

Evaluation of Facial Reconstructive Surgery on Patients with Facial Palsy using Optical Strain

Matthew Shreve¹, Neeha Jain¹, Dmitry Goldgof¹, Sudeep Sarkar¹,
Walter Kropatsch², Chieh-Han John Tzou³, and Manfred Frey³

¹ University of South Florida, Department of Computer Science and Engineering,
Tampa Florida

[mshreve, neehajain, goldgof, ssarkar]@cse.usf.edu,

² Vienna University of Technology, Vienna Austria **

krw@prip.tuwien.ac.at

³ Medical University of Vienna, Division of Plastic and Reconstructive Surgery,
Vienna Austria

[chieh-han.tzou,manfred.frey]@meduniwien.ac.at

Abstract. We explore marker-less tracking methods for the purpose of evaluating the efficacy of facial re-constructive surgery on patients with facial palsies. After experimenting with several optical flow methods, we choose an approach that results in less than 2 pixels in tracking error for 15 markers tracked on the face. A novel method is presented that utilizes the non-rigid deformation observed on facial skin tissue to visualize the severity of facial paralysis. Results are given on a dataset that contains three videos of an individual recorded using a standard definition camera both before and after undergoing facial reconstructive surgery over a period of three years.

Keywords: Optical Flow; Optical Strain; Facial Palsy; Facial Reconstructive Surgery

1 Introduction

Accurately estimating and quantifying the extent of facial paralysis in patients with facial palsy without the need of manually applied markers would be a benefit to patients, researchers, and the medical community at large. In this paper, we propose methods that can be used to measure the severity of facial paralysis using non-invasive tracking methods and motion analysis tools.

The experimental flow is as follows: first, a patient is recorded in front of a video camera mirror system [2] and is asked to perform several standardized expressions multiple times (ex., lifting of eyebrows, smile, close eyes, frown, whistle) [4]. Next, a dense optical flow method is used that tracks all points (pixels) of the face over the entire length of the expressions. These optical flow vectors are then used to calculate optical strain, a feature that is used for two purposes:

** W.K. acknowledges the support of the Austrian Science Fund (FWF), grant no. P20134-N13.

(i) the magnitude of optical strain is utilized in order to detect key moments of an expression (contract, peak, compression) [6]. Finding these moments in an expression allows strain maps to be calculated at the maximal point of facial deformation, so a valid comparison can be done over time; (ii) strain maps are used to represent and quantize the deformation of the soft-skin tissue on the face, which is directly correlated with expansion and contraction of underlying facial muscles that have been surgically altered.

Evaluating the efficacy of facial reconstructive surgery has been the main goal of Frey *et al.* [2]. In their experimental setup, a patient is asked to sit between two angled mirrors ($\sim 90^\circ$). Hand placed markers are applied to the face and are tracked in 3-D as the patient performs expressions. In their setup, the process of applying markers and tracking them takes roughly five hours. In this paper, we use a video dataset from their collection and hope to expand on their initial work firstly by eliminating the need for markers, thus significantly reducing the time needed for data acquisition. Secondly, we suggest a method that provides a denser correspondence and a more detailed visual representation and quantization.

2 Background

When calculating optical strain there are typically two main approaches: either (i) integrate the strain definition into the optical flow equations, or (ii) derive strain directly from the flow vectors. The first approach requires the calculation of high order derivatives, hence is sensitive to image noise. The second approach allows us to post-process the flow vectors before calculating strain, possibly reducing the effects of any errors incurred during the optical flow estimation. We use the second approach in this paper.

2.1 Optical Flow

Optical flow is an established motion estimation technique that is based on the brightness conservation principle [1]. In general, it assumes that the intensity at a point remains constant over a pair of frames, and that the pixel displacement is relatively smooth within a small image region. It is typically represented by the following equation:

$$(\nabla I)^T \mathbf{p} + I_t = 0 \tag{1}$$

where $I(x, y, t)$ represents the temporal image intensity function at point x and y at time t , and ∇I represents the spatial and temporal gradient. The horizontal and vertical motion vectors are represented by $\mathbf{p} = [p = dx/dt, q = dy/dt]^T$ s.

Since large intervals over a single expression can often cause failure in tracking (due to the smoothness constraint), we implemented a vector linking (or stitching) process that combines small, local pairs of small intervals (1-3 frames) into larger pairs to expand over the entire sequence of frames. In section 3.1, we discuss three separate implementations of optical flow.

2.2 Optical Strain

The displacement of any deformable object projected on a 2-D plane can be expressed by a vector $\mathbf{u} = [u, v]^T$. Assuming a small enough motion, a finite strain tensor can be defined:

$$\varepsilon = \frac{1}{2}[\nabla\mathbf{u} + (\nabla\mathbf{u})^T], \quad (2)$$

which can be expanded to:

$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} = \frac{\partial u}{\partial x} & \varepsilon_{xy} = \frac{1}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) \\ \varepsilon_{yx} = \frac{1}{2}(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}) & \varepsilon_{yy} = \frac{\partial v}{\partial y} \end{bmatrix} \quad (3)$$

where $(\varepsilon_{xx}, \varepsilon_{yy})$ are normal strain components and $(\varepsilon_{xy}, \varepsilon_{yx})$ are shear strain components.

Since (u, v) are displacement vectors that over a continuous space, we approximate the strain components using the optical flow data (p, q) :

$$p = \frac{\delta x}{\delta t} \approx \frac{\Delta x}{\Delta t} = \frac{u}{\Delta t}, u = p\Delta t, \quad (4)$$

$$q = \frac{\delta y}{\delta t} \approx \frac{\Delta y}{\Delta t} = \frac{v}{\Delta t}, v = q\Delta t \quad (5)$$

where Δt is the change in time between two image frames. Setting Δt to a fixed interval length, we can estimate the partial derivatives of (4) and (5):

$$\frac{\partial u}{\partial x} = \frac{\partial p}{\partial x} \Delta t, \frac{\partial u}{\partial y} = \frac{\partial p}{\partial y} \Delta t, \quad (6)$$

$$\frac{\partial v}{\partial x} = \frac{\partial q}{\partial x} \Delta t, \frac{\partial v}{\partial y} = \frac{\partial q}{\partial y} \Delta t, \quad (7)$$

The second order derivatives are calculated using the central difference method. Hence,

$$\frac{\partial u}{\partial x} = \frac{u(x + \Delta x) - u(x - \Delta x)}{2\Delta x} \approx \frac{p(x + \Delta x) - p(x - \Delta x)}{2\Delta x} \quad (8)$$

$$\frac{\partial v}{\partial y} = \frac{v(y + \Delta y) - v(y - \Delta y)}{2\Delta y} \approx \frac{q(y + \Delta y) - q(y - \Delta y)}{2\Delta y} \quad (9)$$

where $(\Delta x, \Delta y) \approx 2-3$ pixels.

Finally, each of these values corresponding to low and large elastic moduli are summed to generate the strain magnitude. Each value can also be normalized to 0-255 for a visual representation (strain map).

3 Experiment

In this section, we explore several potential uses of optical flow and optical strain for the marker-less tracking and visualization of expressions for patients with facial palsies. Our dataset consists of three videos from the Medical University of Vienna. Each video corresponds to a different year of the patient undergoing facial reconstructive surgery. The first video records the patient before the surgery (1998), and the next two videos (1999 and 2000) were recorded post surgeries. For each video, there are roughly 30 expressions made. Expressions include raising the eyebrows, smiling, smiling and closing eyes, bunching lips together, and frowning.

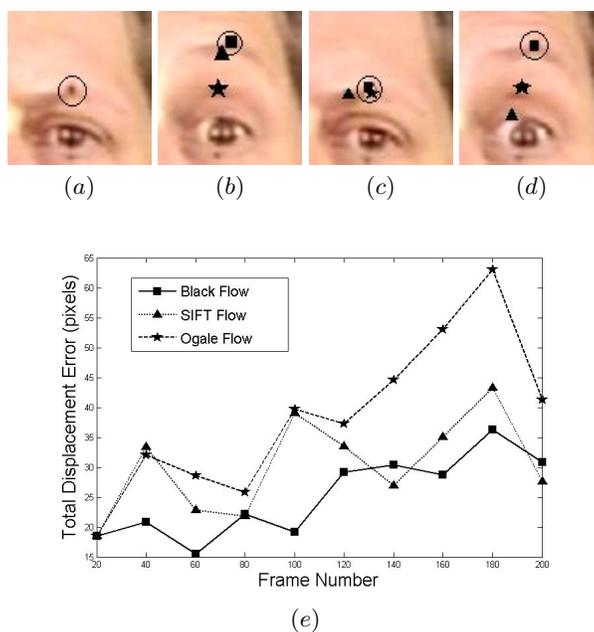


Fig. 1: Example tracking results of point given in circle (a). In (b) - (d), results for black flow (square) Ogale flow (triangle) and SIFT flow (star) at during two 'raise eyebrows' expressions (frame numbers 30, 120, 150). In (e) the actual error for all 15 points (see Fig. 2) is shown for every 20 frames.

3.1 Optical Flow and Tracking

In this paper, the primary purpose of optical flow is to calculate a dense correspondence between pixels over video sequences that contain expressions, a task that is important for the accurate calculation of strain maps. Hence, we explored several implementations of optical flow, including Ogale flow [5], SIFT flow [3],

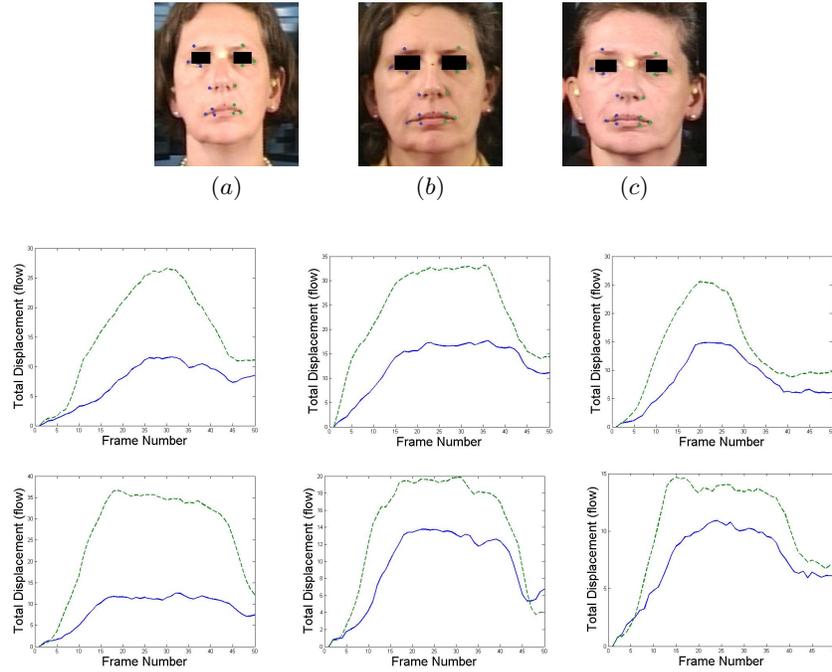


Fig. 2: Comparison of total flow displacement values between the left (solid blue line) and right (dotted green line) sides of face, after re-constructive surgery over 3 years (using Black flow). The images (a), (b) and (c) are the starting frames from each video and show the tracked points. The first row of graphs corresponds to the raised eyebrows expression and the second row corresponds to the smile expression.

and Black Flow [1]. To determine the best implementation choice, we inspected the tracking performance over several expressions at specific points. The points selected were the physical markers placed on the face, since these areas have texture information which aids in optical flow estimation. An example sequence containing the raised eyebrows expression can be seen in Fig. 1. This figure also shows the total summed error that was calculated for the same expression every 20 frames, at all fifteen points given in Fig. 2.

To further analyze the tracking results of each flow algorithm on this sequence, we calculated the average error (see Table 1) for all fifteen points and also for a subset of three select points near the right eye (see Fig. 2) where there was large eye / eyebrow motion. A few observations were made: Ogale flow occasionally showed sporadic tracking by jumping several pixels off and then back again. Overall, it resulted in average error rates of 2 pixels (for fifteen points) and 4.3 pixels (for three points). On the other hand, SIFT flow performed poorly even with small non-rigid movements of the eyebrow, since such local motion was

dominated back the lack of motion in surrounding regions. It had average error rates of 2.5 pixels (all points) and 6.1 pixels (three points). Black flow performed the best of all three and led to the most consistent results, with average error rates of 1.6 (all points) and 2.5 (three points).

Table 1: Average error (in pixels) for all 15 points tracked on face and a subset of 3 points that have relatively large motion.

Flow Type	All Points	Three Points
Black Flow	1.67	2.58
SIFT Flow	2.01	4.35
Ogale FLOW	2.55	6.14

Next, we show the tracking results using black flow for two expressions (raised eyebrows, smile) for each year. For this, points on each side of the face were tracked over two expressions and the displacements were summed to generate total summed displacement. As expected, for both expressions, the difference between the total displacement observed for each half of the face is largely reduced between 1998 and 1999, and even more between the 1998 and 2000 (see Fig. 2). This indicates that optical flow is successfully capturing the motion caused from the facial expressions. Next, we will use these flow vectors to calculate strain magnitude and optical strain maps.

3.2 Optical Strain Maps

Since strain maps represent the non-rigid deformation observed on the face during an expression, it is important that we capture the peak of the expression. We automatically get this frame using an expression spotting algorithm [6]. In summary, the algorithm utilizes the strain magnitude calculated over the entire video sequence, and correlates spatio-temporal regions that contain high strain values as segments containing expressions. It is particularly robust to expressions that occur in small regions or one side of the face, making it ideal for patients with facial palsy. The algorithm returns the frame number in a expression sequence that has the highest summed strain magnitude. These frames are then used for calculating final strain maps. Fig. 3 shows the strain maps calculated for all five expressions, over all three years. It is important to note here that eye regions and mouth regions have been masked due to common flow failure in these regions, due to self-occlusion (eyelids, inside mouth). For areas outside of the masked regions, large intensity values correspond to regions of the patients soft-skin tissue that have deformed significantly due to muscular contraction.

Quantization using Strain Difference Subtracting two strain maps from different years but the same expression allows us to gain a representation of the

change in deformation, or the change in active regions of the face. As can be observed in Fig. 4, the strain maps showing the difference between the years 1998 and 1999 suggest a large amount of improvement (first row), while the gain between 1999 and 2000 (second row) appears to be less.



Fig. 3: Optical strain maps for five expressions, over three years. Strain maps were generated between the start and peak of each expression. Intensity values correspond to amount of deformation observed.

4 Conclusions

In this paper, we explore the use of marker-less tracking methods for the purpose of evaluating the improvement gained from facial re-constructive surgery on patients with facial palsies. We have explored several tracking methods that allow us to create the dense correspondence necessary for strain map calculation and have concluded the Black flow leads to the most consistent and reliable results,

with less than 2 pixel average tracking error. Using these optical flow fields, we have proposed a method that quantizes the non-rigid deformation observed on facial skin tissue into strain maps. Strain maps can then be used to highlight the (a)symmetries between each side of the face, while also providing a useful measure of the changes at each point on the face over time, thus potentially allowing surgeons to quickly evaluate the efficacy of facial reconstructive surgeries.

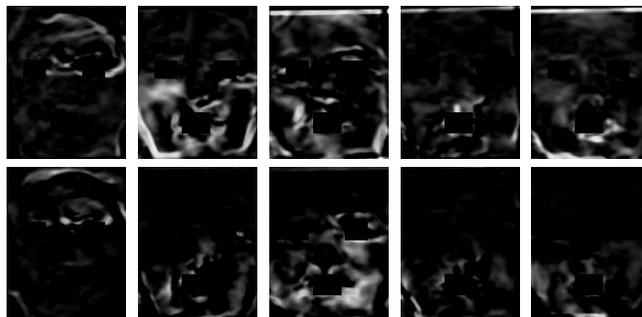


Fig. 4: Quantization using strain map difference between years recorded, post surgery. Each column contains one of the five expressions and shows the difference between strain map rows in Fig 3. The first row shows the change from 1998 and 1999, and the second row shows the change from 1999 to 2000.

References

1. M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. In *Computer Vision and Image Understanding*, volume 63, pages 75–104, New York, NY, USA, 1996. Elsevier Science Inc.
2. M. Frey, P. Giovanoli, M. Slameczka, and E. Stussi. Three-dimensional Video Analysis of Facial Movements: A New Method to Assess the Quantity and Quality of the Smile. In *Journal of Otology and Neurotology*, volume 23, pages 1531–7129, 1999.
3. C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: dense correspondence across different scenes. In *The Tenth European Conference on Computer Vision*, Heraklion, Greece, 2008.
4. M. Frey, A. Jenny, P. Giovanoli, and E. Stussi. Development of a new documentation system for facial movements as a basis for the international registry for neuromuscular reconstruction in the face. In *Plastic and Reconstructive Surgery*, volume 93, pages 1334–1350, 1994.
5. A. S. Ogale and Y. Aloimonos. Shape and the stereo correspondence problem. In *International journal on Computer Vision*, volume 65, pages 147–162, Hingham, MA, USA, 2005. Kluwer Academic Publishers.
6. M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar. Macro- and micro-expression spotting in long videos using spatio-temporal strain. In *The Ninth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011)*, 2011.