Unconstrained Vision Guided UAV Based Safe Helicopter Landing

Arindam Sikdar[1], Abhimanyu Sahu[1], Debajit Sen[2], Rohit Mahajan[2], Ananda S. Chowdhury[1]
[1]Imaging Vision and Pattern Recognition Group, Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India; [2]Elmax Systems and solutions Kolkata, India.

Problem Statement/Contribution

Goal : To identify safe zones for helicopter landing for possible relief distribution in disaster stricken areas from aerial videos captured by UAVs in an unconstrained manner.

The major contributions of this work are:

- Proposed a epipolar constrained based clustering approach to extract stereo pairs from single camera
- Minimum Spanning tree (MST) based graph clustering is applied where the constraint is used to prune the edges of graph
- The constrained in applied over pair-wise frames collected during flight
- Additionally proposed publicly available AHL dataset containing several landing/no-landing zones along with GPS and camera information.
- Also provided ground truth annotations as foreground masks for frames with safe landing zone.

Stereo-Pair Generation using Constrained Clustering

Formulation of Epipolar Constraint:

- Two image planes captured from well calibrated cameras will form a stereo pair if they satisfy the epipolar constraint
- This is equivalent to saying the three vectors in fig 1:

$$\overrightarrow{C_0p_0}\cdot[\overrightarrow{C_0C_1}\times\overrightarrow{C_1p_1}]=0$$

- An overall fitness value of two given frames to be a stereo pair can be represented at iteration t in form of a matrix operation as: $F_{epi}=\frac{1}{M}Tr(\mathbf{Q}'^{T}\mathcal{E}^{t}\mathbf{Q})$
- The fundamental matrix E is estimated using M-point algorithm while minimizing Fepi at iteration t.

$$\frac{1}{M}Tr(\mathbf{Q}'^{T}\mathcal{E}^{t}\mathbf{Q})<\epsilon$$
$$\implies Tr(\mathbf{Q}'^{T}\mathcal{E}^{t}\mathbf{Q})<\epsilon M$$
$$\implies Tr(\mathbf{Q}'^{T}\mathcal{E}^{t}\mathbf{Q})<\delta$$

- The physical interpretation of imposing such constraint is that two frames can only be in the same group (cluster) if and only if their fitness are bounded by δ.

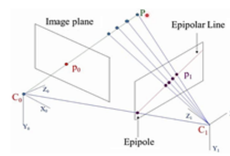

Fig 1. Epipolar geometry between two frames

Stereo-Pair Generation using Constrained Clustering

Construction of a Video Representation Graph:

- A complete weighted graph is initially constructed in the 64- dimensional feature space with all frames in a video segment.
- We term this as a video representation graph (VRG). Each frame is deemed as a vertex.
- Edges are established between each pair of vertices. The edge weight wij between the frames i and j with respective feature vectors fi and fj is given by:

$$w_{ij}=\sum_{k=1}^{64}|f_{ik}-f_{jk}|$$

- Now, the epipolar constraint in equation (7) is used to prune the complete graph to form a sparse graph. So, for edge weights, we write:

$$e_{ij}=\begin{cases} w_{ij} & Tr(\mathbf{Q}_i'^{T}\mathcal{E}_{ij}^{t}\mathbf{Q}_j)<\delta \\ 0 & otherwise \end{cases}$$

- Typically, a video segment with 230 frames would consists of $230C_2\approx 25000$ edges which reduces to a sparse graph with ≈ 5000 vertices.

Dataset Generation

This work contributes a dataset especially for developing and testing models for vision guided helicopter landing.

As per our knowledge there are no publicly available datasets of such kind and we are the first to propose and render real aerial dataset for such a problem.

We have named our dataset as Autonomous Helicopter Landing (AHL) dataset.

AHL dataset with all the ground-truth annotations are made available at https://sites.google.com/site/ivprgroup/helicopter-landing.

Fig 3. Various scenes of our dataset with and without safe-zones. Frames in first rows are without any safe landing zones while in second row the different safe zones are marked with red-mask

Experimental result (Quantitative)

TABLE I
CLUSTERING RESULTS FOR DIFFERENT APPROACHES

Video set	Methods	DI(↑)	DBI(↓)
Set01	k-means	0.123	0.952
	Agglomerative	0.155	0.781
	COP-Kmeans	0.164	0.744
	CSP	0.181	0.627
	CciMST	1.110	0.576
	Ours	**1.280**	**0.405**
Set02	k-means	0.151	1.810
	Agglomerative	0.181	1.760
	COP-Kmeans	0.183	1.300
	CSP	0.226	1.290
	CciMST	0.346	0.924
	Ours	**0.570**	**0.869**
Set03	k-means	0.190	1.280
	Agglomerative	0.201	0.934
	COP-Kmeans	0.188	0.871
	CSP	0.210	0.884
	CciMST	0.287	0.713
	Ours	**0.307**	**0.634**
Set04	k-means	0.217	1.240
	Agglomerative	0.235	1.160
	COP-Kmeans	0.296	1.140
	CSP	0.422	0.940
	CciMST	0.525	0.649
	Ours	**0.853**	**0.370**
Set05	k-means	0.614	0.553
	Agglomerative	0.649	0.518
	COP-Kmeans	0.677	0.525
	CSP	0.714	0.453
	CciMST	0.797	0.438
	Ours	**0.867**	**0.357**

- **Comparisons based on Cluster Validity Measures:** We have chosen standard k-means and hierarchical agglomerative clustering as baseline approaches along with two constrained clustering algorithms, namely, constrained k-means clustering and constrained-spectral-clustering as well as one recently proposed MST based clustering technique, namely, CciMST for comparison.

TABLE II
FRAME-LEVEL SAFE ZONE COMPARISON WITH VARIOUS COMPETING CLUSTERING METHODS

Methods	Criteria					
	TPR	TNR	Precision	Recall	F-score	Accuracy
k-means	0.125	0.952	0.438	0.125	0.194	0.761
Agglomerative	0.196	0.947	0.524	0.196	0.286	0.774
COP k-means	0.250	0.952	0.609	0.250	0.354	0.790
CSP	0.304	0.925	0.548	0.304	0.391	0.782
CciMST	0.339	0.941	0.633	0.339	0.442	0.802
Our Proposed	**0.554**	**0.952**	**0.775**	**0.554**	**0.646**	**0.860**

Experimental result (Qualitative)

- **Comparisons for Frame Level Safe-zone Detection with Ground Truth:**

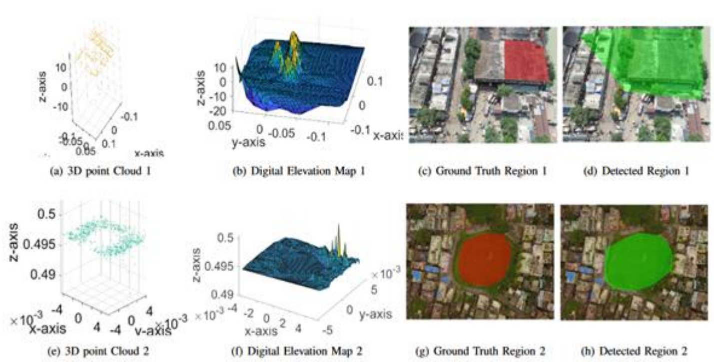

Fig 4. 3D Point Cloud, Digital Elevation Map, Ground Truth and Detected safe-zone in a frame from Set02 (top row) and Set04 (bottom row)