Computer Vision and Image Understanding **78,** 84–98 (2000)
doi:10.1006/cviu.1999.0829, available online at http://www.idealibrary.com on IDEAL®

Robust Image Matching under Partial Occlusion and Spatially Varying Illumination Change

Shang-Hong Lai

Imaging and Visualization Department, Siemens Corporate Research, Princeton, New Jersey 08540
E-mail: lai@scr.siemens.com

Received February 12, 1999; accepted November 5, 1999

Image matching is one of the most important tasks in computer vision. Most existing methods cannot achieve precise pattern matching under spatially varying illumination variations and partial occlusions. In this paper, we explicitly model spatial illumination variations as low-order polynomial functions in an energy minimization framework. Data constraints for the alignment and illumination parameters are derived from the first-order Taylor series approximation of the generalized brightness assumption with low-order polynomials used for modeling spatial illumination variations. We formulate the parameter estimation problem in a weighted least-squares framework by incorporating the influence function from robust estimation to derive an iterative reweighted least-squares algorithm. A dynamic weighting scheme, which combines the factors from the influence function, a measure of consistency between image gradients, and nonlinear image intensity sensing characteristics is used to improve the robustness of the image matching. In addition, a selective constraint sampling and an estimation-warping alternating strategy are used in the proposed algorithm to improve the efficiency and accuracy of the estimation. We have successfully applied the proposed algorithm to estimate affine transformations under partial occlusion and spatially varying illumination change for various industrial inspection tasks. Experimental results are shown to demonstrate the robustness, efficiency, and accuracy of the algorithm.

1. INTRODUCTION

Image matching or image alignment is one of the most fundamental tools in many areas of computer vision, including target tracking, motion analysis, object recognition, and automated visual inspection. For target tracking, image matching is used to accurately locate target objects in temporal image sequences. Various image matching techniques provide point or region correspondence in an image sequence for image motion analysis. An essential step for object recognition is pose determination, which can be accomplished

1077-3142/00 $35.00

by image alignment techniques. In automated visual inspection, image alignment between the inspection image and the reference image is usually the first and the most crucial step for most inspection problems.

A large number of image matching techniques [1] have been reported in literature to improve the accuracy, generality, robustness and speed of image matching. They can be classified into two categories, the feature-based matching approach [2–3] and the intensity-based matching approach [4–7]. The feature-based matching approach requires reliable feature extraction as well as robust feature correspondence to overcome missing feature and outlier problems due to partial occlusion. Its main advantage is robustness against illumination changes. The intensity-based matching approach is primarily based on the SSD (sum of squared differences) formulation [4], which does not require feature extraction or direct correspondence between two sets of features. However, this approach is more sensitive to illumination changes than the feature-based approach. In addition, the conventional SSD-based formulation is not robust against occlusion.

Recently, Black and Jepson [6] presented an object matching algorithm based on robust estimation and eigenspace projection to recover affine transformations for an object under different views. The eigenspace is computed from a collection of images for the same object under different views. Robust estimation is used in their work to allow for partial occlusion. It is not clear if this method can work under different illumination conditions since they did not take the issue of illumination variations into account. Hager and Belhumeur [7] proposed an efficient region matching and tracking algorithm based on a similar robust framework. They modeled the illumination changes into the SSD formulation by using a low-dimensional linear subspace determined from several images of the same object under different illumination conditions. The main disadvantage of this algorithm is the need of several images of the same object under different illumination conditions to compute the linear subspace before the tracking process. Viola and Wells [8] proposed a very general image alignment technique based on maximization of mutual information. However, this method is not robust against partial occlusions or spatially varying illumination changes.

In this paper, we propose a robust and efficient image matching algorithm by explicitly modeling the spatially varying illumination multiplication and bias factors as low-order polynomials. The problem of estimating the global geometric transformation and the polynomial illumination variations is formulated in a weighted least-squares framework. We incorporate the influence function from robust estimation to derive an iterative reweighted least-squares algorithm. A dynamic weighting scheme is used in our iterative algorithm to improve the robustness of the image matching by combining the factors from the influence function, a measure of consistency between image gradients, and the nonlinear image intensity sensing characteristics. In addition, we applied a selective constraint sampling scheme and an estimation-warping alternating strategy to improve the efficiency and accuracy of the proposed algorithm. The affine image matching is accomplished by an iterative numerical solution derived from the reweighted least-squares method for the minimization of the energy function from robust estimation. This framework can generally be extended to other global transformations, such as planar transformation or perspective transformation. In addition, we can always partition the image region into several subregions and assume an affine transformation for each subregion to achieve a piecewise affine transformation model for the recovery of nonrigid transformations.

The remainder of this paper is organized as follows. In the next section, we formulate the image matching as an energy minimization problem using polynomial models for

spatially varying illumination changes. This energy minimization formulation is modified by incorporating robust estimation and an additional weighting scheme in Section 3. Subsequently, a modified iterative reweighted least-squares algorithm based on a selective constraint sampling scheme, a hierarchical nearest-neighbor search method for providing an initial guess, and an estimation-warping alternating strategy for improving accuracy is described in Section 4. We demonstrate the accuracy, efficiency, and robustness of the proposed algorithm in Section 5. Finally, we conclude this paper in the last section.

2. ENERGY MINIMIZATION FORMULATION

The image matching algorithm proposed in this paper is an intensity-based approach. Conventional intensity-based methods derived from the SSD formulation, which is based on the brightness constancy model given as

$$I_0(x+u, y+v) = I_1(x, y), \tag{1}$$

where I_0 and I_1 are the image intensity functions at time t_0 and t_1, respectively, and (u, v) is the displacement vector at a location (x, y). Some previous work [9, 10] used a modified brightness model to account for uniform photometric variation; i.e.,

$$I_0(x+u, y+v) = \alpha I_1(x, y) + \beta. \tag{2}$$

Note that the constants α and β are the illumination multiplication and bias factors. However, this model cannot account for spatially varying illumination variations. To overcome this restriction, a generalized dynamic image model [11] is proposed by assuming the illumination multiplication and bias factors (α and β) to be functions of location (x, y). In this paper, we assume that these two illumination factors are slowly varying functions of location (x, y); thus they can be well approximated by low-order polynomials of (x, y),

$$I_0(x+u, y+v) = \alpha_p(x, y) I_1(x, y) + \beta_p(x, y), \tag{3}$$

where $\alpha_p(x, y)$ and $\beta_p(x, y)$ are the low-order polynomial functions with the coefficients represented by $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})$, respectively. The resulting polynomial illumination model can account for spatially varying smooth illumination variations with an additional small number of polynomial coefficients to be determined together with the geometric transformation. The polynomial illumination model should be suited for objects with smooth surfaces, but it is not appropriate to account for shadows in the images.

The geometric transformation in the image matching for planar objects can be well represented by an affine transformation when orthographic projection is a good approximation for the situation. There are a number of previous methods [6, 7, 9] proposed to solve the affine image matching problem. In this paper, we also focus on the recovery of affine transformations for the image matching. However, this framework can be extended to other global transformations. For an affine transformation, the displacement vector (u, v) at the location (x, y) can be written as

$$u(x, y, \Delta\mathbf{a}) = \Delta a_0 + \Delta a_1 x + \Delta a_2 y \tag{4}$$

$$v(x, y, \Delta\mathbf{a}) = \Delta a_3 + \Delta a_4 x + \Delta a_5 y, \tag{5}$$

where $\Delta\mathbf{a} = (\Delta a_0, \Delta a_1, \ldots, \Delta a_5)$ is the parameter vector for the affine transformation.

By assuming the displacement vector (u, v) is small, we take the first-order Taylor series approximation on I_0 in Eq. (3) to obtain the data constraint

$$I_0(x, y) + I_x(x, y)u(x, y, \Delta\mathbf{a}) + I_y(x, y)v(x, y, \Delta\mathbf{a}) - \alpha_p(x, y)I_1(x, y) - \beta_p(x, y) = 0, \tag{6}$$

where $I_x(x, y)$ and $I_y(x, y)$ are the partial derivatives of I_0 with respect to x and y. Replacing the displacement components u and v in Eq. (6) with the affine model given in Eq. (4) and (5) yields the constraint equation

$$\begin{aligned} &I_0 + I_x(x, y)(\Delta a_0 + \Delta a_1 x + \Delta a_2 y) + I_y(x, y)(\Delta a_3 + \Delta a_4 x + \Delta a_5 y) \\ &\quad - \alpha_p(x, y)I_1(x, y) - \beta_p(x, y) = 0. \end{aligned} \tag{7}$$

For ease of exposition, we assume linear models for the illumination multiplication and bias functions, i.e., $\alpha_p(x, y) = \alpha_0 + \alpha_1 x + \alpha_2 y$ and $\beta_p(x, y) = \beta_0 + \beta_1 x + \beta_2 y$, in the following derivation of our method. Thus, the above constraint equation at a location (x_i, y_i) can be rewritten as

$$\mathbf{f}_i \mathbf{c} = g_i, \tag{8}$$

where $g_i = -I_0(x_i, y_i)$ and

$$\mathbf{c} = (\Delta a_0 \quad \Delta a_1 \quad \Delta a_2 \quad \Delta a_3 \quad \Delta a_4 \quad \Delta a_5 \quad \alpha_0 \quad \alpha_1 \quad \alpha_2 \quad \beta_0 \quad \beta_1 \quad \beta_2)^{\mathrm{T}}, \tag{9}$$

$$\begin{aligned} \mathbf{f}_i = (&I_{x,i} \quad I_{x,i}x_i \quad I_{x,i}y_i \quad I_{y,i} \quad I_{y,i}x_i \quad I_{y,i}y_i \quad -I_1(x_i, y_i) \quad -I_1(x_i, y_i)x_i \\ &-I_1(x_i, y_i)y_i \quad -1 \quad -x_i \quad -y_i). \end{aligned} \tag{10}$$

Note that $I_{x,i} = I_x(x_i, y_i)$, $I_{y,i} = I_y(x_i, y_i)$. To determine the unknown vector $\mathbf{c}$, i.e., the geometric transformation parameters and illumination transformation parameters, we can simply apply the weighted least-squares method to minimize the energy function

$$E(\mathbf{c}) = \sum_i w_i(\mathbf{f}_i \mathbf{c} - g_i)^2, \tag{11}$$

where w_i is the weight associated with the data constraint computed at the ith location (x_i, y_i). This weighting can provide more accurate estimation when an appropriate reliability measure for each data constraint is available.

The above linear data constraint equation is derived based on the assumption that the displacement between the two images, I_0 and I_1, is small. This derivation can be generalized to the case of large displacement with a good initial guess for the geometric transformation as follows. Assume an approximate geometric transformation $T_0(x, y)$ is available. Thus, the generalized brightness constraint equation given in Eq. (3) can be modified to be

$$I_0(x + u, y + v) = \alpha_p(x, y)I_1\big(T_0^{-1}(x, y)\big) + \beta_p(x, y). \tag{12}$$

Following the same derivation, we can obtain a modified linear constraint $\bar{\mathbf{f}}_i \mathbf{c} = g_i$, where

$$\begin{aligned} \bar{\mathbf{f}}_i = \big(&I_{x,i} \quad I_{x,i}x_i \quad I_{x,i}y_i \quad I_{y,i} \quad I_{y,i}x_i \quad I_{y,i}y_i \quad -I_1\big(T_0^{-1}(x_i, y_i)\big) \quad -I_1\big(T_0^{-1}(x_i, y_i)\big)x_i \\ &-I_1\big(T_0^{-1}(x_i, y_i)\big)y_i \quad -1 \quad -x_i \quad -y_i\big). \end{aligned}$$

Then the modified energy function can be written as

$$E(\mathbf{c}; T_0) = \sum_i w_i (\bar{\mathbf{f}}_i \mathbf{c} - g_i)^2. \tag{13}$$

In general, the above weighted least-squares method is not robust against partial occlusion or outlier noise. To improve the robustness of the estimation, we will modify the weighted least-squares energy function by incorporating the robust regression into our formulation in the next section.

3. MINIMIZATION OF ROBUST ENERGY FUNCTION

Robust estimation has been successfully used to allow partial occlusion and remove outliers in several image matching methods [6, 7, 12]. In robust estimation, the quadratic function in the energy function for the least-squares estimation is replaced by a ρ-function, which assigns smaller weights for constraints with larger residues. Two ρ-functions commonly used in computer vision are the Lorentzian (or Cauchy) function and the Geman and McClure function [12],

$$\rho_{\mathrm{LO}}(x) = \log\left(1 + \frac{x^2}{2}\right), \quad \rho_{\mathrm{GM}}(x) = \frac{x^2}{1 + x^2},$$

where x is the normalized residue of the data constraint. The Lorentzian function was used in our implementation. Other ρ-functions commonly used in statistics, such as Tukey's biweight function [13], can also be used in our algorithm. When the ρ-function is used for robust estimation in model fitting, the influence for each data constraint on the solution is characterized by an influence function ψ, which is the derivative of the ρ-function. If we take the derivatives of the above two ρ-functions, we can see the influence functions decrease as the magnitude of the residue increases [12]. For the least-squares estimation, the influence function is linearly increasing as the magnitude of the residue increases. Therefore the least-squares estimation is more sensitive to outliers than the robust estimation. To use robust estimation in our energy-minimization framework, we can simply replace the quadratic function by a ρ-function and normalize the residue $g_i - \bar{\mathbf{f}}_i \mathbf{c}$ with a scale parameter σ in the energy function given in Eq. (13). This yields the new energy function [13]

$$E'(\mathbf{c}; T_0) = \sum_i w_i \rho\left(\frac{g_i - \bar{\mathbf{f}}_i \mathbf{c}}{\sigma}\right), \tag{14}$$

where σ denotes an estimate of the scale parameter. The minimization of the energy function $E'(\mathbf{c}, T_0)$ assumes the scale parameter σ is constant. A commonly used scale parameter estimator is the median absolute deviation (MAD) estimate [13] given by

$$\sigma^{(k)} = \frac{1}{0.6745} \operatorname*{med}_i \left\{ \left| g_i - \bar{\mathbf{f}}_i \mathbf{c}^{(k)} - \operatorname*{med}_j \{ g_j - \bar{\mathbf{f}}_j \mathbf{c}^{(k)} \} \right| \right\}, \tag{15}$$

where med denotes a median operator and $\mathbf{c}^{(k)}$ is the estimate of the vector $\mathbf{c}$ at the kth iteration. The main difference between the energy function in Eq. (14) and the standard energy function for robust estimation is the additional weighting w_i associated with each data constraint in our energy function. The minimization of this energy function can be

accomplished by a generalized M-estimator, which is to solve the following system of equations:

$$\sum_i w_i \psi\left(\frac{g_i - \bar{\mathbf{f}}_i \mathbf{c}}{\sigma}\right)\bar{\mathbf{f}}_i^{\mathrm{T}} = \mathbf{0}. \tag{16}$$

A reweighted least-squares method [13] has been successfully used to approximate the solution of the standard M-estimator. We apply a generalized reweighted least-squares method to update the estimate for the solution $\mathbf{c}$ in the above system of nonlinear equations with σ fixed with its current estimate as follows,

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \left(\mathbf{F}^{\mathrm{T}}\mathbf{W}^{(k)}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{W}^{(k)}\left(\mathbf{g} - \mathbf{F}\mathbf{c}^{(k)}\right), \tag{17}$$

where k is the index of the iteration, $\mathbf{g} = (g_1 \;\; g_2 \;\; \cdots \;\; g_n)^{\mathrm{T}}$, $\mathbf{F}$ is the concatenation of the vectors $\bar{\mathbf{f}}_i$, $i = 1, 2, \ldots, n$, i.e., $\mathbf{F} = (\bar{\mathbf{f}}_1^{\mathrm{T}} \;\; \bar{\mathbf{f}}_2^{\mathrm{T}} \;\; \cdots \;\; \bar{\mathbf{f}}_n^{\mathrm{T}})^{\mathrm{T}}$, and $\mathbf{W}^{(k)}$ is a diagonal matrix with diagonal elements given by

$$W_{i,i}^{(k)} = w_i \frac{\psi\left[\left(g_i - \bar{\mathbf{f}}_i \mathbf{c}^{(k)}\right)/\sigma^{(k)}\right]}{\left(g_i - \bar{\mathbf{f}}_i \mathbf{c}^{(k)}\right)/\sigma^{(k)}}, \quad i = 1, 2, \ldots, n. \tag{18}$$

Note that the weighting $W_{i,i}^{(k)}$ for the ith data constraint at the kth iteration is the multiplication between the original weight w_i and the factor $\psi(r_i^{(k)}/\sigma^{(k)})/(r_i^{(k)}/\sigma^{(k)})$, where $r_i^{(k)} = g_i - \bar{\mathbf{f}}_i \mathbf{c}^{(k)}$. For the Lorentzian function used in this paper, the factor $\psi(r_i^{(k)}/\sigma^{(k)})/(r_i^{(k)}/\sigma^{(k)})$ is given by $2\sigma^{(k)^2}/(2\sigma^{(k)^2} + r_i^{(k)^2})$. Note that the update equation in (17) involves computing the inverse of the $p \times p$ matrix $(\mathbf{F}^{\mathrm{T}}\mathbf{W}^{(k)}\mathbf{F})^{-1}$. To achieve nice numerical stability, we can apply the SVD-based method to compute the pseudo inverse [14] by neglecting the eigenvalues that are very close to 0 in the computation.

The additional weighting factor w_i associated with each data constraint is used in this paper to augment the robustness of the estimation by taking a measure of consistency between the gradients at the corresponding locations in two images and the nonlinearity of the brightness function into account. This is accomplished by setting w_i to be the multiplication between an exponential function of relative discrepancy between the gradient vectors and two sinusoidal functions of the intensity values at corresponding locations in two images, i.e.,

$$w_i = e^{-\frac{\|(I_{x,i}, I_{y,i}) - (\partial I_1(x_i, y_i)/\partial x, \partial I_1(x_i, y_i)/\partial y)\|^2}{\lambda \|(I_{x,i}, I_{y,i})\|^2}} w(I_0(x_i, y_i)) w(I_1(x_i, y_i)), \tag{19}$$

where λ is a constant that was empirically set to 5 in our implementation and the intensity weighting function $w(\,)$ is defined as

$$w(x) = \begin{cases} \sin\left(\frac{x\pi}{2h}\right) & 0 \le x < h \\ 1 & h \le x < 255 - h \\ \cos\left(\frac{(x+h-255)\pi}{2h}\right) & 255 - h \le x \le 255. \end{cases}$$

The parameter h is a constant chosen between 0 and 255/2. This intensity weighting function is plotted in Fig. 1. The first term in Eq. (19) penalizes large relative discrepancies between the gradients at the corresponding locations in two images for each data constraint. The

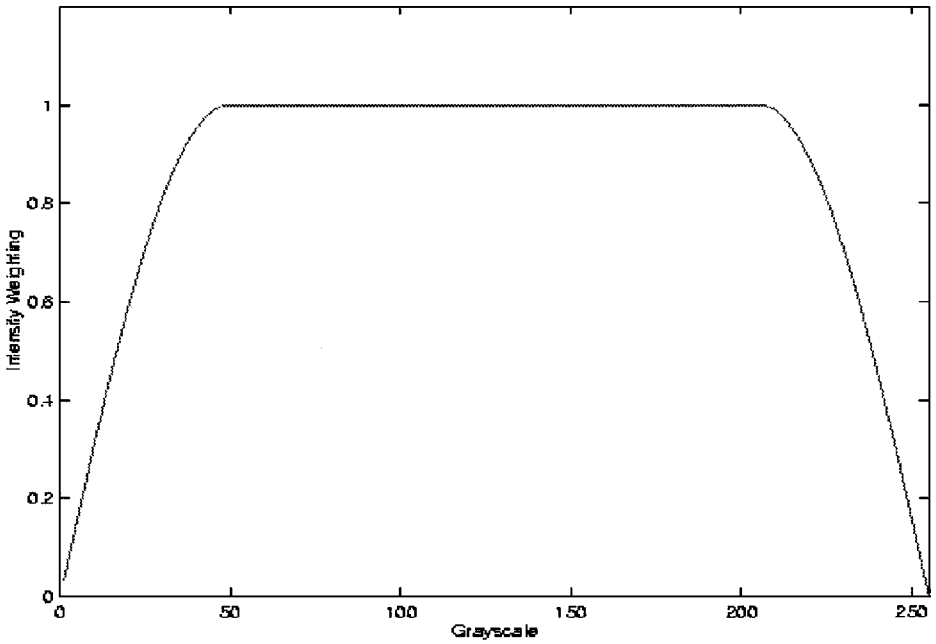

FIG. 1. The plot of the intensity weighting function with $h = 50$.

addition of this weighting can resolve the confusion of matching two similar intensity values with completely different image gradient directions. This confusion is due to the matching purely based on the intensity values, which is also a weakness of the SSD formulation. The inclusion of this weighting factor more or less overcomes this problem. The intensity weighting functions are used to alleviate errors due to the nonlinear characteristics of the sensors acquiring image intensity values, especially near brightness saturation or cutoff regions. By incorporating this intensity weighting, we disincline to use constraints with the corresponding intensity values near the saturation or cutoff regions, i.e., 0 or 255, for the update of parameters in the iterative reweighted least-squares algorithm. Thus we have a very robust scheme to adaptively adjust the weight for each data constraint based on its residue, the gradient discrepancies, and the intensity values in an iterative reweighted least-squares algorithm.

4. EFFICIENT ITERATIVE NUMERICAL SOLUTION

The energy minimization formulation of using a generalized brightness model and the robust regression technique for the image matching problem was described in the previous section. In this section, we present an efficient iterative numerical algorithm using the modified reweighted least-square method for estimating the geometric transformation parameters and illumination transformation parameters. In addition, we propose three additional schemes to augment the robustness, efficiency, and accuracy of the proposed algorithm. One is a selective data sampling scheme to reduce the computational cost. The other is a hierarchical nearest-neighbor search method for finding a good initial guess for the geometric transformation parameters to improve the robustness of the algorithm. Finally, we apply an estimation-and-warping strategy to constantly update the energy function to be minimized based on the newly updated geometric transformation. The Taylor approximation errors in

the data constraints can be greatly reduced by incorporating this scheme into the iterative reweighted least-squares algorithm, thus leading to very accurate estimation results. The complete image alignment algorithm is given in Section 4.3.

4.1. Selective Data Sampling

The energy function to be minimized is the weighted sum of the ρ-functions of the data constraints computed at the locations within the region of interest. Of course, we can take the sum all over the pixel locations within the region of interest. However, this may cause very high computational cost for large regions. Especially for the above optimization problem of a small number of parameters, it is not very crucial to use a very large number of constraints. Thus we propose to select a sparse set of locations to form the energy function. Our selection scheme is designed based on the consideration of efficiency as well as reliability.

In our selection scheme, we first partition the template image into $m \times n$ uniform blocks. From our experiments, we found 400 constraints are quite sufficient to provide accurate affine alignment with a bilinear model for illumination multiplication variations. So we set $m \times n$ to be roughly 400. Note that we can use more constraints to obtain higher accuracy or use less constraints to reduce the computational cost. Then, we select a location with a reliable constraint in each block to form the energy function. This scheme provides a diverse spatial distribution of the selected data constraints in the region. The selection of the data constraint in each block is described in the following.

Since the data constraint is derived from the first-order Taylor series approximation of the brightness function, approximation errors in computing the partial derivatives are inevitable due to inaccurate numerical approximation as well as the temporal and spatial aliasing in the brightness function. These approximation errors are the main source of errors in the data constraints [15, 16]. Since linear numerical approximation methods are usually used for the partial derivatives, our feature point selection is primarily determined by comparing the degree of local linearity for all the locations in each block. We define a reliability measure $\delta(x, y)$ associated with the data constraint at location (x, y) as the inverse of the sum of minimum distance errors in a local linear model fitting to the brightness function I_0, i.e.,

$$\delta(x, y) = \frac{I_x^2(x, y) + I_y^2(x, y)}{\text{sse}(x, y) + \varepsilon}, \tag{20}$$

where $\text{sse}(x, y)$ is the sum of squared errors from the first-order linear fitting in the local neighborhood of location (x, y), and ε is a small positive number to prevent overamplification of very small values of $\text{sse}(x, y)$ to the reliability measure. In our implementation, the local neighborhood was chosen to be a 3×3 window, the derivatives I_x and I_y are computed in the same window, and the constant ε was set to 10. Note that the numerator in Eq. (20) is mainly for normalizing the sum of squared errors to approximate a minimum distance measure. Thus, our selection of a reliable data constraint in each block is simply to find the location with the largest reliability measure.

4.2. Hierarchical Nearest-Neighbor Search

A hierarchical nearest-neighbor search method [17] is used for approximate pattern localization, thus providing a rough initial guess for the iterative robust image alignment algorithm. This method is based on the learning-from-examples principle. A hierarchical

nearest-neighbor network is constructed from the training examples synthesized from the geometric transformation of the reference image with the transformation parameters uniformly sampled in the parameter search space. The transformation parameter vector of the nearest-neighbor sample in the hierarchical network provides an approximate object localization. To improve the efficiency of the search algorithm, we find the best match based on the comparisons between feature vectors computed from the images, instead of comparing directly between images.

The feature vector used in this paper contains normalized wavelet features computed from the image. The feature generation involves the following steps. At first, the image inside the template window is resampled to a size with both width and height being powers of 2. This is primarily used in preparing the image data for wavelet transformation. Subsequently, we compute the wavelet transform of the resampled image and concatenate the wavelet coefficients at the lowest resolution level into a vector. Finally, the difference of this vector from its mean value is normalized by its standard deviation to obtain the feature vector of the image.

There are two phases in the proposed hierarchical nearest-neighbor search algorithm, the training phase and execution phase. The training phase consists of the generation of image templates from the reference image under different transformation parameter vectors $\mathbf{a}_1, \ldots, \mathbf{a}_N$, the computation of representative feature vectors for all the transformed image templates, and the training of hierarchical competitive layer neural networks. The execution phase consists of feature extraction from the input image and hierarchical nearest-neighbor search for the transformation parameter vector of the best match.

The hierarchical nearest-neighbor search is used to drastically reduce the number of comparisons between feature vectors by hierarchically organizing all the feature vectors in the entire training data set. The hierarchical clustering of the entire feature vectors is accomplished by using a hierarchical competitive network [17], which is a modification of a competitive network. The modification involves outputting multiple best matches instead of a single best-match output in the clustering network.

4.3. Iterative Numerical Algorithm

We described the iterative reweighted least-squares solution update equations for minimizing the energy function for robust image alignment given an approximate geometric transformation in Section 3. The final solution is a compromise of the selected data constraints, which are derived from the first-order Taylor approximation of the generalized brightness assumption given in Section 2. To alleviate the errors due to this approximation, we can apply an estimation-and-warping strategy in the iterative reweighted least-squares algorithm as follows.

When an approximate estimate for the geometric transformation $T(x, y)$ is available, the generalized brightness assumption is given in Eq. (12). The energy function to be minimized, as given in Eq. (14), is a weighted sum of the ρ-function of the residues of the linear data constraints, which are derived from the first-order Taylor series approximation computed at the approximate geometric transformation $T_0(x, y)$. To achieve more accurate estimation, the derived constraints should be refined based on the Taylor approximation computed at the updated geometric transformation $T_k(x, y)$, where k is the index of iterations. Thus, the energy function to be minimized can be generalized from a fixed one, i.e., $E'(\mathbf{c}; T_0)$, as given in Eq. (14), to a dynamic energy function $E'(\mathbf{c}; T_k)$ for the kth iteration.

Now we consider the affine transformation assumed for the geometric transformation in this paper. By modifying the generalized brightness equation given in Eq. (12) with the

affine transformations $T_{\Delta\mathbf{a}}$ for $(x+u, y+v)$ and $T_{\mathbf{a}^{(k)}}$ to replace $T_0(x, y)$, we have

$$I_0(T_{\Delta\mathbf{a}}(x, y)) = \alpha_p(x, y) I_1\left(T^{-1}_{\mathbf{a}^{(k)}}(x, y)\right) + \beta_p(x, y). \tag{21}$$

Note that $T_{\Delta\mathbf{a}}$ is the estimated affine transformation computed with the data constraints derived at an approximate affine transformation $T_{\mathbf{a}^{(k)}}$. We can convert the generalized brightness equation in the form of Eq. (21) to the following form:

$$I_0(x, y) = \alpha'_p(x, y) I_1\left(T^{-1}_{\mathbf{a}^{(k+1)}}(x, y)\right) + \beta'_p(x, y). \tag{22}$$

It is obvious that $T^{-1}_{\mathbf{a}^{(k+1)}} = T^{-1}_{\mathbf{a}^{(k)}} \circ T^{-1}_{\Delta\mathbf{a}}$, $\alpha'_p(x, y) = \alpha_p(T^{-1}_{\Delta\mathbf{a}}(x, y))$ and $\beta'_p(x, y) = \beta_p(T^{-1}_{\Delta\mathbf{a}}(x, y))$. The functions $\alpha'_p(x, y)$ and $\beta'_p(x, y)$ are still polynomial functions of degree p. The updated affine transformation $T_{\mathbf{a}^{(k+1)}}$ is a composite transformation of $T_{\Delta\mathbf{a}}$ and $T_{\mathbf{a}^{(k)}}$. Let the affine transformation $T_{\Delta\mathbf{a}}(x, y)$ be written as

$$T_{\Delta\mathbf{a}} : \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} 1+\Delta a_1 & \Delta a_2 \\ \Delta a_4 & 1+\Delta a_5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \Delta a_0 \\ \Delta a_3 \end{pmatrix}, \tag{23}$$

with $\Delta\mathbf{a} = (\Delta a_0, \Delta a_1, \ldots, \Delta a_5)$ being the affine parameter vector. Similarly, the affine transformation $T_{\mathbf{a}^{(k)}}$ is characterized by the parameter vector $\mathbf{a}^{(k)} = (a_0^{(k)}, a_1^{(k)}, \ldots, a_5^{(k)})$. Then the affine parameters can be updated as follows:

$$\begin{bmatrix} a_1^{(k+1)} & a_2^{(k+1)} \\ a_4^{(k+1)} & a_5^{(k+1)} \end{bmatrix} = \begin{bmatrix} 1+\Delta a_1 & \Delta a_2 \\ \Delta a_4 & 1+\Delta a_5 \end{bmatrix} \begin{bmatrix} 1+a_1^{(k)} & a_2^{(k)} \\ a_4^{(k)} & 1+a_5^{(k)} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{24}$$

$$\begin{bmatrix} a_0^{(k+1)} \\ a_3^{(k+1)} \end{bmatrix} = \begin{bmatrix} 1+\Delta a_1 & \Delta a_2 \\ \Delta a_4 & 1+\Delta a_5 \end{bmatrix} \begin{bmatrix} a_0^{(k)} \\ a_3^{(k)} \end{bmatrix} + \begin{bmatrix} \Delta a_0 \\ \Delta a_3 \end{bmatrix}. \tag{25}$$

Finally, the modified iterative reweighted least-squares algorithm is given as follows:

1. Find an initial guess for $\mathbf{a}^{(0)}$ by using the hierarchical nearest-neighbor search method.
2. Set the iteration index $k = 0$ and initialize $\mathbf{c}^{(0)}$ to be (0 0 0 0 0 0 0 1 0 0 1 0 0).
3. Update $\sigma^{(k)}$ via Eq. (15).
4. Update $\mathbf{c}^{(k+1)}$ via Eq. (17).
5. Update the affine parameters in $\mathbf{a}^{(k+1)}$ via Eqs. (24) and (25) with the vector $\Delta\mathbf{a}$ replaced by $\Delta\mathbf{a}^{(k+1)}$.
6. Reset the values of $\Delta\mathbf{a}^{(k+1)}$ in $\mathbf{c}^{(k+1)}$ to be zero.
7. Set $k = k+1$.
8. If $\mathbf{c}^{(k+1)} \approx \mathbf{c}^{(k)}$, stop; else go back to step 3.

The algorithm is converged when $\|\mathbf{c}^{(k+1)} - \mathbf{c}^{(k)}\| < \text{TOL}$. The error tolerance TOL was empirically set to 0.05 in our experiments.

5. EXPERIMENTAL RESULTS

We have applied the proposed image matching algorithm to align images for industrial inspection under different illumination conditions and partial occlusions. Image alignment is a very important problem in industrial inspection. The image reference approach that is

popularly used in industrial inspection requires very precise image alignment to proceed further inspection tasks [18]. Since the iterative reweighted least-squares algorithm requires a good initial guess to converge to the true solution, we apply a hierarchical nearest-neighbor search algorithm [17] as the first step in providing a rough initial guess for the proposed image matching algorithm. In our experiments, the hierarchical nearest-neighbor search algorithm can provide a rough estimate for the geometric transformation within ± 5 pixels in both horizontal and vertical directions and $\pm 5^\circ$ for rotation angles. We employed a bilinear model for the illumination multiplication function and a constant model for the illumination bias. Figure 2 shows an example of very precise marker localization on PCB images under different spatially varying illumination changes. This task is very common for pick-and-place applications. For most cases, our iterative algorithm converged reasonably well within 20 iterations. It took less than 0.03 s for each image matching task on a Pentium II 450MHz PC with 128 Mb RAM running Windows NT.

The second experiment is the pattern alignment for PCB inspection under spatially varying illumination conditions. Figure 3a shows the reference PCB image with the specified region of interest and reference point. The templates in the inspection images under different

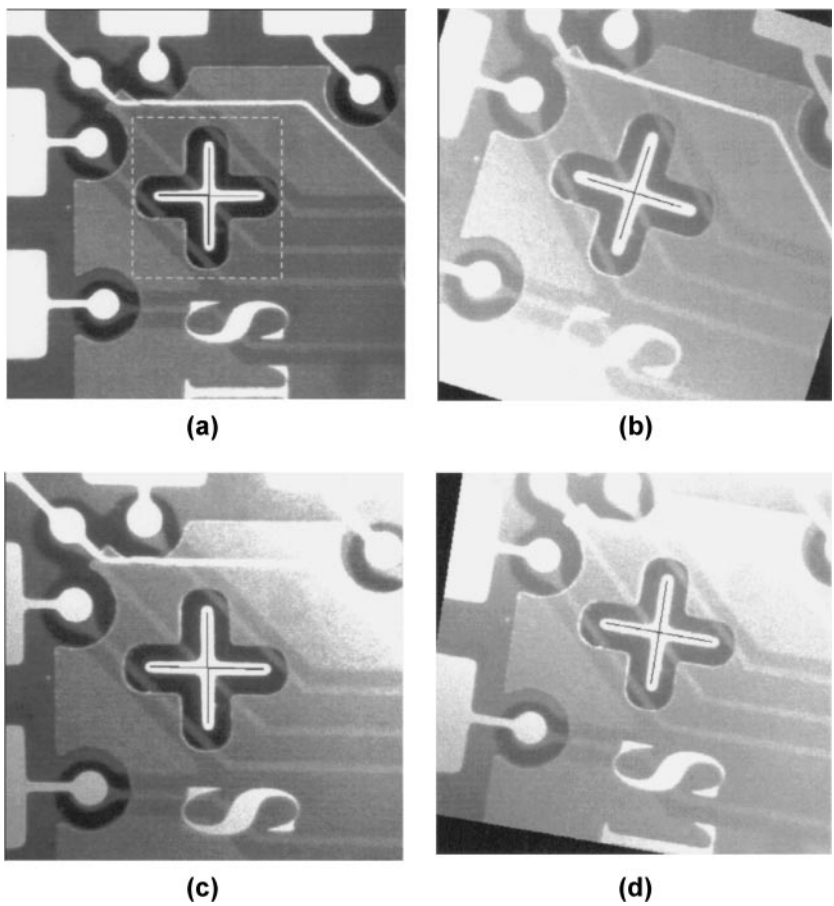

FIG. 2. Examples of image alignment under spatially varying illumination variations. (a) A reference image with the region of interest specified in the dashed window and the reference position with a cross sign. Example images from (b) to (d) show accurate affine image alignment with the cross signs specified the estimated reference locations.

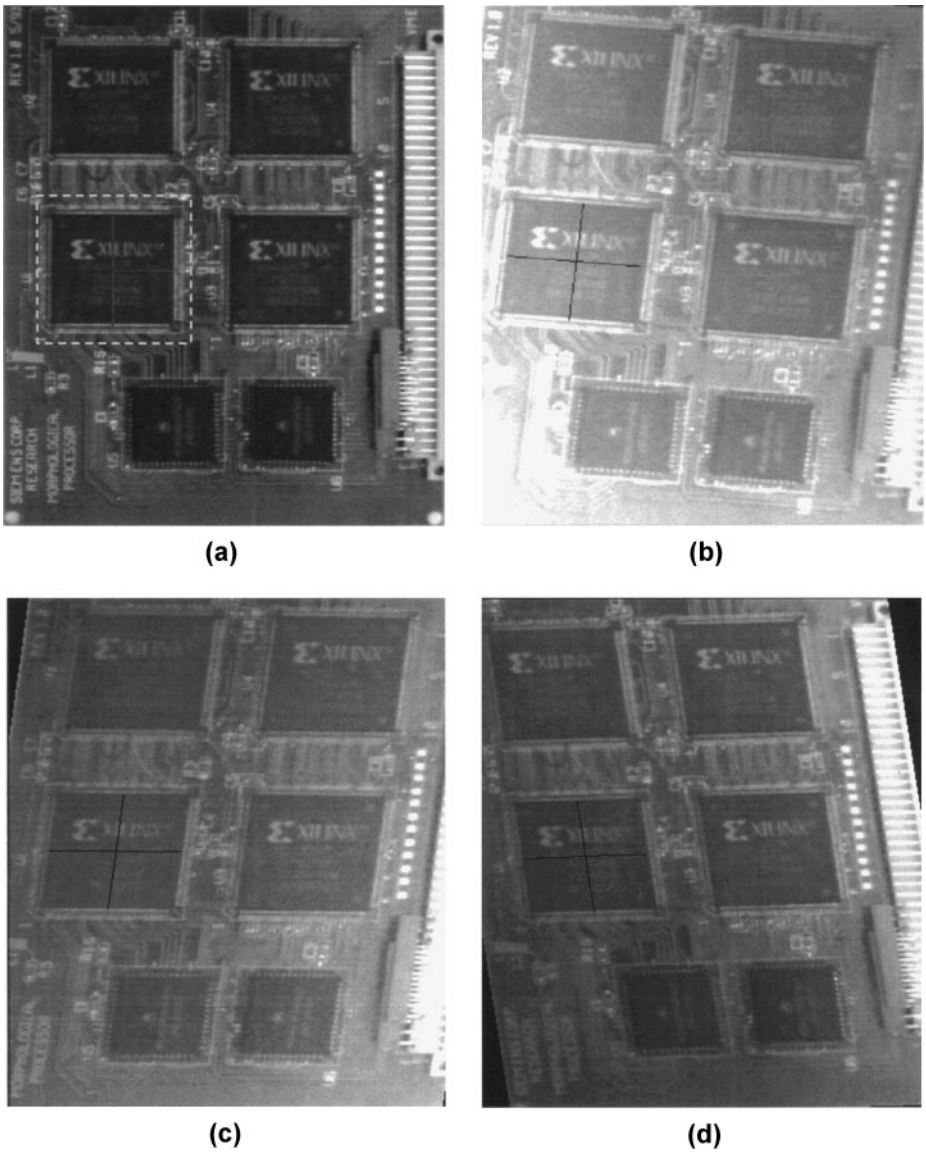

FIG. 3. Object alignment for PCB under different illumination conditions. (a) A reference PCB image with the region of interest specified in the dashed window and the reference position with a cross sign. Example images from (b) to (d) show accurate affine image alignment with the cross signs specified the estimated reference locations.

illumination conditions are accurately aligned as illustrated in Figs. 3b–3d. The alignment errors for these examples are within 0.1 pixels and 0.05°.

Figure 4 depicts some examples of image alignment with partial occlusions and spatially varying illumination changes. The image occlusions are simulated by randomly placing a small component pattern on the inspection image. From our experiments, our algorithm can tolerate up to roughly a 30% region of interest being occluded. For the images with very large occlusion areas, our algorithm may not converge to the correct solution. Two examples with large percentages of occlusion areas in the template window are shown in Fig. 5. The proposed image alignment algorithm fails to converge to the correct solutions for these two cases.

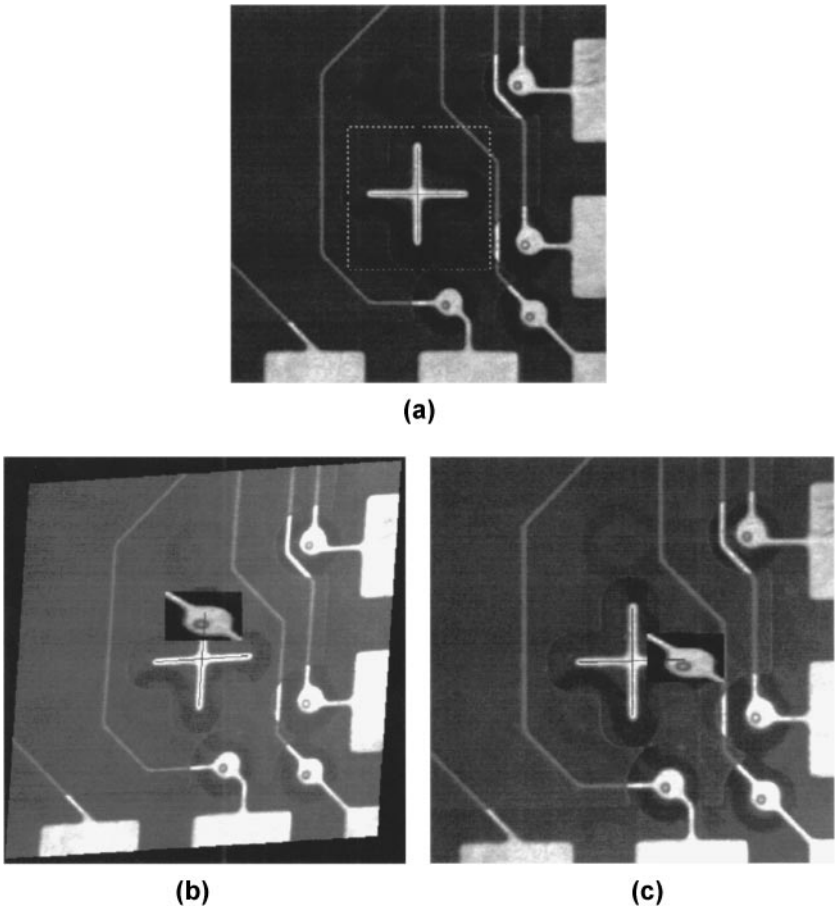

FIG. 4. Examples of image alignment under partial occlusion and spatially verying illumination changes. (a) A reference image with the region of interest specified by the dashed window and the reference position specified with a cross sign. Two examples in (b) and (c) show accurate affine image alignment with the cross signs specifying the estimated reference locations.

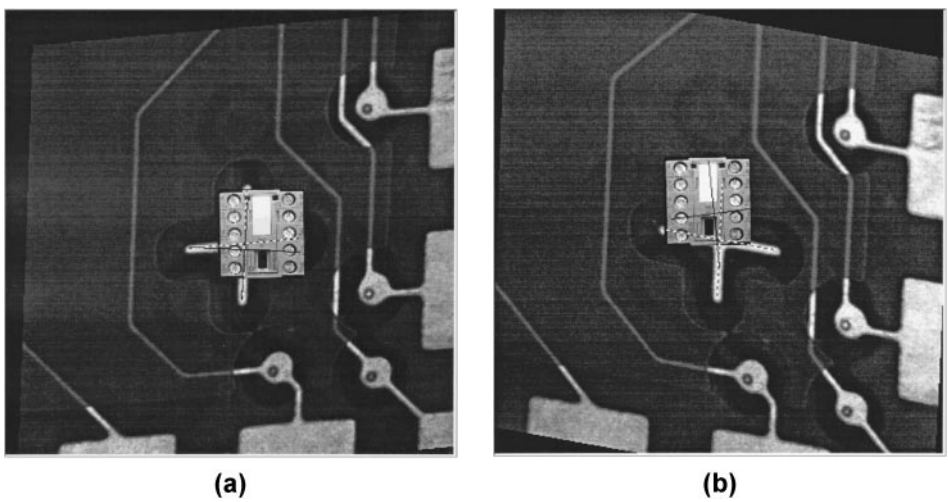

FIG. 5. Examples of misalignment of the pattern specified in Fig. 4a under large areas of occlusion. The dot crosses indicate the correct reference locations while the solid crosses indicate the estimated locations.

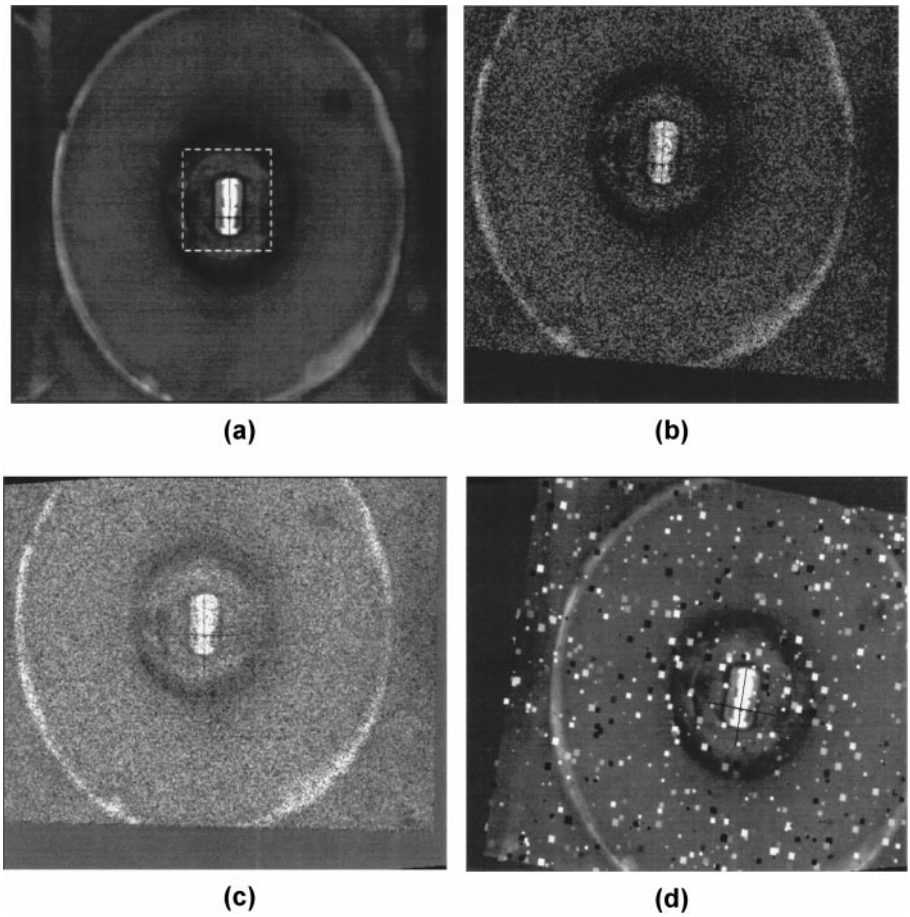

FIG. 6. (a) A reference image with the region of interest specified by the dashed window and a reference cross sign. Example images from (b) to (d) show very robust image alignment using the proposed algorithm in very noisy environments. The cross sign in each image indicates the estimated reference location.

Due to the use of robust estimation, our algorithm can also tolerate various kinds of noise. Figure 6 shows examples of accurate image alignment for very noisy images. It is evident from these experiments that our algorithm can accurately align the images containing very fine-structured noises and large block noises with different illumination conditions.

6. CONCLUSIONS

In this paper, we propose a robust, accurate, and efficient image alignment algorithm for robust image matching under partial occlusion and spatially varying illumination variations. The image matching is formulated as an robust estimation problem with the use of an affine transformation for the geometric transformation and polynomial functions for illumination factors. We derived the data constraints from the first-order Taylor series approximation of the generalized brightness assumption with low-order polynomial models for modeling spatially varying illumination changes. We incorporated the influence function from robust statistics into the new energy function that is a combination of the data constraints. A modified iterative reweighted least-squares algorithm was derived to solve the optimization

problem. We used a dynamic weighting scheme, which combines the factors from influence function, consistency of image gradients, and nonlinear image intensity sensing characteristics, to improve the robustness of the image alignment. In addition, we applied a selective constraint sampling and an estimation-warping alternating strategy to improve efficiency and accuracy of our algorithm. Experimental results on the image alignment for industrial inspection and pick-and-place applications demonstrated the robustness, efficiency, and accuracy of our algorithm.

ACKNOWLEDGMENTS

The author thanks the reviewers and associate editors for their valuable comments and suggestions, which make this revised paper more readable.

REFERENCES

1. L. G. Brown, A survey of image registration techniques, *ACM Surv.* **24**, No. 4, 1992, 325–376.
2. W. M. Wells III, Statistical approaches to feature-based object recognition, *Int. J. Comput. Vision* **21**, No. 1–2, 1997, 63–98.
3. J. W. Hsieh, H. Y. M. Liao, K. C. Fan, M. T. Ko, and Y. P. Hung, Image registration using a new edge-based approach, *Comput. Vision Image Understanding* **67**, 1997, 112–130.
4. P. Anandan, A computational framework and an algorithm for the measurement of visual motion, *Int. J. Comput. Vision* **2**, No. 3, 1989, 283–310.
5. R. Brunelli and T. Poggio, Template matching: Matched spatial filters and beyond, *Pattern Recognition* **30**, No. 5, 1997, 751–768.
6. M. J. Black and A. D. Jepson, Eigen Tracking: Robust matching and tracking of articulated objects using a view-based representation, *Int. J. Comput. Vision* **26**, No. 1, 1998, 63–84.
7. G. D. Hager and P. N. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, *IEEE Trans. Pattern Anal. Mach. Intelligence* **20**, No. 10, 1998, 1025–1039.
8. P. Viola and W. M. Wells III, Alignment by maximization of mutual information, *Int. J. Comput. Vision* **24**, No. 2, 1997, 137–154.
9. R. Szeliski and J. Coughlan, Spline-based image registration, *Int. J. Comput. Vision* **22**, No. 3, 1997, 199–218.
10. S.-H. Lai and M. Fang, An accurate and fast pattern localization algorithm for automated visual inspection, *Real Time Imaging* **5**, 1999, 3–14.
11. S. Negahdaripour, Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, No. 9, 1998, 961–979.
12. M. J. Black and P. Anandan, The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields, *Comput. Vision Image Process.* **63**, No. 1, 1996, 75–104.
13. G. Li, Robust regression, in *Exploring Data Tables, Trends, and Shapes* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, Eds.), pp. 281–343, Wiley, New York, 1985.
14. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge Univ. Press, Cambridge, UK, 1992.
15. J. Weber and J. Malik, Robust computation of optical flow in a multiscale differential framework, *Int. J. Comput. Vision* **14**, No. 1, 1995, 67–81.
16. S.-H. Lai and B. C. Vemuri, Reliable and efficient computation of optical flow, *Int. J. Comput. Vision* **29**, No. 2, 1998, 87–105.
17. S.-H. Lai and M. Fang, A FLASH system for fast and accurate pattern localization, in *Proceedings of SPIE Conference on Machine Vision Applications in Industrial Inspection VII, San Jose, CA, Jan. 1999*, Vol. 3652, pp. 164–173.
18. T. S. Newman and A. K. Jain, A survey of automatic visual inspection, *Comput. Vision Image Understanding* **61**, 1995, 231–262.