Multiple Non-Rigid Surface Detection and Registration

Yi Wu
University of California, Merced
ywu29@ucmerced.edu

Yoshihisa Ijiri
OMRON Corporation
ykoni@ari.ncl.omron.co.jp

Ming-Hsuan Yang
University of California, Merced
mhyang@ucmerced.edu

Abstract

Detecting and registering nonrigid surfaces are two important research problems for computer vision. Much work has been done with the assumption that there exists only one instance in the image. In this work, we propose an algorithm that detects and registers multiple nonrigid instances of given objects in a cluttered image. Specifically, after we use low level feature points to obtain the initial matches between templates and the input image, a novel high-order affinity graph is constructed to model the consistency of local topology. A hierarchical clustering approach is then used to locate the nonrigid surfaces. To remove the outliers in the cluster, we propose a deterministic annealing approach based on the Thin Plate Spline (TPS) model. The proposed method achieves high accuracy even when the number of outliers is nineteen times larger than the inliers. As the matches may appear sparsely in each instance, we propose a TPS based match growing approach to propagate the matches. Finally, an approach that fuses feature and appearance information is proposed to register each nonrigid surface. Extensive experiments and evaluations demonstrate that the proposed algorithm achieves promising results in detecting and registering multiple non-rigid surfaces in a cluttered scene.

1. Introduction

Detecting and registering deformable surfaces are of great importance for tasks such as non-rigid structure-from-motion, medical imaging, and augmented reality. This is a difficult problem since the appearance of imaged surfaces varies due to many factors such as camera pose, surface deformation and lighting conditions. In addition, the high dimensionality of parameters in the deformation model makes the problem far more complicated than the rigid object detection so that it is difficult to directly employ a robust estimator, such as RANSAC [14]. When multiple deformed instances of an object appear in an image, this makes the detection and registration task even more challenging.

Much work has been done with the assumption that there exists only one instance in the image and significan-

t progress has been made in recent years. Pilet et al. [23] propose a feature based approach that employs an iterative semi-implicit optimization scheme to refine the matches. In [29], Zhu et al. propose a method to detect and register a surface, which fuses feature and appearance information. Both algorithms are based on the triangulated mesh model and achieve impressive performance when only one instance is in the image. However, it is not clear how these approaches can be extended to detect and register multiple surfaces in a scene.

In this paper, we address the multiple non-rigid surface detection and registration problem (See Figure 1). Given the template image representing the surface of interest in a canonical shape, the goal is to locate the deformed instances of the surface and estimate the deformation parameters of each instance. To tackle this challenging problem, we use local feature detectors (e.g., SIFT [18]) to obtain the feature points in both template and input image, and then the correspondences between these two sets are constructed based on the similarity of local appearance around feature points. To locate all instances, we adopt a hierarchical clustering approach in which the affinity matrix is constructed on a high-order graph. After rejecting outliers in each correspondence cluster, the sparsely distributed matches are propagated by a Thin Plate Spline (TPS) based match growing algorithm. The deformation parameters of each nonrigid surface are estimated by an approach that fuses feature and appearance information based on the TPS model.

The contributions of this work are as follows: (1) A novel high-order affinity graph is constructed to model the consistency of local topology and then a hierarchical clustering approach is employed to locate the nonrigid surfaces; (2) A deterministic annealing algorithm based on the TPS model is proposed to remove the matching outliers; (3) A TPS based match growing scheme is developed to propagate the matches; (4) An approach that combines feature and appearance information based on the TPS model is presented to estimate the deformations.

2. Related Work and Problem Context

Extensive work has been done on the image registration, which can be divided into two categories: feature-based
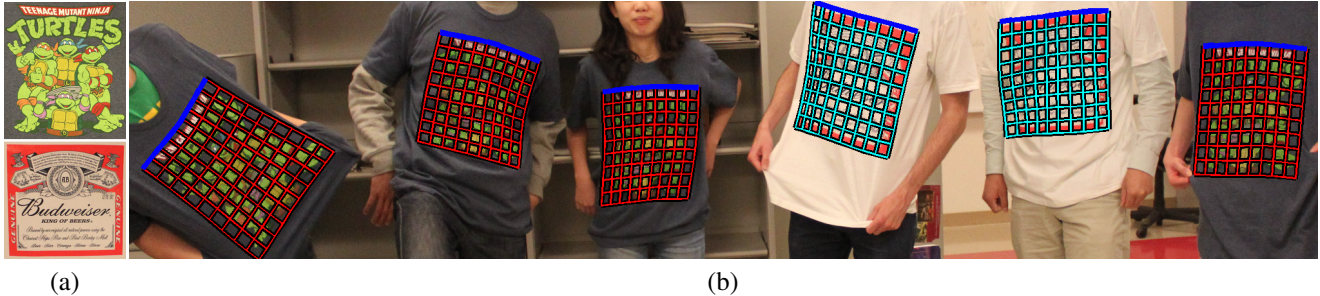
<div align="center">(a)   (b)</div>

Figure 1. (a) Two templates; (b) The detection and registration results overlaid on the input image, where the estimated non-rigid warps are illustrated in deformable meshes with different colors for different objects. The blue curves indicate the top of the instances in accordance with the corresponding templates. The input image is cropped for better illustration.

and appearance-based. The feature-based approaches rely on establishing correspondences between features of a template and an input image. For regions lacking texture or having large deformations, the correspondences may not be well established so that the performance of such methods is likely to degrade significantly. On the other hand, the appearance-based approaches do not rely on features and directly minimize the intensity discrepancy between the template and an input image warped back onto the coordinate frame of the template, which could achieve better registration accuracy. However, it requires good initialization to avoid local minima and is computationally expensive.

Feature-based Approaches. The feature-based approaches first establish correspondences by feature matching, then after eliminating outliers the transformation is estimated. The correspondences can be directly established by comparing the local feature descriptors, such as SIFT [18], which ignores the geometric relationship between features. Many inaccurate correspondences (outliers) are inevitably included and thus a matching refinement step is required. Due to the high dimensionality of parameters in the deformation model, it is difficult to directly employ robust estimators, e.g., RANSAC [14], to estimate the deformation model. With 25% of outliers and 100 transformation parameters, which is a common case for nonrigid surface registration, a RANSAC approach requires 10^{12} samples to guarantee that, with high probability, at least one sample is included without outliers [15]. Recently, for nonrigid surface registration, Pilet et al. [23] adopt a robust estimator to measure the outliers and employ an iterative semi-implicit optimization scheme to refine the matches. In [29], Zhu et al. use a similar strategy to reject outliers but with a different robust estimator, which can be formulated as an unconstrained quadratic optimization problem. In [10], a robust point set matching approach, TPS-RPM, is proposed to match points and remove outlier points iteratively. The TPS-RPM method solves the joint optimization problem iteratively by deterministic annealing and soft assignment with one-to-one correspondence assumption, which is computationally expensive. In this paper, we adopt the deterministic annealing approach, but use it only to remove spurious matches.

We also propose a bootstrap method to further increase the robustness and accuracy. Our method achieves high accuracy even when the percentage of inliers is less than 5%.

Recent feature matching methods formulate visual correspondence as a graph matching problem by considering pairwise geometric distortion of objects between images [16, 4, 25]. The constant distance constraint is adopted in [16] which is only invariant to rotation. In [4, 25], both distance and angle constraints are used for rotation and scale invariance. Zass and Sashua [28] formulate correspondences as a hypergraph matching problem by using higher order geometric constraints. By comparing the angles of triangles, a high-order scale invariant matching method is proposed by Chertok and Keller [7]. In [12], Duchenne et al. use higher order potentials to increase the geometric invariance of image features. The above mentioned methods commonly require a one-to-one feature mapping constraint and cannot deal with multiple object matching. Cho et al. [8] propose a hierarchical correspondence clustering approach by adopting pairwise geometric constraint without the one-to-one assumption. However, it heavily depends on the affine-covariant feature detectors and thus not applicable to other features. As shown in [22], affine shape determination is not accurate and thus shape adaptation can be noisy. Our method for correspondence clustering aims to overcome the above drawbacks. It operates on any kind of feature points and different types of features can be combined. A novel high-order affinity graph is constructed to model the consistency of local topology, and a hierarchical clustering approach is used to locate the nonrigid surfaces.

To establish more matches of feature points, Ferrari et al. [13] propose an approach that gradually explores the surrounding area of each existing match. Cho et al. [9] adopts this strategy to detect and segment multiple identical objects in an unsupervised way. Both approaches depend on the affine transformation from the affine feature detector [20] for match growing. In this work, to deal with the case that the matches are too sparsely distributed, we propose an approach to propagate matches locally by using the TPS model. Furthermore, a stability criterion is proposed to select the best local region to grow matches.

Appearance-based Approaches. Appearance-based approaches utilize the texture information of an object to estimate the deformation. Most appearance-based methods can be viewed as extensions of the original Lucas-Kanade algorithm [19, 1] by directly minimizing the intensity discrepancy between the template image and the input image warped back onto the coordinate frame of the template. In the active appearance model [11], a combination of shape and texture is used to estimate shape deformation, where a training process is needed. Matthews and Baker [21] improves the active appearance model with an inverse composition algorithm [1] in which triangulation meshes are used and piecewise affine warps are estimated. Recent work shows that the TPS warp can also be estimated by direct methods [3, 17]. In addition, the inverse compositional approach [1] is incorporated in a direct TPS-based method [6].

In [29], a fusion approach based on the triangulation mesh is proposed which takes advantage of both appearance information and local features. The piecewise affine warp is used to map points, where the mapping of one point only considers three surrounding control points. In this paper, we propose a fusion approach using the TPS warp which maps the points in a holistic way. The fusion approach can deal with large distortions in which correct feature correspondences are difficult to obtain.

3. Surface Detection

In this work, local feature points are detected by the difference of Gaussian operator and their SIFT descriptors [18] are extracted. The initial feature correspondences between a template and an input image are obtained by comparing the SIFT descriptors based on Euclidean distance using the k nearest-neighbor algorithm. Note that each feature point in the template may be matched to multiple points in the input image. We then obtain N feature correspondences, $\{M_n = (\mathbf{p}_n, \mathbf{q}_n)\}_{n=1,\ldots,N}$, where M_n denotes a match that consists of a feature point pair $(\mathbf{p}_n, \mathbf{q}_n)$ detected from a template and an input image, respectively. Given these initial feature correspondences, our goal is to partition $\{M_n\}_{n=1,\ldots,N}$ to several clusters.

The critical part of clustering is to measure the similarity or distance between matches. We note that it is ineffective to model the affine geometric constraint based only on one pair of feature correspondence. As it requires at least three correspondences to determine one affine transformation, we construct a high order graph, $G = (V, E)$, of triple correspondences. Each node in the graph is a triple-correspondence, namely a corresponding triangle pair (CTP), which determines an affine transformation. We can also consider G as a graph of affine transformations and each node is an affine matrix.

Given N pairs of corresponding points, the number of CTPs could be in the order of N^3. After we use the De-launay triangulation method to construct neighborhood triangles in the model, the number of CTPs is still large as one-to-many correspondence is allowed. However, most of these CTPs contain false correspondences and we propose a filter based on local topology to reject most outlier CTPs before constructing the high order graph.

The proposed filter is based on the insight that when a CTP contains an outlier correspondence, it is seldom topologically consistent with other correspondences. Formally, the j-th CTP is denoted by $\mathcal{P}_j = (\mathcal{T}_j^m, \mathcal{T}_j^t)$, which contains feature correspondences $M_{j_1}, M_{j_2}, M_{j_3}$, where $\mathcal{T}_j^m = (\mathbf{p}_{j_1}, \mathbf{p}_{j_2}, \mathbf{p}_{j_3})$ denotes the triangle in the model and $\mathcal{T}_j^t = (\mathbf{q}_{j_1}, \mathbf{q}_{j_2}, \mathbf{q}_{j_3})$ is the corresponding triangle in the test image. Based on this triangle correspondence, we obtain the affine transformation R_j that maps $\mathcal{T}_j^m$ to $\mathcal{T}_j^t$. Then we use R_j to map the feature points $\{\mathbf{p}_n\}_{n \neq j_1, j_2, j_3}$, which are not collinear with points $\mathbf{p}_{j_1}$, $\mathbf{p}_{j_2}$ and $\mathbf{p}_{j_3}$ in the model image, to the test image and compute the distance between the mapped points with their matched points in the test image. If all the distances are large, the CTP is classified as an outlier. With this filtering, more than 90% CTPs are rejected and thus the constructed graph is small. The number of CTPs can be reduced from tens of thousands order to hundreds order and then the final clustering process is very efficient. This filter can also be used to facilitate other hypergraph matching methods [12, 7, 28].

We define the similarity between CTPs, namely, the attribute of edge, E, to construct the graph, G. As each CTP defines one affine transformation, the distance between nodes is defined as the distance between the affine matrices. However, it is not easy to compute the distance in the affine group/manifold. We compute the attribute by using the affine matrix to transform each triangle in the template to the input image plane and compute the Euclidean distances between the transformed points and the corresponding points in the input image. For each CTP, we obtain three values and use the maximal one as the distance between the nodes. The distance between $\mathcal{P}_j$ and $\mathcal{P}_k$ is defined as

$$d_{j,k} = \max_{k=k_1,k_2,k_3} \|R_j(\mathbf{p}_k) - \mathbf{q}_k\|.$$

To get a symmetric metric, we set $d(j, k) = \frac{1}{2}(d_{j,k} + d_{k,j})$. If two triplets have similar affine transformations, the distance is small. Note that the transformed points have already been computed and stored during the filtering process, so the edge attribute can be obtained efficiently.

From the affinity matrix, we use a hierarchical clustering algorithm [8] and obtain the correspondence clusters: $\theta = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K\}$. Each match cluster $\mathcal{C}_i$ is composed of feature matches M_n^i, $\mathcal{C}_i = \{M_n^i = (\mathbf{p}_n^i, \mathbf{q}_n^i)\}_{n=1,\ldots,N_i}$. A few trivial clusters that contain small number of matches are classified as outliers, and others correspond to instances of the template object.

4. Surface Registration

After obtaining object correspondences, we register the detected instance to the object model (template). As the correspondences in a cluster may still contain some outliers, we propose a fast determinative annealing method based on the TPS deformation model to further refine the matches.

4.1. Deformation Model

We adopt the parametrization of TPS [5] proposed in [17], which is further formulated in [2]. Given m control points $\mathbf{c}_i \in \mathbb{R}^2$ and their function values $\alpha_i \in \mathbb{R}, i = 1, \ldots, m$, the TPS mapping function $f(\mathbf{p})$ is formulated by minimizing the bending energy,

$$E_b = \int_{\mathbb{R}^2} \left\| \frac{\partial^2 f(\mathbf{p})}{\partial \mathbf{p}^2} \right\|_F^2 d\mathbf{p}. \tag{1}$$

The function $f(\mathbf{p})$ is parameterized by an $m + 3$ coefficient vector $\boldsymbol{\eta}^\top = (\mathbf{w}^\top, \mathbf{a}^\top)$,

$$f(\mathbf{p}) = \sum_{i=1}^{m} w_i \rho(\|\mathbf{p} - \mathbf{c}_i\|) + \mathbf{a}^\top \hat{\mathbf{p}} = \boldsymbol{\ell}_{\mathbf{p}}^\top \boldsymbol{\eta}, \tag{2}$$

where the element of $\mathbf{b}$ is $b_i = \|\mathbf{p} - \mathbf{c}_i\|^2 \log(\|\mathbf{p} - \mathbf{c}_i\|^2)$, $\boldsymbol{\ell}_{\mathbf{p}}^\top = (\mathbf{b}^\top, \hat{\mathbf{p}}^\top)$, and $\hat{\mathbf{p}}^\top = (\mathbf{p}^\top, 1)$. The coefficients in $\mathbf{w}$ must satisfy $\mathbf{C}^\top \mathbf{w} = 0$, where the i-th row of $\mathbf{C}$ is $(\mathbf{c}_i^\top, 1)$. Therefore, we have the following equation:

$$\begin{pmatrix} \mathbf{K}_\lambda & \mathbf{C} \\ \mathbf{C}^\top & \mathbf{0} \end{pmatrix} \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix}. \tag{3}$$

where $\mathbf{K}_\lambda = \mathbf{K} + \lambda \mathbf{I}_m$, λ ($\lambda = 1$ in this paper) is the regularization parameter to control the amount of smoothness, and $\mathbf{I}_m$ is an $m \times m$ identity matrix to handle noise. The solution of (3) is

$$\boldsymbol{\eta} = \boldsymbol{\varepsilon}_\lambda \boldsymbol{\alpha} = \begin{pmatrix} \mathbf{K}_\lambda^{-1}(\mathbf{I}_m - \mathbf{D}) \\ \mathbf{D} \end{pmatrix} \boldsymbol{\alpha} = \begin{pmatrix} \hat{\boldsymbol{\varepsilon}}_\lambda \\ \mathbf{D} \end{pmatrix} \boldsymbol{\alpha}, \tag{4}$$

and $\mathbf{D} = \mathbf{C}(\mathbf{C}^\top \mathbf{K}_\lambda^{-1} \mathbf{C})^{-1} \mathbf{C}^\top \mathbf{K}_\lambda^{-1}$. Thus, the TPS mapping function can be written as:

$$f(\mathbf{p}) = \boldsymbol{\ell}_{\mathbf{p}}^\top \boldsymbol{\varepsilon}_\lambda \boldsymbol{\alpha}, \tag{5}$$

and the TPS bending energy E_b can be written as

$$E_b = 8\pi \mathbf{w}^\top \mathbf{K}_\lambda \mathbf{w} = 8\pi \boldsymbol{\alpha}^\top \hat{\boldsymbol{\varepsilon}}_\lambda \boldsymbol{\alpha}. \tag{6}$$

By stacking two functions for two coordinates, we have the TPS warp,

$$\mathcal{W}(\mathbf{p}; \mathbf{h}) = \begin{pmatrix} \boldsymbol{\ell}_{\mathbf{p}}^\top \boldsymbol{\varepsilon}_\lambda & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\ell}_{\mathbf{p}}^\top \boldsymbol{\varepsilon}_\lambda \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_x \\ \boldsymbol{\alpha}_y \end{pmatrix} = \mathbf{M}\mathbf{h}. \tag{7}$$

By extending (6) from TPS mapping to TPS warp, we obtain the smoothness energy which is used as the smoothness regularization for estimating the deformation:

$$E_s = \|\mathbf{Z}\mathbf{h}\|^2, \tag{8}$$

where $\mathbf{Z}^\top \mathbf{Z} = 8\pi \begin{pmatrix} \hat{\boldsymbol{\varepsilon}}_\lambda & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\varepsilon}}_\lambda \end{pmatrix}$. Note that in our parametrization, $\mathbf{h}$ is a vector while a matrix is used to denote the deformation parameters in [2]. The adopted vector parameterization facilitates solving the fusion optimization problem which will be discussed later.

4.2. Outlier Rejection

To remove matching outliers, both [23] and [29] adopt a condition function as the robust estimator to assess the penalties for matching residuals of different levels. In this work, we propose a deterministic annealing-based approach and show that, even directly adopting ℓ_2 norm in the cost function, we can easily solve it with least squares minimization and successfully remove outliers.

In the correspondence cluster i, the sum of squared residuals of N_i correspondences based on the TPS model is:

$$E_f = \frac{1}{N_i} \sum_{n=1}^{N_i} \|\mathcal{W}(\mathbf{p}_n; \mathbf{h}) - \mathbf{q}_n\|^2,$$

where $\mathbf{p}_n = (x_n, y_n)$ and $\mathbf{q}_n = (u_n, v_n)$. The corresponding matrix form is

$$E_f = \frac{1}{N_i} \|\mathbf{A}\mathbf{h} - \mathbf{u}\|^2, \tag{9}$$

where

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\varepsilon}_\lambda^\top \boldsymbol{\ell}_{\mathbf{p}_1} & \cdots & \boldsymbol{\varepsilon}_\lambda^\top \boldsymbol{\ell}_{\mathbf{p}_{N_i}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\varepsilon}_\lambda^\top \boldsymbol{\ell}_{\mathbf{p}_1} & \cdots & \boldsymbol{\varepsilon}_\lambda^\top \boldsymbol{\ell}_{\mathbf{p}_{N_i}} \end{pmatrix}^\top,$$

and $\mathbf{u} = (u_1, \cdots, u_{N_i}, v_1, \cdots, v_{N_i})^\top$.

By assembling two terms in (9) and (8), we have following regularized cost function:

$$E = \frac{1}{N_i} \|\mathbf{A}\mathbf{h} - \mathbf{u}\|^2 + \lambda_s \|\mathbf{Z}\mathbf{h}\|^2. \tag{10}$$

The optimal warp parameters $\mathbf{h}$ are given in closed form as:

$$\mathbf{h} = (\mathbf{A}^\top \mathbf{A} + N_i \lambda_s \mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{A}^\top \mathbf{u}. \tag{11}$$

The correspondence set, $\mathcal{C}_i$, may contain some outliers, which likely degrades the solution got from (11). Also, in (10) it is difficult to determine the weight, λ_s, for the smoothness constraint. We propose an approach based on the deterministic annealing algorithm [27] to gradually reduce λ_s. At the beginning of the annealing process, λ_s is large, which greatly limits the range of deformation and avoids overfitting due to the outliers. On the other hand, when λ_s is small, more local non-rigid deformation is allowed. Furthermore, during the annealing process, the threshold, d, to determine the outliers is also gradually decreased. When d is larger, it tolerates more outliers. The whole process ends when d is below a small value, d_{final}.

Specifically, λ_s is set to be $T \cdot \lambda_0$, where λ_0 is a constant and the temperature T is gradually reduced. Thus, the cost function (10) is written as

$$E = \frac{1}{N_i} \|\mathbf{A}\mathbf{h} - \mathbf{u}\|^2 + T \cdot \lambda_0 \|\mathbf{Z}\mathbf{h}\|^2. \tag{12}$$

After obtaining the estimated $\mathbf{h}$ by (11), we map each $\mathbf{p}_n$ to the input image, $\mathcal{W}(\mathbf{p}_n; \mathbf{h})$, and compute the distance between $\mathcal{W}(\mathbf{p}_n; \mathbf{h})$ and $\mathbf{q}_n$ by $d_n = \|\mathcal{W}(\mathbf{p}_n; \mathbf{h}) - \mathbf{q}_n\|$. If d_n is larger than the threshold d, it is classified as an outlier where d is also gradually reduced via the annealing procedure by setting $d = T \cdot d_0$.
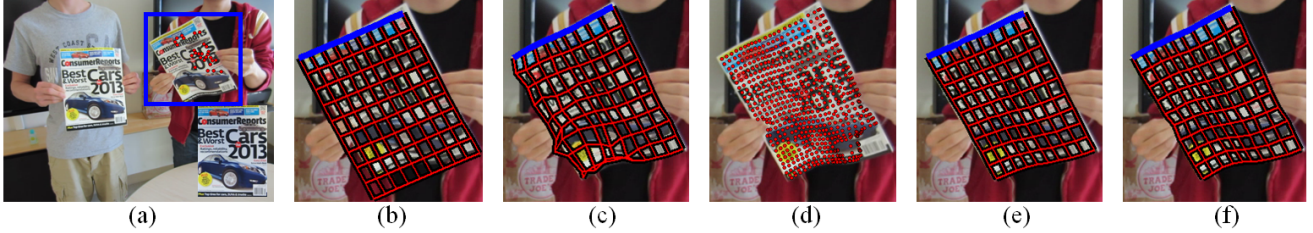
| (a) | (b) | (c) | (d) | (e) | (f) |

Figure 2. In (a) a template is shown in the bottom right of the input image. One instance is taken as an example and the matched feature points are marked as red. The non-rigid warps estimated by feature-based approach and fusion-based approach based on the feature matches shown in (a) are shown in (b) and (c), respectively. In (d) the matched feature points after match growing are marked as red points. The warps estimated by feature-based and fusion-based approaches based on the feature matches are shown in (e) and (f), respectively.

In the beginning, $T_{init} = T_0$, and then the temperature parameter T is gradually reduced $T_{new} = T_{old} \cdot r_T$, where $r_T \in (0, 1)$ is the annealing rate. Furthermore, determining T_0 is not straightforward. If T_0 is small and the correspondence set contains too many outliers, the initially estimated deformation parameters are far from the ground truth and the threshold d is also small. Thus a large amount of points will be classified to be outliers which would include many inliers. Therefore, in the beginning we set the maximum percentage of points that are allowed to be classified to outliers so that few inliers are falsely classified. If the number exceeds the threshold, T_0 is replaced by T_0/r_T. In the experiments, we show that with this bootstrap process, the recall of inliers is highly improved. On the other hand, T cannot be reduced to a too small value so we set a threshold T_{final}. When T reaches T_{final}, it remains a constant. After all the outliers are removed, we obtain good estimates of the deformation parameters $\mathbf{h}$ via (11).

4.3. Fusion Optimization

For regions lacking texture or having large deformations, the correspondences may not be well established and thus deformation parameters may not be well estimated by (11). In this paper, we propose a fusion approach based on the TPS warp by exploiting both appearance and local features.

The energy for the appearance, E_a, is formulated as minimizing the intensity discrepancy between the template, $\mathbf{T}$, and the input image, $\mathbf{I}$, warped back onto the coordinate frame of the template:

$$E_a = \|\mathbf{I}(\mathcal{W}(\mathbf{p}; \mathbf{h})) - \mathbf{T}\|^2. \qquad (13)$$

By using all the energy terms in (13), (10) and (8), we have the fusion energy function:

$$E(\mathbf{h}) = E_a + \lambda_f E_f + \lambda_s E_s$$
$$= \|\mathbf{I}(\mathcal{W}(\mathbf{p}; \mathbf{h})) - \mathbf{T}\|^2 + \frac{\lambda_f}{N_i}\|\mathbf{A}\mathbf{h} - \mathbf{u}\|^2 + \lambda_s\|\mathbf{Z}\mathbf{h}\|^2.$$
$$(14)$$

The minimization of (14) is a non-convex optimization problem in that the first term is nonlinear. Inspired by the Lucas-Kanade algorithm [19], (14) can be linearized by using the first order Taylor expansion w.r.t. the deformation parameters $\mathbf{h}$, and thus the solution can be iteratively improved. Thus, the fusion energy is linearized to

$$E(\Delta\mathbf{h}) = \|\mathbf{I}(\mathcal{W}(\mathbf{p}; \mathbf{h})) + \mathbf{J}\Delta\mathbf{h} - \mathbf{T}\|^2$$
$$+ \frac{\lambda_f}{N_i}\|A(\mathbf{h} + \Delta\mathbf{h}) - \mathbf{u}\|^2 + \lambda_s\|\mathbf{Z}(\mathbf{h} + \Delta\mathbf{h})\|^2,$$

where $\mathbf{J} = \nabla\mathbf{I}\frac{\partial\mathcal{W}}{\partial\mathbf{h}}$ is the Jacobian of $\mathbf{I}$ w.r.t. $\mathbf{h}$. Minimizing (15) w.r.t $\Delta\mathbf{h}$ is a regularized least squares problem and the closed-form solution is available:

$$\Delta\mathbf{h} = (\frac{\lambda_f}{N_i}\mathbf{A}^\top\mathbf{A} + \lambda_s\mathbf{Z}^\top\mathbf{Z} + \mathbf{J}^\top\mathbf{J})^{-1}$$
$$\cdot(\mathbf{J}^\top\mathbf{G} - \frac{\lambda_f}{N_i}\mathbf{A}^\top(\mathbf{A}\mathbf{h} - \mathbf{u}) - \lambda_s\mathbf{Z}^\top\mathbf{Z}\mathbf{h}), \qquad (15)$$

where $\mathbf{G} = \mathbf{T} - \mathbf{I}(\mathcal{W}(\mathbf{p}; \mathbf{h}))$ is intensity discrepancy between the template, $\mathbf{T}$, and the input image, $\mathbf{I}$, warped back onto the coordinate frame of the template. The warp parameters are then iteratively updated by

$$\mathbf{h} \leftarrow \mathbf{h} + \Delta\mathbf{h}. \qquad (16)$$

4.4. Match Growing

As shown in the blue bounding box of Figure 2(a), the feature matches marked as red are concentrated in small regions which significantly degrades the registration accuracy as illustrated in Figure 2(b)-(c). Therefore, it is crucial to improve the matching quality to enhance the registration performance. We propose a match growing approach based on the TPS model, which propagates the matches locally to establish more correspondences.

In this work, the template is partitioned to several non-overlapped rectangular cells. Each feature point in the template belongs to one cell, and four adjacent cells form one group, $\mathcal{G}$. Each time we pick one group for matching propagation and two scores are defined for selection: (1) S_a: the number of cells that contain matches in $\mathcal{G}$; (2) S_b: the total number of matches contained in $\mathcal{G}$. We choose the group that is with the maximum of S_b among the groups that are with the maximum of S_a. In the selected group, we use the matches to estimate the local TPS warp (9 control points are used) based on the feature-based approach. Then, the fusion approach is used to refine the warp, $\mathcal{W}$. The cells that have not been processed by the match growing process before are selected for propagation. Four matches estimated by $\mathcal{W}$ in each cell are added to the correspondence set. Note that after each time of propagation, the scores for the groups that contain the cells with added matches should be updated. In case that the propagated matches contain some outliers,

after obtaining the propagated matches, the proposed outlier rejection algorithm is used again to refine the matches. Finally, we use the feature-based approach to estimate the TPS warp (100 control points are used) for the deformed instance, which is further refined by the fusion approach.

As shown in Figure 2(d) the matched feature points after match growing are uniformly distributed. Thus, the estimated warp based on feature-based approach (Figure 2(e)) is better than that without growing (Figure 2(b)). When the initial warp is far from the ground truth (Figure 2(b)), the refined warp obtained by the fusion approach is still not accurate (Figure 2(c)). On the other hand, if the initial warp is close to the ground truth (Figure 2(e)), the fusion approach facilitates obtaining a more accurate warp (Figure 2(f)).

5. Experiments

We present the detection and registration results on both synthetic and real images. Our method is implemented in MATLAB and the source code will be released. All the experiments are carried out on an Intel i7 3770 (3.4 GHz) machine with the following settings. In the initial matching step, the SIFT detector from the VLFeat package [26] with the default parameters is used to obtain initial feature points and descriptors (the matching threshold is set to 200 to obtain more correspondences). The outlier rejection is performed with $T_0 = 10$, $T_{final} = 2$, $r_T = 0.5$, $\lambda_0 = 1$, $d_0 = 30$, and $d_{final} = 3$.

For the quantitative evaluation of non-rigid surface detection and outlier rejection, the synthetic deformed surfaces are generated by TPS warp with 900 control points and these correspondences are considered as the ground truth. For the inlier matches, small noises are added to the points in the synthetic image. For each outlier ratio, the random test is carried out 10 times.

Non-Rigid Surface Detection. We evaluate the detection performance of our approach using synthetic images. Four images with different levels of deformation are generated and stitched together to form one image (i.e., four deformed instances in the synthetic image as shown in Figure 3(e)). We randomly sample 100 points in the template from the ground truth as the feature points denoted by F_t. We further sample N_m points from F_t to construct the inlier matches for each instance. Denote the number of instances by N_s and thus the total number of inlier matches is $N_t = N_s \times N_m$. Let the ratio of outliers to inliers be r_o and thus the number of outliers is $N_o = r_o \times N_t$. To generate these outlier matches, we randomly sample F_t for N_o times, and the corresponding points in the input image are randomly sampled to construct spurious correspondences. We compute the precision P_d and recall R_d of the surfaces to measure the detection performance. Moreover, we also measure the precision P_m and recall R_m of the matches in the detected truth surfaces, which is critical for the registra-

tion task. To evaluate the robustness to outliers, r_o is varied from 0.1 to 2 with interval of 0.1. To measure the performance with different number of inliers, N_m is varied from 40 to 80 with interval of 10.

As shown in Figure 3, our approach is robust to outliers. With the increase of outliers, the precision and recall of both detection and matching do not vary much. On the other hand, our method is affected by the number of inliers. When N_m is larger than 40, our approach performs well. However, the recall of matches decreases significantly when the number inliers is decreased (Figure 3(d)), which can be addressed by the proposed match growing algorithm.

Outlier Rejection. In this experiment, we perform a comparative evaluation on the outlier rejection algorithms. In the following, we refer the proposed approach as AOR (Annealing-based Outlier Rejection). Four images with different TPS deformation energies are generated. Some inliers are randomly selected from the ground truth. To evaluate the performance dependence on inliers, the number of inliers is varied from 20 to 110 with interval of 10. The number of control points is set to be the same as the inliers but the locations are different. In order to assess the robustness to the noise levels, the outlier correspondences are incrementally added. The points in the template are randomly generated and the corresponding points in the synthetic images are also randomly generated. The ratio of outliers is varied from 0.1 to 19.9 with interval of 0.2, and thus we have 100 outlier ratios.

We use the precision and recall of matches to measure the performance. Figure 4(a) shows the performance of AOR with different numbers of inliers. The precision and recall rates are monotonously increased with the increase of inliers, and the recall is always less than the precision. The increases of precision and recall are similar when the number of inliers is small (less than 50). However, when the number of inliers is further increased, the increase rate of recall is a little higher than precision. Figure 4(b) shows the performance of AOR with different outlier ratios. The precision and recall rates are decreased with the increase of outliers. The proposed algorithm is more robust to outliers when more inliers are obtained.

We compare our approach with the outlier rejection algorithm proposed in [24], LPOR (Local Topology based Outlier Rejection) for short, with the provided MATLAB implementation (http://isit.u-clermont1.fr/~ab/Research/index.html). This method assumes that the surface is locally smooth and that its local topology must be preserved. In contrast, our approach adopts a global model with an annealing process to handle outliers. As shown in Figure 4(c), the LPOR method can only deal with small amount of outliers. Its performance decreases significantly with increase of outliers and its precision is always lower than recall. This can be attributed
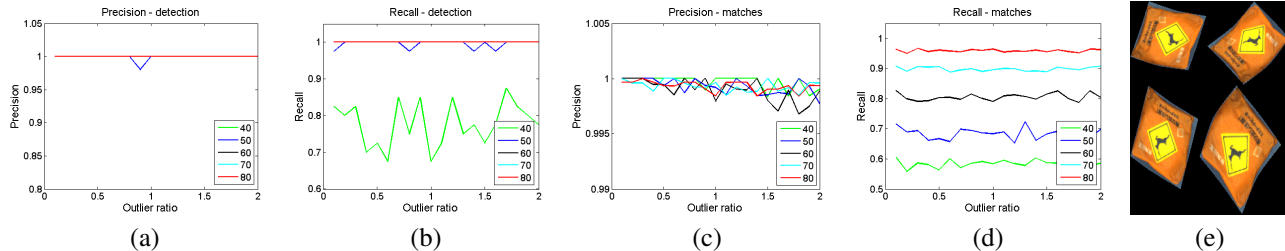
Figure 3. (a) Precision of detection. (b) Recall of detection. (c) Precision of matches. (d) Recall of matches. (e) Synthetic image.

outlier ratio	0.1	0.3	0.5	0.7	0.9	1.1
precision	0.97	0.91	0.54	0.12	0.18	0.06
recall	0.99	0.97	0.59	0.13	0.20	0.06

Table 1. Precision and recall of TPS-RPM with different levels of outliers.

to the fact that more outliers in the template will degrade the effect of the local topology structure assumed by the LPOR method. On the contrary, the performance of our approach is only slightly decreased with increase of outlier level. Even when the number of outliers is 19 times more than inlier, the precision is still above 0.95 and the recall is around 0.9. Furthermore, our approach is also much faster than the LPOR method. With 110 inliers and 2200 outliers, the average speed of the LPOR method is 148 seconds per frame while our approach takes only 0.25 seconds. The average iteration number of our approach for this challenging case is 33. One example of matches before and after outlier rejection is shown in the Figure 4(e-f).

We quantitatively evaluate the effect of proposed bootstrap process for deterministic annealing. The approach without bootstrap only takes 0.05 seconds to process the correspondence set with 110 inliers and 2200 outliers, and we refer it as fast AOR (FAOR for short). The experimental setup is the same as that with the LPOR method. As shown in Figure 4(d), when the outlier ratio is low (e.g., 3), the performance of the FAOR method is similar to that of the AOR approach. However, with increase of outliers, the recall of the FAOR method decreases more significantly than that of the AOR approach. On the other hand, the precision of the FAOR method is slightly lower than that of AOR approach after the ratio is larger (e.g., 17).

The TPS-RPM method [10] also uses the annealing for point matching. For comparison, we use it for the outlier rejection problem by initializing its assignment matrix in accordance with the correspondences. As shown Table 1, when the outlier ratio is low (< 0.5), the performance of the TPS-RPM method is similar to our algorithm. However, when the ratio is larger (≥ 0.5), both the precision and recall rates are very low.

Non-Rigid Surface Detection and Registration. Figure 1 shows the detection and registration results on the image, *T-shirt*. Our approach successfully detects all the deformed

surfaces of the given templates and we also get good registration results. Figure 5 and 6 illustrate some results on images of *home* and *ID* magazines, respectively. The results show that our approach obtains good estimates of the non-rigid warps even with large deformations. More results are available in the supplemental material.

6. Conclusions

In this paper, we present an algorithm to detect and register all the nonrigid instances of given templates from noisy observations. After obtaining the initial matches between the template and the input image, a novel high-order affinity graph is constructed to model the local topology consistency and then a hierarchical clustering approach is used to detect the nonrigid instances. We propose a deterministic annealing approach based on the TPS model to remove the spurious matches in each cluster. The proposed fusion approach exploits both appearance information and local features to register each nonrigid instance to the corresponding template. To improve the registration performance, a TPS based match growing scheme is developed to propagate the matches. Extensive experiments demonstrate the effectiveness and efficiency of the proposed algorithm for multiple non-rigid surface detection and registration.

References

[1] S. Baker and I. Matthews. Lucas-Kanade 20 Years On: A Unifying Framework. *IJCV*, 56(3):221–255, 2004.

[2] A. Bartoli. Maximizing the Predictivity of Smooth Deformable Image Warps through Cross-Validation. *JMIV*, 31(2-3):133–145, 2008.

[3] A. Bartoli and A. Zisserman. Direct Estimation of Non-Rigid Registrations. In *BMVC*, 2004.

[4] A. C. Berg, T. L. Berg, and J. Malik. Shape Matching and Object Recognition using Low Distortion Correspondences. In *CVPR*, 2005.

[5] F. L. Bookstein. Principal Warps: Thin-Plate Splines and The Decomposition of Deformations. *PAMI*, 11(6):567–585, 1989.

[6] F. Brunet, V. Gay-Bellile, A. Bartoli, N. Navab, and R. Malgouyres. Feature-Driven Direct Non-Rigid Image Registration. *IJCV*, 93(1):33–52, 2011.

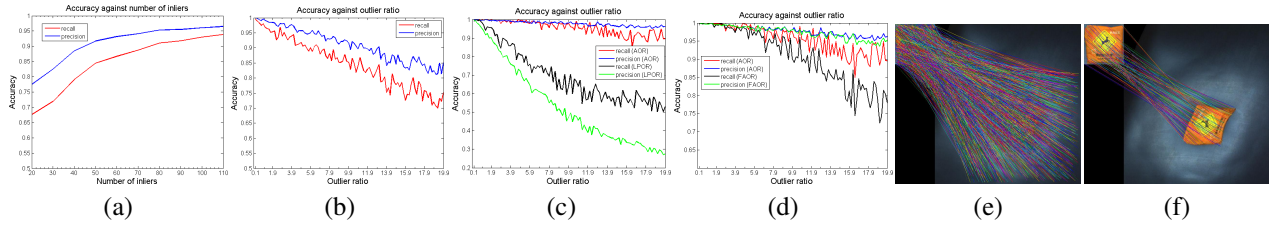[7] M. Chertok and Y. Keller. Efficient High Order Matching. *PAMI*, 32(12):2205–15, 2010.

Figure 4. (a) Performance of AOR with different numbers of inliers. (b) Performance of AOR with different outlier ratios. (c) Comparison of AOR and LPOR. (d) Comparison of AOR and FAOR. (e) Matches before outlier rejection. (f) Refined matches by AOR.
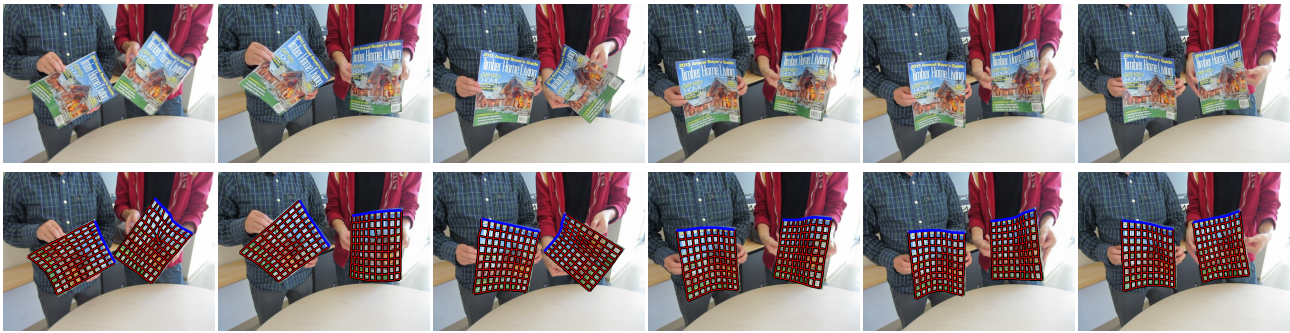

Figure 5. Detection and registration results on the *home* magazine images.

Figure 6. Detection and registration results on the *ID* magazine images.

[8] M. Cho, J. Lee, and K. M. Lee. Feature Correspondence and Deformable Object Matching via Agglomerative Correspondence Clustering. In *CVPR*, 2009.

[9] M. Cho, Y. M. Shin, and K. M. Lee. Unsupervised Detection and Segmentation of Identical Objects. In *CVPR*, 2010.

[10] H. Chui and A. Rangarajan. A New Point Matching Algorithm for Non-Rigid Registration. *CVIU*, 89(2-3):114–141, 2003.

[11] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. *PAMI*, 23(6):681–685, 2001.

[12] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce. A Tensor-Based Algorithm for High-Order Graph Matching. *PAMI*, 33(12):2383–2395, 2011.

[13] V. Ferrari, T. Tuytelaars, and L. V. Gool. Simultaneous Object Recognition and Segmentation from Single or Multiple Model Views. *IJCV*, 67(2):159–188, 2006.

[14] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6):381–395, 1981.

[15] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. 2000.

[16] M. Leordeanu and M. Hebert. A Spectral Technique for Correspondence Problems Using Pairwise Constraints. In *ICCV*, 2005.

[17] J. Lim and M.-H. Yang. A Direct Method for Modeling Non-Rigid Motion with Thin Plate Spline. In *CVPR*, 2005.

[18] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004.

[19] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *IJCAI*, 1981.

[20] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *ECCV*, 2002.

[21] I. Matthews and S. Baker. Active Appearance Models Revisited. *IJCV*, 60(2):135–164, 2004.

[22] D. Ok, R. Marlet, and J.-Y. Audibert. Efficient and Scalable 4th-order Match Propagation. In *ACCV*, 2012.

[23] J. Pilet, V. Lepetit, and P. Fua. Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation. *IJCV*, 76(2):109–122, 2008.

[24] D. Pizarro and A. Bartoli. Feature-Based Deformable Surface Detection with Self-Occlusion Reasoning. *IJCV*, 97(1):54–70, 2012.

[25] L. Torresani, V. Kolmogorov, and C. Rother. Feature Correspondence via GraphMatching: Models and Global Optimization. In *ECCV*, 2008.

[26] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms, 2008.

[27] A. L. Yuille. Generalized Deformable Models, Statistical Physics, and Matching Problems. *Neurul Computation*, 24(1987):1–24, 1990.

[28] R. Zass and A. Shashua. Probabilistic Graph and Hypergraph Matching. In *CVPR*, 2008.

[29] J. Zhu, M. R. Lyu, and T. S. Huang. A Fast 2D Shape Recovery Approach by Fusing Features and Appearance. *PAMI*, 31(7):1210–1224, 2009.