

Connected Vibrations Method for Non-rigid Motion Tracking

Hai Tao , Thomas S. Huang

Beckman Institute
University of Illinois, Urbana, IL 61801, USA
E-mail: {tao, huang}@ifp.uiuc.edu

Abstract

This paper introduces a new algorithm for tracking non-rigid motion in image sequences. Non-rigid motions are modeled as the deformations of connected 2D membrane patches. Local smoothness constraints for each patch are the low-frequency vibration modes obtained from modal analysis. Hinges are imposed between patches to maintain the global topological structure. The resulted over-constrained linear system is solved efficiently using a least square estimator. Promising results on both natural and synthetic facial motion sequences are demonstrated in this paper.

1. Introduction

Non-rigid motion modeling and analysis in image sequences is an important research topic with many applications in computer graphics, biomedical image analysis and human computer interaction [2]. Three major families of non-rigid motion models are geometric models [5], stochastic models [1], and physics-based models [3]. Each type of models govern themselves with certain type of constraints and demonstrate flexibility as well as regularity.

In this paper, the proposed method is inspired by research works on modal analysis approach in which the deformations of an object fall into a subspace spanned by its low-frequency vibration modes [3]. In our algorithm, the model consists of multiple deformable 2D patches. The simple movements of each patch are determined by a set of constraints corresponding to different vibration modes. These patches are connected with each other through hinges with desired strength. As a result, complicated non-rigid motion is decomposed into simpler local deformations and the global structural information is embedded in the connections.

This work was supported in part by Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0003, and in part by Joint Services Electronics Program Grant ONR N00014-96-1-0129.

2. Approach

The system block diagram is shown in Figure 1. First, interested region is decomposed by user into patches. Each patch is modeled as a polygonal mesh and its deformation resembles the behavior of an elastic membrane. For each patch, its vibration modes are derived in step 2 using finite element method based on its physical properties. In the tracking process, each patch only changes its shape in principal modes to maintain its spatial smoothness and reduce computational cost. For each patch, the best tracking result is the deformation that achieves maximum image intensity correlation under this local constraint. To maintain the relative positions of patches in the tracking process, a global linear system combining correlation function of individual patch and penalty for patch drifting is proposed. For each time instance, the deformation is represented by the magnitudes of principal vibration modes. For smooth motions, these magnitudes show strong temporal continuity. This property is exploited by Kalman filtering in vibration magnitude parameter space. It provides both estimation and prediction of deformation parameters. Details of these modules are described in the next section.

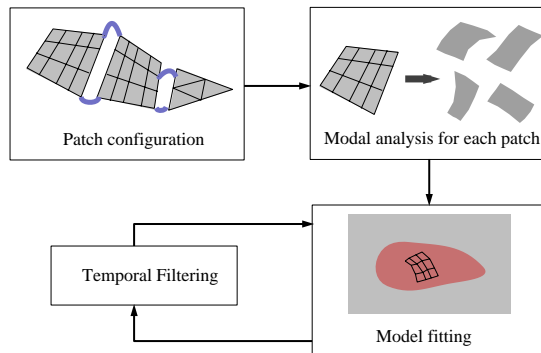


Figure 1: Components of connected vibration tracking system.

3. Connected vibrations method

3.1. Membrane patch and modal analysis

Standard finite element method [4] is used to derive patch vibration modes. The in-plate deformation of this patch is described by nodal displacement vector $\mathbf{u} = [u_{1x}, u_{1y}, \dots, u_{nx}, u_{ny}]^t$, where u_{ix} , u_{iy} are x and y components of displacement vector for i th node. The dynamics of this patch is governed by Lagrange's equation

$$[\mathbf{M}]\{\ddot{\mathbf{u}}\} + [\mathbf{C}]\{\dot{\mathbf{u}}\} + [\mathbf{K}]\{\mathbf{u}\} = \{\mathbf{Q}\}, \quad (1)$$

where \mathbf{M} is the inertia (mass) matrix, \mathbf{K} is the stiffness matrix, and \mathbf{C} is the damping matrix. In our case, it is assumed that $\mathbf{C} = 0$, $\mathbf{Q} = 0$. Each vibration mode Φ_i satisfies the following condition

$$[\mathbf{K} - \omega_i^2 \mathbf{M}]\{\Phi\}_i = 0, \quad (2)$$

where ω_i is the vibration frequency of that mode. It is trivial to see that each Φ_i is a eigen-vector of matrix $\mathbf{M}^{-1}\mathbf{K}$. Normalized form of Φ_i is used through this paper.

For a patch with n nodes, $2n$ vibration modes exist. We only use those of low frequencies, i.e., those with small eigen-values. The first 3 modes are rigid motions of the patch; two for translation and one for rotation. The rest modes are in-plate free vibrations. To confine the deformation of a patch in a displacement space spanned by m mode vectors $\Phi_1, \Phi_2, \dots, \Phi_m$, a deformation \mathbf{u} is written as $\mathbf{u} = \sum_{i=1}^m f_i \Phi_i$. We define $\mathbf{f}_i = [f_1, f_2, \dots, f_m]^t$ as the deformation parameters for i th patch. If there are totally p patches, deformation parameters for all these patches are denoted as $\mathbf{F} = [\mathbf{f}_1^t, \mathbf{f}_2^t, \dots, \mathbf{f}_p^t]^t$.

The derivation of \mathbf{M} and \mathbf{K} for a 2D membrane patch is given in [4]. Examples of some low-frequency vibration modes are shown in Figure 2.

3.2. Model fitting

Deformation \mathbf{F} is derived through an over-determined linear system. In a membrane patch, the 2D motion vector of each node is first estimated by standard image template matching technique from two consecutive video frames. For i th patch, we denote the nodal displacement vector as $\hat{\mathbf{V}}_i = [\hat{v}_{1x}, \hat{v}_{1y}, \hat{v}_{2x}, \hat{v}_{2y}, \dots, \hat{v}_{n_x}, \hat{v}_{n_y}]^t$ and the vibration modes as $\mathbf{U}_i = [\Phi_1, \Phi_2, \dots, \Phi_{m_{j_1}}]$. Then a least square estimation problem for this patch is formulated as finding the deformation \mathbf{f}_i that minimizes $\|\mathbf{U}_i \mathbf{f}_i - \hat{\mathbf{V}}_i\|$. The solution is $\mathbf{f}_i = (\mathbf{U}_i^t \mathbf{U}_i)^{-1} \mathbf{U}_i^t \hat{\mathbf{V}}_i$.

To maintain the global structure, hinges are introduced between patches. If there is a hinge between n_1 th node of patch j_1 and n_2 th node of patch j_2 . The

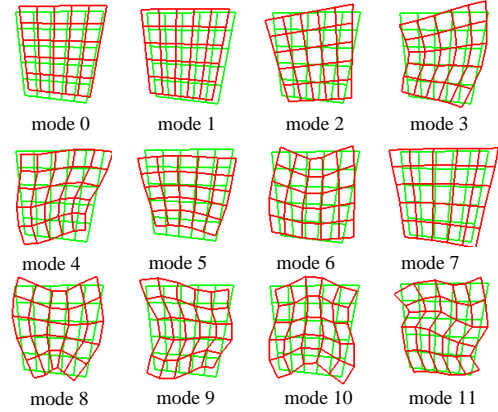


Figure 2: The first 12 vibration modes of a patch. Mode 1, 2, and 3 correspond to rigid motions.

distance between these two nodes should remain unchanged during tracking. To integrate this constraint into the system, we have

$$\begin{bmatrix} \mathbf{U}_{j_1} & 0 \\ 0 & \mathbf{U}_{j_2} \\ \lambda \mathbf{S}_{j_1} \mathbf{U}_{j_1} & -\lambda \mathbf{S}_{j_2} \mathbf{U}_{j_2} \end{bmatrix} \begin{bmatrix} \mathbf{f}_{j_1} \\ \mathbf{f}_{j_2} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{V}}_{j_1} \\ \hat{\mathbf{V}}_{j_2} \\ 0 \end{bmatrix}, \quad (3)$$

where λ represents the strength of the link and

$$\mathbf{S}_{j_k} = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & 0 & \dots \\ 0 & \dots & 0 & 0 & 1 & 0 & \dots \\ \dots & 2n_k - 1, & 2n_k & \dots \end{bmatrix}$$

is $2 \times m_{j_k}$ matrix for patch j_k that selects its n_k th node.

A larger LSE problem needs to be solved for the whole connected patch system. We write that system as $\mathbf{A}\mathbf{F} = \hat{\mathbf{V}}$. The LSE solution is $(\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \hat{\mathbf{V}}$. $(\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t$ needs to be computed only once.

4. Results

The proposed algorithm has been tested on video sequences of natural and synthetic facial motions. In a two-second synthetic sequence, a texture-mapped face model performs certain global motions including translations, in-depth and in-plane rotations. Non-rigid motions presented in the sequence are two expressions, smile and surprise, and various mouth/lip movements (Figure 3). Video sequences with real face motions are also tested. The results are evaluated subjectively (Figure 4).

At the initialization stage, fourteen quadrilateral patches are constructed based on the locations of manually picked facial feature points. Hinges are also imposed between these patches by users (Figure 3(a)). 8

to 15 of vibration modes are derived for each patch to describe the local deformation. The strength of inter-patch connections is adjusted by λ . In our system, $\lambda = 10$. Each connection reduces the degree of freedom by 2. In our implementation, there are totally 458 nodes, 153 vibration modes, and 55 connections. The input motion estimation is of dimension 916. The deformation is confined in a 43 dimensional sub-space if $\lambda = \infty$. To alleviate the error accumulation problem in template matching method, templates from both the previous frame and the initial frame are applied randomly with equal probabilities. Figure 3(d)-(f) and Figure 4(d)-(f) show the tracking results. For the synthetic sequence, the average tracking error for all nodes is 1.21 pixels. For real face motion sequences, the tracking results are good for sequence with in-plane motions. Small in-depth rotations can also be tolerated.

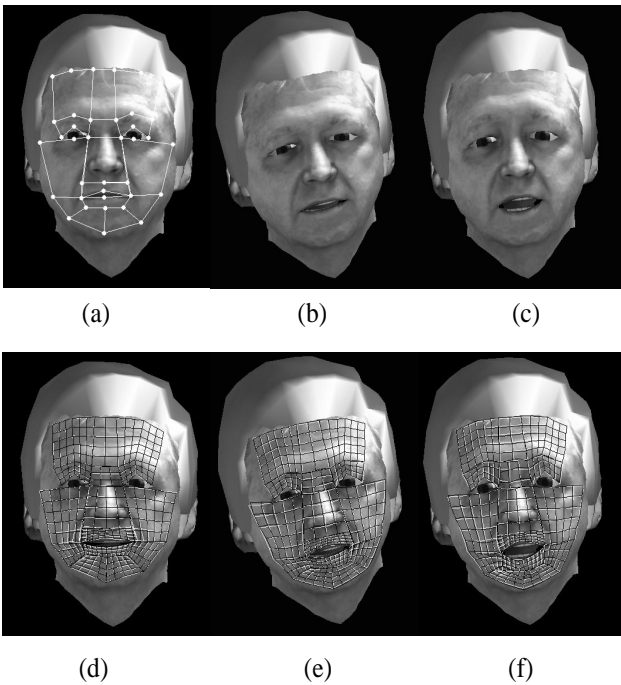


Figure 3: (a) The 14 membrane patches. (b)-(f) an image sequence with synthetic face/head motions (white mesh) and tracking results (black mesh). (b)(e) frame 30. (c)(f) frame 50. (d) frame 10.

5. Conclusions

The key idea of the proposed algorithm is to integrate simple local deformations into complicated system through connections. Instead of vibration modes, other local constraints such as free-form deformation could be used

as well if they are smooth and linear. Domain knowledge is embedded in connections. We may also call this framework connected linear deformation method.

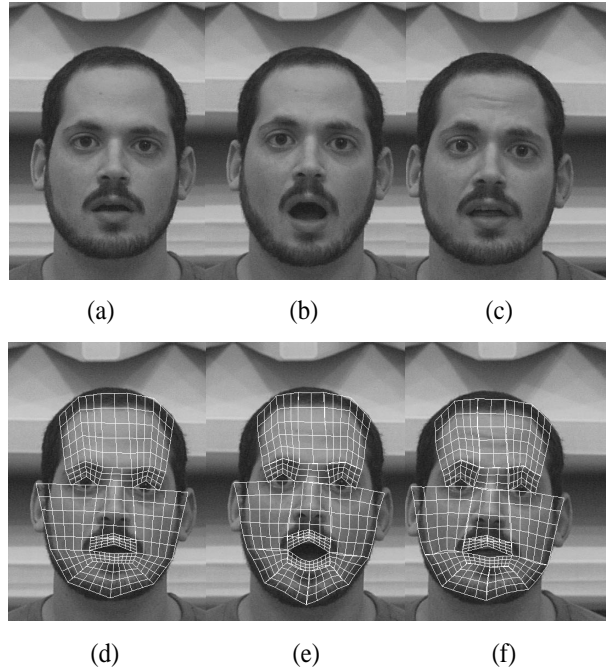


Figure 4: A real video sequence with face/head motions and its tracking results. (a)(d) frame 0. (b)(e) frame 30. (c)(f) frame 50.

6. References

- [1] F. Heitz C. Kervrann and P. Perez. Statistical model-based estimation and tracking of non-rigid motion. In *ICPR'96*, pages 244–248, 1996.
- [2] C. Kambhamettu, D. B. Goldgof, D. Terzopoulos, and T. S. Huang. Nonrigid motion analysis. In Tzay Young, editor, *Handbook of PRIP: Computer Vision*, volume 2. Academic Press, 1993.
- [3] Alex Pentland and Stan Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(7):715–729, July 1991.
- [4] M. Petyt. *Introduction to Finite Element Vibration Analysis*. Cambridge University Press, 1990.
- [5] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.