha} d_{(k)}^2 ,$$
$$mae_\alpha = \frac{1}{K-2\alpha}\sum_{k=\alpha+1}^{K-\alpha} |d_{(k)}| .$$

where $d_{(k)} = d_l$ if $d_l = |a_l - b_l|$ is the kth largest absolute difference and α is a small number. We often set α equal to $\lfloor K/20 \rfloor$.

14.3.2 Cross-Correlation Coefficient

The cross-correlation coefficient is defined as follows:

$$\rho = \frac{\sum_k a_k b_k}{\sqrt{\sum_k a_k^2 \sum_k b_k^2}} . \tag{14.2}$$

Sometimes we use a zero-mean cross correlation coefficient. This is defined as

$$\rho_Z = \frac{\sum_k (a_k - \bar{A})(b_k - \bar{B})}{\sqrt{\sum_k (a_k - \bar{A})^2 \sum_k (b_k - \bar{B})^2}} ,$$

where $\bar{A}$ and $\bar{B}$ are, respectively, the mean gray-levels of A and B.

The cross-correlation coefficients are more robust to changes of illumination than the mse and mae. The cross-correlation coefficient should be used when the images are captured by the same sensor and any changes in illumination may be approximated with a linear transformation. Many changes in illumination are not however linear. In this case we must use mutual information and other ordinal similarity measures.

The cross-correlation coefficients may be easily made robust against outliers [2]. For example, a robust version of (14.2) is

$$S'_{CC} = \frac{\sum_k \rho_k a_k b_k}{\sqrt{\sum_k \rho_k^A a_k^2 \sum_k \rho_k^B b_k^2}} .$$

where

$$\rho_k^A = \begin{cases} a_k & \text{if } a_k < 1.345\sigma_A , \\ 1.345\sigma_A \mathrm{sgn}(a_k) & \text{otherwise} , \end{cases}$$
$$\rho_k^B = \begin{cases} b_k & \text{if } b_k < 1.345\sigma_B , \\ 1.345\sigma_B \mathrm{sgn}(b_k) & \text{otherwise} , \end{cases}$$
$$\rho_k = \sqrt{\rho_k^A \rho_k^B} ,$$

and σ_A, σ_B are the standard deviations of the a_k and b_k values.

14.3.3 Mutual Information

The mutual information [3] between two input images A and B is defined as follows:

$$MI(A,B) = \int\int p_{AB}(a,b)\log_2 \frac{p_{AB}(a,b)}{p_A(a)p_B(b)} dxdy ,$$

where $p_A(a)$ is the probability a pixel (x,y) in A has a gray-level a, $p_B(b)$ is the probability a pixel (x,y) in B has a gray-level b and $p_{AB}(a,b)$ is the probability a pixel (x,y) in A has a gray-level a and the same pixel in B has a gray-level b.

In multi-modal applications no direct relationship between the input image intensities can be assumed. In this case, similarity measures which rely on the probabilistic relation and the distribution of the intensities in the input images is used. If the input images have been captured by different sensors or by different spectral bands, then the mutual information between two images A and B is used. Further details on MI and how it is calculated is given in Sect. 4.6.

14.3.4 Ordinal Global Similarity Measures

Ordinal global similarity measures are based on order statistics. They do not use the pixel gray-levels in A and B, but use instead the ordered gray-levels. In general, these measure are insensitive to changes in illumination if the order of the gray-levels is preserved. They are often used in applications involving change detection or in applications where the images have been captured with two different sensors.

Two classical ordinal dissimilarity measures are the Spearman ρ measure and Kendall's τ measure [10]. If A,B each contain K pixels with gray-levels $a_k, b_k, k \in \{1,2,\ldots,K\}$, then these dissimilarity measures are defined, respectively, as

$$\rho = 1 - \frac{\sum_{k=1}^{K} |r_A(k) - r_B(k)|^2}{6K(K-1)} , \tag{14.3}$$

$$\tau = \sum_{k=1}^{K}\sum_{l=1}^{K} \frac{\text{sgn}(a_k - a_l)\text{sgn}(b_k - b_l)}{K(K-1)} , \tag{14.4}$$

where $r_A(k)$ and $r_B(k)$ denote, respectively, the rank of the kth pixel in A and B [4] and

$$\text{sgn}(u) = \begin{cases} -1 & \text{if } u < 0 , \\ 0 & \text{if } u = 0 , \\ 1 & \text{if } u > 0 . \end{cases}$$

Note: The definitions given in (14.3) and (14.4) assume no ties. For corrections necessary if ties are present see e. g. [11]. Two additional ordinal dissimilarity measures are the Kemeny-Snell d_{KS} [15] and the Bhat-Nayar [5] d_{BN} distance measures.

[3] Mutual information is a similarity measure which reaches its maximum value when $A = B$

[4] The ranks $r_A(k)$ and $r_B(k)$ are defined as follows. Suppose A and B each contain K pixels with gray-levels a_k, b_k. Then $r_A(k) = l$ if a_k is the lth largest gray-level in A and $r_B(k) = l$ if b_k is the lth largest gray-level in B.

The Kemeny-Snell distance d_{KS} compares the relative ranking of each ordered pair of locations in one image with its relative ranking in the other image. Smaller values of d_{KS} indicate more agreement between the images.

Suppose A and B both contain K pixels with gray-levels $a_k, k \in \{1,2,\ldots,K\}$, and b_k. Mathematically, d_{KS} is defined as follows:

$$d_{KS}(A,B) = \sum_{k=1}^{K} \sum_{l=1}^{K} |\phi_{kl} - \psi_{kl}| ,$$

where

$$\phi_{kl} = \begin{cases} 1 & \text{if } a_k > a_l , \\ \frac{1}{2} & \text{if } a_k = a_l , \\ 0 & \text{otherwise} , \end{cases} \qquad \text{and} \qquad \psi_{kl} = \begin{cases} 1 & \text{if } b_k > b_l , \\ \frac{1}{2} & \text{if } b_k = b_l , \\ 0 & \text{otherwise} . \end{cases}$$

A normalized form of d_{KS} is

$$\hat{d}_{KS} = \frac{d_{KS}}{\bar{d}_{KS}} ,$$

where $\bar{d}_{KS}$ is the value of d_{KS} if the pixel gray-levels occurring in A and B were randomly distributed among the pixel locations in the two images.

Example 14.10. Kemeny-Snell Distance [15]. Given two one-dimensional image patches

$$A = (24,12,14,7,50)^T \qquad \text{and} \qquad B = (30,14,13,40,4)^T .$$

The corresponding ϕ_{kl} and ψ_{kl} maps are

$$\phi_{kl} = \begin{pmatrix} 0.5 & 1 & 1 & 1 & 0 \\ 0 & 0.5 & 0 & 1 & 0 \\ 0 & 1 & 0.5 & 1 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 1 & 1 & 1 & 1 & 0.5 \end{pmatrix} \qquad \text{and} \qquad \psi_{kl} = \begin{pmatrix} 0.5 & 1 & 1 & 0 & 1 \\ 0 & 0.5 & 1 & 0 & 1 \\ 0 & 0 & 0.5 & 0 & 1 \\ 1 & 1 & 1 & 0.5 & 1 \\ 0 & 0 & 0 & 0 & 0.5 \end{pmatrix} ,$$

and the Kemeny-Snell distance is

$$d_{KS} = \sum_{k=1}^{5} \sum_{l=1}^{5} |\psi_{kl} - \phi_{kl}| = 16 .$$

The Kemeny-Snell distance measure has proven efficient and useful for content-based image retrieval applications (cf. Ex. 14.1).

14.4 Local Similarity Measures

In this section we consider the family of local similarity measures. By definition, these measures require the spatial alignment of the two input images A and B. In Ex 14.3 we explained how we may convert a global similarity measure to a local similarity measure. We may use the sliding window procedure to generate local mse, mae, correlation coefficient, mutual information and d_{KS} similarity measures. In general the windows required for these similarity measures should be at least 20×20.

We now consider the Bhat-Nayer distance measure which is, by definition, a local ordinal similarity measure. It therefore does not require such a large window size: windows of 3×3 to 13×13 are common.

14.4.1 Bhat-Nayar Distance Measure

Let $W(m,n)$ denote a $L \times L$ window centered at the pixel (m,n). If $\widetilde{A}$ and $\widetilde{B}$ denote, respectively, the image pixels which lie in $W(m,n)$ in images A and B. Then the Bhat-Nayar (BN) distance measure computes the similarity of the two windows $\widetilde{A}$ and $\widetilde{B}$ by comparing the rank permutations of their pixel gray-levels as follows.

Given the two windows $\widetilde{A}$ and $\widetilde{B}$, we rewrite them as image vectors $\widetilde{\mathbf{a}} = (\widetilde{a}(1), \widetilde{a}(2), \ldots, \widetilde{a}(K))^T$ and $\widetilde{\mathbf{b}} = (\widetilde{b}(1), \widetilde{b}(2), \ldots, \widetilde{b}(K))^T$, where $K = L^2$. The corresponding rank vectors are:

$$\widetilde{\mathbf{r}}_A = (\widetilde{r}_A(1), \widetilde{r}_A(2), \ldots, \widetilde{r}_A(K))^T \qquad \text{and} \qquad \widetilde{\mathbf{r}}_B = (\widetilde{r}_B(1), \widetilde{r}_B(2), \ldots, \widetilde{r}_B(K))^T \ .$$

Let $k = \widetilde{r}_A^{-1}(h)$ if $h = \widetilde{r}_A(k)$. Then we may define a composite rank vector $\widetilde{\mathbf{s}}$ as:

$$\widetilde{\mathbf{s}} = (\widetilde{s}(1), \widetilde{s}(2), \ldots, \widetilde{s}(K))^T \ ,$$

where

$$\widetilde{s}(k) = \widetilde{r}_B(h) = \widetilde{r}_B\left(\widetilde{r}_A^{-1}(k)\right) \ .$$

Informally, $\widetilde{\mathbf{s}}$ is the ranking of B with respect to the ranks of A.

The BN distance measure is then

$$\lambda_{BN} = 1 - 2\frac{\max_k(d_k)}{\lfloor K/2 \rfloor} \ ,$$

where

$$d_k = k - \sum_{h=1}^{k} J(\widetilde{s}(h), k) \ ,$$

$$J(a,b) = \begin{cases} 1 \text{ if } a \le b \ , \\ 0 \text{ otherwise} \ . \end{cases}$$

A modified BN distance measure due to Scherer, Werth and Pinz (SWP) [22] is

$$\lambda_{SWP} = 1 - \left(\frac{\max_k(d_k)}{\lfloor K/2 \rfloor} + \frac{\sum_k d_k}{\lfloor K^2/4 \rfloor} \right) .$$

The following example illustrates the calculation of the BN and the modified BN distance measures.

Example 14.11. Bhat-Nayar Distance Measure [5, 22]. Given two one-dimensional input images

$$A = (10,20,30,50,40,70,60,90,80)^T ,$$
$$B = (90,60,70,50,40,80,10,30,20)^T ,$$

the corresponding rank vectors are

$$\mathbf{r}_A = (1,2,3,5,4,7,6,9,8)^T , \qquad \mathbf{r}_B = (9,6,7,5,4,8,1,3,2)^T .$$

The composition permutation vector $\widetilde{\mathbf{s}}$ is

$$\widetilde{\mathbf{s}} = (9,6,7,4,5,1,8,2,3)^T ,$$

and the corresponding distance vector is

$$\mathbf{d} = (1,2,3,3,3,2,2,1,0)^T .$$

The BN and the modified BN distance measures are:

$$\lambda_{BN} = 1 - 2\frac{\max_k(d_k)}{\lfloor K/2 \rfloor} = 1 - 2 \times 3/4 = -0.5 ,$$
$$\lambda_{SWP} = 1 - \left(\frac{\max_k(d_k)}{\lfloor K/2 \rfloor} + \frac{\sum_k d_k}{\lfloor K^2/4 \rfloor} \right) = 1 - (3/4 + 17/20) = -0.60 .$$

The following matlab code may be used to calculate λ_{BN} and λ_{SWP}.

Example 14.12. Matlab Code for λ_{BN} and λ_{SWP}. Let A and B be two input vectors containing M gray-levels $A_k, k \in \{1,2,\ldots,K\}$, and $B_k, k \in \{1,2,\ldots,K\}$.

```
[junk, invrA] = sort(A); [junk, rA] = sort(invrA);
[junk, invrB] = sort(B); [junk, rB] = sort(invrB);
s = rB(invrA);
S = ones(K,1) * s(:)';
G = (1:K)' * ones(1,K);
```

$d = (1:K)' - sum(tril(S < G), 2)$;
$\lambda_{BN} = 1 - 2 * max(d)/floor(K/2)$;
$\lambda_{SWP} = 1 - (\max(d)/floor(K/2) + \sum(d)/floor(K^2/4))$;

14.4.2 Mittal-Ramesh Ordinal Measure

Although the ordinal similarity measures are robust to monotonic changes in intensity, they are not very robust to Gaussian noise. Even a small amount of Gaussian noise can completely change the rankings between pixels that are not far from each other in gray-level. Such a drawback occurs because the ordinal similarity measure do not take into account the pixel gray-levels at all. In the Mittal-Ramesh ordinal measure we take into account the pixel gray-levels. The similarity measure has a very good performance but is computationally very expensive.

14.5 Binary Image Similarity Measure

Special similarity measures are used for binary images. Given two binary images A and B we may define a local distance measure [4] as follows: Let d_A and d_B be the corresponding distance transform images [8]:

$$d_A(m,n) = \min_{(u,v)} \widetilde{A}(u,v)\sqrt{(m-u)^2+(n-v)^2} ,$$
$$d_B(m,n) = \min_{(u,v)} \widetilde{B}(u,v)\sqrt{(m-u)^2+(n-v)^2} ,$$

where

$$\widetilde{A}(u,v) = \begin{cases} 1 & \text{if } A(u,v) = 1 , \\ \infty & \text{otherwise} , \end{cases} \quad \text{and} \quad \widetilde{B}(u,v) = \begin{cases} 1 & \text{if } B(u,v) = 1 , \\ \infty & \text{otherwise} . \end{cases}$$

Then, the local distance measure is defined as:

$$L(m,n) = |A(m,n) - B(m,n)| \max(d_A(m,n), d_B(m,n)) .$$

The following example illustrates the calculation of the local distance measure

Example 14.13. Local distance map. Given two binary images

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} ,$$

the corresponding distance transforms are

$$d_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad d_B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & \sqrt{2} \\ 0 & 1 & \sqrt{5} \end{pmatrix} .$$

and the local similarity measure distance map is

$$L = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & \sqrt{2} \\ 0 & 0 & \sqrt{5} \end{pmatrix} . \qquad (14.5)$$

We now describe how we may convert the local binary distance measure map $L(m,n)$ into global Hausdorff measures.

14.5.1 Hausdorff Metric

The Hausdorff distance [9] between two $M \times N$ binary images A and B is defined as

$$H(A,B) = \max_{(m,n)} (L(m,n)) , \qquad (14.6)$$

where $L(m,n)$ is the local distance measure defined in (14.5). According to (14.6), $H(A,B)$ is the maximum distance from a point in one image to the nearest point in the second image. It is therefore very sensitive to noise and for this reason we often use robust variants of the Hausdorff distance where we replace the maximum operator in (14.6) by robust alternatives. Some examples are:

Partial Hausdorff distance

$$H_k(A,B) = L_{(k)} ,$$

where $L_{(k)} = L(m,n)$ if $L(m,n)$ is the kth largest local distance value.

Mean Hausdorff distance

$$H_{AVE}(A,B) = \sum_{(m,n)} L(m,n)/(MN) .$$

Median Hausdorff distance

$$H_{MED}(A,B) = \text{med}_{(m,n)} (L(m,n)) .$$

The following example illustrates the calculation of the original Hausdorff distance and its variants for the two binary images A and B.

Example 14.14. Hausdorff distances. Given two binary images

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix},$$

the corresponding local distance map $L(m,n)$ (see Ex. 14.13) is

$$L = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & \sqrt{2} \\ 0 & 0 & \sqrt{5} \end{pmatrix}.$$

The original, partial, mean and median Hausdorff distance measures are, respectively, $H(A,B) = \sqrt{5}$, $H_k(A,B) = \sqrt{2}$, $H_{AVE}(A,B) = (3+\sqrt{2}+\sqrt{5})/9 \approx 0.75$, and $H_{MED}(A,B) = 1$, where $H_k(A,B)$ was calculated assuming $k = 8$.

14.6 Software

END-L1. A fast matlab routine for the earth movers distance assuming an L^1 metric. Authors: Haibin Ling and Kazunori Okada [14].

COMP STATS TOOLBOX. A computational statistics toolbox. Authors Wendy Martinez and Angel Martinez [16].

14.7 Further Reading

In this chapter we have concentrated on similarity measures which are fixed in the sense that they are not learnt from training data. Recently the training of such measures has received increasing interest. A comprehensive survey of the subject is given in [24].

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Patt. Anal. Mach. Intell. 28, 2037–2041 (2006)
2. Arya, K.V., Gupta, P., Kalra, P.K., Mitra, P.: Image registration using robust *M*-estimators. Patt. Recogn. Lett. 28, 1957–1968 (2007)
3. Baggerly, K.A.: Probability binning and testing agreement between multivariate immunofluorescence histogram extending the chi-squared test. Cytometry 45, 141–150 (2001)
4. Baudrier, E., Millon, G., Nicolier, F., Ruan, S.: The adaptive local Hausdorff-distance map as a new dissimilarity measure. Patt. Recogn. 41, 1461–1478 (2008)

5. Bhat, D.N., Nayar, S.K.: Ordinal measures for image correspondence. IEEE Trans. Patt. Anal. Mach. Intell. 20, 415–423 (1998)
6. Birge, L., Rozenholc, Y.: How many bins should be put in a regular histogram. Euro. Series in Appl. and Indust. Math. Prob. Stat. 10, 24–45 (2006)
7. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: ideas, influences and trends of the new age. ACM Comp. Surveys 40, 1–60 (2008)
8. Fabbri, R., Da, L., Costa, F., Torelli, J.C., Bruno, O.M.: 2D Euclidean distance transform algorithms: A comparative survey. ACM Comp. Surveys 40 (2008)
9. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance. IEEE Trans. Patt. Anal. Mach. Intell. 15, 850–863 (1993)
10. Kendall, M.: Rank Correlation Methods. Griffin (1962)
11. Kumar, S.S., Chatterji, B.N.: Robust similarity measures for stereo correspondence IE(I). Journal CP 85(2), 44–51 (2004)
12. Levina, E., Bickel, P.: The earth mover's distance is the Mallow's distance: some insights from statistics. In: Proc. IEEE Int. Conf. Comp. Vis., pp. 251–256 (2001)
13. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: state of the art and challenges. ACM Trans. Multimedia Comp. Comm. Appl. 2, 1–19 (2006)
14. Ling, H., Okada, K.: An efficient earth mover's distance algorithm for robust histogram comparison. IEEE Trans. Patt. Anal. Mach. Intell. 29, 840–853 (2007)
15. Luo, J., Etz, S.P., Gray, R.T.: Normalized Kemeny and Snell distance: a novel metric fo quantitative evaluation of rank-order similarity of images. IEEE Trans. Patt. Anal. Mach. Intell. 24, 1147–1151 (2002)
16. Martinez, W., Martinez, A.: Computational Statistics Handbook with Matlab, 2nd edn. CRC Press, Boca Raton (2008)
17. Mignotte, M.: Segmentation by fusion of histogram-based K-means clusters in different color spaces. IEEE Trans. Image Process. 17, 780–787 (2008)
18. Rabin, J., Delon, J., Gousseau, Y.: Circular Earth Mover's Distance for the comparison of local features. In: Proc. Int. Conf. Patt. Recogn. (2008)
19. Roederer, M., Treister, A., Moore, W., Herzenberg, L.A.: Probability binning comparison: a metric for quantitating univariate distribution differences. Cytometry 45, 37–46 (2001)
20. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comp. Vis. 40, 99–121 (2000)
21. Santini, S., Jain, R.: Similarity Measures. IEEE Trans. Patt. Anal. Mach. Intell. 21, 871–883 (1999)
22. Scherer, S., Werth, P., Pinz, A.: The discriminatory power of ordinal measures - towards a new coefficient. In: Proc. IEEE Conf. Comp. Vis. Patt. Recogn. (1999)
23. Serratosa, F., Sanroma, G.: A fast approximation of the earth-movers distance between multidimensional histograms. Int. J. Patt. Recogn. Art. Intell. 22, 1539–1558 (2008)
24. Yang, L.: Distance metric learning: A comprehensive survey. Department of Computer Science and Engineering. Michigan State University, May 19 (2006)
25. Yu, J., Amores, J., Sebe, N., Tian, Q.: Ranking metrics and evaluation measures. Adv. Imaging Electron Phys. 144, 291–316 (2006)
26. Yu, J., Amores, J., Sebe, N., Redeva, P., Tian, Q.: Distance learning for similarity estimation. IEEE Trans. Patt. Anal. Mach. Intell. 30, 451–462 (2008)