Statistical Reach Feature Method And Its Application To Template Matching

Ryushi Ozaki
Tsukuba University
OZAKI.Ryushi@gmail.com

Yutaka Satoh
Information Technology Research Institute, AIST

Kenji Iwata
Information Technology Research Institute, AIST

Katsuhiko Sakaue
Information Technology Research Institute, AIST

Abstract

A novel method for image template matching which gives robust results for various disturbances in the real world, including local and/or global variations of illumination, occlusions, and noises, is proposed. The authors defined the similarity index from the set of pixel pairs with distinct intensity-difference in the given template image. The 'distinctness' is defined from the statistical view point. The authors develop the mathematical model of the inverting-ratio, which is the quantity strongly related to the similarity index, for the Gaussian-disturbances. The authors also verified the model by numerical experiments. The mathematical model, enforced by the numerical results, gives the theoretical backbone for the robustness of the proposed method.

1 Introduction

The image template matching techniques, which give the criterion of similarity of two images, are the most fundamental pattern recognition method, and also wide-applied core technologies in the industry e.g. appearance inspection. In application of template matching, sometimes following properties are needed:

- robust matching for occlusion e.g. by gloss
- robust matching for local and/or global illumination variation
- robust matching for various random noises

In this paper, we will develop *the Statistical Reach Feature (SRF)* method of the template matching, as an answer to these requirements. The key idea of the method is very simple : roughly speaking, for a given pair of points, two signs of intensity-differences, one from the first image, and the other from the second image, tend to coincide if the two given images are sufficiently similar and the absolute value of the difference taken from one of the images is sufficiently large.

Like SRF, some of existing methods of template matching use the intensity-based information. Above those, we mention SSD(sum of squared difference) [1], CC(the normalized cross-correlation) [2], and ISC(increment sign correlation) [3]. SSD uses the criterion based on the sum of squared difference of the intensity of the given images and provides fast matching. However, it is well known to fail for the illumination variation and the occlusion, even for rather weak ones. CC uses the criterion based on the normalized cross correlation of the intensity of the given images. This method is robust under the noise and the uniform variation of illumination. However, this method is weak for the local variation of illumination and occlusion. ISC uses the vector of Boolean values which corresponds to the sign of the difference of the intensity of right-adjacent pixels of the given image. The criterion of ISC is correlation coefficient of the Boolean value vectors made from the given two images. This method is robust for local and/or global variation of illumination, noise, and occlusion. Some improvements for this method are proposed, e.g.[4]. However, the methods based on ISC are weak especially for the regions with low mean-contrast, because the robustness of the methods depend on the intensity-difference of adjacent pixels.

SRF is a method to extract noise-robust feature from the given image, by sign of the intensity-difference of the selected pixel-pairs of the images. The selection of the pairs is done by a statistical criterion. SRF method is already applied to background subtraction [6] and showed the high performance for the subject. In this paper, we show SRF template matching provides the robust matching for illumination variation, random (e.g. Gaussian) noise, and other various disturbances. The authors also performed the experimental comparison with CC and ISC, which shows the superiority of the SRF template matching over CC and ISC.

This paper is organized as follows. In section 2, the definition of the similarity index is given. In section 3, we show the experimental results, in the real-world images with various disturbance, including illumination variation and partial-occlusion. In Section 4, we develop the mathematical model of the statistical behavior of inverting ratio and examine it numerically. In section 5, we give the concluding remarks.

2 Template Matching by SRF Method

2.1 The Definition Of Similarity Index

Now we show the definition of the similarity index $c(\boldsymbol{I}, \boldsymbol{J})$ of given pair of same-sized images $\boldsymbol{I}$ and $\boldsymbol{J}$. Suppose the size of input images are of W×H. By grid

Γ, we mean a set of the coordinates of pixels in image $\boldsymbol{I}$ (and hence $\boldsymbol{J}$). Namely,

$$\Gamma := \{(i,j) \,|\, i = 0, \ldots, W-1,\ j = 0, \ldots, H\}. \quad (1)$$

In what follows, we regard the images of size W×H as the intensity function defined on Γ. For arbitrary pair $(\boldsymbol{p}, \boldsymbol{q})$ of grid-points in Γ, we define the value $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}\,; T_P)$ as follows:

$$\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}\,; T_P) := \begin{cases} 1 & \boldsymbol{I}(\boldsymbol{p}) - \boldsymbol{I}(\boldsymbol{q}) \geq \ \ T_P, \\ 0 & \boldsymbol{I}(\boldsymbol{p}) - \boldsymbol{I}(\boldsymbol{q}) \leq -T_P, \\ \phi & \text{otherwise.} \end{cases} \quad (2)$$

Here, $T_P (> 0)$ is the threshold of intensity difference. For the role and how to set the value, see next subsection. We call the grid-point pair $(\boldsymbol{p}, \boldsymbol{q})$ *the reach*, if $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}\,; T_P) \neq \phi$. In the following, we write $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q})$ rather to $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}\,; T_P)$, unless there is any confusion. For the given reach $\rho = (\boldsymbol{x}, \boldsymbol{y})$, we call the grid-point $\boldsymbol{x}$ and $\boldsymbol{y}$ *arising point of* ρ and *end point of* ρ, respectively. The length $\|(\boldsymbol{x}, \boldsymbol{y})\|$ of a reach $(\boldsymbol{x}, \boldsymbol{y}) = ((i_1, j_1), (i_2, j_2))$ is defined as Chebychev (or chess board) distance as follows:

$$\|(\boldsymbol{x}, \boldsymbol{y})\| := \max\{\, |i_1 - i_2|,\ |j_1 - j_2| \,\}. \quad (3)$$

By the *local reach graph* $\mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p}; \boldsymbol{I}, T_P, T_D)$, we mean the set of reaches which arises from the given grid-point $\boldsymbol{p} \in \Gamma$ with length $\leq T_D$. Namely,

$$\mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p}; \boldsymbol{I}, T_P, T_D) := \{(\boldsymbol{p}, \boldsymbol{y}) \in \Gamma \times \Gamma \mid \mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{y}) \neq \phi,\ \|(\boldsymbol{p}, \boldsymbol{y})\| \leq T_D \,\}.$$

Note that $\mathrm{RG}_{\mathrm{loc}}$ is possibly nullset. The role of the threshold T_D and how to set it is explained in the next subsection. So far, we treated only image $\boldsymbol{I}$. Now we define the incremental sign $\mathrm{b}(\boldsymbol{p} \succ \boldsymbol{q})$ for image $\boldsymbol{J}$ as follows:

$$\mathrm{b}(\boldsymbol{p} \succ \boldsymbol{q}) := \begin{cases} 1 & \boldsymbol{J}(\boldsymbol{p}) \geq \boldsymbol{J}(\boldsymbol{q}), \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

For a grid-point $\boldsymbol{p} \in V(\boldsymbol{I}, T_P, T_D)$, $r(\boldsymbol{p}, \boldsymbol{J})$ is defined as follows:

$$r(\boldsymbol{p}, \boldsymbol{J}) := \frac{|\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p}) \mid \mathrm{srf}(\boldsymbol{x} \succ \boldsymbol{y}) = \mathrm{b}(\boldsymbol{x} \succ \boldsymbol{y})\}|}{|\mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p})|}. \quad (5)$$

Here, $\mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p})$ above is just the abbreviation of $\mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p}; \boldsymbol{I}, T_P, T_D)$. The *local similarity* $c_{loc}(\boldsymbol{p}, \boldsymbol{J})$ is defined as follows:

$$c_{loc}(\boldsymbol{p}, \boldsymbol{J}) := \begin{cases} 1 & r(\boldsymbol{p}, \boldsymbol{J}) \geq T_B, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Here, the value $T_B \in [0, 1]$ is the threshold. The role of threshold T_B and how to set it is explained in the next subsection. We define $V(\boldsymbol{I}, T_P, T_D)$ to be the set of arising points of reaches in image $\boldsymbol{I}$. Namely,

$$V(\boldsymbol{I}, T_P, T_D) := \{\boldsymbol{p} \in \Gamma \,|\, \mathrm{RG}_{\mathrm{loc}}(\boldsymbol{p}; \boldsymbol{I}, T_P, T_D) \neq \phi\}. \quad (7)$$

The similarity coefficient $c(\boldsymbol{I}, \boldsymbol{J})$ is defined as follows:

$$c(\boldsymbol{I}, \boldsymbol{J}) := \frac{\sum_{\boldsymbol{p} \in V(\boldsymbol{I}, T_P, T_D)} c_{loc}(\boldsymbol{p}, \boldsymbol{J})}{|V(\boldsymbol{I}, T_P, T_D)|}. \quad (8)$$

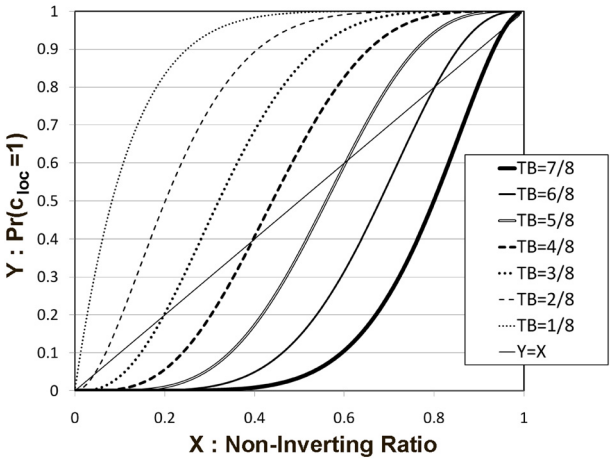

Figure 1: The effect of T_B.

2.2 The Roles and Selection of Thresholds

Now we explain the threshold T_P. Suppose the two images $\boldsymbol{I}$ and $\boldsymbol{J}$ is given to be similar (in intuitive sense). Then, we can assume as if the image $\boldsymbol{J}$ is generated from the image $\boldsymbol{I}$ by adding a noise with some (unknown, generally) probability model. Then, from the elementary probability argument, we know that if a pair of grid-points has sufficiently large (in absolute value) intensity difference in the image $\boldsymbol{I}$, the sign of the difference is highly expected to coincide with that of the grid-point pair in the image $\boldsymbol{J}$. So, if we take $T_P > 2\sigma$, where σ is the SD of the noise, the *inverting ratio* which is defined as the probability of discrepancy between $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}\,; T_P)$ and $\mathrm{b}(\boldsymbol{p} \succ \boldsymbol{q})$, shall be low. For the quantitative treatment of inverting ratio, see section 4. In general, the bigger we set T_P, the fewer reaches we obtain, which causes the unstable matching by the over quantization of similarity indeces.

Next we explain the threshold T_D. Since the local correlation $c_{loc}(\boldsymbol{p}, \boldsymbol{J})$ at the given grid-point $\boldsymbol{p}$ is determined by the reaches which arise at $\boldsymbol{p}$ and the reaches are chosen to be of length $\leq T_D$, the magnitude of T_D determines the rate of contribution of low spatial-frequency components of the given image pair. Hence, the threshold T_D should be chosen in consultation with the level of detail of the given images. In what follows, we always use $T_D = 1$ for the simplicity of the arguments.

Now we explain about threshold T_B in the definition of the local similarity. Let us consider the probability that $\Pr(c_{loc}(\boldsymbol{p}, \boldsymbol{J}) = 1)$, provided we have eight reaches for the given $\boldsymbol{p} \in \boldsymbol{I}$. The graph of Fig.1 shows the the relationship between the constant inverting ratio X and $\Pr(c_{loc}(\boldsymbol{p}, \boldsymbol{J}) = 1)$, for $T_B = 1/8$ to $T_B = 7/8$. Except both extreme cases, each curve is sigmoid and we see $\Pr(c_{loc}(\boldsymbol{p}, \boldsymbol{J}) = 1) \gtrless X$ if X is sufficiently large or small, respectively. In the case $X = 0.5$ which means locally uncorrelated case, the correspoinding Y value is < 0.5 if $T_B \geq 5/8$. Hence, when we set $T_B \geq 5/8$, the contribution of the locally uncorrelated point to $c(\boldsymbol{I}, \boldsymbol{J})$ is stochastically low. When $T_B = 7/8$, the most part of the corresponding curve is below the line $Y = X$, which means $c(\boldsymbol{I}, \boldsymbol{J})$ does not goes high even non-inverting ratios is rather high. We used $T_B = 6/8$ for every experiment in the following.

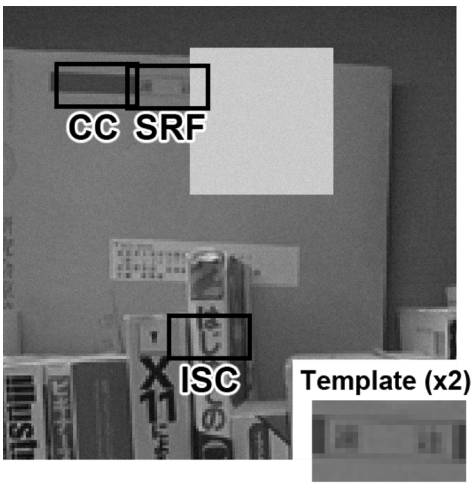

Figure 2: Experiment (stamp).

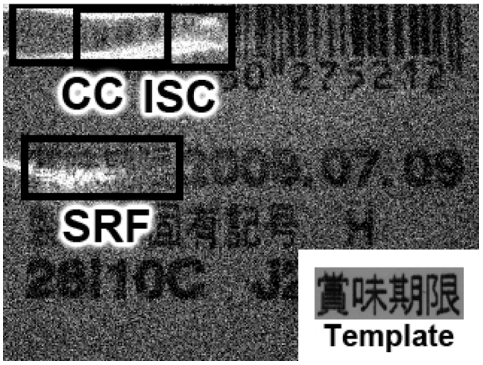

Figure 3: Experiment (food label).

3 Experiments

3.1 Experiment（Stamp）

Fig.2 shows the experimental results which demonstrate the difference of the features among ISC, CC and SRF. The template image is cut out from the source image. The matching target image is obtained by adding artificial occlusion and Gaussian noise with mean=0 and SD=5 to source image. The artificial occlusion is managed to cut in the region which has the high similarity with the template image. In this experiment, ISC fails in the template matching, though it is known to be robust for the variations of illumination and occlusions. The reason of the failure comes from the low contrast of the target image. The failure of CC comes from the occlusion. On the other hand, SRF showed the good results, which shows the effectiveness of the selection mechanism of noise-robust point pairs.

3.2 Experiment (Food Label)

Next, we performed the experiment in the harder situation. The source image of the target is the picture of a glossy food label, which has the highlight occlusion caused by mal-illumination. The target image for matching is generated by adding Gaussian noise with mean=0 and SD=38. The template image (100×35) for matching was prepared not to have highlight occlusion. The assumed situation for this experiment: handling a large number of labels in a short time. In this situation, high-speed shuttering is needed, which brings on the low SNR image input, caused by poor light intensity and large gain. But, the effort of intensity compensation by high-intensity illumination causes the strong highlight occlusion for the input images. So, we must make it for the noise-rich inputs.

Fig.3 shows the experimental results for SRF, CC, and ISC. The thresholds for the SRF are : $T_P = 50, T_D = 1,$ and $T_B = 0.75$. With this threshold, we obtained 1,824 reaches from the template. Both CC and ISC failed by disturbances and SRF provides successful matching, regardless of the disturbances such as noise and highlights in the target image. Fig.4 shows the correlation map of CC(a), ISC(b), and the similarity index map of SRF(c), respectively. The correlation map of CC has so many nearing local maxima. The correlation map of ISC is narrow-ranged and it is hard to find the maximum poit. The similarity index map of SRF, on the other hand, has sharp spike with overwhelming magnitude, which provides clear maximum across the whole similarity index map.

4 The Statistical Analysis of Similarity Index

In this section, we develop the mathematical model of inverting ratio, because from the definition, it is strongly related to the similarity index and we can expect the model to profile the rough behavior of the index. In what follows, we render the theoretical inverting ratio with some assumptions. Suppose the template image $\boldsymbol{I}$ is generated from $N(\mu, \tau^2)$ and the target image $\boldsymbol{J}$ from $\boldsymbol{I}$ by adding the noise $N(0, \sigma^2)$. Fix $\boldsymbol{p}$, $\boldsymbol{q}$ in the grid and let Ω be the set of all 4-tuples of value $(\boldsymbol{I}(\boldsymbol{p}), \boldsymbol{I}(\boldsymbol{q}), \boldsymbol{J}(\boldsymbol{p}), \boldsymbol{J}(\boldsymbol{q}))$ such that $|\boldsymbol{I}(\boldsymbol{p}) - \boldsymbol{I}(\boldsymbol{q})| \geq T_P$. The set Ω is assumed to be the event $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}) \neq \phi$. Let N to be the sub-event $\mathrm{srf}(\boldsymbol{p} \succ \boldsymbol{q}) \neq \mathrm{b}(\boldsymbol{p} \succ \boldsymbol{q})$. The inverting ratio r_{inv} is now formulated as the value $\Pr(N)$. Now we split Ω into two sub-event:

$$\Omega^+ := \{(x\,, y\,, x', y') \in \Omega \mid x - y \geq\ T_P\,\},$$

and

$$\Omega^- := \{(x\,, y\,, x', y') \in \Omega \mid x - y \leq -T_P\,\}.$$

Since $\Omega = \Omega^+ \cup \Omega^-$ and $\Omega^+ \cap \Omega^- = \phi$, $r_{inv} = \Pr(N|\Omega^+)\Pr(\Omega^+) + \Pr(N|\Omega^-)\Pr(\Omega^-)$. The symmetries of the distributions for the generation of the images lead us $\Pr(\Omega^+) = \Pr(\Omega^-) = 0.5$ and $\Pr(N|\Omega^+) = \Pr(N|\Omega^-)$. Hence, now we have $r_{inv} = \Pr(N|\Omega^+)$. Now we split the event Ω^+ into small pieces like $\Omega^+ = \bigcup_{k>T_P} \Omega_k^+$, where

$$\Omega_k^+ := \{(x\,, y\,, x', y') \in \Omega^+ \mid x - y = k\,\}.$$

Now then, we have

$$r_{inv} = \Pr(N|\Omega^+) = \sum_{k \geq T_P} \Pr(N|\Omega_k^+)\ \Pr(\Omega_k^+|\Omega^+). \tag{9}$$

Since $\boldsymbol{I}(\boldsymbol{p}) - \boldsymbol{I}(\boldsymbol{q}) = k$, we know that $\boldsymbol{J}(\boldsymbol{p}) - \boldsymbol{J}(\boldsymbol{q})$ follows $N(k, 2\sigma^2)$, and that

$$\Pr(N|\Omega_k^+) = \frac{1}{\sqrt{\pi}\,2\sigma} \int_{-\infty}^{-k} \exp\left(-\frac{u^2}{4\sigma^2}\right) du = \frac{1}{2}\,\mathrm{erfc}\left(\frac{k}{2\sigma}\right). \tag{10}$$

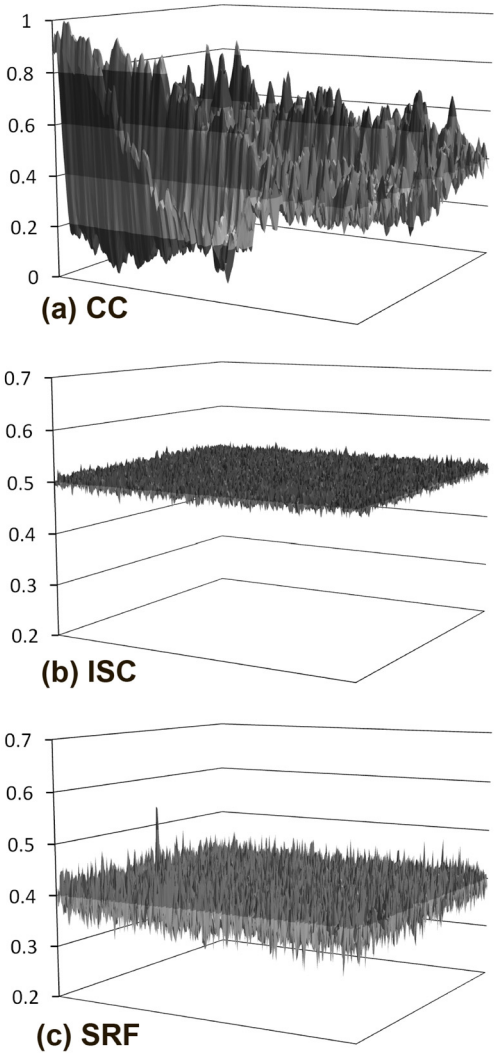

Figure 4: The correlation maps and the similarity map.

The probability $\Pr(\Omega_k^+|\Omega^+)$ is given by:

$$\Pr(\Omega_k^+|\Omega^+) = \frac{\int_k^{k+1} \exp\left(-\frac{u^2}{4\tau^2}\right) du}{\int_{T_P}^{\infty} \exp\left(-\frac{v^2}{4\tau^2}\right) dv} = \frac{\mathrm{erfc}\left(\frac{k}{2\tau}\right) - \mathrm{erfc}\left(\frac{k+1}{2\tau}\right)}{\mathrm{erfc}\left(\frac{T_P}{2\tau}\right)}. \quad (11)$$

Combining (9), (10) and (11), we obtain:

$$r_{inv} = \Pr(N|\Omega^+) = \frac{1}{2} \sum_{k \geq T_P} \mathrm{erfc}\left(\frac{k}{2\sigma}\right) \left(\frac{\mathrm{erfc}\left(\frac{k}{2\tau}\right) - \mathrm{erfc}\left(\frac{k+1}{2\tau}\right)}{\mathrm{erfc}\left(\frac{T_P}{2\tau}\right)}\right) \quad (12)$$

$$\fallingdotseq \frac{1}{2\tau\,\mathrm{erfc}\left(\frac{T_P}{2\tau}\right)\sqrt{\pi}} \int_{T_P}^{\infty} \mathrm{erfc}\left(\frac{k}{2\sigma}\right) \exp\left(-\frac{k^2}{4\tau^2}\right) dk. \quad (13)$$

Before the last expression, we introduced two approximations: first, replaced the difference by its derivative, and, secondly, replaced the summation by an integralation. Fig.5 shows the non-inverting ratio to SD of the added noise, with the constraint $T_P = 2\sigma$, for both approximating expression (13) and experimental plots. The plots perform quite good fits to the theoretical line. The exact form (12) of r_{inv} leads us its major estimate : $\frac{1}{2}\,\mathrm{erfc}\left(\frac{T_P}{2\sigma}\right)$. When we take $T_P = 2\sigma$, the upper bound of r_{inv} is nearly 0.079. Fig.5 also shows the non-inverting ratio of ISC. Compared to ISC, the non-inverting ratios of SRF are kept high even for hard noises, and this fact is the source of superiority of SRF over ISC.

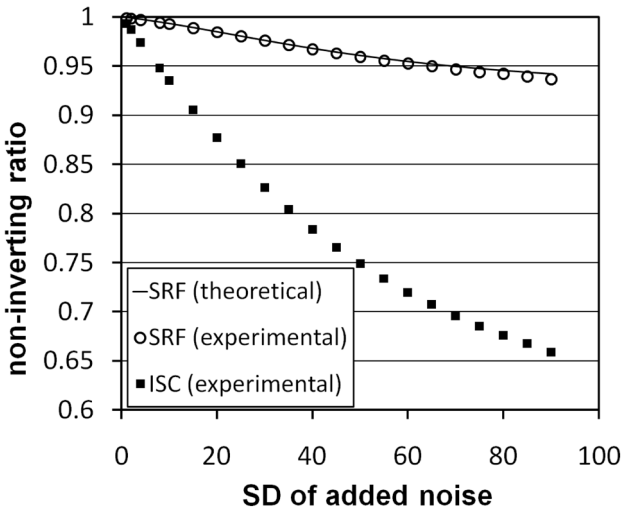

Figure 5: Non-inverting ratio to added noise SD.

5 Conclusion

We have proposed the novel method for noise-robust template matching technique. The proposed method is shown to be robust, under the strong noise, occlusions including highlights. We also showed the experimental superiority of the method over some of the existing techniques, namely CC and ISC. We also developed the mathematical model for the inverting ratio, and revealed its relationship to the noise.

References

[1] E. I. Barnea and H.F.Silverman.: "A class of Algorithms for Fast Digital Image Registration" *IEEE Transactions on Computers*, vol.C-21, pp.179-186, 1972.

[2] R.L. Lillestrand.: "Techniques for change detection" *IEEE Transactions on Computers*, vol.C-21, no.7, pp.654-659, 1972.

[3] S. Kaneko, et al.: "Robust image registration by increment sign correlation" *Pattern Recognition*, vol.35, no.10, pp.2223-2234, 2002

[4] K. Tsukiyama, F. Saitoh: "Image Template Matching Using Directable Increment Sign Correlation" *IEEJ Trans.*, vol.124-D, No.10, pp.994-1000, 2004.

[5] Y. Satoh, et al.: "Robust Picture Matching Using Selective Correlation Coefficient" *IEICE Tech. Rep.*, vol.100, no.134, PRMU2000-31, pp.57-62, 2000

[6] K. Iwata, et al.: "Robust Background Subtraction by Statistic Reach Feature" *IEICE Tech. Rep.*, vol. 108, no. 263, PRMU2008-105, pp.97-102, 2008