MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Sensitivity Characteristics of Cross-Correlation Distance Metric and Model Function

Porikli, F.

TR2003-146 March 25, 2004

Abstract

We present a 3-fold distance metric and a transfer function to evaluate the similarity of two finite length sequences. We analyze the sensitivity characteristics of the proposed metrics for Gaussian shape functions. Our method is based on cross-correlation matrix analysis and extrapolation of a minimum cost path using dynamic programming. Unlike the existing sequential (bin-by-bin) and non-sequential (cross-bin) approaches that compute a single scalar as a result of the measurement, we calculate the distance as well as determine how two sequences are correlated with each other in terms of a non-parametric transfer function. We shown that the proposed metrics provide better discrimination than conventional metrics do. Furthermore, we show that we can reduce our metric to any one of sequential metrics with suitable simplification.

Conference on Information Sciences and Systems (CISS) 2004

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Sensitivity Characteristics of Cross-Correlation Distance Metric and Model Function

F. Porikli
Mitsubishi Electric Research
Laboratories
558 Central Avenue
Murray Hill, New Jersey 07974
e-mail: fatih@merl.edu

Abstract —

We present a 3-fold distance metric and a transfer function to evaluate the similarity of two finite length sequences. We analyze the sensitivity characteristics of the proposed metrics for Gaussian shape functions. Our method is based on cross-correlation matrix analysis and extrapolation of a minimum cost path using dynamic programming. Unlike the existing sequential (bin-by-bin) and non-sequential (cross-bin) approaches that compute a single scalar as a result of the measurement, we calculate the distance as well as determine how two sequences are correlated with each other in terms of a non-parametric transfer function. We shown that the proposed metrics provide better discrimination than conventional metrics do. Furthermore, we show that we can reduce our metric to any one of sequential metrics with suitable simplification.

I. Introduction

Distance between two sequences is one of the most common measures used in computer algorithms for sequence analysis. From image retrieval in multimedia databases to comparison of amino acid sequences for DNA pattern recognition, it is used in various areas. However, past studies have shown that most distance metrics are neither robust to small shape deformations of sequences nor nonlinear shifts on the indexing axis. Furthermore, there is no metric that can compute the distance and evaluate the alignment of sequences in terms a transfer function at the same time.

A finite-length sequence, h, is a vector $[h[0], \ldots, h[M]]$ in which each bin $h[m]$ is the value of the vector at the index number m. In case h represent an image histogram, $h[m]$ contains the number of pixels corresponding to the color range of m in the image $\mathcal{I}$ where M is the total number of the bins. In order words, it is a mapping from the set of color vectors to the set of positive real numbers $\mathcal{R}^+$. In this paper, we assume that bins are identical i.e. sampling frequency of the indexing axis is constant $m_i - m_{i-1} = m_j - m_{j-1}$, and the sequences are normalized such that $\sum_{m=0}^{M} h[m] = 1$.

II. Cross-Correlation Distance (CCD)

We define a cross-correlation matrix C between two sequences as the set of positive real numbers that represent the bin-wise distances. Let $h_1[m]$ and $h_2[m]$ be two sequences with $m = 1, \ldots, M$ and $m = 1, \ldots, N$ i.e. the lengths are not necessarily same. The cross-correlation matrix is

$$C_{M\times N} = h_1 \otimes h_2 = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1N} \\ c_{21} & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ c_{M1} & & \ldots & c_{MN} \end{bmatrix} \quad (1)$$

where each element is a positive real number, and

$$c_{mn} = d(h_1[m], h_2[n]) \quad (2)$$

where $d(\cdot) \geq 0$ is a distance norm which satisfies the triangle-inequality. As a matter of fact, this definition stands for the dissimilarity of sequences instead of their correlation. The correlation can be easily established by defining $c_{mn} = 1 - d(h_1[m] - h_2[n])$.

Theorem 1 *The sum of the diagonal elements of C represents the bin-by-bin distance with given norm $d(\cdot)$ if the sequences have equal number of bins $M = N$.*

For example, by choosing the distance norm as L_1, the sum of the diagonals becomes the magnitude distance between a pair of sequences

$$\sum_{m}^{M} c_{mm} = \sum_{m}^{M} |h_1[m] - h_2[m]| = d_{L1}(h_1, h_2). \quad (3)$$

Let $p : \{(m_0, n_0), ..., (m_i, n_i), ..., (m_I, n_I)\}$ represents a minimum cost path (defined in the next section) from the c_{11} to c_{MN} in the matrix C, i.e. the sum of the matrix elements on the connected path p gives the minimum score among all possible routes. The total length of the path cannot be more than the sum of the lengths of the sequences

$$\sqrt{M^2 + N^2} \leq I \leq M + N \quad (4)$$

We define a cost function for the path as $g(p_i) = c_{m_i, n_i}$ where p_i denotes the path element (m_i, n_i). We define a mapping $i \rightarrow j$ from the path indices to the projection onto the diagonal of the matrix C, and an associated transfer function $f(j)$ that gives the distance from the diagonal with respect to the projection j. The transfer function is a mapping from the matrix indices to real numbers

$$(m_i, n_i) \quad \xrightarrow{t} \quad f(j) \quad (5)$$

where $j = 1, \ldots, J$ and $J = \sqrt{M^2 + N^2}$. Depending on the shape of the path, these mappings may not be one-to-one.

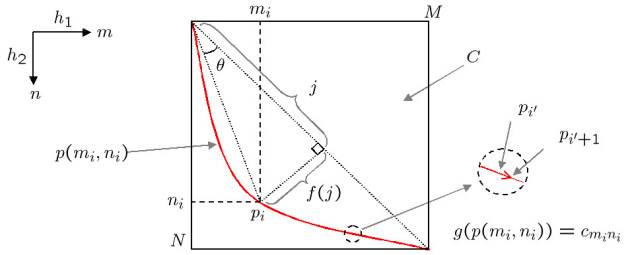

Figure 1: The figure shows the relation between the minimum cost path and $f(j)$.

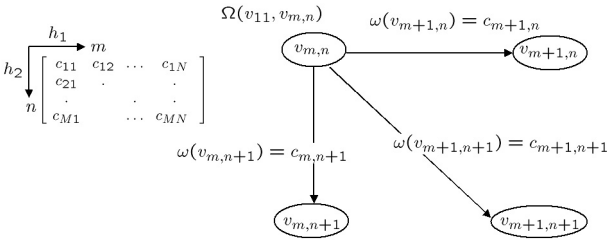

Figure 2: Each vertex represents a matrix index combination and each edge is the corresponding matrix element for that index.

From Fig.1, the angle between the diagonal and the current path index is

$$\theta = \tan^{-1}\left(\frac{M}{N}\right) - \tan^{-1}\left(\frac{m_i}{n_i}\right) \quad (6)$$

Without loss of generality, we may assume $M = N$, i.e. $\tan^{-1}(\frac{M}{N}) = \frac{\pi}{4}$. Then, the magnitude of the projection j is

$$\begin{aligned} j &= |p_i| \cdot \cos\theta & (7)\\ &= \sqrt{m_i^2 + n_i^2}\cos\left(\frac{\pi}{4} - \arctan\left(\frac{m_i}{n_i}\right)\right) & (8)\\ &= \frac{m_i + n_i}{\sqrt{2}} & (9) \end{aligned}$$

Thus, the transfer function $f(j)$ becomes

$$\begin{aligned} f(j)^2 &= -j^2 + (m_i^2 + n_i^2) & (10)\\ &= \frac{1}{2}\left(m_i^2 + n_i^2\right) + m_i n_i & (11) \end{aligned}$$

The $f(j)$ is negative if $m_i < n_i$. The mapping t in equation 5 is decomposed into two functions $t_m(m_i) = n_i$ and $t_n(n_i) = m_i$ such that they give the minimum cost path as a function of sequence index.

Their derivatives with respect to both indices represent the amount of warping of the bins

$$\begin{aligned} \partial t_m(m_i) &= t_m(m_i) - t_m(m_i - 1) & (12)\\ \partial t_n(n_i) &= t_n(n_i) - t_n(n_i - 1) & (13) \end{aligned}$$

It is straightforward to derive the following properties

$$\begin{aligned} f(j) = 0 &\Rightarrow m_i = n_i & (14)\\ f(j) > 0 &\Rightarrow m_i > n_i & (15)\\ f(j) < 0 &\Rightarrow m_i < n_i & (16)\\ \partial f(j) = 0 &\Rightarrow \partial t_m(m_i) = \partial t_n(n_i) & (17)\\ \partial f(j) > 0 &\Rightarrow \partial t_m(m_i) < \partial t_n(n_i) & (18)\\ \partial f(j) < 0 &\Rightarrow \partial t_m(m_i) > \partial t_n(n_i) & (19) \end{aligned}$$

where the derivative of $f(j)$ with respect to j is limited in range $-\frac{\pi}{2} \le \partial f(j) \le \frac{\pi}{2}$

Definition The cross-correlation distance (CCD) is the total cost along the transfer function (CCF)

$$d_{CC}(h_1, h_2) = \sum_{i=0}^{I} |g(m_i, n_i)| \quad (20)$$

An alternative definition of the above distance metric weights the transfer function with the current cost

$$d_{CC}(h_1, h_2) = \sum_{j=0}^{J} |f(j)|g((m_i, n_i)) \quad (21)$$

The distance can be measured as the length of the minimum cost path as well

$$d_{CC}(h_1, h_2) = J. \quad (22)$$

III. Dynamic Programming

Dynamic programming is an approach developed to solve sequential, or multi-stage, decision problems [4]. Basically, what dynamic programming approach does is that it solves a multi-variable problem by solving a series of single variable problems. The essence of dynamic programming is Richard Bellman's Principle of Optimality. This principle is intuitive: from any point on an optimal trajectory, the remaining trajectory is optimal for the corresponding problem initiated at that point.

Given two sequences, the question is what is the best alignment of their shapes and how can the alignment be determined? We reduce the comparison of two sequences to finding the minimum cost path in a directed weighted graph. A minimum cost path from a vertex to another vertex in a directed graph is a path that has the smallest total edge-weights among all paths from the same source vertex to the same destination vertex. Let v be a vertex and e be an edge between the vertices of a directed weighted graph. We associate a cost to each edge $\omega(e)$. We want to find the minimum cost path by moving from an origin vertex v_0 to a destination vertex v_S. The cost of a path $p(v_0, v_S) = \{v_0, .., v_S\}$ is the sum of its constituent edges

$$\Omega(p(v_0, v_S)) = \sum_{s}^{S} \omega(v_s) \quad (23)$$

Suppose we already know the costs $\Omega(v_0, v_*)$ from v_0 to every other vertex. Let's say v_* is the last vertex the path goes through before v_S. Then, the overall path must be formed by concatenating a path from v_0 to v_*, i.e. $p(v_0, v_*)$, with the edge $e(v_*, v_S)$. Further, the path $p(v_0, v_*)$ must itself be a minimum cost path since otherwise concatenating the minimum cost path with edge $e(v_*, v_S)$ would decrease the cost of the overall path. Another observation is that $\Omega(v_0, v_*)$ must be equal or less than $\Omega(v_0, v_S)$, since $\Omega(v_0, v_S) = \Omega(v_0, v_*) + \omega(v_*, v_S)$ and we are assuming all edges

have non-negative costs, i.e. $\omega(v_*, v_S) \geq 0$. Therefore if we only know the correct value of $\Omega(v_0, v_*)$ we can find a minimum cost path.

We modified Dijkstra's algorithm is modified to find the shortest paths between one source vertex and all the other vertices which are the destinations. To find all minimum cost paths between all pairs of vertices we need to apply it to each of the vertices as a source vertex. Let Q be the set of active vertices whose minimum cost paths from v_0 have already been determined, and $\vec{p}(v)$ is a back pointer vector that shows the neighboring minimum cost vertex of v. Then the iterative procedure is given as

1. Set $u_0 = v_0$ $Q = \{u_0\}$, $\Omega(u_0) = 0$, $\vec{p}(v_0) = v_0$, and $\omega(v) = \infty$ for $v \neq u_0$.
2. For each $u_i \in Q$: if v is a connected to u_i, assign $\omega(v) \leftarrow \min\{\omega(u_i), \Omega(u_i) + \omega(v)\}$. If $\omega(v)$ is changed, assign $\vec{p}(v) = u_i$ and update $Q \leftarrow Q \cup v$.
3. Remove u_i from Q.
4. If $Q \neq \emptyset$ go to step 2.

Then the minimum cost path $p(v_0, v_s) = \{v_0, ..., v_S\}$ is obtained by tracing back pointers by starting from the destination vertex v_S as $v_{s-1} = \vec{p}(v_s)$. The algorithm takes time $O(S^2)$. As shown in figure 2, the graph that is converted from the cross-correlation matrix is directed such that a vertex v_{mn} has directional edges to vertices $v_{m+1,n}, v_{m,n+1}, v_{m+1,n+1}$ only. Therefore, we do not allow overlaps of the bin indices, and eliminate cyclic paths.

The dynamic programming can be applied to obtain the partial matches between two sequences. To find the best match for the part $[m_a, ..., m_b]$ of the first sequence in the second sequence, we modify the initial conditions such that the initial vertex is iteratively assigned to (m_a, n_1), where $n_1 = 1, .., N$, and the target vertex is chosen as (m_b, n_2) where $n_2 = \frac{N}{M} m_a, .., N$. The above process is repeated for every combination and the minimum cost path is chosen.

IV. Case Study: Illumination Compensation

We tested the proposed transfer function to recover distorted color histograms. The intensity values of an input image Fig.3-a is distorted non-linearly by hand to obtain its over-exposed version Fig.3-b. We extracted histograms (Fig.3-c, upper graphs) of the input and over-exposed images. We computed the cross-correlation matrix using these histograms. As the distance kernel, we used the L_2 norm. Then, we found the minimum cost path within the cross-correlation matrix (Fig.3-d) by starting from the lower-right end of the matrix and tracking towards to upper-left corner as explained in the dynamic programming section. We remapped the intensity values of the over-exposed image using transfer function that is obtained by projecting the minimum cost path (Fig.3-c, lower graph) on the main diagonal as explained in the cross-correlation section. Note that, the distortion is not linear, and it is not parametric either. The result of the compensation is given in Fig.3-e. We observed that the transfer function matches with the non-linear distortion characteristics. As visible in the histogram graph (Fig.3-c), the nonparametric transfer function successfully compensated for the non-linear distortions by taking the advantage of the non-parametric transfer function.

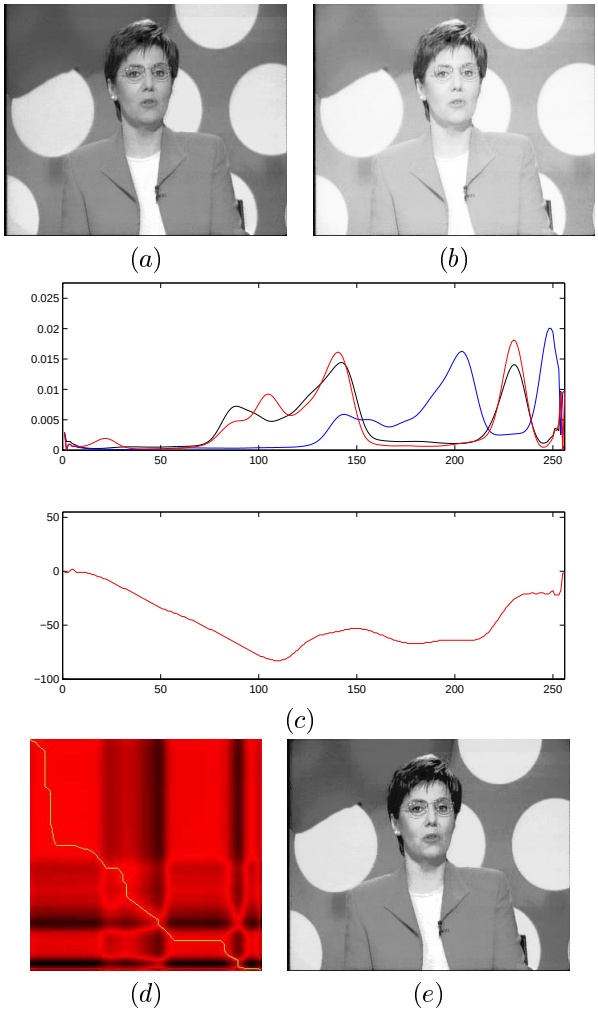

Figure 3: (a) A sample image, (b) its over-exposed version. (c) The upper graph shows the histograms of the input image (black), over-exposed image (blue), and the compensated image (red). The lower graph is the transfer function. (d) The cross-correlation matrix and the minimum cost path (yellow). Higher red values indicate smaller distances. (e) The compensated image.

We repeated the same analysis using several other color/gray-level images. We observed that the corrected images visually are much similar to the originals after the compensation. Their color histograms are accurately aligned as well. The improvement is substantial even though histogram operations are invariant to spatial transformations, and thus have only limited impact. We computed the distance of two histograms such that the distance score allows the amount of non-linear, non-parametric but sequential alignment of the two histograms. Note that, no other distance metric can give such a compensated distance.

V. Sensitivity Analysis for Gaussian Functions

We analyzed the sensitivity characteristics of the proposed metric for Gaussian shaped sequences. We generated a reference Gaussian sequence with zero mean and unit variance $\mathcal{N}(0, 1)$, and compared it with a set (Set-1) of Gaussians se-

quences $\mathcal{N}(k,1)$ where $k : 1,..,10$, i.e. their variances are same but the means are different, as plotted in Fig.5-a. We also tested another set (Set-2) of zero mean Gaussians with different variances $\mathcal{N}(0,k)$, $k : 1,..,10$ (Fig.5-b). We computed distances between the original Gaussian and the Set-1 for the metrics that are described in the Appendix and also the corresponding CDD distances using both the total cost and the total length norms as defined in equations 20 and 22. We presented these results in Fig. 5-c. As visible, the total cost norm is shift invariant. Then, we computed distances for the Set-2. The corresponding graph is given in Fig.5-d.

We observed that the Kolmogorov-Smirnov, Lorentzian and Intersection distances have almost identical responses, and the Bhattacharyya and Kullback-Leibler distances have similar results for the Set-1. For same-mean shifted-variance case (Set-2), the Lorentzian and magnitude distances have similar responses.

As visible, one norm of the CCD metric (total cost) can identify the same shape sequences while another norm (total length) can effectively detect the mean differences for the Gaussian shape functions. An ideal metric is supposed to have linear response for linearly varying means and variances of the input sequences in our case. Most of the above metrics satisfies this constraint. The graphs obtained using the CCD show that it is linearly proportional to the linear changes of the input sequence. The graphs justifies that the proposed metrics are sensitive to the changes of the mean and variance values for Gaussian shape functions.

We also evaluated each distance metric described in the Appendix and our CCD metrics for varying mean and variance values as given in Fig.4. We observed that the CCD metrics are very sensitive to the shape changes of the input sequences. Even the mean value of the sequences are diverges, our metrics accurately identify variance deviations. On the other hand, most other conventional metrics lose their sensitivity and become inversely proportional to the variance changes in case of severe mean shifts.

VI. Discussion on Distance Measures

A major drawback of the bin-by-bin distance measures (Minkowski, Intersection, Lorentzian, Chi, Bhattacharyya, etc.) is that they account only for the correspondence between bins with the same index, and do not use information across bins. A shift of the bin index may result in larger distances although the two sequences otherwise have the same values. For image histograms, quantization is yet another consideration; a slight change in lighting conditions may result in a corresponding shift in the color sequence, causing these metrics to misjudge similarity completely. Contribution of the empty bins is also important. Weighted versions of the Minkowski metric may underestimate distances because they tend to accentuate the similarity between color sequences presenting many nonempty bins. Furthermore, not all sequences have the same number of bins. Yet, the bin size may not be identical within the same sequence either. The bin-by-bin measures do not allow matching different size sequences, while the CCD does.

The Hausdorff distance provides the best mechanism to handle partial matches, as well as the sequence intersection, the quadratic distance, the EMD, and the CCD. Since the K-L divergence evaluates only the relative distance between the given sequences by using one of them as a reference, it is not symmetric, thus it does not qualify as a metric. For most of

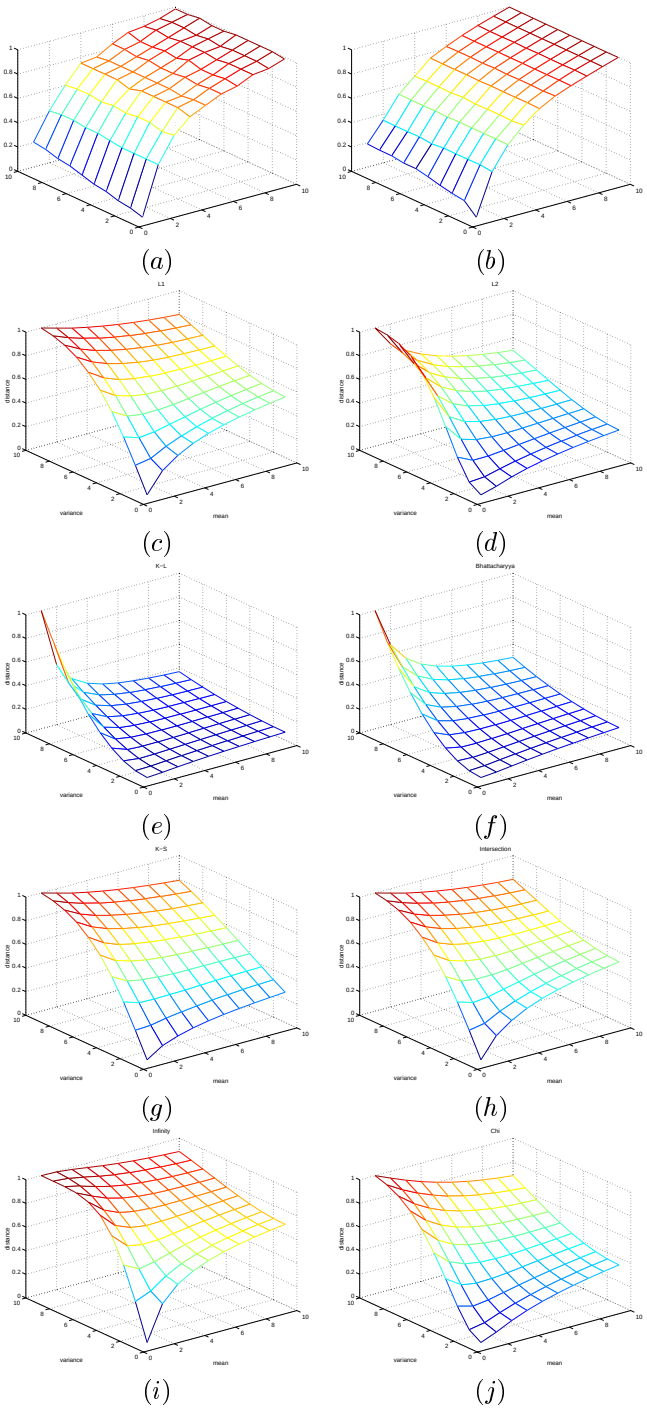

Figure 4: (a) The CCD, which is computed using the total length definition, between $\mathcal{N}(0,1)$ and $\mathcal{N}(\mu,\sigma^2)$'s where $\mu = 0..10, \sigma^2 = 1..10$, and (b) the CCD distances that are computed using total cost definition. (c) The magnitude distances, (d) the Euclidean distances, (e) the Kullback-Leibler distances, (f) the Bhattacharyya distances (g) the Kolmogorov-Smirnov distances, (h) the intersection distances, (i) the Minkowski distances for L_∞, and (j) the χ^2 distances. Except the CCD, most metrics lose sensitivity and become inversely proportional to the variance changes in case of severe mean shifts.

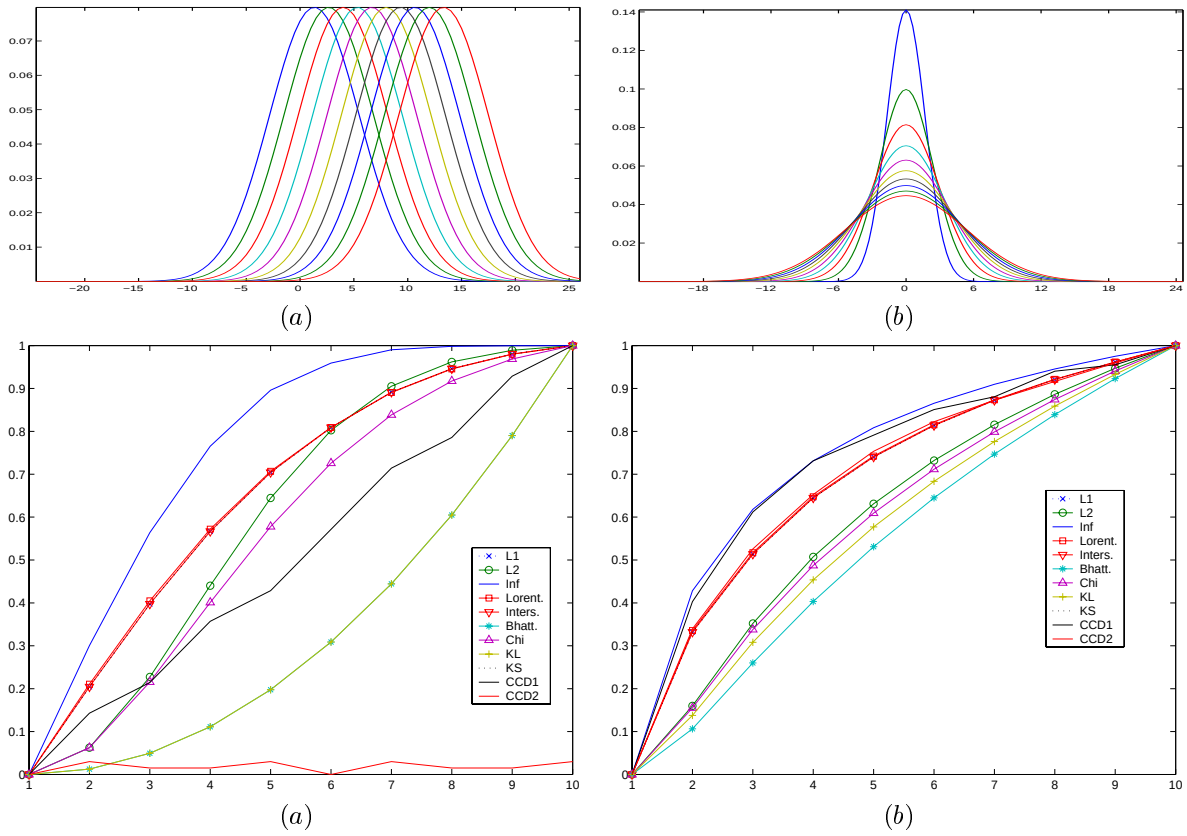

Figure 5: (a) Set of Gaussian shape functions (Set-1) with different mean values, (b) with different variances (Set-2). (c) The graphs of the normalized distances between the original Gaussian $\mathcal{N}(0,1)$ and other Gaussian sequences in Set-1. The cross-correlation distances are computed by Eq.20 (blue) and Eq.22 (red). The horizontal axis is the mean value. (d) The graph of the normalized distances computed for Set-2. The horizontal axis is the variance.

the measures, the triangle inequality, which is important for image retrieval, holds only for specific cases. Only the CCD has the ability to find a non-linear, non-parameterized model of the color warping between the sequences. This property is especially important in prediction of lighting changes.

The quadratic distance requires an ambiguous covariance matrix that states the perceptive relation between the color bins. The choice of the covariance metric effects the qualification of the quadratic distance as a metric. The Hausdorff distance does not qualify as a metric, and it overestimates the similarity of two sequences if there is a partial match.

Not all measures can be extended to the multi-dimensional sequences, e.g. the Kolmogrov-Smirnov statistic. Computational complexity of the cross-bin measures are higher than the bin-by-bin measures. In cases of computing the distance where the number of bins is large, or sequences are multi-dimensional, the EMD, the Hausdorff distance, and the quadratic form become infeasible. Although cross-bin matching is possible for EMD, the Hausdorff, and the quadratic, these methods do not have any mechanism to preserve the ordering of the color bins. Obviously, changing the order of the color bins may significantly deteriorate the accuracy of the image distance since a sequence is already a marginal. The CCD, on the other hand, preserves the order of bins while matching. None of the distance measures has the ability to recover a mapping function that transfers one sequence to other except the CCD.

VII. Conclusion

In this contribution, we investigated the sensitivity properties of our cross-correlation matrix and dynamic programming based distance metric for Gaussian shape functions. We showed that the proposed metric is sensitive to the mean and variance variations of the input sequences. Our metric evaluates the similarity of two finite length sequences and determines a non-parametric transfer function that accurately aligns the input sequences. The distance may be computed using one of the three proposed definitions which are the total cost on the minimum cost path, the length of the minimum cost path, and the total area under the transfer function. The total cost norm is invariant to mean changes, and it measures the shape divergence of the sequences. The length and total area norms react the shape mismatches as well as they detect the mean changes. The transfer function compensates for the non-linear warping of the sequences. This is an additional functionality which is crucial in histogram matching applications. Our metrics also provide better discrimination than conventional metrics, and allow the matches of the empty bins which can not be done by the most other bin-by-bin metrics. Furthermore, our metric can reduce to any one of the bin-by-bin metrics by suitable simplification.

References

[1] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions", Bull. Calcutta Math. Soc., Vol. 35, 99109, 1943

[2] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection and robust statistics with application in early vision", Journal Computer Vision, Vol 19, 57-91, 1996.

[3] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack. "Efficient color histogram indexing for quadratic form distance functions," *IEEE. Transactions on Pattern Analysis and Machine Intelligence*, July 1995.

[4] R. Keeney and H. Raiffa, "Decisions with multiple objectives", Wiley, 1976.

[5] S. Kullback, "Information theory and statistics", Dover, 1968.

[6] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. "Empirical evaluation of dissimilarity measures for color and texture". Computer Vision and Image Understanding, Vol 84, 25-43, 2001.

[7] M. Swain and D. Ballard, "Color Indexing", International Journal of Computer Vision, Vol. 7, pp.1132, 1991.

Appendix

The Minkowski distance [7] is a generalized form of common spatial distance norms such as the magnitude L_1, the Euclidean L_2, and the maximum L_∞. It is defined as

$$d_{Lp}(h_1,h_2) = \left(\sum_{m=0}^{M} |h1[m]-h_2[m]|^p\right)^{1/p} \quad (24)$$

The higher order norms ($p>1$) exponentially weight the absolute difference, thus they are more sensitive to the mismatches.

The Lorentzian distance [2] is frequently used in robust estimators to minimize the effect of the outliers. It is defined as

$$d_R(h_1,h_2) = \sum_{m=0}^{M} \log\left(1+|h_1[m]-h_2[m]|\right) \quad (25)$$

Usually, a scaling factor is used to normalize the absolute difference term.

The sequence intersection is defined by the area of the overlap between two sequences

$$d_\cap(h_1,h_2) = 1 - \frac{\sum_{m=0}^{M}\min\left(h_1[m],h_2[m]\right)}{\min\left(\sum_{m=0}^{M}h_1[m],\sum_{n=0}^{M}h_2[n]\right)} \quad (26)$$

The Bhattacharyya distance [1] is a separability measure between two Gaussian distributions. We adapted it for sequence comparison as

$$d_B(h_1,h2) = -\ln\sum\sqrt{h_1(m)h_2(m)} \quad (27)$$

The χ^2 distance weights inversely the squared differences between color bins by the expected frequency, and tends to equalize the contributions of rare and frequent color values to the metric structure of the space

$$d_\chi(h_1,h_2) = \sum_{m=0}^{M}\frac{(h_1[m]-h_2[m])^2}{h_1[m]+h_2[m]} \quad (28)$$

The Kullback-Leibler (K-L) distance is perhaps the most frequently used to evaluate the distance between two sequences of random variables that have the same Markovian dependence order [5] because of its geometrical importance.

$$d_{KL}(h_1,h_2) = \sum_m h_1[m]\log\frac{h_1[m]}{h_2[m]} \quad (29)$$

However, the K-L distance is non-additive and non-symmetric, besides it requires identical bins.

The quadratic distance [3] is given by

$$d_Q(h_1,h_2) = \sum_{m=0}^{M}\sum_{n=0}^{M} h_{12}[m,n]a_{mn}h_{12}[m,n] \quad (30)$$

where the coefficient $h_{12}[m,n] = |h_1[m]-h_2[n]|$, and the covariance matrix element a_{mn} is based on the perceptual similarity of the colors m and n, which is expressed as

$$a_{mn} = 1 - \frac{h_{12}[m,n]}{\max h_{12}} \quad (31)$$

When a ground distance that matches perceptual dissimilarity is available for single features, incorporating this information results in perceptually more meaningful dissimilarity measures for distributions of features.

The Earth Movers distance (EMD) is defined as

$$d_E(h_1,h_2) = -\frac{\sum_m\sum_n d(h_1[m]h_2[n])f_{mn}}{\sum_m\sum_n f_{mn}} \quad (32)$$

where f_{mn} stands for the flow between $h_1[m]$ and $h_2[n]$ that minimizes an overall cost function. Given two distributions, one can be seen as piles of earth in feature space, the other as a collection of holes in that same space [6]. The distance between two color distributions is defined as the minimum amount of work needed to transform one color distribution into the other.

The Hausdorff distance computes the degree of mismatch between two sequences as the maximum distance between the colors

$$d_H(h_1,h_2) = \max\left[\max_m\left(\min_n|h_1[m]-h_2[n]|\right),\ \max_n\left(\min_m|h_1[m]-h_2[n]|\right)\right] \quad (33)$$

The Kolmogorov-Smirnov statistic determines the greatest distance between two cumulative distributions. This statistic can be expressed in terms of the significance level of an observed value of the statistic, giving the probability for the null hypothesis that both data sets are drawn from the same distribution. Despite these advantages, the K-S test has several important limitations: It only applies to continuous distributions. It tends to be more sensitive near the center of the distribution than at the tails. We define this statistic by the cumulative sequences as

$$d_{KS}(H_1,H_2) = \max_m|H_1[m]-H_2[m]| \quad (34)$$