ICARCV 2014

Vision-based Detection and Pose Estimation for Formation of Micro Aerial Vehicles

Mengmi Zhang, Feng Lin, Ben M Chen
Unmanned Systems Research Group, Dept of Electrical & Computer Engineering
National University of Singapore
Outline of the Presentation

- Introduction
- System Overview
- Methodology
- Results and Discussion
- Conclusion and Future Work
A micro aerial vehicle (MAV) is a class of unmanned aerial vehicles that has a size restriction and the modern craft can be as small as 15 cm.

Applications of MAVs:

- Search for snipers
- High risk indoor inspection
- Surveillance
- Others
Problem statement:

- MAVs are required to complete certain missions collaboratively.
- In extreme situations, there are no inter-vehicle communications and computer vision can be one alternative solution.
- Each MAV has limited computation power, information, and payload capabilities; hence, our research focuses on pose estimation using one monocular camera.
Our previous solution to pose estimation:

- The relative distance and orientation of the target with respect to the follower can be computed by using the known camera model and geometry information of the target when it is identified and its location is determined in the image.

Problem occurs when the follower sees the leader from different aspect views where its maximum length varies.
Computer Vision Based Sensing System

- Collected Model Images In Database
- Follower Pan/tilt Camera
- Leader Pan/tilt Camera
- Proposed Solution 1
  - Shape Context Descriptor
  - Cost Matrix
  - Hungarian Algorithm
  - Affine Transformation
- Proposed Solution 2
  - Shape Context Descriptor
  - Cost Matrix
  - Relaxation Labeling
  - Thin Plate Spline Transformation

- Relative Position and Orientations
- Object detection
- Image tracking
- Shape Matching
- Best Match with Models
- Pose Estimation
- Relative Position and Orientation
Object Detection

Various Haar wavelet features form weak classifiers and are then cascaded based on their weightage to separate the object of interest from the background.
Image Tracking using the Kalman Filter

An effective combination of detection and motion estimator is able to speed up the processing, as well as deal with moving targets and uncertainties of outdoor environments, such as variations in lighting and backgrounds.

\[
\begin{align*}
    x(k|k-1) & = \Phi x(k-1) + F_{\text{cam}} \Delta u_{\text{cam}} + \Lambda w(k-1) \\
    z(k) & = H x(k) + v(k),
\end{align*}
\]

- Predict next state based on current state
- Update current state based on observations
- Measurement noise taken into account
- Pan/tilt camera motion taken into account
Randomly resample both contours to equal #points

Point matching between two point sets:
1. Hungarian algorithm
2. Relaxation labeling

Point set transformation:
1. Affine transformation
2. TPS (Thin Plate Spline) transformation
A circular histogram: 12 sectors for angles and 5 sectors for radius
A circular histogram using log-polar space for radius and tangential angle as the reference angle.
A one-to-one bipartite graph matching problem. We want to minimize the objective function:

\[
H(\pi) = \sum_{i} C(p_i, q_{\pi(i)}),
\]

Hungarian Algorithm
Faster

Relaxation Labeling Algorithm
More accurate
**ICARCV 2014**

**Shape Matching - Point Set Transformation**

**Affine Transformation** of Point Set Q to Point Set P based on the correct correspondence between P and Q. The objective function is to minimize

\[ S_{A,t} = \sum_{k=1}^{m} \| p_{ki} - A q_{ki} - t \|^2. \]

Compared with affine transformation, it can deal with complex and deformed shape. The aim is to find the **TPS** interpolate \( f(x, y) \) to minimize the bending energy

\[ I_f = \int \int_{\mathbb{R}^2} (\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial y^2})^2 \, dx \, dy \]

**TPS can deal with situations when the shape is deformed (not simple rotation, scaling, and translation but distorted) with extra computation costs.**
Evaluation between the warped shapes and the prototypes

\[ D_{sc}(P, Q) = \frac{1}{n} \sum_{p \in P} \arg_{q \in Q} \min C(p, T(q)) + \frac{1}{m} \sum_{q \in Q} \arg_{p \in P} \min C(p, T(q)) \]
Shape Matching using Hungarian and Affine Algorithms
Shape Matching using Relaxation Labeling and TPS Algorithms
Results and Discussion

Depth Estimation Based On Flight Test Images

Our result performs better compared with our previous one in terms of its average error percentage.
Conclusion

The research work contributes towards the vision-based approach for MAV swarm formation in the following aspects:

1. An innovative approach for MAV detection and pose estimation using only one monocular camera
2. An integrated object detection, tracking, and pose estimation computer vision system
3. Relatively high depth accuracy in terms of average error percentages

Future Work

- Use additional shape descriptors to extract shape features
- Search for other computation efficient matching algorithms
- Use segmented sampled points with large curvature or histogram discrepancies as feature points
- Investigate on the performances of our depth estimation algorithms regarding the stability of our signal
Thank you!