On Optimizing Template Matching via Performance Characterization

Tony X. Han∗ Visvanathan Ramesh †
∗Beckman Institute and ECE Department
University of Illinois at Urbana-Champaign
405 N. Mathews Ave., Urbana, IL 61801
{xuhan, huang}@ifp.uiuc.edu

Ying Zhu† Thomas S. Huang∗
†Real-Time Vision and Modeling Department
Siemens Corporate Research
755 College Road East, Princeton, NJ 08540
{Visvanathan.Ramesh, Ying.Zhu}@siemens.com

Abstract

Template matching is a fundamental operator in computer vision and is widely used in feature tracking, motion estimation, image alignment, and mosaicing. Under a certain parameterized warping model, the traditional template matching algorithm estimates the geometric warp parameters that minimize the SSD between the target and a warped template. The performance of the template matching can be characterized by deriving the distribution of warp parameter estimate as a function of the ideal template, the ideal warp parameters, and a given noise or perturbation model. In this paper, we assume a discretization of the warp parameter space and derive the theoretical expression for the probability mass function (PMF) of the parameter estimate. As the PMF is also a function of the template size, we can optimize the choice of the template/block size by determining the template/block size that gives the estimate with minimum entropy. Experimental results illustrate the correctness of the theory. An experiment involving feature point tracking in face video is shown to illustrate the robustness of the algorithm in a real-world problem.

1. Introduction

Template/Block matching and its variations have been widely used in tasks such as visual tracking [1, 2, 3], video coding [4], layered motion estimation [5], image alignment [6], 3D reconstruction [7, 8], medical imaging [9] and image mosaicing[10]. By assuming a parameterized warping model, the traditional template matching algorithm determines the optimal warping parameter that minimizes the SSD (sum of square differences) between the target and the warped template. These methods tackle template matching as an optimization problem and solve it by various hill climbing methods such as gradient descent, Newton, Gauss-Newton, steepest descent and Levenberg-Marquardt. Practical application of template matching requires the selection of appropriate scales and/or features for which the matching can happen reliably. This is typically done through the use of a multi-scale approach and/or choice of textured features with minimal tracking errors. For instance, in Shi and Tomasi’s widely influential work [11], they discuss template matching under affine warping assumption. Dissimilarity, measured as the SSD between back warped target and template, is used to evaluate the matching performance. Image patches with sufficient texture produce reliable solution and thus treated as good feature. More recently, Kaneko and Hori [12] provide an upper bound of average matching error for translation and suggest the use of features that have a small upper bound.

A number of papers have been published on evaluation of motion estimation algorithms. Most of them have been in the context of study and influence of errors in motion estimation on Structure-from-Motion. Fermuller and Aloimonos [13] study the bias and variance in flow estimation and relate it to explain many of the illusions that are perceived by the human visual system. Aguiar and Moura [14] study the stability of motion estimates as a function of the brightness pattern in the image and show the relationship between the image gradient magnitudes and the stability/convergence of the motion estimation algorithm. Both these papers study how the bias and covariance of estimated parameters relate to the image measurements and therefore can be viewed as studying Gaussian approximations to the error statistics of the motion estimate.

Our paper is concerned with deriving the distribution of the estimated motion/warping parameters as a function of the image brightness values, noise parameter, and true warping parameter. The parameterized warping model in our framework can be any one-to-one warping (bijection). The warping parameter estimate is treated as one of possible set of warping parameters with a probability mass function (PMF). The PMF of the warp estimate is derived as a function of the ideal image brightness values, the noise model parameters, and the true warp parameters. In a given block/template matching problem, this PMF is also a function of the size of the block/template used. The entropy of the estimate PMF is a measure of uncertainty of the estimate. We use this fact to choose the block/template size that gives the minimum entropy across the range of choices...
for the size parameter.

The main contribution of this paper is providing a complete probabilistic characterization for template matching under any bijection warping model. Unlike past work where typically a Gaussian approximation is used for modeling the error distribution for geometric warp parameters (i.e. translation, affine parameter estimates for example), we derive the complete distribution of the estimate (Our analysis however uses the assumption of discretization of the warp parameter space). A major difference with other papers is the explicit modeling of underlying correlation between evaluated SSD scores for each target vs. image template comparison. To enable faster computation of the PMF we provide an approximation that exploits this correlation structure of the random variables involved in the estimation algorithm. The entropy of the estimate’s PMF is used as a measure to decide the right scale at which the template matching should be performed. The theory is validated on simulated examples and its application to feature tracking in a face video is shown to illustrate its robustness.

This paper is organized as follows. We first provide the performance characterization of template matching in section 2. Section 3 explains in detail about how to apply the performance characterization framework to the data-driven template selection. Experiment results and the discussion are given in section 4. And section 5 concludes the paper.

2. Performance Characterization of Template Matching

Suppose \( f_1(x) \) and \( f_2(x) \) are two frames of observed image in the input image sequence, where \( x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d \). If \( d = 2 \), \( f_1(x) \) and \( f_2(x) \) are 2D images. When \( d = 3 \), \( f_1(x) \) and \( f_2(x) \) are volumetric images with voxel \( x \in \mathbb{R}^3 \).

The template or image patch we want to match between frames is denoted as \( g(x) \) for \( x \in \Omega \). Here \( \Omega = \{ x \in \mathbb{R}^d : \text{The pixel } x \text{ lies inside the template} \} \) and the size of the template is \( W = |\Omega| \), i.e. the cardinality of \( \Omega \). We assume that the template is contained in a “clean” image, \( f(x) \), defined as:

\[
f(x) = \begin{cases} 
g(x) & x \in \Omega, \\
0 & \text{otherwise.} 
\end{cases}
\]

By modeling the background clutter, illumination, and the camera noise as the AWGN, the first frame of the observed image sequence is

\[
f_1(x) = f(x) + \eta(x),
\]

where \( \eta(x) \) is the zero mean Gaussian random variable with the variance denoted as \( \sigma^2 \). The \( \eta(x) \)s are independent to each other, i.e.

\[
Cov[\eta(x_1), \eta(x_2)] = \begin{cases} 
\sigma^2 & x_1 = x_2 \in \mathbb{R}^d, \\
0 & \text{otherwise.}
\end{cases}
\]

We assume the parameterized template warping model between two images is a bijection \( \tilde{\alpha} : \mathbb{R}^d \rightarrow \mathbb{R}^d \). If the warping model is pure translation, we have \( \tilde{\alpha}(x) = x + t \). Here \( t \in \mathbb{R}^d \) is the translation vector. If the template undergoes affine warping, we have \( \tilde{\alpha}(x) = \tilde{A}x + \tilde{b} \), where \( \tilde{A} \in GL(d, \mathbb{R}) \) and \( \tilde{b} \in \mathbb{R}^d \). Here \( GL(d, \mathbb{R}) \) denote the real general linear group on \( \mathbb{R}^d \). So the second frame observed is:

\[
f_2(x) = f(\tilde{\alpha}(x)) + \xi(x).
\]

Here \( \xi(x) \) is the AWGN in the second frame of the image. We assume the noise is stationary and uncorrelated to each other among frames. Thus \( \xi(x) \) has exactly the same distribution as \( \eta(x) \) and their cross correlation

\[
E[\eta(x_1)\xi(x_2)] = 0.
\]

for \( x_1, x_2 \in \mathbb{R}^d \). For convenience, we slightly abuse the notation \( \tilde{\alpha} \). We use it to represent both the “template warping model” and the “parameters” of the model. This should cause no confusion because the meaning of \( \tilde{\alpha} \) is clear within context.

Given \( f_1(x) \) and \( f_2(x) \), the traditional template matching is to estimate the parameters of the template warping model \( \tilde{\alpha} \) minimizing SSD between warped template and target. In our framework, we are trying to estimate the \( \tilde{\alpha} \) with the highest matching probability. We first give some preliminary definitions for this formulation. And the definition of the matching probability will be given accordingly. From equation (1,2), we get

\[
\begin{align*}
\eta & = f(x) - f(\tilde{\alpha}(x)) \\
\tilde{\alpha}(x) & = f_1(x) - \eta(x) \\
f_2(x) & = f_1(x) - \eta(x) + \xi(x)
\end{align*}
\]

Suppose one possible estimate of \( \tilde{\alpha} \) is \( \alpha \). We can rectify the “clean” image in the second frame, \( f(\tilde{\alpha}(x)) \), by warping it back using \( \alpha^{-1} \), i.e. \( f(\tilde{\alpha}^{-1}(x)) = f_2(\alpha^{-1}(x)) - \xi(\alpha^{-1}(x)) \). If the estimate is good, the rectified “clean” image \( f(\tilde{\alpha}^{-1}(x)) \) is close to the original “clean” image \( f(x) \). Ideally, if \( \alpha = \tilde{\alpha} \) the rectified version is the same as the “clean” image in the first frame, i.e. \( f(\tilde{\alpha}^{-1}(x)) = f(x) \). The similarity between \( f(x) \) and \( f(\tilde{\alpha}^{-1}(x)) \) indicates the goodness of the estimation. Therefore we define the square of the \( l_2 \) distance between the “clean” image \( f(x) \) and its rectified version, \( f(\tilde{\alpha}_0^{-1}(x)) \) as a function depends on \( \alpha \):

\[
d(\alpha) = \sum_{x \in \Omega} (f(x) - f(\tilde{\alpha}_0^{-1}(x)))^2
\]

\[
= \sum_{x \in \Omega} ((f_1(x) - \eta(x)) - (f_2(\alpha^{-1}(x)) - \xi(\alpha^{-1}(x))))^2,
\]

(5)

Here \( \alpha \) can be all possible parameterized warping of the image patch between two frames. If \( \alpha = \tilde{\alpha} \), \( d(\alpha) = 0 \).
Reader should also notice that since we only have access to \(f_1(x)\) and \(f_2(\alpha^{-1}(x))\), \(d(\alpha)\) is a random variable. It is different from SSD since the SSD is a determined number. We denote the set of all possible warping parameters that \(\alpha\) can take as \(\tilde{\alpha} = \{\alpha^{(j)}\}\), where \(j = 1, \ldots, A\). Here \(A = |\mathcal{A}|\).

We denote the matching probability of “the true warping parameters \(\tilde{\alpha} = \alpha^{(j)}\)” as \(Pr[\alpha^{(j)}]\). From equation (5), we have \(d(\tilde{\alpha}) < d(\alpha^{(j)})\). This means the event “the true warping parameters \(\tilde{\alpha} = \alpha^{(j)}\)” is equivalent to the event “\(d(\tilde{\alpha}) < d(\alpha^{(j)})\)” for all \(\alpha\).

\[
Pr[\tilde{\alpha} = \alpha^{(j)}] \triangleq Pr[\alpha^{(j)}] = Pr\left[\bigcap_{i \neq j} (d(\alpha^{(j)}) < d(\alpha^{(i)}))\right] \quad (6)
\]

The matching probability \(Pr[\alpha^{(j)}]\) is determined by not only the similarity between the template and the warped-back target under warping \(\alpha^{(j)}\), but also their dissimilarity under other possible warpings.

Therefore the template matching result, i.e. the estimation of the warping parameters \(\tilde{\alpha}\) is

\[
\tilde{\alpha} = \arg \max_{\alpha^{(j)}} Pr[\alpha^{(j)}], \quad (7)
\]

where \(j = 1, \ldots, A\). Since \(\eta(x)\) and \(\xi(x)\) are IID Gaussian random variables, \(d(\alpha)\) is a random variable distributed as \(\chi^2(W; \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2)\) (non-central Chi-squared distribution with \(W = |\Omega|\) degrees of freedom). In order to estimate the warping parameters, we need to evaluate equation (6) to determine the \(\alpha^{(j)}\) with highest matching probability. However, it is extremely hard to evaluate equation (6) when \(d(\alpha)\)'s are non-central Chi-squared distributed and are dependent on each other. Also, as the cardinality of \(\mathcal{A}\), \(A\) is usually very large (> 100). This makes the evaluation problem even more difficult due to the curse of dimensionality. By noticing that the non-centrality parameter \(\lambda = \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2\) is usually large enough, we can approximate the joint distribution of \(d(\alpha)\)'s as a multivariate Gaussian with the uniform approximation error \(O(\lambda^{-1/2})\) (Refer to page 466 of [15]). Let \(D_\alpha = (d(\alpha^{(1)}), d(\alpha^{(2)}), \ldots, d(\alpha^{(A)}))^T\), we approximate \(D_\alpha\) as \(G_\alpha \sim N(\mu_\alpha, \Sigma_\alpha)\) such that \(\mu_\alpha = E[D_\alpha]\) and \(\Sigma_\alpha = Cov[D_\alpha]\). For all \(\alpha \in \mathcal{A}\), i.e. all possible warpings, the expectation, variance, and covariance of \(d(\alpha)\) can be approximated as:

\[
E[d(\alpha)] = 2W\sigma^2 + \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2 \quad (8)
\]

\[
Var[d(\alpha)] = 8W\sigma^4 + 8\sigma^2 \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2 \quad (9)
\]

When \(\alpha \neq \beta\) and \(\alpha, \beta \in \mathcal{A}\)

\[
Cov[d(\alpha), d(\beta)] = 2W\sigma^2 + 2|\Omega_\alpha \cap \Omega_\beta|\sigma^4
\]

\[
+ 4\sigma^2 \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x))) (f_1(x) - f_2(\beta^{-1}(x)))
\]

\[
+ 4\sigma^2 \sum_{x \in \Omega, \alpha \neq \beta} (f_1(\alpha(x)) - f_2(x)) (f_1(\beta(x)) - f_2(x)),
\]

(10)

where we define the \(\alpha\)-warping set of \(\Omega\) as \(\Omega_\alpha \triangleq \{x \in \mathbb{R}^d : \alpha^{-1}(x) \in \Omega, \alpha \in \mathcal{A}\}\). Similarly \(\Omega_\beta \triangleq \{x \in \mathbb{R}^d : \beta^{-1}(x) \in \Omega, \beta \in \mathcal{A}\}\). Readers can refer to the appendix for the derivation of equation (8)(9)(10). Through equations (8,9,10), we fit the Gaussian random vector \(G_\alpha\) to \(D_\alpha\) by granting \(G_\alpha\) the same mean and covariance matrix as \(D_\alpha\). For small \(A (A < 10)\), we can compute equation (6) according to equation (36) at page 51 of [16]. But usually \(A \gg 10\). We have three possible ways to evaluate equation (6): 1. **Dimension reduction.** and 2. **Sampling based method.** We can approximate \(G_\alpha\) by a low dimensional Gaussian random vector \(G_L\) (\(L \ll A\)) by eigen-decomposition of the covariance matrix \(\Sigma_\alpha\). Unfortunately, since \(A\) is very big (> 100), simple dimension reduction by projection to low dimension space won’t have a good approximation. The sampling based methods use Monte Carlo techniques to evaluate equation (6). However, sampling methods are infeasible when \(A\) is large.

3. **Factorization based method.** Due to the limitations of the previous two approaches, we further add another assumption to make the evaluation of equation (6) more tractable. The assumption, enlightened by the Naive Bayesian method [17], is that the matching probability in equation (6) can be pairwise factorized as

\[
Pr[\alpha^{(j)}] = Pr\left[\bigcap_{i \neq j} (d(\alpha^{(j)}) < d(\alpha^{(i)}))\right]
\]

\[
\simeq K \prod_{i \neq j} Pr[d(\alpha^{(j)}) < d(\alpha^{(i)})] \triangleq Q[\alpha^{(j)}], \quad (11)
\]

where \(K\) is a normalizing constant to ensure \(\sum_{\alpha} Pr[\alpha^{(j)}] = 1\). We can easily compute this probability by knowing the expectation and covariance of \(d(\alpha)\) computed from equations (8,9,10).

Let the random variable \(Y_{ij} = d(\alpha^{(j)}) - d(\alpha^{(i)})\), \(\tilde{Y}_{ij}\) is the difference of two dependent non-central Chi-square random variable. We use a Gaussian random variable \(\tilde{Y}_{ij}\) to approximate \(Y_{ij}\) such that \(E[Y_{ij}] = E[\tilde{Y}_{ij}]\) and \(Var[Y_{ij}] = Var[\tilde{Y}_{ij}]\). Therefore we have

\[
E[Y_{ij}] = E[d(\alpha^{(j)})] - E[d(\alpha^{(i)})]
\]

\[
Var[Y_{ij}] = E\left[(d(\alpha^{(j)}) - d(\alpha^{(i)}) - E[d(\alpha^{(j)})] + E[d(\alpha^{(i))}]^2\right] - Var[d(\alpha^{(j)})] + Var[d(\alpha^{(i)})] - Cov[d(\alpha^{(i)}), d(\alpha^{(j)})]
\]

3
Finally the matching probability in equation (6) is approximated as:
\[ \Pr[\alpha(i)] \simeq Q[\alpha(i)] = K \prod_{i \neq j} \Pr[Y_{ij} < 0] \]
\[ = K \prod_{i \neq j} \Phi\left(\frac{-E[Y_{ij}]}{\sqrt{\text{Var}[Y_{ij}]}},\right), \quad (12) \]
Here \( \Phi(x) \) is the Cumulative Distribution Function (CDF) of \( \mathcal{N}(0,1) \).

In order to apply the above template matching algorithm to the 2D (image) or 3D data (volumetric image), a reasonable run-time is required for the algorithm. Thus the factorization approach is a good computable implementation of our formulation. Experiments results in section 4 show that the approximation is reasonable and facilitate the evaluation of equation (6) a lot. Our matlab code for approach 3 runs 0.04 second for 1d signal in an PIII 1.1GHz laptop. And our implementation in C++ code runs more than 100 times faster.

Monte Carlo simulation to validate the factorizable assumption
To validate the assumption that the matching probability \( \Pr[\alpha(i)] \) can be approximated as \( Q[\alpha(i)] \) defined in equation (11) by pairwise factorization, we run Monte Carlo simulations and measure the similarity between \( \Pr[\alpha(i)] \) and \( Q[\alpha(i)] \). We use the Bhattacharyya coefficient to measure the similarity of two distributions. For a discrete random variable, the Bhattacharyya coefficient between the two probability mass function (PMF) is defined as \( \rho = \sum_{j=1}^{M} \sqrt{\Pr[\alpha(i)] Q[\alpha(i)]} \). Therefore \( 0 \leq \rho \leq 1 \) and the closeness of \( \rho \) to 1 reflects the similarity between the two PMFs. Particularly, when the two PMFs are identical, \( \rho = 1 \). For various kinds of 1d signal such as rectangular wave, triangle pulse, sinusoidal signal, circular wave, etc., we run 10000 times of Monte Carlo simulation under different SNR \( \frac{20\log_{10}(E[f(x)/\eta(x)])}{} \). Figure 1(a) and 1(b) shows 1 simulation for a three pulse triangle wave with SNR = 5.4921db. It can be seen that \( \Pr[\alpha(i)] \) and \( Q[\alpha(i)] \) shown in figure 1(b) are very close. Their Bhattacharyya coefficient is 0.9817. The average Bhattacharyya coefficient over all signals with different SNR is shown in figure 1(c). The two PMFs are very close when the SNR is above 2db and are identical when SNR is above 6db. This simulations show that our approximation is very good for SNR larger than 2db. From our experience, template matching is almost impractical when SNR is lower than 2db.

3. Data-Driven Template Selection
For template matching, one of the most important questions that we need to answer is “How to choose a template that produces reliable matching results?” Suppose we have many available templates that contain the image content we want to match, we need a criterion to select the best template which gives the most reliable matching result. Denote the family of all available templates containing the content of interest as \( \Omega = \{\Omega_k\} \), where \( k = 1, \ldots, T \) and \( T = |\Omega| \). Here \( \Omega_k \) is an image patch and a template candidate.

For each template \( \Omega_k \), the distribution of the warp parameter estimate can be computed by equation (11). When we use \( \Omega_k \) as template to match, the matching probability \( \Pr[\alpha(i)] \) for warping parameters \( \alpha(i) \) is denoted as \( P_k^{(i)} \). For template \( \Omega_k \), its entropy of the estimate is defined as

\[ H(\Omega_k) = -\sum_{j=1}^{A} p_k^{(j)} \log(p_k^{(j)}) \quad (13) \]

Thus the optimal template \( \hat{\Omega} \), which gives the least entropy of the estimate i.e. the least matching ambiguity can be selected as:

\[ \hat{\Omega} = \arg \min_{\Omega_k} H(\Omega_k), \quad (14) \]

where \( k = 1, \ldots, T \). By finding the template that minimizes the entropy of the estimate, the data driven template selection is achieved. The optimal template chosen by this criterion gives the matching result with least ambiguity.

4. Experiment Results
Figure 2 shows the matching performance evaluation and template selection for a 1d signal consists of two pulses of width 1, apart from each other by 3 units. The two frames of 1d signal are shown in figure 2(b). The original 1d signal translation 2/3 unit between two frames. The translation is very precisely estimated using the template selected by the proposed framework. The matching probability is shown in figure 2(c). The original signal is contaminated by AWGN with \( \sigma = 0.04 \). The variance is accurately estimated by frequency domain method. Since the energy in
the high frequency is mainly from AWGN, we estimate the Power Spectrum Density (PSD) of the high frequency and solve the variance using Parseval’s equation. In this experiment, the templates set contains the 1d window starting at origin with width ranging from 0.1 to 9. In order to estimate the translation of the pulse, the optimal template should be the template big enough to contain all the discriminative feature of the signal and tight enough to exclude the non-discriminative feature. In this case, since the AWGN is zero mean, all the signal value that is around 0 is non-discriminative feature. The bigger the signal value, the more discriminative. In order to include all the discriminative features while excluding non-discriminative features, the optimal template should only contain the two pulses and rejects other parts of the signal beyond the pulses. Thus the optimal template should be the window just wide enough to contain the 2-pulse signals, i.e. \( W = 1 + 3 + 1 = 5 \). Figure (2(a)) shows that our estimation of the block size is 4.9, which is much better than the template selected by minimizing normalized SSD. The template minimizing normalized SSD has the width \( W_{SSD} = 0.1 \). Figure (2(d)) shows that for template with width less than 4, there are matching errors. This is quite reasonable because the two pulses are similar. If the template contains only 1 pulse, matching algorithm cannot tell which of the two pulses in the second frame corresponds to the pulse contained in the template. The big drop in the entropy of the estimate (shown in figure 2(a)) for template with width larger than 4.2 shows that our method suggests to take the template with width larger than 4 to disambiguate the confusion, which is a very nice property. In this case the choice of template really affects the matching result a lot. The template selected by minimizing the entropy of the estimate is a template big enough to avoid matching error and small enough to exclude the non-discriminative feature.

Figure 2: Scale selection for signal containing 2 pulse with width 5.0

Figure 3 shows the performance evaluation and template selection for matching a \( 16 \times 16 \) face patch that is inserted into the picture of a background scene. The patch is translated by \( (25, 15) \) units. The variance of the assumed AWGN is estimated by 2d Parseval’s equation. Our template family in this experiment is all rectangular templates that are centered in the image. In order to include all the discriminative features while exclude non-discriminative features, the optimal template should be the one that only contains the face and rejects other smooth background parts. Thus the size of the optimal template (the ground truth) should be \( 16 \times 16 \). The entropy of the estimate shown in figure 3(c) and 3(d) suggests size of the optimal template is \( W_y = 13 \) and \( W_x = 15 \). When the optimal template is used, the matching probability shown in figure 3(e) gives the accurate estimation of the translation: \( (26, 14) \). Reader should notice that the choice of the template affects the matching results a lot. Figure 3(f), shows us that only the template with size from \( 10 \times 10 \) to \( 16 \times 16 \) gives correct matching result. If we use the template of other sizes, there will be matching error. Figure 4 shows the results of matching a \( 24 \times 8 \) image patch encapsulated in the same picture. The image patch is formed by 2 heavily down sampled face patches (\( 8 \times 8 \)) and a \( 8 \times 8 \) black square in figure 4(a). The template family contains rectangular templates with various sizes. The template lies at the center of the first image. The entropy of the estimate shown in figure 4(c) and 4(d) suggests template size \( W_y = 5, 7 \) and \( W_x = 7 \). This is not consistent with patch size \( 24 \times 8 \). After we carefully examine the patch, we find that due to heavy downsampling, the \( 8 \) by \( 8 \) face patch is very smooth and similar to background intensities, therefore the pixel in the face patches are not discriminative features anymore. So our algorithm actually suggests that one should use the \( 8 \times 8 \) template containing the black square, which is very discriminative for template matching. Figure 4(e) shows that the suggested template of size \( 5 \times 7 \) or \( 7 \times 7 \) gives tiny matching error of 1 unit. From figure (4(e)), shows that if we choose the image patch of \( 24 \times 8 \) as the template, there will be a big matching error. These results show that the template selected by minimizing the entropy of the estimate gives accurate matching results with least ambiguity.

Figure 5 and 6 show the results of matching performance evaluation and template selection by applying our framework to the standard talking face video. The video can be downloaded at http://www.isbe.man.ac.uk/~bim/data/talking_face/talking_face.html

The talking face video consists of 5000 frames taken from a video of a person engaged in conversation. 68 points are selected as feature points on the face. For each frame, the ground truth positions of the feature points are labeled semi-automatically by AAM[18] and human checking. We will compare the template matching results with the ground truth. The entropy of the estimate is used to evaluate the
In this paper we have derived the expression for the probability mass function (pmf) of motion parameter estimates as a function of the input images, the true motion parameters, and additive white gaussian noise model for sensor perturbations. The optimal choice of scale parameters for point. The images in the left column show the matching results using prefixed $25 \times 25$ template around each feature point. The templates drift away after 150 frames if we use the prefixed template around each feature point. In contrast, none of the entropy-minimizing template drifts away for all the feature points. (Please refer to the video file) Figure 7 shows the template matching errors (i.e. the Euclidean distance from the matched template locations to the ground truth of 68 points) for various templates. Figure 7 substantiate our assumption that choosing the template sizes to minimize entropy of the estimate gives the best matching results. These results show the advantage of our probabilistic characterization framework for template selection and template matching evaluation.

5. Conclusion

Figure 3: Template selection and matching evaluation for 2d image patch encapsulated in real picture.

Figure 4: Template selection and matching evaluation for heavily downsampled image patch.
Appendix

Deriving the expectation, variance and covariance of $d(\alpha)$ as functions of the observation

$$E[d(\alpha)] = E \left[ \sum_{x \in \Omega} (f(x) - f(\alpha^{-1}(x)))^2 \right]$$

$$= \sum_{x \in \Omega} E \left[ (f_1(x) - \eta(x)) - (f_2(\alpha^{-1}(x)) - \xi(\alpha^{-1}(x)))^2 \right]$$

$$= 2W\sigma^2 + \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2$$

$$\text{Var}[d(\alpha)]$$

$$= E \left[ (d(\alpha) - E[d(\alpha)])^2 \right]$$

$$= E \left[ -2W\sigma^2 + \sum_{x \in \Omega} (f_1(x) - \eta(x) - f_2(\alpha^{-1}(x)) + \xi(\alpha^{-1}(x)))^2 \right]$$

$$= E \left[ -2W\sigma^2 + \sum_{x \in \Omega} -2(f_1(x) - f_2(\alpha^{-1}(x)))(\eta(x) - \xi(\alpha^{-1}(x))) \right.$$  

$$+ (\eta(x) - \xi(\alpha^{-1}(x)))^2 \bigg]$$

$$= 4W^2\sigma^4 - 8W^2\sigma^2 + E \left[ \left( \sum_{x \in \Omega} 2(f_1(x) - f_2(\alpha^{-1}(x)))(\eta(x) - \xi(\alpha^{-1}(x))) \right)^2 \right]$$

$$+ E \left[ \left( \sum_{x \in \Omega} \eta(x) - 2(\eta(x)\xi(\alpha^{-1}(x)) + \xi^2(\alpha^{-1}(x))) \right)^2 \right]$$

$$= -4W^2\sigma^4 + 8W^2\sigma^2 \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2 + 4W(W - 1)\sigma^4$$

$$+ 3W^2\sigma^4 + 3W\sigma^2 + 2W^2\sigma^2 + 4W^4\sigma^4$$

$$= 8W^2\sigma^4 + 8W^2\sigma^2 \sum_{x \in \Omega} (f_1(x) - f_2(\alpha^{-1}(x)))^2$$

We compute the covariance for two different cases. If $\alpha = \beta$, $\text{Cov}(d(\alpha), d(\beta)) = \text{Var}(d(\alpha)) = \text{Var}(d(\beta))$. If $\alpha \neq \beta$, ...
References


