Hierarchical Structure and Nonrigid Motion Recovery from 2D Monocular Views

Abstract

Inferring both 3D structure and motion of nonrigid objects from monocular images is an important problem in computational vision. The challenges stem not only from the absence of point correspondences but also from the structure ambiguity. In this paper, a hierarchical method which integrates both local patch analysis and global shape descriptions is devised to solve the dual problem of structure and nonrigid motion recovery by using an elastic geometric model—extended superquadrics. The nonrigid object of interest is segmented into many small areas and local analysis is performed to recover small details for each small area, assuming that each small area is undergoing similar nonrigid motion. Then, a recursive algorithm is proposed to guide and regularize local analysis with global information by using an appropriate global shape model. This local-global hierarchy enables us to capture both local and global deformations accurately and robustly. Experimental results on both simulation and real data are presented to validate and evaluate the effectiveness and robustness of the proposed approach.

1 Introduction

Estimation of 3D structure and motion from 2D monocular image sequences has been a central problem in computer vision for many years. Many studies have focused on recovering structure from feature correspondences or optic flow by using rigidity assumption. However, the rigidity assumption is inadequate for representing motions in many real-world examples. Nonrigid motion is ubiquitous. For instance, most biological objects are flexible and articulated: fingers bend, cheeks stretch, neck and body twist, fish swim, and trees sway. Nonrigid motion analysis has drawn significant attention, as it is important in many computer vision problems, and is motivated by numerous applications. In this paper, our goal is to deal with the recovery of structure and nonrigid motion from 2D monocular views when the object of interest is undergoing a deformation.

Several efforts have been made to solve the problem of structure and nonrigid motion recovery. In general, the existing methods can be classified into two categories: the local patch based methods and the model based methods. The local patch based methods simply abandon the idea of recovering a whole body description of the motion, and recover structure on a patch-by-patch basis in order to cope with nonrigidity [16, 20]. Unfortunately, using a local description limits the methods themselves to using only noise-sensitive local measurements. Consequently, such patch-by-patch recovery of structure is not likely to be either very meaningful or robust. On the contrary, the model based methods utilized predefined 3D active shape models to recover a global description of nonrigid motion. Essentially, the predefined shape models provide extra constraints which incorporate prior knowledge about a shape's smoothness and its resistance to deformation. A number of different 3D deformable model formulations have been proposed and actually used for the task of structure and nonrigid motion recovery [19, 15, 13, 14, 8]. Terzopoulos *et.al.* [19] proposed a physically based modeling framework and recovered limited global descriptions (e.g. symmetry axis shape) from nonrigid motion. Pentland *et.al.* [15] used finite element method (FEM) models which incorporate elastic properties of real materials for the recovery of structure and nonrigid motion. Huang and Goldgof [8] proposed the adaptive-size physically-based models for nonrigid shape and motion analysis. Many applications of model-based structure and nonrigid motion recovery were also presented [9, 11, 17, 18, 7, 3, 12]. For example, Kakadiaris and Metaxas [9] discussed the human body tracking, DeCarlo and Metaxas [7] estimated the shape and motion of human faces with a deformable model. However, the major limitation of such methods is that only overconstrained global shape descriptions of nonrigid motion are recovered and small details may be ignored or smoothed out. With the absence of local measurements, it would be very hard to obtain dense and accurate structure information and 3D correspondences, by only using global shape models. Thus, hierarchical methods which are capable of capturing both local patch information and global shape descriptions are very desirable. Authors

in [2] combined local and global analysis and estimated nonrigid motion and 3D structure of hurricanes (with subpixel accuracy) from a sequence of 2D satellite images. Local analysis within every small area was first performed to get rough initial results, then global smoothing was applied to obtain accurate hurricane top heights and motion. However, due to the lack of the shape model for describing hurricane's fluid motion, their method suffered from the overconstrained smoothness assumption and had larger errors in the areas of hurricane eye and hurricane edge.

In this paper, a novel hierarchical method which integrates local patch analysis and global shape descriptions is devised to solve the dual problem of structure and nonrigid motion recovery by using an elastic geometric model—extended superquadrics (ESQ). The nonrigid object of interest is segmented into many small areas and local analysis is performed to recover small details for each small area, assuming that each small area is undergoing similar nonrigid motion. Then, a recursive algorithm is proposed to guide and regularize local analysis with global information by using an appropriate global ESQ model. This local-global hierarchy enables us to capture both local patch and global shape deformations accurately and robustly. The rest of the paper is organized as follows: Section 2 formulates the local patch analysis and Section 3 introduces a set of global shape constraints based on extended superquadric (ESQ) models. In Section 4, local patch analysis is integrated with global shape constraints by a recursive algorithm. Experimental results on both simulation and real image sequences are demonstrated to validate and evaluate the effectiveness and robustness of the proposed approach in Section 5. Finally, conclusions and future work are presented in Section 6.

2 Tracking Local Patches

Due to the extremely varying nature of nonrigid motion, different kinds of motion such as bending, expansion, and contraction, can be found locally within one nonrigid object. For instance, it was found that in facial motion, the motion of cheeks, lips, and so on, correspond to different kinds of nonrigid motion [5]. Consequently, local patch tracking is very necessary in order to capture small details of objects' deformations by some local measurement techniques.

Ideally, one would like to segment the nonrigid objects appropriately and track consistent nonrigid motion locally within each segmented region. However, the segmentation itself is an extremely hard problem. In this paper, we segment the whole nonrigid object evenly into several small areas. For each small area, if the area is small enough, say 3×3 pixels, we can assume that this small region is undergoing nonrigid motion according to the same nonrigid motion model.

2.1 Local Nonrigid Motion Model

A key component of the local patch analysis is defining a good nonrigid motion model for each small area. Different motion models such as *isometric motion*, *homothetic motion*, *conformal motion* etc., have been employed in the past. In this paper, since there is no any prior knowledge about the object's motion behaviors, affine motion model is chosen because affine motion model is a general nonrigid motion model and has more power in describing nonrigid motion.

With the affine motion model, we have

$$\mathbf{P}_l{}^{i+1} = \mathbf{M}^i \mathbf{P}_l{}^i, \tag{1}$$

where $\mathbf{P}_l{}^i$ is the point l in the frame i,

$$\mathbf{P}_l{}^i = \begin{pmatrix} x_l{}^i & y_l{}^i & z_l{}^i & 1 \end{pmatrix}^T, \tag{2}$$

and $\mathbf{M}^i$ is the affine transformation matrix,

$$\mathbf{M}^i = \begin{pmatrix} a_1{}^i & b_1{}^i & c_1{}^i & d_1{}^i \\ a_2{}^i & b_2{}^i & c_2{}^i & d_2{}^i \\ a_3{}^i & b_3{}^i & c_3{}^i & d_2{}^i \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Since the dual problem of structure and nonrigid motion recovery is a very ill-posed problem and it is under-constrained in terms of unknowns that needed to be estimated, more constraint equations are necessary to make the algorithm robust. In this paper, we assume that the nonrigid motion within each small area is consistent not only spatially but also temporally. Thus, we have

$$\mathbf{M}^i = \delta^i \begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4}$$

where δ^i is a scaling factor between successive frames in order to compensate for possible deviations of temporal scale.

Eq. 2 describes objects' motion behaviors in the 3D space. When $\mathbf{P}_l{}^i$ is projected to the 2D image coordinate under perspective projection, the following equations can be easily derived:

$$\begin{aligned} X_l{}^{i+1} &= \frac{a_1 X_l{}^i + b_1 Y_l{}^i + c_1 + d_1 Z_l{}^i}{a_3 X_l{}^i + b_3 Y_l{}^i + c_3 + d_3 Z_l{}^i}, \\ Y_l{}^{i+1} &= \frac{a_2 X_l{}^i + b_2 Y_l{}^i + c_2 + d_2 Z_l{}^i}{a_3 X_l{}^i + b_3 Y_l{}^i + c_3 + d_3 Z_l{}^i}, \\ Z_l{}^{i+1} &= \frac{\delta^i Z_l{}^i}{a_3 X_l{}^i + b_3 Y_l{}^i + c_3 + d_3 Z_l{}^i}, \end{aligned} \tag{5}$$

where $({X_l}^i, {Y_l}^i)$ is in the image coordinate and

$$ {Z_l}^i = \frac{k}{{z_l}^i}. \tag{6} $$

Eq. 5 represents the constraint equation for tracking a point across a sequence of images using affine motion model.

2.2 Local Motion Model Fitting

Given the affine motion formulation for each small area, the remaining problem for local patch tracking is to fit all the unknowns in Eq. 5 by using 2D optical flow measurements, $\mathbf{V}_o({u_l}^i, {v_l}^i)$, as input data. In this paper, Levenberg-Marquardt non-linear optimization method is utilized to accomplish this task.

With optical flow vectors $\mathbf{V}_o({u_l}^i, {v_l}^i)$ as input data, a least-square EOF function can be defined by the sum of the square distance between the optic flow vectors and the 2D perspective projection of the affine flow vectors obtained by Eq. 5:

$$ {\chi^2}_1 = \sum_{i=1}^{m \text{ frames}} \sum_{l=1}^{n \text{ points}} \|\mathbf{V}_o - \mathbf{V}_a\|^2, \tag{7} $$

where

$$ \mathbf{V}_a = \begin{pmatrix} {X_l}^{i+1} - {X_l}^i \\ {Y_l}^{i+1} - {Y_l}^i \end{pmatrix}. \tag{8} $$

By minimizing Eq. 7, the object's nonrigid deformation within each local patch can be initially estimated, given the observed optical flow vectors.

In this paper, we will only analyze the case where the object's range data at time $t = 0$ is known. Then, the depth unknowns in the first frame can be eliminated by fixing them with the available range data. Actually, this is a very important step to ensure our algorithm's robustness and convergence. When performing 3D nonrigid motion tracking based on only 2D information with no prior knowledge of the nonrigid objects, very limited depth information or constraints are available. Without a good initial depth at the first frame, it is almost impossible to get a robust and meaningful tracking. On the other hand, in practical monocular vision systems, the determination of initial object 3D information can be achieved by some specific system initialization routines. For example, during the system initialization, the object of interest is asked to rotate rigidly or the system camera rotates around the object in order to recover the object's initial 3D information. Such techniques have been successfully utilized in some commercial vision systems and research has been done on this particular issue of structure from controlled motion estimation [4, 6]. Since we are concerned with the estimation of nonrigid motion and structure from monocular images. In this paper, we do not deal with the determination of initial object 3D information. Rather, we use the initial depth assumption for our particular purpose.

3 Global Shape Constraints

With local patch analysis, our algorithm is capable of capturing small details of the object's deformation. However, local measurements are very noise-sensitive and there may be heavy over-fitting involved in the local motion model optimization. It is clear that it will not be enough to robustly and precisely track structure and nonrigid motion only by local patch analysis. Global constraints are necessary in order to limit possible nonrigid behaviors and regularize the locally tracked nonrigid motion and structure.

In the literature of computer vision, different kinds of shape models have been proposed as global descriptions of nonrigid motion [10, 19, 15, 13, 14, 8]. With global shape models, specific global constraints can be defined to guide the local nonrigid motion analysis by rejecting bad hypotheses (point with "wrong" structure or "wrong" motion). A well-suited global shape model will be very beneficial for the problem of structure and nonrigid motion analysis. In this section, we first shortly review the existing nonrigid shape models. Then, a set of global shape constraints is proposed based on a novel geometric model, extended superquadrics (ESQ), for our purpose of structure and nonrigid motion recovery.

3.1 Nonrigid Shape Models

Most of the existing shape models can be classified into three major families: physics-based models, stochastic models, and geometric models. These models govern themselves with particular types of constraints and demonstrate both flexibility and regularity. Physics-based models exploit the physical properties of objects to describe their nonrigid behaviors. Deformations are the results of applying external and/or internal forces to these models. Due to their underlying consistency with real-world motion dynamics, promising results can be obtained by using these models. However, the measurement of the physical properties is a very hard problem and physical dynamics usually involves heavy computational expense. For stochastic models, correlation between motions of different parts is extracted to reduce the dimension of searching space. But these models usually require extensive training to acquire accurate knowledge of the motion. Geometric models model both global and local deformations by adjusting a set of control points or control parameters. Nonrigid motion tracking can be

accomplished by some simple and efficient representations. However, geometric models deform only according to their own geometric properties and realistic deformations can not be modeled precisely. In this paper, since our purpose is to have a global description of nonrigid motion in order to regularize the locally recovered details, it is not necessary or important to have the global shape model capture all the motion details. In this sense, we prefer geometric models due to their simplicity and efficiency.

Authors in [1] proposed a novel geometric shape model called extended superquadrics (ESQ). ESQ representations can model non-symmetrical shapes because they extend the exponents of superquadrics to functions of latitude angle and longitude angle in the spherical coordinate system. They also maintain many desired properties of superquadrics such as compactness, controllability, and intuitive meaning, which are all advantageous for shape modeling, reconstruction, and motion analysis. In this paper, a set of global shape constraints are defined based on the extended superquadric model for the dual problem of structure and nonrigid motion analysis.

3.2 ESQ Based Global Constraints

Extended superquadrics are an extension of superquadric models. They are defined by a set of points $\mathbf{x}(x, y, z)$ satisfying the equations as following.

$$\mathbf{x} = \begin{pmatrix} a_1 sign(cos\theta_s cos\phi_s) \|cos\theta_s\|^{f_2(\theta_p)} \|cos\phi_p\|^{f_1(\phi_p)} \\ a_2 sign(sin\theta_s cos\phi_s) \|sin\theta_s\|^{f_2(\theta_p)} \|cos\phi_s\|^{f_1(\phi_p)} \\ a_3 sign(sin\phi_s) \|sin\phi_s\|^{f_1(\phi_p)} \end{pmatrix}, \tag{9}$$

where θ_p , ϕ_p represent the latitude and longitude angles respectively in the spherical coordinate system, and θ_s, ϕ_s represent the superquadric angles. The exponents, $f_1(\phi_p), f_2(\theta_p)$ are Bézier functions of ϕ_p and θ_p:

$$f_1(\phi_p) = \sum_{i=0}^{n} P_{1i} B_i^n \left(\frac{\phi_p + \frac{\pi}{2}}{\pi} \right),$$
$$f_2(\theta_p) = \sum_{i=0}^{n} P_{2i} B_i^n \left(\frac{\theta_p + \pi}{2\pi} \right), \tag{10}$$

where P_{1i} and P_{2i} are the control points, and the B_i^n are the Bernstein polynomials of degree n,

$$B_i^n(t) = \frac{n!}{(n-i)!i!} t^i (1-t)^{n-i}, \quad i = 0, 1, ..., n. \tag{11}$$

Whereas the exponents of superquadrics are constants, the extended superquadrics have exponents changing according to θ_p and ϕ_p. Since the exponents, $f_1(\phi_p)$ and $f_2(\theta_p)$, are two relative shape parameters ($f_1(\phi_p)$ is the squareness parameter along the z axis and $f_2(\theta_p)$ is the squareness parameter in the x-y plane), local deformations can be easily obtained by adjusting the corresponding control points in both exponent functions. Furthermore, it is even not necessary for the control points to be evenly distributed in the exponent functions. Instead of using one Bézier curve for each exponent function, several Bézier curve segments with common endpoints can be utilized. More control points can be defined for the area with a lot of local details, while less control points can be defined for other areas in an adaptive way. This property is extremely important for modeling complex nonrigid objects with economical number of parameters. Fig. 1 shows a sample of shapes that can be achieved by deforming an initial spherical shape randomly using its 19 control points. As can be seen, a wide range of nonrigid motions and their resulting shapes can be produced with relatively few control points.

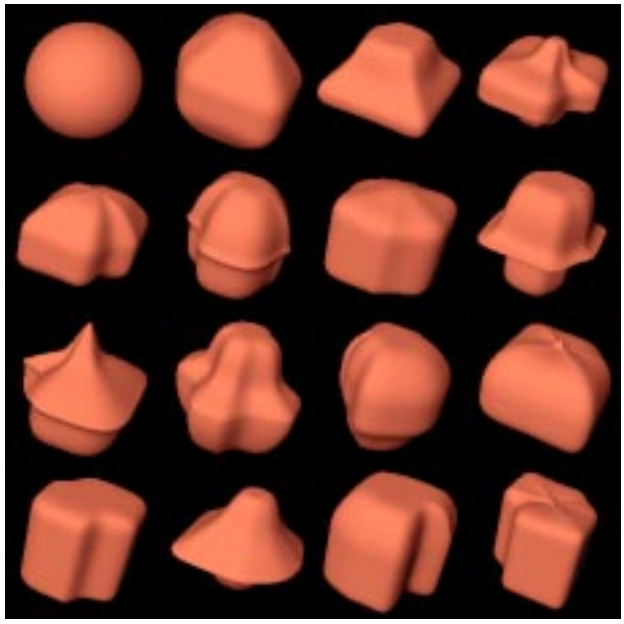


Figure 1: A sample of shapes that can achieved by deforming an initial spherical shape randomly using its 19 control points.

Based on ESQ global shape models, a set of global shape constraints can be defined to further guide the local patch tracking process. In this paper, a model force is simply defined as

$$EM_l{}^i = \left\| 1 - F^i(\mathbf{P}_l{}^i) \right\|^2, \tag{12}$$

where F is the inside-outside function of ESQ models, as shown in Eq. 13.

Clearly, with Eq. 12, over-fitted or local noise re-

$$F(\mathbf{x}) = \left[\left(\frac{x}{a_1}\right)^{\frac{2}{f_2\left(tan^{-1}\left(\frac{y}{x}\right)\right)}} + \left(\frac{y}{a_2}\right)^{\frac{2}{f_2\left(tan^{-1}\left(\frac{y}{x}\right)\right)}} \right]^{\frac{f_2\left(tan^{-1}\left(\frac{y}{x}\right)\right)}{f_1\left(tan^{-1}\left(\frac{z}{\sqrt{x^2+y^2}}\right)\right)}} + \left(\frac{z}{a_3}\right)^{\frac{2}{f_1\left(tan^{-1}\left(\frac{z}{\sqrt{x^2+y^2}}\right)\right)}}. \quad (13)$$

sults will be rejected since they may be "far" away from those predicted by the global shape models. In the next section, we will discuss the methodology to integrate the global shape constrains into the local patch analysis.

4 Integrating Local Patch Analysis with Global Shape Constraints

Having defined a model force by using the global ESQ models, the remaining problem is to integrate local patch tracking with global shape constraints. However, the ESQ shape models for every frame are unknown initially. An ESQ shape model should be recovered for every frame in the first stage.

4.1 ESQ Shapes Fitting with An Elastic Force

Since a set of initial results will be available after local patch tracking, they can be utilized to fit the global ESQ models for every frame. Authors in [1] have discussed the methodology for ESQ model fitting. An error-of-fit function was first defined using the inside-outside function of ESQ. Then, the fitting of ESQ parameters was accomplished by a hierarchical nonlinear optimization. However, since the initial locally tracked results usually consist a lot of noise in z field due to the depth ambiguity, the fitting techniques discussed in [1] will be insufficient. More constraints are necessary in order to alleviate the influence of the noise data. In this paper, an elastic force is defined according to the small elastic motion assumption.

As we know, for elastic objects, the neighborhood relationship between points will never change during motion, i.e. local elements can only deform smoothly and continuously. Thus, there exists no abrupt change in the latitude and longitude angles for all the points on elastic objects during small elastic motion. For instance, when elastic objects uniformly expand or contract, the latitude and longitude angles of all the points will not change at all. Based on this observation, an elastic force can be defined as

$$EE_l{}^i = \left\|\theta_l{}^{i+1} - \theta_l{}^i\right\|^2 + \left\|\phi_l{}^{i+1} - \phi_l{}^i\right\|^2, \quad (14)$$

where $\theta_l{}^i$, $\phi_l{}^i$ represents the latitude and longitude angles for point l in the ith frame. With the elastic force, a new EOF function for the ESQ fitting (in frame i) can be easily defined by adding $EE_l{}^i$ to the

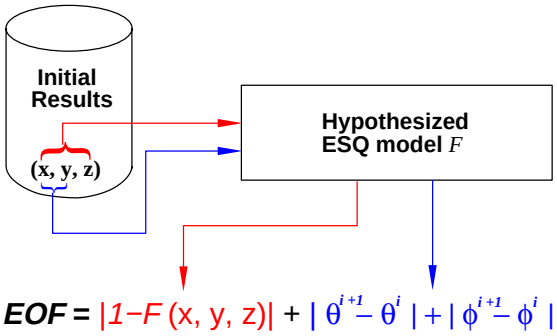


Figure 2: Fitting global ESQ with elastic force for frame i.

EOF function defined in [1],

$$\chi^2{}_{esq} = \sum_{j=1}^{N\ \text{data}} \left\|1 - F(\mathbf{x}_j)^{f_1}\right\|^2 + EE_l{}^i. \quad (15)$$

The advantage of Eq. 15 is that the latitude and longitude angles can be computed only from the x, y information and the hypothesized ESQ model, which means that the new defined EOF function puts more weights on the locally recovered x, y information than the z information. Fig. 2 illustrates the strategy. In this paper, we experimentally found that Eq. 15 works very well for our particular purpose. Since only 2D optical flow measurements are utilized as input data, the x, y information in the initial results are more meaningful and useful than the initially obtained z information. It would be, of course, unfair to treat them equally. Eq. 15 actually provides a reasonable way to capture the difference.

4.2 Adaptive Weights

Once the global ESQ models are fitted for every frame, the model force defined in Eq. 12 can be utilized to guide and regularize the local nonrigid motion tracking. However, simply including the model force into Eq. 7 with a constant weight will be biased because the fitted global ESQ models cannot describe every detail of the nonrigid object perfectly. At those places where the global shapes cannot provide a good estimation, the model force will be less meaningful. Hence, the model force must have an adaptive weight which should be selected according to the exactness of the global shape models. In this paper, the residual errors after the ESQ model fitting are utilized to construct such an adaptive weight because they actually

reflect the difference between the fitted ESQ shape models and the initial results. Although the residual errors may not provide a very accurate assessment, they are very simple and found to work very well in our particular situation.

With the residual errors, the global shape constraints can be integrated into local patch analysis by modifying Eq. 7 as

$$\chi^2{}_2 = \sum_{i=1}^{m \text{ frames}} \sum_{l=1}^{n \text{ points}} \|\mathbf{V}_o - \mathbf{V}_a\| + {w_l}^i E {M_l}^i, \quad (16)$$

where

$${w_l}^i = \frac{d}{{e_l}^i}, \quad (17)$$

and ${e_l}^i$ is the residual error for point l in frame i when minimizing Eq. 15. It should be noted that the adaptive weight scheme essentially relaxes the global shape constraints at places where the objects have small details, thus giving the local patch analysis more flexibility.

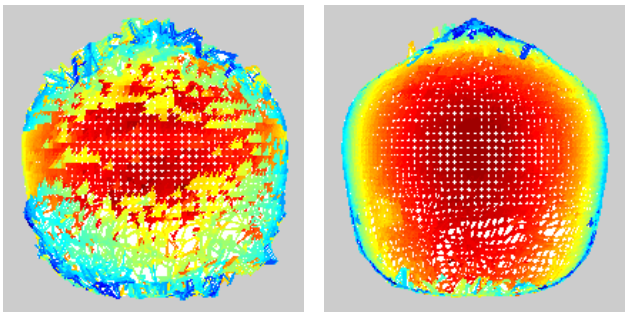
4.3 A Recursive Algorithm

In the previous scheme, we presented a scheme to guide local patch tracking with global shape constraints. Since the optimal results often cannot be achieved in just one step, we propose a recursive algorithm which estimates global ESQ parameters and performs local tracking iteratively until the recovered structure and nonrigid motion converge to a stable solution. A single step of the recursive algorithm includes local patch tracking according to the current global ESQ shapes, updating the previous motion and structure results, and fitting a new set of ESQ shape models to the updated results. Based on the difference between the current results and the previous results, it is very easy to decide whether the algorithm converges to a stable solution. The complete algorithm is shown in Algorithm 1.

Intuitively, the recursive algorithm retrieves useful information (global ESQ shapes) under some proper constraints from current results and then uses the information as global constraints for the next iteration's optimization. A solid theoretical study of the convergence properties of the recursive algorithm is yet to be made, but the experimental results in Section 6 show the desired convergence in both the simulation and real-world applications. Fig. 3 presents the initial results (without global shape constraints) and the results after nine iterations for frame 10 in the simulation experiments. It is clear that the initial recovered structure and motion have a lot of noise and change dramatically from one small area to the other while

Algorithm 1: A Recursive Algorithm for Hierarchical Structure and Nonrigid Motion Analysis

begin
  for $i := 1$ **to** n Areas **step** 1 **do**
    minimize Eq. 7 and do local patch tracking to get initial results for small area i;
  choose arbitrary initial **s** of large magnitude;
  while $\|\mathbf{s}\|$ is greater than some threshold **do**
    fit ESQ parameters for every frame from the current results;
    for $i := 1$ **to** n Areas **step** 1 **do**
      minimize Eq. 16 (with global shape constraints) to get a new set of results for small area i
    calculate the difference **s** between the current results and the previous results;
end



(a) (b)

Figure 3: Comparison of the results without and with global shape constraints: (a) locally tracked initial results; (b) results after nine iterations with global shape constraints.

the results after incorporating global constraints look very smooth and reasonably correct.

5 Experimental Results

Experiments on both simulation and real image sequences have been performed. The simulation image sequences are generated by using Open Inventor in SGI workstation. The main purpose of the simulation experiments is to quantitatively evaluate the effectiveness and accuracy of our algorithm since the ground truth for the simulation data is available. Fig. 4(a) shows a simulation image sequence in which a sphere deforms nonrigidly. In order to test the capability of our algorithm in tracking small local details, some small artificial pits and bumps are added to the sphere surface. Nine iterations were performed with our recursive algorithm. The recovered dense structure and

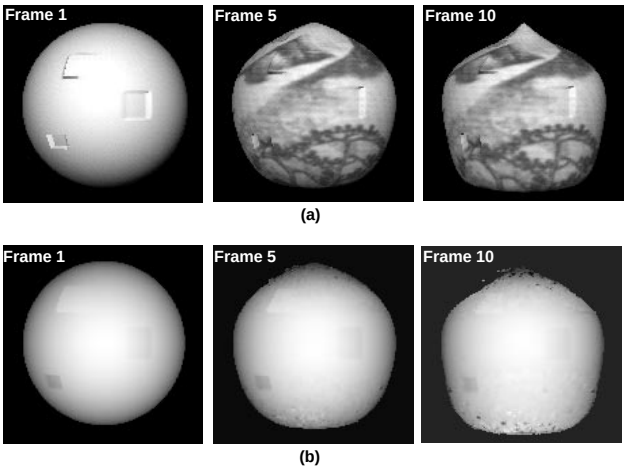


Figure 4: (a): A simulation image sequence generated by Open Inventor (Note that the texture of the first frame is removed in order to show the artificial pits and bumps clearly); (b) The estimated dense structure for the sphere.

nonrigid motion (2D projection in the $x-y$ plane) are shown in Fig. 4(b) and Fig. 5 respectively. As can be seen, the structure and motion of the nonrigid sphere was successfully tracked. Fig. 6 shows the mean errors of our results when compared with the ground truth at every frame. Our results are very encouraging, with the mean errors of around 0.57 pixel for the motion and 0.02–0.1 pixel for the structure. Also, it can be noted that although the global ESQ models cannot model the sphere well at the small pits and bounds, local analysis enable us to capture these small details and track them reliably.

We now present experiments on real image sequences. The real image sequences were captured in a poorly illuminated environment because we want to test our algorithm with a noisy optical flow as input data. Fig. 7(a) shows a real sequence, in which the subject made an expression and large motion can be seen at the lips and the chin. Fig. 7(b) shows the estimated dense structure of the face and Fig. 7(c) shows the rendered images of the recovered 3D faces. Unfortunately, a quantitative validation cannot be performed for the real image sequence since the ground truth is not available. However, verification can be performed via visual inspection and qualitative evaluation. When manually comparing the facial motion presented in the input 2D images and the recovered 3D motion correspondences, it is found that the facial motion were tracked quite correctly and the estimated 3D face structure is similar to the face in the original imagery. Particularly, the motion and structure at the lips and the chin are also tracked successfully despite the presence of the large motion. To a substantial extent, this accuracy is attributable to the local-global hierarchy presented in our algorithm. The global shape constraints enable us to refine and regularize the nonrigid motion and structure even with poor quality input data while local patch tracking captures all the small details of the nonrigid objects.

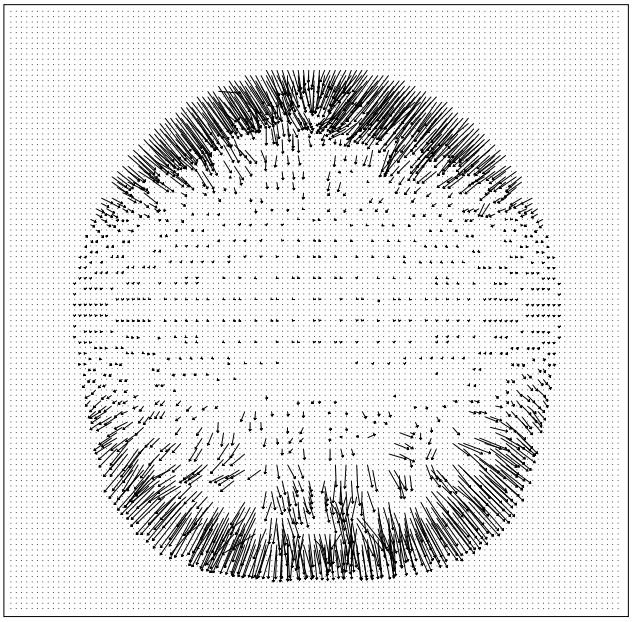


Figure 5: 2D projection of the recovered nonrigid motion for the simulation sequence.

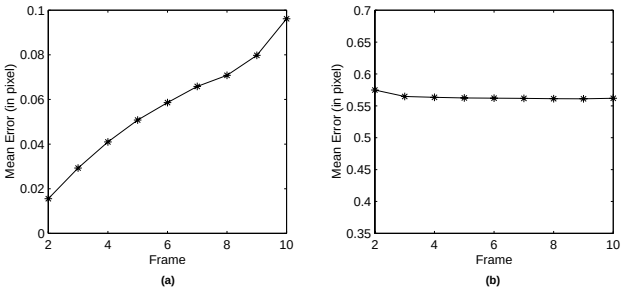


Figure 6: (a): Mean errors of the estimated structure compared with the ground truth at every frame; (b): Mean errors of the estimated motion compared with the ground truth at every frame.

For more results and animations, please refer to our video demo (410-1.mpg). This demo includes six animations. The first part is a simulation sequence, followed by the estimated 3D nonrigid motion and dense structure, presented in the second and the third part. Then, we show the real experimental results. The fourth animation presents a real sequence. The

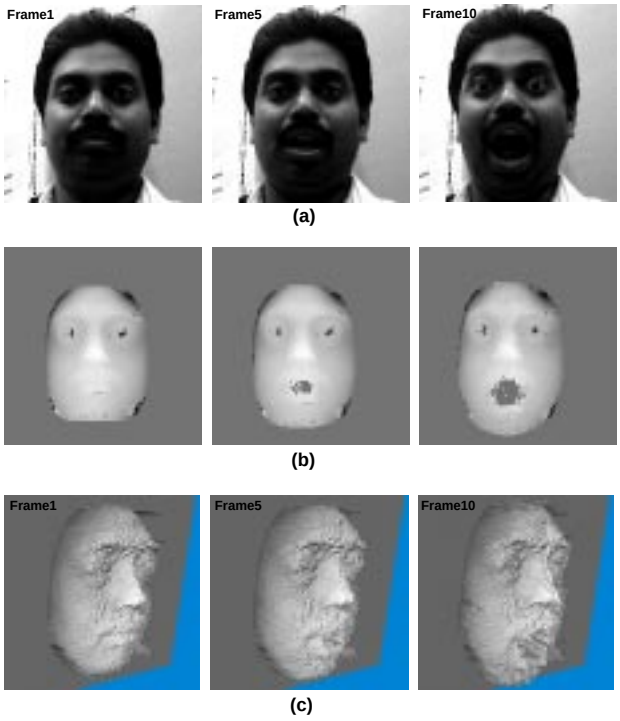


Figure 7: (a): A real image sequence; (b): Estimated dense structure of the face; (c): Rendered images of the recovered 3D faces.

fifth one shows the recovered 3D motion while the last one shows the estimated dense structure.

6 Conclusions and Future Work

This paper motivates and implements a new hierarchical method for the problem of structure and nonrigid motion recovery by using 2D monocular images. This method has shown promise in capturing both global and local nonrigid deformations. Dense structure and 3D correspondences are estimated to subpixel accuracy. Experiments on simulation and real image sequences confirm the effectiveness and robustness of the proposed algorithm. Our future work will focus on parallel implementations in order to construct a real-time structure and nonrigid motion tracking system.

Acknowledgments

Research funding was provided by the National Science Foundation.

References

[1] Extending superquadrics with exponent functions: Modeling and reconstruction. *IEEE Conference on Computer Vision and Pattern Recognition*, pages II–73–78, 1999.

[2] Extracting nonrigid motion and 3D structure of hurricanes from satellite image sequences without correspondences. In *CVPR99*, pages II:280–285, 1999.

[3] M.J. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *IJCV*, 25(1):23–48, October 1997.

[4] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin. Structure from controlled motion. *PAMI*, 18(5):492–504, May 1996.

[5] Kai Chen, Chandra Kambhamettu, and Dmitry Goldgof. Extraction of mpeg-4 fap parameters from 3d face data sequences. *International Workshop on Very Low Bitrate Video Coding*, October 8-9, 1998.

[6] A.K. Dalmia and M. Trivedi. High-speed extraction of 3d structure of selectable quality using a translating camera. *CVIU*, 64(1):97–110, July 1996.

[7] D. DeCarlo and D. Metaxas. Deformable model-based shape and motion analysis from images using motion residual error. In *ICCV98*, pages 113–119, 1998.

[8] W. C. Huang and D. B. Goldgof. Adaptive-size meshes for rigid and nonrigid shape analysis and synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):611–616, June, 1993.

[9] I.A. Kakadiaris and D. Metaxas. Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection. In *CVPR96*, pages 81–87, 1996.

[10] Chandra Kambhamettu, Dmitry B. Goldgof, Demetri Terzopoulos, and Thomas S. Huang. Nonrigid motion analysis. In Tzay Young, editor, *Handbook of PRIP: Computer vision*, volume II, pages 405–430. Academic Press, San Diego, California, 1994.

[11] R. Koch. Dynamic 3-d scene analysis through synthesis feedback control. *PAMI*, 15(6):556–568, June 1993.

[12] H. Li, P. Roivainen, and R. Forchheimer. 3-d motion estimation in model-based facial image coding. *PAMI*, 15(6):545–555, June 1993.

[13] T. McInerney and D. Terzopoulos. A finite element model for 3d shape reconstruction and nonrigid motion tracking. In *ICCV93*, pages 518–523, 1993.

[14] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *PAMI*, 15(6):580–591, June 1993.

[15] A.P. Pentland and B. Horowitz. Recovery of nonridid motion and structure. *PAMI*, 13(7):730–742, July 1991.

[16] M. Subbarao. Interpretation of image flow: A spatio-temporal approach. *PAMI*, 11(3):266–278, March 1989.

[17] H. Tao and T.S. Huang. Explanation-based facial motion tracking using a piecewised bezier volume deformation model. In *CVPR99*, pages I:611–617, 1999.

[18] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.

[19] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3D shape and nonrigid motion. *Artificial Intelligence*, 36(1):91–123, 1988.

[20] P. Werkhoveh, A. Toet, and J.J. Koenderink. Displacement estimates through adaptive affinities. *PAMI*, 12(7):658–663, July 1990.