A Probabilistic Approach to Fast and Robust Template Matching and its Application to Object Categorization

Takeshi Mita  Toshimitsu Kaneko  Osamu Hori
Multimedia Laboratory, Corporate Research & Development Center, Toshiba Corporation
1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa 212-8582, Japan
E-mail: takeshi.mita@toshiba.co.jp

Abstract

This paper presents a new statistic, called Probabilistic Increment Sign Correlation (Probabilistic ISC), for evaluating similarity between images of objects which have intra-class variation such as individual differences of human faces. The new statistic evaluates similarity between an input image and object classes, whereas most conventional methods, such as normalized cross-correlation, calculate correlation between an input image and a template. The new statistic is defined as a log-likelihood based on probabilities of observing the increment signs. Probabilistic ISC provides two advantages over conventional correlation-based methods: 1) robustness against the intra-class variation because it gives larger weights to stable features which are commonly observed in reference images and 2) robustness against noise and change in illumination. It yields higher performance even if a small number of reference images are given, whereas other methods such as the subspace method and AdaBoost cannot maintain their accuracy. We show these advantages through several experiments of face detection and face orientation estimation.

1 Introduction

Template matching is a fundamental task in the field of image processing because it is applicable to numerous different tasks such as object detection and categorization due to its simplicity of implementation [1]. Robustness against change in illumination is required because a reference image (template) and an input image are captured in different illumination environments in many real world tasks. Normalized cross-correlation [2] (referred to as CC in this paper) is one of the most popular techniques. However, it cannot deal with local brightness changes or occlusions. In recent years, several robust techniques have been proposed in order to realize more flexible vision systems which work in the real world. For example, orientation code matching [3] and increment sign correlation (ISC) [4] improve robustness against these outliers because unlike CC they do not depend on absolute pixel values but use simple features based on spatial gradient. In particular, ISC has many advantages over conventional methods, such as small computational cost, simple implementation and discriminating power. However, most of these correlation-based methods encounter difficulties in matching objects which have large variation within each class such as human faces. The matching accuracy depends on one reference image which represents the object class. There are two approaches to this problem. One is to use multiple templates. This, however, increases computational cost. Another method is to use an average image as the reference image. However, it is insufficient for object classes with large variation because it cannot evaluate differences between reference images which belong to the same object class. There is a limit for dealing with the variation as noise, which is added to the average image.

The subspace method [5] has been widely used for handling the variation. However, it requires more computational cost for matching than conventional correlation-based methods because it needs multiple inner-product operations. Furthermore, if raw pixel data is used as feature vectors, it is easily affected by occlusions and local illumination changes.

In this paper, we propose a new statistic which realizes fast and robust template matching even if the target object contains large variation within the object class. The original inspiration of the idea is drawn from ISC, which uses simple but robust features. The idea is extended by incorporating a probabilistic approach in order to deal with the intra-class variation. The statistic is called Probabilistic Increment Sign Correlation (referred to as Probabilistic ISC in this paper), which is defined as a log-likelihood based on probabilities of observing the increment signs in multiple reference images. It can be easily extended to deal with multi-class classification by formulating the classification
function as a log-likelihood ratio test.

The main contributions of this paper are: 1) unlike conventional correlation-based methods, Probabilistic ISC yields higher matching performance when the object class has large variation. This is achieved by giving larger weight to stable features; 2) unlike conventional methods using raw pixel data as features, Probabilistic ISC is robust against noise and change in illumination; 3) the computational cost is equivalent to ISC and much smaller than the subspace method or other non-linear classification algorithms; 4) it maintains its performance even if a small number of reference images are given, whereas other methods such as the subspace method and AdaBoost cannot maintain their accuracy with such small training data. We show this in an experiment comparing face detectors based on the new statistic with the ones proposed by Viola and Jones [6].

The rest of the paper is organized as follows: Section 2 presents the definition of Probabilistic ISC and discusses robustness against intra-class variation and outliers; Section 3 extends it to multi-class classification; Section 4 gives the results of several experiments.

2 Probabilistic ISC

This section defines Probabilistic ISC and discusses robustness against intra-class variation and outliers.

2.1 Definition

For simplicity, images are regarded as one-dimensional vectors in this section. Actual implementation for two-dimensional images is described in 4.1. Assume that \( N \) reference images are given and let \( I_n(i) \) be the value of the \( i \)-th pixel in the \( n \)-th reference image \( I_n \).

The increment sign sequence \( B_n \) is generated by comparing adjacent pixel values in each reference image \( I_n \). Each increment sign \( B_n(i) \) assigned to the \( i \)-th pixel is calculated by [4]

\[
B_n(i) = \begin{cases} 
1 & I_n(i+1) > I_n(i) \\
0 & \text{otherwise}
\end{cases}
\]  

(1)

where the sequence length of \( B_n \) is \( M \) when the length of \( I_n \) is \( M + 1 \). Then, probability values \( P(v, i) \) of observing the increment sign \( v \) in \( i \)-th pixel are calculated by (2). \( P(v, i) \) represents the ratio of number of images in which \( v \) is observed to the total number of reference images \( N \). These probability values are stored in a probability map.

\[
P(v, i) = \frac{1}{N} \sum_{n=1}^{N} \delta(v, B_n(i)),
\]

(2)

where

\[
\delta(v, z) = \begin{cases} 
1 & z = v \\
0 & z \neq v
\end{cases}
\]

(3)

For an unknown input image \( I' \), an increment sign sequence \( B' \) is generated by (1). Then, the probability values \( P(B'(i), i) \) stored in the probability map are referred to. If the input image belongs to the object class, the probability values referred to from the probability map are expected to be large. The correlation between the input image and the object class is defined as a log-likelihood \( C \) by assuming that all increment signs in an image are observed independently.

\[
C = \sum_{i=1}^{M} \log P(B'(i), i)
\]

(4)

For the actual implementation, the probability map stores log values of \( P(v, i) \). By doing so, the correlation is calculated by only simple operations: 1) comparing pixel values; 2) referring to the probability map; 3) summing up the referred values. The computational complexity for these operations is equivalent to ISC and can be easily implemented on hardware.

2.2 Robustness against Variation within the Object Class

We now introduce the main contribution of the paper. This section discusses the robustness of Probabilistic ISC against variation within the object class. Consider the probability values \( P(v, i) \) of observing \( v \) in the \( i \)-th pixel. \( P(v, i) \) represents stability of \( v \) in the object class. For example, if the same increment sign is observed at the same pixel in many reference images, the feature is stable and \( P(v, i) \) becomes large with a maximum value of 1. If different increment signs are observed among the reference images, the feature is not stable in the object class. When the reference images are uncorrelated with each other, \( P(0, i) \) and \( P(1, i) \) become 0.5. Probabilistic ISC gives larger weight to the stable features for calculating \( C \) by (4). In this way, it achieves robust matching without being affected by the intra-class variation. On the other hand, most of the conventional correlation-based methods give the same weight to all features without evaluating the feature stability.

Figure 1 shows an example of a stable and an unstable feature. A histogram of differential values between the pixel A and the next pixel is shown in (a) and probabilities of increment signs are shown in (b). From the pixel B, (c) and (d) were calculated. The pixel A is located between hair region and the right eyebrow. The pixel A is not a good feature because the hair length is different between individuals. The histogram (a) is distributed around 0 and \( P(0, A) \cong P(1, A) \). The pixel B located near left eye is a stable feature. The weights given to the pixel A and B are about 0.5 and 0.9, respectively.
2.3 Robustness against Noise and Illumination Changes

The increment sign used as a feature for Probabilistic ISC has robustness against noise and change in illumination. When these outliers are not large and do not invert the inequality in (1), the increment sign remains unchanged. It is invariant to a uniform offset added to the whole image. Even if an offset is added to a local region, the increment sign remains unchanged except for the boundary of the local region.

3 Multi-Class Classification based on Probabilistic ISC

This section formulates classification functions based on Probabilistic ISC for a multi-class problem. For classifying \( K \) classes, \( K \) probability maps are generated by (2). Given an unknown input image \( I' \) and corresponding increment sign sequence \( B' \), \( K \) correlation values \( C_k \) are calculated by (4). We first deal with a two-class problem and then discuss a multi-class problem.

3.1 Two-Class Problem

Object detection can be viewed as a two-class classification problem in which an image region is classified as an object or a non-object. The classification function is formulated by

\[
k^* = \begin{cases} +1 & C_{+1} - C_{-1} \geq \theta \\ -1 & \text{otherwise} \end{cases}
\]  

(5)

where \( C_{+1} \) and \( C_{-1} \) are correlation values for the object class and the non-object class respectively. \( \theta \) is a threshold for adjusting sensitivity. According to (4), the classification function is interpreted as a log-likelihood ratio test as follows:

\[
C_{+1} - C_{-1} = \sum_{i=1}^{M} \log \frac{P(B'(i), i|k = +1)}{P(B'(i), i|k = -1)} \geq \theta.
\]  

(6)

This function compares similarities to two classes by giving larger weights to stable and discriminative features, whereas (4) evaluates similarity only to the object class.

3.2 Multi-Class Problem

The above classification method is easily extended to a multi-class problem. The class label is determined by choosing the class which provides the highest correlation value. This is equivalent to a naive Bayesian classifier, which classifies two arbitrary classes among \( K \) by the above log-likelihood test with \( \theta = 0 \).

\[
k^* = \arg \max_k C_k.
\]  

(7)

In this way, a classifier is constructed by a discriminative approach. One advantage over other discriminative methods such as the linear discriminant analysis (LDA) is that the proposed method requires small computational cost for updating a classification function when one or more classes are added. Although LDA requires re-computation of the function when a new class is added, the proposed method only generate a probability map for the new class.

4 Experimental Results

This section shows the effectiveness of the new statistic through three experiments using face images. The reason that we use face images in the experiments is that a face class is a typical example which has large variation. Building a face detector is not our main goal.

Exp.1: Face detection
Exp.2: Face detection with a small number of reference images
Exp.3: Face orientation estimation (9-class problem)

Exp.1 deals with a face detection task and compares performance between the proposed method and conventional correlation-based methods. Exp.2 deals with the same task but compares performance between classifiers trained using a smaller number of training examples than Exp.1. It shows that the new statistic maintains its performance despite the small number of reference images, whereas other classifiers trained by the subspace method and AdaBoost cannot maintain their performance. Exp.3 compares performance in face orientation estimation. The task is to choose one orientation from nine classes which consist of different rotation angles and poses. Rowley et al. constructed a router network for similar tasks and built a rotation invariant face detection system [7]. The router processes each input window to determine its orientation and then this information is used to prepare the window for one or more detectors.
4.1 2D Increment Sign Sequence

Two-dimensional increment sign sequences are used through all experiments. The sequences are calculated by comparing horizontally and vertically adjacent pixel values using

\[ BH(x, y) = \begin{cases} 
1 & I(x + 1, y) > I(x, y) \\
0 & \text{otherwise}
\end{cases}, \quad (8) \]

and

\[ BV(x, y) = \begin{cases} 
1 & I(x, y + 1) > I(x, y) \\
0 & \text{otherwise}
\end{cases}. \quad (9) \]

The correlation value is calculated by

\[
C = H \sum_{y=1}^{H-1} \sum_{x=1}^{W-1} \log PH(BH'(x, y), x, y) \\
+ \sum_{y=1}^{H-1} \sum_{x=1}^{W} \log PV(BV'(x, y), x, y), \quad (10)
\]

where \( BH' \) and \( BV' \) are increment sign sequences generated from an input image. \( PH \) and \( PV \) are probability maps. \( W \) and \( H \) are width and height of the reference images, respectively. The total number of evaluated features is \((W - 1)H + W(H - 1)\).

4.2 Exp.1: Face Detection

Exp.1 compares face detection performance between three methods, which are CC (normalized cross-correlation) [2], ISC (increment sign correlation) [4] and Probabilistic ISC. ISC and Probabilistic ISC are spatial gradient-based methods, whereas CC uses raw pixel data as features. Only Probabilistic ISC can deal with intra-class variation.

We collected 1,000 face images and 1,000 nonface images as reference images. They are of size 20 × 20 pixels. Figure 2 shows the average image (a) of the face images, the probability maps (b) and examples of the test set (c). The average image is used as a reference image for CC and ISC. The probability map indicates the probability values of observing the increment signs. There are four images corresponding to \( PH(0, x, y) \), \( PH(1, x, y) \), \( PV(0, x, y) \) and \( PV(1, x, y) \), respectively. Higher probability values indicated by brighter pixel values are observed around eyes and a nose because there are strong brightness changes. We use a subset of the XM2VTS database [8] as a test set for this experiment. The test set contains 1,176 illuminated faces. Examples of the test set are shown in (c). Positions of both eyes and nostrils are manually input for evaluation.

Figure 3 shows ROC curves obtained from the three methods. The better method pushes the ROC curve to the upper left. Probabilistic ISC yields higher performance than ISC. This indicates that the proposed statistic is more robust against intra-class variation. The processing speed of Probabilistic ISC is 1.8 times faster than CC.

4.3 Exp.2: Face Detection with a Small Number of Training Examples

Figure 4 shows ROC curves obtained from twelve different face detectors, which are trained by three methods using different numbers of training examples. Subspace indicates a classifier based on the subspace method [5]. Eigenvectors are selected so that the cumulative proportion of the vectors becomes more than 0.99. The number of the eigenvectors is between 12 and 25. AB is a face detector based on Haar-like features and AdaBoost, proposed by Viola and Jones [6]. The number of Haar-like features selected by AdaBoost is 760, which is the same as the number of features evaluated by Probabilistic ISC. For example, a curve indicated by AB100 shows performance of a detector trained by Viola and Jones’ method using 100 face images and 100 nonface images. As the number of training examples decreases...
4.4 Exp.3: Face Orientation Estimation

Exp.3 compares the performance in a task of face orientation estimation on a test set which contains illuminated and partially occluded face images shown in Figure 5. There are nine classes which consist of seven face classes rotated in plane (RIP) and two face classes rotated out of plane (ROP). The number of the reference images of each class is 1,000. The RIP images are generated by rotating the original upright frontal faces, which are indicated by 0 degree. All test images are partially occluded by random patterns. The occluded area is 20% of the whole image. The total number of test images is 9,401. Table 1 shows the number of misclassifications and the correct estimation rates. The proposed method classifies 97.5% of the test images correctly with the comparable computational cost to CC and ISC. The processing time of Subspace is 13 times as long as that of the proposed method. The standard AdaBoost algorithm is not applicable to multi-class problems.

5 Conclusions

In this paper, we have proposed a new statistic for fast and robust template matching. The proposed method evaluates the feature stability within the object class and gives larger weight to the stable features, which leads to robustness against variation within the object class. Additionally, the new statistic has the advantage in that it maintains its performance even if a small number of reference images is given, whereas the accuracy of other methods such as the subspace method and AdaBoost decreases.

References