A framework for ordinal-based image correspondence

Ilya Shmulevich, Bogdan Cramariuc, Moncef Gabbouj
Signal Processing Laboratory
Tampere University of Technology
P.O. Box 553, Tampere, Finland
{ilya,crama,moncef}@cs.tut.fi

ABSTRACT

We propose a general framework for ordinal-based image correspondence. This framework not only contains Kendall's τ and Spearman's ρ correlation measures as special cases, but allows one to design other correspondence measures that can potentially incorporate region-based spatial information. We consider one such possible correspondence measure and evaluate its performance on a set of test images.

1 Introduction

In many contexts, such as stereo matching [1], image retrieval [2], and motion estimation, a measure of association or similarity between two images is needed. It is desirable for such a measure to be robust in the presence of outliers while being invariant under reasonable image transformations. Outliers are generally caused by various kinds of noise, such as impulsive and bit-error noise. Additionally, different lighting conditions and camera calibrations can distort image intensity values. Thus, it is natural to expect that making an image brighter or even applying a monotonically increasing function to its intensity values should not alter its similarity with another image. Because of such considerations, traditional image matching measures such as correlation or squared Euclidean distance, which are essentially based on pixel intensity values, fail to satisfy our requirements.

To illustrate this, let us consider the normalized correlation coefficient

$$NCC = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

as well as the squared Euclidean distance

$$SSD = \sum_{i=1}^{n}\left(X_i - Y_i\right)^2$$

between images $X = \{X_1, \cdots, X_n\}$ and $Y = \{Y_1, \cdots, Y_n\}$, where $\overline{X}$ and $\overline{Y}$ are the sample means of X and Y and n is the total number of pixels in each image [3]. Suppose that the intensities of image X become distorted by some nonlinear monotonically increasing transformation, say, $f(x) = x^2$. Such a situation could arise when different cameras used for stereo have different responses to image irradiance [3]. It is easy to see that both the NCC and SSD will become affected by such a transformation. Moreover, suppose one pixel in image X is an outlier. Once again, both measures will become arbitrarily distorted.

In order to overcome these problems, so-called ordinal measures of association [4], [5] have been used. The ordinal measures operate on the ranks of pixels rather than directly on the pixel values. Thus, only relative ordering between data values is of consequence in determining the distance or correlation between two images. Two well-known ordinal measures, often used in psychological measurements, are Kendall's τ and Spearman's ρ [5]. For the reader's convenience, we briefly review these two measures.

Kendall's τ is given by

$$\tau = 2\binom{n}{2}^{-1} \cdot \left(\sum_{i,j} \delta\left[\left(X_i - X_j\right)\left(Y_i - Y_j\right) > 0\right]\right) - 1 \tag{1}$$

and δ is the indicator function:

$$\delta(A) = \begin{cases} 1, & \text{if event } A \text{ is true} \\ 0, & \text{otherwise} \end{cases}.$$

Spearman's ρ is given by

$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left(\pi_X^i - \pi_Y^i\right)^2}{n^3 - n}, \tag{2}$$

where π_X and π_Y are the rank vectors of images X and Y.

Another ordinal correlation measure is κ, introduced in [6] and [3] for image correspondence. Informally speaking, this measure is based on the ranking of one image with respect to the ranks of the other. This measure is robust in the presence of outliers, is invariant under monotonically increasing transformations, and insensitive to rank distortions [3].

In the present paper, we propose a general framework for performing ordinal-based image correspondence. This framework not only contains Kendall's τ and Spearman's ρ as special cases, but allows one to design other correspondence measures that can potentially incorporate region-based spatial information. We consider one such possible correspondence measure and evaluate its performance on a set of test images. The present idea was inspired by an original application of Kendall's τ measure for comparing pitch contours in music [7], [8], [9]. In these papers, Kendall's measure was computed using a non-traditional matrix representation.

2 The proposed general structure

The general structure of the proposed method is illustrated in Figure 1. We start with two images, X and Y, which are assumed to be of equal size. In a practical setting, images of different sizes can be resized and/or reshaped by an appropriate application dependent method. Let $\{X_1, \cdots, X_n\}$ and $\{Y_1, \cdots, Y_n\}$ be the pixels belonging to image X and Y, respectively. We select a number of areas $\{R_1, R_2, \cdots, R_m\}$ and extract the pixels from both images that belong to these areas. Consequently, R_j^X and R_j^Y contain the pixels from image X and Y, respectively, which belong to area R_j.

The goal is to compare the two images using a region-based approach. To this end, we will be comparing R_j^X and R_j^Y, for each $j = 1, \cdots, m$. Because of the considerations mentioned in the Introduction, our approach is an ordinal one and hence, only the ranks of the pixels are to be utilized. For every pixel X_k, we construct a so-called *slice* which is defined as follows:

$$S_k^X = \left\{S_{k,l}^X : l = 1, \cdots, n\right\},$$

where

$$S_{k,l}^X = \delta\left(X_k > X_l\right).$$

As can be seen, slice S_k^X corresponds to pixel X_k and is a binary image of size equal to image X. Slices are built in a similar manner for image Y as well.

With the goal of comparing regions R_j^X and R_j^Y, we first combine the slices from image X, corresponding to all the pixels belonging to region R_j^X. The slices are combined using the operation $OP_1(\cdot)$ into a so-called *metaslice* M_j^X. More formally,

$$M_j^X = OP_1\left(\left\{S_k^X : X_k \in R_j^X\right\}\right), \; j = 1, \cdots, m.$$

Similarly, we combine the slices from image Y to form M_j^Y. It should be noted that the metaslices are equal in size to the original images and could be multi-valued, depending on the operation $OP_1(\cdot)$. Each metaslice represents the relation between the region it corresponds to and the entire image itself.

The next step is a comparison between all pairs of metaslices M_j^X and M_j^Y by using operation $OP_2(\cdot)$, resulting in the *metadifference* D_j. That is,

$$D_j = OP_2\left(M_j^X, M_j^Y\right), \; j = 1, \cdots, m.$$

We thus construct a set of metadifferences

$$D = \{D_1, D_2, \cdots, D_m\}.$$

The final step is to extract a scalar measure of correspondence from set D, using operation $OP_3(\cdot)$. In other words,

$$\lambda = OP_3(D).$$

2.1 Modeling of known measures using the proposed structure

In this subsection, we shall illustrate how the above structure can be used to model the well-known Kendall's τ and Spearman's ρ measures. The recently proposed κ measure can also be modeled by using this structure, but we do not discuss it in this paper.

2.1.1 Kendall's τ

In order to model Kendall's correlation measure, the region size must be equal to one pixel. That is, $R_j^X = X_j$ and $R_j^Y = Y_j$, for $j = 1, \cdots, m$. Thus, it can be seen that $m = n$. Operation OP_1 is just the identity operation so that metaslice $M_j^X = S_j^X$ and $M_j^Y = S_j^Y$. Operation OP_2 counts the number of matching elements, out of a total of m elements, in the corresponding metaslices:

$$OP_2\left(M_j^X, M_j^Y\right) = m - \sum_i \left(M_j^X \oplus M_j^Y\right),$$

where the symbol $\oplus$ stands for component-wise addition mod 2 (XOR) and the summation is over all the elements of the metaslice. Finally, operation OP_3 performs the addition of all the metadifferences and the necessary normalization:

$$\lambda = OP_3(D) = 2 \cdot \binom{n}{2}^{-1} \left(\sum_j D_j\right) - 1.$$

In this way, the proposed structure implements equation (1).

2.1.2 Spearman's ρ

In order to model Spearman's correlation measure, we use only one region with size equal to the entire image. In other words, $R_1^X = X$ and $R_1^Y = Y$. Thus, it can be seen that $m = 1$. Since Spearman's measure directly uses the ranks of the pixels, operation OP_1 is chosen so that the values in the metaslices are equal to the ranks:

$$M_1^X = OP_1\left(\left\{S_1^X, \cdots, S_n^X\right\}\right) = (n+1) - \sum_k S_k^X \quad (3)$$

$$M_1^Y = (n+1) - \sum_k S_k^Y. \quad (4)$$

In the above equation, all operations, such as addition and subtraction, are understood to be component-wise. Operation OP_2 is selected to be

$$D_1 = OP_2\left(M_1^X, M_1^Y\right) = \left\|M_1^X - M_1^Y\right\|_2^2,$$

which is the sum of the squared component-wise differences between the two metaslices. As there is only one metadifference, $D = \{D_1\}$, operation OP_3 is simply used for purposes of normalization:

$$\lambda = OP_3(D) = 1 - \frac{6D_1}{n^3 - n}.$$

It is worth noting that the normalization could also have been performed by OP_2, in which case, OP_3 would have been an identity operation.

2.2 Constructing new measures using the proposed structure

Consider again Kendall's and Spearman's measures as viewed within the proposed framework. Both of these measures make inter-image comparisons on the basis of individual pixels. However, in certain types of applications such as image matching, it may be more reasonable to perform region-based inter-image comparisons instead. The proposed measure, described below, is able to achieve this goal while still maintaining the general characteristics of both Kendall's and Spearman's approaches.

As discussed in Section 2, both images, X and Y, are partitioned into a number of regions. This partition can be obtained by segmenting one of the two images or simply by splitting them into blocks of equal size. Thus, each block in one image is compared to the corresponding block in the other image in an ordinal fashion. In order to achieve this, the proposed structure offers several degrees of freedom by allowing one to choose operations OP_1, OP_2, and OP_3. For example, we have considered the following selection of these operations that extends Kendall's and Spearman's measures to the region-based approach. Operation OP_1 is simply the component-wise summation operation; that is, metaslice M_j is the summation of all slices corresponding to the pixels in block j or in other words, $M_j^X = \sum_{k:X_k \in R_j^X} \left(S_k^X\right)$. Note that for Spearman's ρ, a similar summation takes place in equations (3) and (4). Thus, metaslice j represents the "ranking" of block j with respect to the entire image in exactly the same way that metaslice j represents the "ranking" of pixel j with respect to the entire image in the case of Kendall's measure.

Next, operation OP_2 is chosen to be the Euclidean distance between corresponding metaslices. That is, $D_j = \left\| M_j^X - M_j^Y \right\|_2$. This is again similar to the operations used for Spearman's and Kendall's methods, where in the latter case, the mod 2 addition or Hamming distance is equal to the squared Euclidean distance (SSD) in the binary domain. Finally, operation OP_3 sums together all metadifferences to produce $\lambda = \sum_j D_j$.

3 Experimental results and discussion

In this section, we illustrate the use of the proposed correspondence measure for image matching. We have used 15 images taken from the Syntim Project [10] which were converted to 8-bit grayscale and resized to 128×128. The block size was chosen to be 8×8. The correspondence λ was calculated for each pair of images. The results can be seen in Figure 2. Those values which are shown in boldface are those which correspond to subjectively similar images. It can be seen that these values are generally smaller than those between visually dissimilar images. While these values are not normalized, as was the case for other correlation measures, this is not a shortcoming for image matching applications, since only the relative values are relevant. Future work should focus on choosing meaningful regions by using image segmentation algorithms.

References

[1] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9, pp. 920-932, 1994.

[2] E. Vicario (Ed.), *Image Description and Retrieval,* Plenum, 1998.

[3] D. N. Bhat and S. K. Nayar, "Ordinal measures for image correspondence," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 4, pp. 415-423, 1998.

[4] R. A. Gideon and R. A. Hollister, "A rank correlation coefficient," *J. Am. Statistical Assoc.*, Vol. 82, No. 398, pp. 656-666, 1987.

[5] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, fifth ed. New York: Edward Arnold, 1990.

[6] D. N. Bhat and S. K. Nayar, "Ordinal measures for visual correspondence," Columbia Univ., Computer Science, tech. rep. CUCS-009-96, Feb. 1996.

[7] E. Marvin and P. Laprade, "Relating musical contours: Extensions of a theory for contour," *Journal of Music Theory*, Vol. 31, pp. 225-267, 1987.

[8] L. Polansky, "Morphological metrics," *Journal of New Music Research*, Vol. 25, pp. 289-368, 1996.

[9] I. Quinn, "The combinatorial model of pitch contour," *Music Perception*, Vol. 16, No. 4, pp. 439-456, 1999.

[10] http://www-syntim.inria.fr/syntim/analyse/images-eng.html

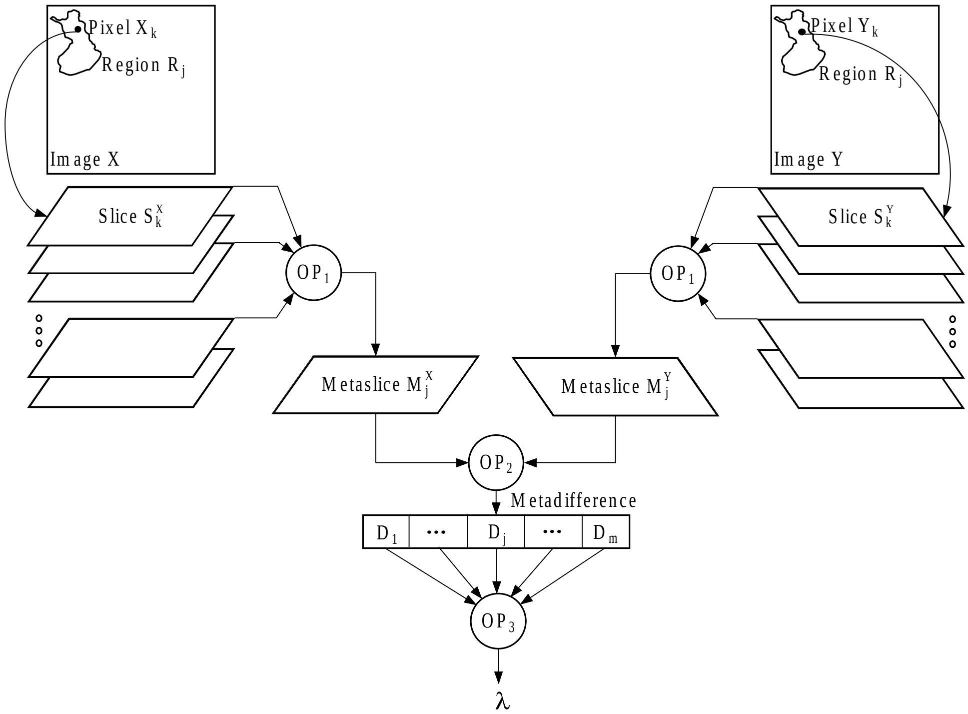

Figure 1: The diagram of the proposed structure for ordinal-based image correspondence.

	0.00	1.59	1.57	1.56	1.04	1.33	1.28	1.25	1.27	1.20	1.18	1.21	1.38	1.20	1.35
	1.59	**0.00**	**0.79**	**0.84**	1.47	1.32	1.51	1.56	1.59	1.56	1.71	1.55	1.23	1.35	1.24
	1.57	**0.79**	**0.00**	**0.58**	1.39	1.41	1.58	1.57	1.63	1.54	1.67	1.53	1.24	1.31	1.22
	1.56	**0.84**	**0.58**	**0.00**	1.40	1.42	1.58	1.59	1.65	1.56	1.67	1.56	1.28	1.34	1.24
	1.04	1.47	1.39	1.40	0.00	1.38	1.30	1.32	1.33	1.30	1.23	1.31	1.42	1.32	1.36
	1.33	1.32	1.41	1.42	1.38	0.00	1.21	1.20	1.25	1.27	1.35	1.27	1.33	1.32	1.36
	1.28	1.51	1.58	1.58	1.30	1.21	**0.00**	**0.80**	**0.80**	**1.09**	**1.04**	**1.09**	1.36	1.27	1.28
	1.25	1.56	1.57	1.59	1.32	1.20	**0.80**	**0.00**	**0.77**	**0.96**	**1.03**	**0.99**	1.34	1.25	1.27
	1.27	1.59	1.63	1.65	1.33	1.25	**0.80**	**0.77**	**0.00**	**0.88**	**0.98**	**0.87**	1.35	1.25	1.26
	1.20	1.56	1.54	1.56	1.30	1.27	**1.09**	**0.96**	**0.88**	**0.00**	**1.00**	**0.40**	1.29	1.16	1.20
	1.18	1.71	1.67	1.67	1.23	1.35	**1.04**	**1.03**	**0.98**	**1.00**	**0.00**	**1.01**	1.38	1.25	1.25
	1.21	1.55	1.53	1.56	1.31	1.27	**1.09**	**0.99**	**0.87**	**0.40**	**1.01**	**0.00**	1.27	1.15	1.18
	1.38	1.23	1.24	1.28	1.42	1.33	1.36	1.34	1.35	1.29	1.38	1.27	0.00	1.18	1.16
	1.20	1.35	1.31	1.34	1.32	1.32	1.27	1.25	1.25	1.16	1.25	1.15	1.18	0.00	1.14
	1.35	1.24	1.22	1.24	1.36	1.36	1.28	1.27	1.26	1.20	1.25	1.18	1.16	1.14	0.00

Figure 2: Correspondence values λ. All values are divided by 10^6.