# Bayesian Network based Reparameterization of Haar-like feature

**Hirotaka NIITSUMA**

Advanced Industrial Science and Technology(AIST)

AIST Tsukuba Central2, 1-1-1 Umezono Tsukuba, Ibaraki 305-8568, Japan

Tel: +81-29-852-0501

E-Mail: niitsuma@mub.biglobe.ne.jp URL: http://www2s.biglobe.ne.jp/ niitsuma

## Abstract

Object detection using Haar-like features is formulated as a maximum likelihood estimation. Object features are described by an arbitrary Bayesian Network (BN) of Haar-like features. We proposed variable translation techniques transform the BN into the likelihood for the object detection. The likelihood is a BN which includes a node that represents the object's position, angle and scale. The object detection can be achieved by inference for the node.

## Introduction

Viola et al. have proposed a multi-stage classification procedure that reduces the processing time substantially while achieving almost the same accuracy as compared to a much slower and more complex single stage classifier(Viola & Jones 2001). This paper extends their rapid object detection framework in two important ways: Firstly, a Bayesian Network (BN) based representation of Haar-like feature is introduced. Secondly, we formulate detecting an object described by the BN as an optimization problem $\arg\max_{\mathbf{u}} p(\mathbf{u})$, where, $\mathbf{u}$ is a transformed representation of an object position, angle and scale. $p$ is a probability of $\mathbf{u}$. The optimization problem can be solved by Bayesian inference for the BN.

## Reparameterization of Haar-like Feature

Haar-like feature is a linear combination of sums of pixels intensities, within few rectangular regions. We approximate Haar-like feature by using the difference of two rectangular regions. The approximated Haar-like feature is showed in Fig. 2. In this figure, Haar-like feature is approximated as the difference between a rectangular region filled with an oblique line pattern and a region filled with an a vertical line pattern. The approximated Haar-like feature is represented by three vertexes $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ and parameters $\alpha_1, \alpha_2$. Here, the parameters $\alpha_1, \alpha_2$ represent relative location of the smaller rectangle from the larger rectangle. Using these 8 parameters $\hat{\mathbf{x}} = (\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_3, \hat{y}_3, \alpha_1, \alpha_2)$ , an object's feature distribution can be formulated as a conditional probability

$$P(z|\hat{\mathbf{x}}) = \mathcal{N}(\mu + \Lambda\hat{\mathbf{x}}, \Psi), \qquad (1)$$
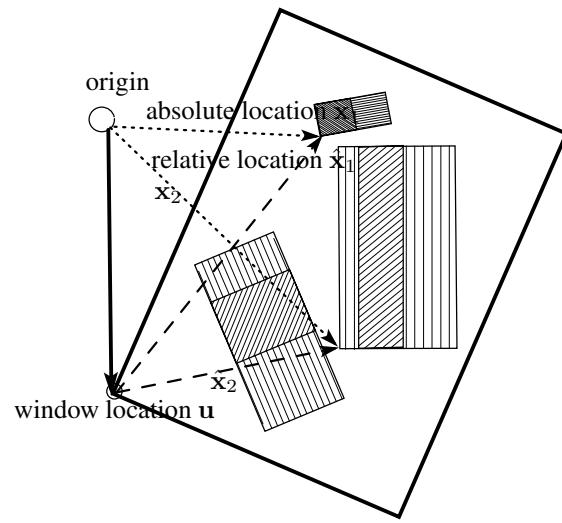
Figure 1: Object Detection

where $\hat{\mathbf{x}}$ is a vector of parameters that determine location and shape of Haar-like feature. $z$ is response of the Haar-like feature. $\mathcal{N}(\mu, \Sigma)$ denotes a normal distribution with mean $\mu$ and covariance $\Sigma$. In this study, we call this Haar-like feature Conditional Haar-like Feature (CHlF).

## Object Detection

Various images can be regarded as combination of CHlF as Fig. 3 . The combination of CHlF can be represented as a BN of CHlF. Let us consider detecting the pattern of Fig. 3 in Fig. 1. Fig. 1 is a rotated, scaled and shifted image of Fig. 3. In order to detect such image, we introduce variable translation techniques for BN. First, we consider the simple model of equation (1). Let us assume, Fig. 1 is an affine transformed image of Fig. 3. Here, the affine transformation is expressed by

$$\mathbf{v} = B\hat{\mathbf{v}} + \mu.$$

$\hat{\mathbf{v}}$ represents coordinate system of an image of a registered original model like Fig. 3. $\mathbf{v}$ represents coordinate system of affine transformed image like Fig. 1. $\mu, B$ are parameters of the affine transformation. When a Haar-like feature on
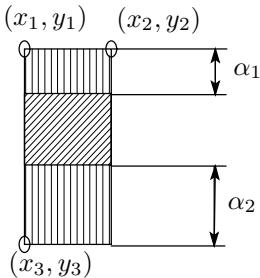
Figure 2: Reparameterization of Haar-like feature
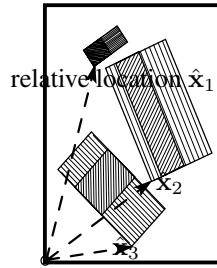


Figure 3: Distribution of Haar-like feature
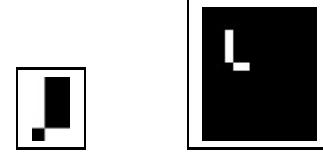


Figure 4: Target image
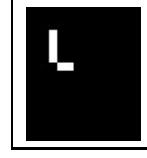


Figure 5: Test image
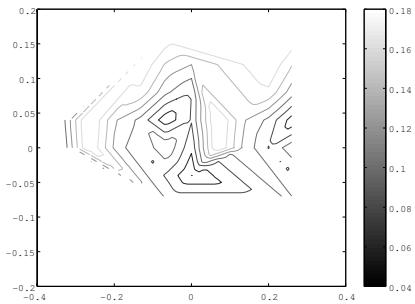


Contour plot of an accuracy of inferred target position vs x and y shift of center of sampling area.

Figure 6: Error of inference

the affine transformed image

$$(x_1, y_1, x_2, y_2, x_3, y_3, \alpha_1, \alpha_2, z)$$

is measured, the measured feature can be compared to the equation (1) after the affine transformation. The affine transformed feature is

$$(\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_3, \hat{y}_3, \alpha_1, \alpha_2, z),$$

where,

$$
\begin{pmatrix} \hat{x}_1 \\ \hat{y}_1 \\ \hat{x}_2 \\ \hat{y}_2 \\ \hat{x}_3 \\ \hat{y}_3 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} =
\begin{bmatrix}
-1 & 0 & x_1 & y_1 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & x_1 & y_1 & 0 & 0 \\
-1 & 0 & x_2 & y_2 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & x_2 & y_2 & 0 & 0 \\
-1 & 0 & x_3 & y_3 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & x_3 & y_3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \alpha_1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_2
\end{bmatrix}
\begin{pmatrix} \lambda_x \\ \lambda_y \\ c_{xx} \\ c_{xy} \\ c_{yx} \\ c_{yy} \\ 1 \\ 1 \end{pmatrix}
$$

$$= \Omega(\mathbf{x})\mathbf{u} \qquad (2)$$

$$C = \begin{pmatrix} c_{xx}, c_{xy} \\ c_{yx}, c_{yy} \end{pmatrix} = B^{-1} \qquad (3)$$

$$\lambda = (\lambda_x, \lambda_y)^t = B^{-1}\mu. \qquad (4)$$

Substituting the affine transformed feature into the equation (1), we obtain the following conditional probability.

$$P(z|\mathbf{u}, \mathbf{x}) = \mathcal{N}(\mu + \Lambda\Omega(\mathbf{x})\mathbf{u}, \Psi) \qquad (5)$$

Note that, equations (3) and (4) are variable translation from the representation of the affine transform $\mu$, $B$ to a translated representation $\mathbf{u}$. The BN of equation (5) have a node of the translated representation $\mathbf{u}$. Thus, the substituted BN can infer most probable affine transformation $\mathbf{u}$ for the case of the measured feature $(x_1, y_1, x_2, y_2, x_3, y_3, \alpha_1, \alpha_2, z)$ is generated from model (1), by inference for the node $\mathbf{u}$. This inference can infer not only shift also but scale and rotation of an image.

Similarly, these variable translation and substitution techniques can apply to arbitrary BNs of CHlF.

## Experiment

We have showed that an image described in an arbitrary BN can be detected by applying the variable translation techniques. The detection is achieved by simple inference for the node $\mathbf{u}$.

In this section, accuracy of the inference on the target position is investigated by an experiment. In this experiment, detection of a target image in Fig. 4 in test image in Fig. 5 was investigated. The target image was modeled by the following BN.

$$P(z|\hat{\mathbf{x}}) = \sum_j \pi_j \mathcal{N}(\mu_{1j} + \Lambda_{1j}\hat{\mathbf{x}}, \Psi_{1j}) \cdot \mathcal{N}(\mu_{2j} + \Lambda_{2j}\hat{\mathbf{x}}, \Psi_{2j})...$$

Parameters of the BN were manually chosen. CHlF in the test image randomly sampled around the target image. Using the proposed technique, target position was inferred. Fig. 6 plot an accuracy of inferred target position vs shift of center of sampling area. This result shows the inferred results are valid when the sampling area is away from the target by less than half of target size.

## Conclusion

Object detection method using CHlF is proposed. The method can apply various images described by an arbitrary BN of CHlF. We showed experimental result that this method can infer target image from away from target by half of target size.

## References

Viola, P., and Jones, M. 2001. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR*.