Sketch Sequence Alignment: Video Stabilization Through Contextual Descriptors

Joseph Geumleak & Ariel Meng

Advisor: Ernie Esser & iCAMP

August 13, 2012
Video Stabilization Motivation

In order to see the relevant changes in a video, it helps to align the frames to remove camera movements.

**Figure:** A Visualization of Frame Alignment

When there are many frames, this takes a lot of work. It would be nice to have it done automatically.
How Are We Different?

- General strategy: find enough point correspondences to determine simple transformation.
- Many methods for video stabilization exist, most using just local features to find matches.
- Our method uses more global information, mainly a point’s context relative to some other points.
Method Outline

1. Point Selection:
   - Preprocessing
   - Corner Detection

2. Descriptors:
   - Global Descriptor
   - Local Descriptors: Color, Edge
   - Cost Function

3. Point Matching:
   - Select matches
   - Calculate Transformation
   - Eliminating bad matches
Original Data Set

Video of Original Data Set
Challenges

Sea shell sketches: Video-specific Challenges

- Changes in contrast
- Self-similarity
- Addition of detail
Frame Preprocessing

1. Apply edge filter – 2. Equalize – 3. Apply median filter
Contextual Descriptor

\[
\text{Cost}(h_i, h_j) = \sum_k \frac{(h_i(k) - h_j(k))^2}{h_i(k) + h_j(k)}
\]
Global Contextual Descriptor

1. Detect Corners
2. Compute histogram: Every point against reference point distribution.
3. Find best match.
4. Compute transformation and align frames.

Figure: Two Frame Alignment

---


Local Descriptors

Use local features to improve matching by redefine the cost function.

\[ \text{Cost} = w_g \text{Cost}_g + w_c \text{Cost}_c + w_e \text{Cost}_e \]  \hspace{1cm} (2)

\begin{align*}
    w_g &= 1 \\
    w_c &= 0.25 \\
    w_e &= 0.5
\end{align*}
Local Descriptor: Color

- Use RGB vector at each point.
- Normalize to compensate for intensity differences.

\[ \text{Cost}(\vec{r}_i, \vec{r}_j) = \left( \frac{\vec{r}_i}{\|\vec{r}_i\|} - \frac{\vec{r}_j}{\|\vec{r}_j\|} \right)^2 \] (3)
Local Descriptor: Edge

- Apply Prewitt Filter
- Separate out edge points
- Compute each reference point against the edge points distribution.

Figure: Applying Edge Filter
Finding Matches

• Ideal matches are one-to-one.
• Frames have different point counts.

Use a simple greedy algorithm
• Find and record best matching pair.
• Consider the remaining points.
• Repeat.
Calculating Transform

Nonreflective Similarity Transform using least squares optimization.

\[
\begin{bmatrix}
  x_1^{out} & y_1^{out} \\
  \vdots & \vdots \\
  x_n^{out} & y_n^{out}
\end{bmatrix} =
\begin{bmatrix}
  x_1^{in} & y_1^{in} & 1 \\
  \vdots & \vdots & \vdots \\
  x_n^{in} & y_n^{in} & 1
\end{bmatrix}
\begin{bmatrix}
  sc & -ss \\
  ss & sc \\
  t_x & t_y
\end{bmatrix}
\]
Eliminating Bad Matches

Aligning $frame_j$ to $frame_i$:

1. Compute Transformation Matrix using all points.
2. Apply Transformation to points in $frame_j$.
3. Compute left-over distance cost and throw away invalid matches.
4. Recompute Transformation Matrix using only the good matches.
Summary

- Global Descriptor with Preprocessed Frame
- Local Descriptor
- Improve Matching

\[
Cost = w_g Cost_g + w_c Cost_c + w_e Cost_e
\]

\[
\begin{align*}
w_g &= 1 \\
w_c &= 0.25 \\
w_e &= 0.5
\end{align*}
\]
Result

origVScombined
Result: Subtraction

origVScombinedSUB
Comparison to RANSAC

RANdom SAmple Consensus (RANSAC):
- Takes Random Samples.
- Compute Transformation.
- Subset: inliers. (Outliers removed)
- Non-deterministic.
Comparison: RANSAC vs. Contextual

MatlabDemoVScombined
Quick Questions?

More time will be allocated for questions after the next presentation.