# Defining Gestures from Optical Flow

Jon Harter

University of Central Florida

July 16, 2009

# Defining Optical Flow

## Optical Flow

- The apparent motion of objects, edges, and surfaces in a visual scene
- Motion may be caused by movement within the scene or movement of the observer

## System Constraints

- Assume most optical flow is due to the motion of the observer
- Assume optical flow data calculated from out system will be be unique for each possible rotation and translation (unable to assume with one camera)

# Calculating Optical Flow

## Lucas-Kanade Method

- Optical flow equation: $\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = I_x V_x + I_y V_y + I_t = 0$
- Between two frames, assume for small $n \times n$ windows that flow $(V_x, V_y)$ is constant
- Assume constant intensity $(I_{x_i}, I_{y_i})$ between two frames
- Under assumptions, solve an over-determined system:

$$\begin{bmatrix} I_{x_1} & I_{y_1} \\ I_{x_2} & I_{y_2} \\ \vdots & \vdots \\ I_{x_n} & I_{y_n} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t_1} \\ -I_{t_2} \\ \vdots \\ -I_{t_n} \end{bmatrix}$$

- Optional representations:

$$\vec{v} = (A^T A)^{-1} A^T (-b)$$

or

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum I_{x_i}^2 & \sum I_{x_i} I_{y_i} \\ \sum I_{x_i} I_{y_i} & \sum I_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{x_i} I_{t_i} \\ -\sum I_{y_i} I_{t_i} \end{bmatrix}$$

# Ego-motion Constaints

### Longuet-Higgins Model

- Theoretical model (not used to analyze frames)
- Calculate optical flow as a function of rotation and translation
- Optical flow equations:

$$u = \frac{R_1 \cdot (x - \mathbf{t}/Z)}{R_3 \cdot (x - \mathbf{t}/Z)} - x$$

$$v = \frac{R_2 \cdot (x - \mathbf{t}/Z)}{R_3 \cdot (x - \mathbf{t}/Z)} - x$$

- Generalized for multiple cameras by incorporating center and orientation of each camera into $R$ and $\mathbf{t}$

# Distinguishing Between Each Method

### Lucas-Kanade

- Used to calculate optical flow between consecutive frames from a camera
- This method is responsible for the arrows on the screen (demoed earlier)
- This creates the raw data that will be analyzed to determine a gesture

### Longuet-Higgins

- Used to calculate artificial optical flow given specific values for rotation and translation ($R$ and $\mathbf{t}$)
- This method is responsible for synthetic data that the camera data will be compared to

# Gesture Representation

Gesture 1: Zorro

| $C$ | $R_x$ | $R_y$ | $R_z$ | $t_x$ | $t_y$ | $t_z$ |
|---|---|---|---|---|---|---|
| 51 | 2 | 0 | 0 | -1 | 0 | -1 |
| 51 | 2 | 1 | -1 | 1 | -2 | -2 |
| 104 | 1 | 0 | 1 | 0 | -1 | -2 |
| 5 | 1 | 0 | 2 | 0 | -1 | -2 |
| 104 | 1 | 0 | 1 | 0 | -1 | -2 |
| 5 | 1 | 0 | 2 | 0 | -1 | -2 |
| 136 | -1 | -1 | 0 | -1 | 0 | 2 |
| 136 | -1 | -1 | 0 | -1 | 0 | 2 |
| 136 | -2 | -1 | 0 | -1 | 0 | 2 |
| 136 | -1 | -1 | 0 | -1 | 0 | 2 |
| 136 | -1 | -1 | 0 | -1 | 0 | 2 |
| 136 | -1 | -1 | 0 | -1 | 0 | 2 |
| 136 | -1 | -1 | 0 | -2 | 0 | 2 |
| 139 | -2 | 0 | -1 | 0 | -1 | 1 |
| 116 | 1 | 1 | -2 | 2 | -2 | -1 |

Gesture 2: S-Shape

| $C$ | $R_x$ | $R_y$ | $R_z$ | $t_x$ | $t_y$ | $t_z$ |
|---|---|---|---|---|---|---|
| 136 | -1 | -1 | 0 | -2 | 0 | 2 |
| 136 | -1 | -1 | 0 | -2 | 0 | 2 |
| 136 | -1 | -1 | 0 | -1 | 0 | 2 |
| 51 | 1 | 0 | 0 | 0 | -1 | -2 |
| 51 | 1 | 1 | 0 | 1 | 0 | -2 |
| 80 | 1 | 0 | 1 | 0 | -2 | -2 |
| 51 | 2 | 1 | 0 | 1 | 0 | -2 |
| 99 | 1 | 1 | 0 | 2 | 0 | -2 |
| 51 | 2 | 1 | -1 | 1 | -2 | -2 |
| 61 | 2 | -1 | 0 | 2 | 0 | 1 |
| 51 | 1 | 1 | 0 | 1 | 0 | -2 |
| 51 | 1 | 1 | 0 | 1 | 0 | -2 |
| 51 | 2 | 1 | 0 | 1 | 0 | -2 |
| 51 | 2 | 1 | 0 | 1 | 0 | -2 |
| 51 | 2 | 1 | 0 | 1 | 0 | -2 |

# Classifier

- Using data from comparing the camera optical flow to synthetic optical flow as well as other features, gestures can be recognized
- Target gestures: Stab, Swing, Zorro, S-Shape, and any other simple arm motions