Progressive Finite Newton Approach To Real-time Nonrigid Surface Detection

Jianke Zhu and Michael R. Lyu
Department of Computer Science & Engineering
Chinese University of Hong Kong, Shatin, Hong Kong
{jkzhu,lyu}@cse.cuhk.edu.hk

Abstract

Detecting nonrigid surfaces is an interesting research problem for computer vision and image analysis. One important challenge of nonrigid surface detection is how to register a nonrigid surface mesh having a large number of free deformation parameters. This is particularly significant for detecting nonrigid surfaces from noisy observations. Nonrigid surface detection is usually regarded as a robust parameter estimation problem, which is typically solved iteratively from a good initialization in order to avoid local minima. In this paper, we propose a novel progressive finite Newton optimization scheme for the nonrigid surface detection problem, which is reduced to only solving a set of linear equations. The key of our approach is to formulate the nonrigid surface detection as an unconstrained quadratic optimization problem which has a closed-form solution for a given set of observations. Moreover, we employ a progressive active-set selection scheme, which takes advantage of the rank information of detected correspondences. We have conducted extensive experiments for performance evaluation on various environments, whose promising results show that the proposed algorithm is more efficient and effective than the existing iterative methods.

1. Introduction

The detection and tracking of the nonrigid objects in images and videos is an interesting and beneficial research issue for computer vision and image analysis [1, 19, 21]. The goal of nonrigid surface detection is to extract the deformable shape's structure from an input image. The difference between nonrigid surface recovery and detection is that the latter does not require any initialization or a priori pose information. An effective nonrigid surface detection technique can be applied in a variety of applications for digital entertainment, medical imaging [1] and augmented reality, such as the re-texturing of images and videos [22, 23].

Nonrigid surface detection can usually be treated as the problem of recovering the explicit surface with a few deformation parameters and finding out the correct correspondences from noisy data simultaneously. Many applications have been investigated for deformable object tracking and registration, such as face tracking and modelling [6, 24, 25, 3, 7], and also more generic and more deformable objects [1]. The major problem of these methods is that they tend to be computationally expensive and mainly aim at object recognition and image segmentation tasks rather than nonrigid surface recovery. However, a real-time and automated solution [19] has recently been proposed, which takes advantage of an iterative robust optimization scheme.

Unlike the rigid object pose estimation, it is difficult to directly employ a robust estimator, such as RANSAC [8] or Hough transform [10], to remove the spurious matches for nonrigid surface detection. Because the nonrigid surface is usually highly dynamic and represented by many deformation parameters, the problem is far more complex than the rigid object detection. Moreover, it requires a sufficient number of correct correspondences in order to obtain high registration accuracy. An alternative strategy is to iteratively solve for both the correspondence and the transformation [2, 19]. However, these methods are either sensitive to initial conditions and parameter choices, or involve too many iterations and a complex optimization procedure. Consequently, they are neither efficient nor effective for real-time applications.

In this paper, we propose a novel progressive finite Newton optimization scheme for nonrigid surface detection, which has the advantage of solving only a fixed number of linear equations. Moreover, a progressive sample scheme far more efficient than RANSAC is proposed to initialize the optimization process. The previous method [19] is currently generally accepted as the most effective methods of solving this kind of problem. It employs an implicit iterative scheme for the first order partial differential equation; however, this requires a large number of iterations to solve the problem and remove the outliers simultaneously. We tackle this critical problem from two angles. First, the nonrigid surface detection is formulated as an unconstrained

| (a) Starbucks pad | (b) T-shirt | (c) Cover | (d) Paper |

Figure 1. Detecting nonrigid surfaces in real-time video (a-d). (a) The contour is overlaid on the Starbucks pad. (b) T-shirt with shadow. (c) The cover of a magazine. (d) A piece of paper with specular reflection.

quadratic optimization problem, which has a closed-form solution for a given set of observations. Thus, it can be efficiently solved through LU factorization. Then, a progressive sample [5] scheme is employed to initialize the optimization scheme, which can decrease the number of trials significantly. Therefore, the present approach requires much fewer iterations than the semi-implicit iterative optimization scheme [20]. Thus, the proposed method is very efficient for real-time nonrigid surface recovery tasks. To evaluate the performance of our proposed algorithm, we conduct extensive experiments on such diverse objects as a Starbucks pad, a T-shirt, and the cover of a magazine, as shown in Fig. 1.

The rest of this paper is organized as follows. Section 2 reviews the previous approaches employed for the nonrigid surface detection and recovery. In Section 3, we present the proposed progressive finite Newton solution. Section 3.1 describes the nonrigid surface model and mapping function for a feature matching-based method. Section 3.2 presents the object function which minimizes the correspondence errors and surface energy. A robust estimator is introduced to deal with the large outliers. In Section 3.3, the nonrigid surface detection is formulated as an unconstrained quadratic optimization problem, which is efficiently solved using the factorization method. Section 3.4 presents the progressive finite Newton optimization scheme to remove the spurious correspondences, and the progressive sampling method to initialize the optimization. Section 4 provides the details of our experimental implementation and describes our experimental results. We discuss limitations and future work in Section 5. Section 6 sets out our conclusion.

2. Related Work

Although nonrigid surface detection in general is not new to researchers in the computer vision domain, only a few approaches are automatic and can achieve real-time results. Some appearance-based approaches directly minimize the residual image between the input image and the synthesized model image [6]. Moreover, optical flow information [7, 1] can be incorporated into the optimization

scheme to obtain better results. However, the major limitation of these methods is that they tend to become stuck at a local minimum and hence require good initialization. In addition, it is usually difficult to handle the partial occlusion for an appearance-based method. Well-designed markers widely used in motion capture are also applied to recover the structure of a nonrigid surface, such as cloth and paper [23, 22]. As these methods rely on the physical markers, they require the placing of pre-defined patterns on the target surface. Nevertheless, they are capable of high accuracy. On the other hand, feature-based methods [2, 19] try to find out the transformation from the correspondences built by feature matching methods. Thus, these methods can benefit from the recent advances in the feature detection and matching. In [19, 20], J. Pilet et al. proposed an iterative approach to attack the fast nonrigid surface recovery problem. Physical constraints based on the Finite Element Model [21] are employed for regularization. A semi-implicit iterative scheme is proposed to solve the optimization problem.

Recently, several sophisticated feature descriptors [16, 18] have been proposed to handle the wide-baseline matching problem, including images with large deformation [15]. In addition, machine learning methods, such as random classification trees [14], are also employed to find the point correspondences. These methods can take advantage of shifting part of the computational load from the matching phase to the training phase.

It is more complex to handle a large amount of deformation parameters for detecting the nonrigid surface rather than only a few pose parameters used in rigid object detection. Therefore, there are several challenges when applying conventional robust estimators, such as RANSAC and M-estimator, for the nonrigid surface detection task. One is the lack of a concise function which can estimate the deformed mesh from the correspondences directly; instead, one may need to use a large number of free variables, which can lead to a high computational cost for each prediction step. Obviously, the semi-implicit iterative approach [19] is not efficient enough to deal with this problem. Another challenge is that the RANSAC-based approach requires a

large number of trials. This makes the problem even more complex. Moreover, to the best of our knowledge, there is still a lack of criteria for selecting the number of samples for each trial in nonrigid surface detection. In rigid object pose estimation, the sample number is usually set according to the number of free parameters. However, the number of deformation parameters for a nonrigid surface may be larger than the total number of observations. We tackle the initialization problem through a modified RANSAC method. The key is to draw from progressively larger sets of top-ranked correspondences [5], rather than to treat all correspondences as equal and draw random samples uniformly from the full set in RANSAC. Thus, our progressive sample scheme affords large computational savings, and the conventional robust estimator can be engaged for initializing the nonrigid surface detection.

In contrast to the previous work, our proposed approach is based on a progressive finite Newton scheme, in which the optimization problem can be solved very efficiently by a factorization method. In addition to offering computationally highly competitive performance, our proposed modified RANSAC initialization method can further reduce the number of Newton optimization steps.

3. Nonrigid Surface Detection

In this section, we describe the present progressive finite Newton optimization scheme for detecting and recovering the nonrigid surface. For tackling the challenges, a mapping function is used to associate the feature correspondences with a mesh model. Therefore, the nonrigid surface detection turns out to be a problem which minimizes the correspondence error and the surface energy. Moreover, we formulate the nonrigid surface detection into an unconstrained quadratic optimization problem. A progressive scheme is proposed to deal with outliers and find out as many correct correspondences as possible. Finally, a modified RANSAC scheme is introduced to select the initial active set for the optimization scheme.

3.1. 2D Nonrigid Surface Model

The nonrigid surface is usually explicitly represented by triangulated meshes. As shown in Fig. 1(c), we employ a triangulated 2D mesh with N hexagonally connected vertices, which are formed into a shape vector $\mathbf{s}$ as below:

$$\mathbf{s} = \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix}^\top$$
$$= \begin{bmatrix} x_1 & x_2 & \dots & x_N & y_1 & y_2 & \dots & y_N \end{bmatrix}^\top$$

where $\mathbf{x}$ and $\mathbf{y}$ are the vectors of the coordinates of mesh vertices. We assume that a point $\mathbf{m}$ lies in a triangle whose three vertices' coordinates are $(x_i, y_i), (x_j, y_j)$ and (x_k, y_k) respectively, and $\{i, j, k\} \in [1, N]$ is the index of each vertex. The piecewise affine transformation is used to map the

image points inside the corresponding triangle into the vertices in the mesh. Thus, the mapping function $T_S(\mathbf{m})$ is defined as below

$$T_{\mathbf{s}}(\mathbf{m}) = \begin{bmatrix} x_i & x_j & x_k \\ y_i & y_j & y_k \end{bmatrix} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \end{bmatrix}^\top \quad (1)$$

where (ξ_1, ξ_2, ξ_3) are the barycentric coordinates for the point $\mathbf{m}$.

3.2. Nonrigid Surface Recovery

In general, the nonrigid surface detection problem approximates a 2D mesh with $2N$ free variables, which is usually ill-posed. One effective way to attack this problem is to introduce regularization, which preserves the regularity of a deformable surface. The following object function is widely used in deformable surface fitting [11, 12, 19, 20] for energy minimization:

$$E(\mathbf{s}) = E_e(\mathbf{s}) + \lambda_r E_r(\mathbf{s}) \quad (2)$$

where $E_e(\mathbf{s})$ is the sum of the weighted square error residuals for the matched points. Also, $E_r(\mathbf{s})$ is the regularization term that represents the surface deformation energy, and λ_r is a regularization coefficient.

A set of correspondences M between the model and the input image can be built through a point matching algorithm. Therefore, a pair of matched points is represented in the form of $\mathbf{m} = \{\mathbf{m}_0, \mathbf{m}_1\} \in M$, where $\mathbf{m}_0$ is defined as the 2D coordinates of a feature point in the training image and $\mathbf{m}_1$ is the coordinates of its match in the input image. Then, the correspondence error term $E_e(\mathbf{s})$ is formulated as below:

$$E_e(\mathbf{s}) = \sum_{\mathbf{m} \in M} \omega_{\mathbf{m}} \mathcal{V}(\delta, \sigma) \quad (3)$$

where $\mathcal{V}(\delta, \sigma)$ is a robust estimator, and $\omega_{\mathbf{m}} \in [0, 1]$ is a weight linked with each correspondence.

The regularization term E_r in Equ. 2, also known as internal force in Snakes [12], is composed of the sum of the squared second-order derivatives of the mesh vertex coordinates. As the mesh is regular, $E_r(\mathbf{s})$ can be formulated through a finite difference:

$$E_r(\mathbf{s}) = \mathbf{s}^\top \begin{bmatrix} K & 0 \\ 0 & K \end{bmatrix} \mathbf{s} \quad (4)$$

where K is a sparse and banded matrix which is determined by the structure of the explicit mesh model [9].

3.3. Finite Newton Formulation

In this paper, we employ a robust estimator $\mathcal{V}(\delta, \sigma)$ with compact support size σ. Moreover, δ is the residual error, which is defined as follows:

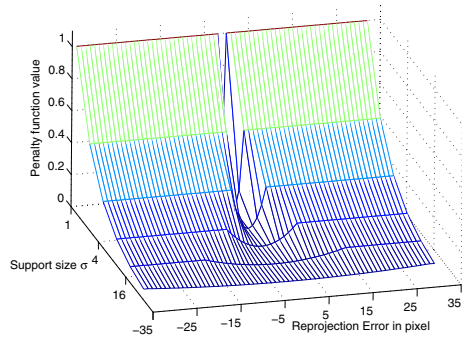$$\delta = \mathbf{m}_1 - T_{\mathbf{s}}(\mathbf{m}_0) \quad (5)$$

Figure 2. The robust estimator that assesses a fixed penalty to residuals larger than a threshold σ.

The robust estimator function $\mathcal{V}(\delta, \sigma)$ that assesses a fixed penalty for residuals larger than a threshold σ is employed in the present work; this approach is relatively insensitive to outliers [4]:

$$\mathcal{V}(\delta, \sigma) = \begin{cases} \frac{\|\delta\|}{\sigma^n}, & M_1 = \{\mathbf{m}|\quad \|\delta\| \le \sigma^2\} \\ \sigma^{2-n}, & M_2 = \overline{M_1} \end{cases} \quad (6)$$

where the set M_1 contains the inlier matches, and M_2 is the set of the outliers. In addition, the order n determines the scale of the residual. As shown in Fig. 2, the most correspondences are included when the support σ is large. As σ decreases, the robust estimator becomes narrower and more selective.

Since the robust estimator function is not convex, the associated penalty function approximation problem becomes a hard combinational optimization problem. We tackle this problem under the finite Newton optimization framework. An augmented vector $\mathbf{t} \in R^N$ containing the barycentric coordinates is defined as below:

$$\mathbf{t}_i = \xi_1 \quad \mathbf{t}_j = \xi_2 \quad \mathbf{t}_k = \xi_3$$

while the remaining elements in the vector $\mathbf{t}$ are all set to zero. Therefore, the residuals for the inlier correspondences can be rewritten as follows:

$$\|\boldsymbol{\delta}\| = (u - \mathbf{t}^\top \mathbf{x})^2 + (v - \mathbf{t}^\top \mathbf{y})^2$$

where (u, v) are the coordinates of $\mathbf{m}_1$. Therefore, the error term in Equ. 3 turns out to be

$$\begin{aligned} E_e &= \sum_{\mathbf{m} \in M_1} \frac{\omega_{\mathbf{m}}}{\sigma^n} \left(u^2 + v^2 - 2 \begin{bmatrix} u\mathbf{t} \\ v\mathbf{t} \end{bmatrix}^\top \mathbf{s} \right. \\ &+ \left. \mathbf{s}^\top \begin{bmatrix} \mathbf{t}\mathbf{t}^\top & 0 \\ 0 & \mathbf{t}\mathbf{t}^\top \end{bmatrix} \mathbf{s} \right) + q\sigma^{2-n} \end{aligned}$$

where q is the number of outliers.

Let $\mathbf{b} \in R^{2N}$ be defined as below:

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_x \\ \mathbf{b}_y \end{bmatrix} = \sum_{\mathbf{m} \in M_1} \frac{\omega_{\mathbf{m}}}{\sigma^n} \begin{bmatrix} u\mathbf{t} \\ v\mathbf{t} \end{bmatrix} \quad (7)$$

and a matrix $A \in R^{N \times N}$ is equal to

$$A = \sum_{\mathbf{m} \in M_1} \frac{\omega_{\mathbf{m}}}{\sigma^n} \mathbf{t}\mathbf{t}^\top \quad (8)$$

Thus, the energy function Equ. 3 is formulated into an unconstrained quadratic optimization problem, which can be solved by the modified finite Newton method [17, 13].

$$\begin{aligned} E &= \mathbf{s}^\top \begin{bmatrix} \lambda_r K + A & 0 \\ 0 & \lambda_r K + A \end{bmatrix} \mathbf{s} - 2\mathbf{b}^\top \mathbf{s} \\ &+ q\sigma^{2-n} + \sum_{\mathbf{m} \in M_1} \frac{\omega_{\mathbf{m}}}{\sigma^n} (u^2 + v^2) \end{aligned}$$

The finite gradient of the energy function E with respect to $\mathbf{s}$ can be derived as below:

$$\nabla = 2 \left(\begin{bmatrix} \lambda_r K + A & 0 \\ 0 & \lambda_r K + A \end{bmatrix} \mathbf{s} - \begin{bmatrix} \mathbf{b}_x \\ \mathbf{b}_y \end{bmatrix} \right)$$

and the Hessian [4] can also be computed by

$$H = 2 \begin{bmatrix} \lambda_r K + A & 0 \\ 0 & \lambda_r K + A \end{bmatrix}$$

Thus the gradient can be rewritten as below:

$$\nabla = H\mathbf{s} - 2\mathbf{b}$$

Each Newton step will perform the following operation:

$$\mathbf{s} \leftarrow \mathbf{s} - \gamma H^{-1} \nabla$$

where γ is the step size. We simply set it equal to one, and no convergence problem occurs in our experiments. Since K is regular, we find that update of the state vector $\mathbf{s}$ can be computed by the following linear equation:

$$\begin{bmatrix} \lambda_r K + A & 0 \\ 0 & \lambda_r K + A \end{bmatrix} \mathbf{s} = \begin{bmatrix} \mathbf{b}_x \\ \mathbf{b}_y \end{bmatrix}$$

Moreover, the problem can be further simplified into two linear equations, which can be efficiently solved via LU decomposition:

$$\mathbf{x} = (\lambda_r K + A)^{-1} \mathbf{b}_x \quad (9)$$
$$\mathbf{y} = (\lambda_r K + A)^{-1} \mathbf{b}_y \quad (10)$$

The overall complexity is thus the complexity of one Newton step.

3.4. Progressive Finite Newton Optimization

Generally speaking, the incorrect matches cannot be avoided in the first stage of the matching process where only local image descriptors are compared. We introduce a coarse-to-fine scheme to deal with those outliers. The support σ of robust estimator $\mathcal{V}(\delta, \sigma)$ is progressively decayed at a constant rate α. Since the derivatives of $\mathcal{V}(\delta, \sigma)$ are inversely proportional to the support σ, the regularization coefficient λ_r is kept constant during the optimization. For each value of σ, the object function E is minimized through the finite Newton step and the result is employed as the initial state for the next minimization. The minimization of E is directly solved through Equ. 9 and Equ. 10 for a given initial state, and one step is enough to achieve convergence. The optimization procedure stops when σ reaches a value close to the expected precision, which is usually one or two pixels. The algorithm reports a successful detection when the number of inlier matches is above a given threshold. Thus, the whole optimization problem can be solved within a fixed number of steps. This is in contrast to the semi-implicit optimization scheme [20], which involves a few iterations for each σ, and at least 40 iterations in total to ensure the convergence.

In order to select most of the correspondences into the initial active set and avoid getting stuck at local minima, the initial value of σ is usually set to a sufficiently large value. However, this requires a fixed initial state. The method is dependent on the object position, and needs a few iterations to compensate for the errors generated by the pose variations. In the present work, we solve this problem through a modified RANSAC approach. Taking advantage of our concise finite Newton formulation and closed-form solution, the explicit mesh can be directly estimated from a given set of correspondences. Moreover, we draw from progressively larger sets of top-ranked correspondences, which decreases the number of trials significantly. In the experiments, the sampling process stopped within 5 trials. In the worst case, such as when an object does not appear in the scene, it still converges towards RANSAC. Therefore, the output of the proposed progressive sample can be employed as the initial state for the finite Newton optimization. Since the result of progressive sample estimation is quite close to the solution, σ is relatively small. Thus, the proposed progressive scheme requires fewer stages, and is somewhat invariant to the initial position.

4. Experimental Results

In this section, we discuss the details of our experimental implementation and report the results of performance evaluation on nonrigid surface detection. We show that the proposed approach is very efficient for real-time tracking, and can be easily employed for augmented reality applications.

4.1. Experimental Setup

In order to register the mesh model conveniently, a model image is acquired when the nonrigid surface contains no deformation. In order to facilitate real-time augmented reality applications, a random-trees based method [14] is used to build the correspondences between the model image and input image.

Since the number of free variables for nonrigid surface recovery is usually quite large (even up to one thousand), the sample size of each RANSAC iteration becomes a tricky issue. We compare the performance with different sample sizes. In our experiments, the support σ is empirically set to 30, and λ_r is set to a large value to ensure the regularity of the nonrigid surface. Interestingly, we find that the best sample size is three. This is because the nonrigid surface degenerates into a rigid one, and only three points are necessary to determine the position of a rigid surface. Moreover, when the sample size increases, the probability of selecting the inlier data is decreased. Thus, three is the best choice for the sample size.

In the finite Newton optimization, the weighting scheme is beneficial for a single step. However, it changes the scale of the error term in the object function, and so the regularization coefficient λ_r is no longer kept constant during the optimization. In our experiments, all weight coefficients ω are set to one. A set of synthetic data is used to select the parameters, and the reference mesh is manually registered. The performance is evaluated by the percentage of mesh vertices within 2 pixels of those in the reference mesh. The best regularization coefficient is found to be around 3×10^{-4} by grid searching. Similarly, the initial support σ_0 is set to 80, and decay rate α is 0.5. Fig. 4 plots the success probability with different orders n of the robust estimator function. Based on these results, n is set to 4.
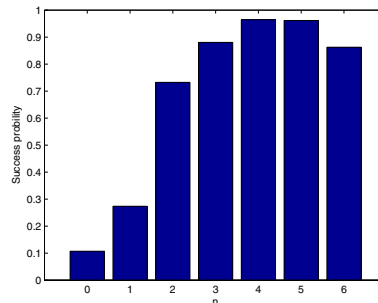

Figure 4. Probability of success with different order n of the robust estimator function.

All the experiments reported in this paper were carried out on a Pentium-4 3.0GHz PC with 1GB RAM, and a DV camera was engaged to capture video. We also implemented a semi-implicit iterative method [20], which is regarded as

| (a) Model image | (b) Result | (c) Result | (d) Result | (e) Plastic cup |

Figure 3. We use a Starbucks pad as the deformable object. The model image is shown in (a) the contour of the model image is extracted using a simple gradient and filling operator, which is overlaid on the input image. (b) to (d) show the results. (e) shows the result where the pad is replaced by a plastic cup. The model contains 120 vertices, and the whole process, including image capturing and rendering, runs around 18 frames per second. More results are included in the supplementary material.

the state-of-the-art approach.

4.2. Computational Efficiency

The complexity of the proposed method is mainly dominated by the order of Equ. 9 and Equ. 10, which is equal to the number of vertices N in the mesh model. Another important factor is the number of inlier matches, which affects the sparseness of matrix A. This usually differs from one frame to another. For the Starbucks pad with 120 vertices, as shown in Fig. 3, the proposed method runs at 18 frames per second on real-time video with size of 720×576. As depicted in Table 1, the proposed optimization scheme requires around 8 iterations and only takes half of the time of the feature matching algorithm, which is the bottleneck of the whole system. Our implementation [1] of semi-implicit iterative approach [20] needs around 40 iterations to reach the convergence, and runs about 9 frames per second. The improvement is more significant for high resolution mesh. Thus, the proposed method requires far less iterations, and is efficient for real-time applications. We also conduct the experiments without using the modified RANSAC initialization, and start the optimization scheme from a sufficiently large support $\sigma = 1000$. This requires 11 iterations, and the fitting accuracy is worse than the proposed method. In addition, the modified RANSAC initialization can also be used for a semi-implicit method, in which case the number of iterations is reduced to around 25.

Table 1. Computational time of proposed method at each step.

Total	Match	Optimization	Iteration	Other
57ms	27ms	14ms	~ 1.9ms	16ms

4.3. Performance of Nonrigid Surface Recovery

We use a Starbucks pad as the deformable object. As shown in Fig. 3, the proposed method is robust to large deformations and perspective distortion. In practice, the whole process runs at around 18 frames per second. Fig. 6 describes the result of detecting a piece of paper, where similar performance is achieved. As another feature-based method, the performance of the proposed method is closely related to the texture of objects. Better results can be obtained for objects with more texture, because it is easy to find more correct correspondences than with those lacking texture. This problem may be solved by incorporating global appearance and edge constraints into the optimization scheme.

4.4. Augmented Reality

Once the nonrigid surface is recovered, an immediate application is to re-texture an image. In order to obtain realistic results, the texture should be correctly relighted. As suggested in [20], a re-textured input image is generated by directly multiplying a blank shaded image, which is the quotient of the input image and the warped reference image. The reference image is acquired when the nonrigid surface is lighted uniformly. Moreover, the quotient image is normalized through multiplying the intensity of white color in the reference image. This relighting procedure is easily

---

[1]We use the same parameters setting as [20]. The convergence condition is set to 0.9995, with at most 5 iterations for each support value σ.

Figure 5. Re-texturing of a shirt print. The first row shows the 720×576 images captured by a DV camera. The second row shows the results of replacing the bunny with the CVPR logo.

done by the GPU and requires only a short OpenGL shading language program; and the whole process runs at about 17 frames per second. Fig. 5 shows the results of re-texturing a T-shirt with a Lambertian surface. It is difficult to estimate a blank shaded image due to dividing near zero intensity values and the use of an uncontrolled optical sensor. However, the visual effect is that the bunny in the input video is re-textured by the CVPR logo. For a specular surface, Fig. 6 describes the results on a piece of paper with a saturated region. In addition, the right two columns of Fig. 6 show the results in a cluttered environment.

5. Discussions and Future Work

We have proposed a novel scheme for non-rigid surface detection by progressive finite Newton optimization. In comparison with semi-implicit optimization methods [20], the proposed method has several advantages. First, we need not solve the optimization iteratively for every σ, because it can be solved in one step directly. Second, the iterative method starts from a sufficiently large support value in order to estimate the location and pose of an object, which leads to a large number of iterations. Thus, the proposed method is far more efficient than the semi-implicit method. Additionally, it is easy to implement the proposed approach, which only involves solving the sparse linear equation, and does not require tuning the viscosity parameters and a sophisticated Levenberg-Marquardt optimization algorithm.

Although promising experimental results have validated the efficiency of our methodology, some limitations and future directions should be addressed. First of all, some jitter may occur due to the point matching algorithm or the lack of texture information. Also, we have focused our attention only on single deformable surface detection, whereas it is also interesting to study the multiple case. In future work, global bundle-adjustment will be introduced to tackle the jittering problem. Furthermore, an efficient octree structure can be used to build a simplified multi-scale mesh model. Finally, we may consider extending the proposed scheme to 3D environments and exploring new regularization methods.

6. Conclusion

This paper presented a novel progressive scheme to solve the non-rigid surface detection problem. In contrast to the previous approaches involving iterative and explicit minimization, we proposed a progressive finite Newton algorithm, which directly solves the unconstrained quadratic optimization problem by an efficient factorization method. Moreover, our modified RANSAC scheme takes advantage of our concise formulation and progressive sampling of the top-ranked correspondences, and can handle high-dimensional spaces with noisy data.

We have conducted extensive experimental evaluations on diverse objects with different materials. The proposed method is very fast and robust, and can handle large deformations and illumination changes. It was tested in several applications, such as real-time Augmented Reality and medical image registration. The promising experimental results showed that our algorithm is more efficient and effective than previous methods.

Figure 6. Re-texturing a picture on a piece of paper. The first row is the 720×576 images captured by a DV camera. The second row is the results of replacing the picture with the CVPR logo.

References

[1] A. Bartoli and A. Zisserman. Direct estimation of non-rigid registration. In *BMVC'04*, Kingston, Sep. 2004. 1, 2

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. PAMI*, 24(4):509–522, 2002. 1, 2

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. PAMI*, 25(9), 2003. 1

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 4

[5] O. Chum and J. Matas. Matching with prosac- progressive sample consensus. In *CVPR'05*, volume 1, pages 220–226, 2005. 2, 3

[6] T. Cootes, G. Edwards, and C. Taylo. Active appearance models. *IEEE Trans. PAMI*, 23(6), June 2001. 1, 2

[7] D. DeCarlo and D. N. Metaxas. Optical flow constraints on deformable models with applications to face tracking. *IJCV*, 38(2):99–127, 2000. 1, 2

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 24(6):381–395, 1981. 1

[9] P. Fua and Y. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *IJCV*, 16(1):35–56, Sep. 1995. 3

[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000. 1

[11] S. Ilic and P. Fua. Implicit meshes for surface reconstruction. *IEEE Trans. PAMI*, 28(2):328–333, February 2006. 3

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, Jan. 1988. 3

[13] S. S. Keerthi and D. DeCoste. A modified finite newton method for fast solution of large scale linear svms. *J. Mach. Learn. Res.*, 6:341–361, 2005. 4

[14] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *IEEE Trans. PAMI*, 2006. 2, 5

[15] H. Ling and D. W. Jacobs. Deformation invariant image matching. In *ICCV'05*, pages 1466–1473, 2005. 2

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

[17] O. L. Mangasarian. A finite newton method for classification. *Optimization Methods and Software*, 17(5):913–929, 2002. 4

[18] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. PAMI*, 27(10):1615–1630, 2005. 2

[19] J. Pilet, V. Lepetit, and P. Fua. Real-time non-rigid surface detection. In *CVPR (1)*, pages 822–828, 2005. 1, 2, 3

[20] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration, and realistic augmentation. *IJCV*, 2006, to appear. 2, 3, 5, 6, 7

[21] L. V. Tsap, D. B. Goldgof, and S. Sarkar. Nonrigid motion analysis based on dynamic refinement of finite element models. *IEEE Trans. PAMI*, 22(5):526–543, 2000. 1, 2

[22] R. White and D. Forsyth. Retexturing single views using texture and shading. In *ECCV*, pages 70–81, 2006. 1, 2

[23] R. White and D. A. Forsyth. Combining cues: Shape from shading and texture. In *CVPR (2)*, pages 1809–1816, 2006. 1, 2

[24] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *IEEE CVPR'04*, volume 2, pages 535–542, 2004. 1

[25] J. Zhu, S. C. Hoi, and M. R. Lyu. Real-time non-rigid shape recovery via active appearance models for augmented reality. In *Proc. European Conf. Computer Vision*, pages 186–197, 2006. 1