25th International Conference on Pattern Recognition
(ICPR 2020)

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]**I**maging **V**ision and **P**attern **R**ecognition Group
Electronics and Telecommunication Engineering
Jadavpur University, Kolkata, India.

[2]Elmax Systems and solutions
Kolkata, India.

Outline of the Presentation

- ❑ Background
- ❑ Contribution
- ❑ Proposed Framework
- ❑ Experimental Results
- ❑ Conclusions
- ❑ Bibliography

Background

- Increasing demand for computer vision based autonomous systems like self-driving cars , vision-guided Robots

- Emerging interests in Vision guided autonomous helicopter control

- Currently active sensor based autonomous helicopter landing systems are bulky, expensive, power hungry and requires specific skill set to operate

- In sharp contrast cameras are simple, cheap, small and require much less power

- Moreover, visual sensors are rich data source for better navigation and close proximity flights

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.

Natural Landmark Detection and Tracking

- Extraction of desired region-of-interests (ROI) due to repeatable nature of natural landmarks
- SURF based key-point extraction detecting and tracking landmarks
  - Hessian matrix is used to detect interest points

$$\mathbf{H}(p,\sigma) = \begin{bmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{yx}(p,\sigma) & L_{yy}(p,\sigma) \end{bmatrix}$$

  - Key-points are tracked over frames using fast approximate nearest neighbor search algorithm
- Constructed a frame representation vector using several detected key-points for robust stereo-pair generation
  - Each detected key-point are a 64 dimensional feature vector
  - Geometrically, median of all key-points in feature space

Stereo-Pair Generation using Constrained Clustering

- Constructing a 3D digital surface model of a real world generally requires a binocular stereo pair.

- Structure from motion is one such approach that uses 2D image sequence to generate a 3D structure.

- Applicable to small local changes in a scene and gives erroneous result with rapid scene changes (due to randomized UAV motion).

- Formulated a constrained graph clustering approach to handle such rapid changes for successful determination of stereo pairs from an unconstrained video capturing by a drone with random motion.

Stereo-Pair Generation using Constrained Clustering (Contd.)

- 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:

$$\overrightarrow{C_0 p_0} \cdot [\overrightarrow{C_0 C_1} \times \overrightarrow{C_1 p_1}] = 0$$

  - 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.

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

  - 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.

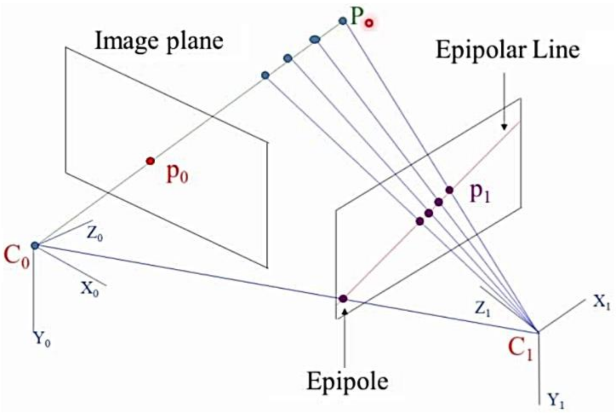


Fig 1. Epipolar geometry between two frames

Stereo-Pair Generation using Constrained Clustering (Contd.)

- 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 w_{ij} 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:

$$e_{ij} = \begin{cases} w_{ij} & Tr(\mathbf{Q}_i^T \mathcal{E}_{ij}^{\tau} \mathbf{Q}_j) < \delta \\ 0 & otherwise \end{cases}$$

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

Stereo-Pair Generation using Constrained Clustering (Contd.)

- MST based Sparse Graph Clustering:
  - We apply a MST based clustering on this sparse graph **G(V,E)**.
  - The MST based clustering algorithm provides some inherent advantages.
  - Firstly, it can also detect irregular shaped clusters including non-convex ones. Secondly, it does not require cluster number in advance.
  - In this clustering approach the inconsistent edges need to be removed. Those edges, whose weights are significantly greater than mean weight of the nearby edges, are deemed as inconsistent.
  - Following the normal distribution of edge weights in MST, the inconsistency measure can be formulated as:

$$w_{ij} > max(\overline{w}_{N_i} + \sigma_{N_i}, \overline{w}_{N_j} + \sigma_{N_j})$$

  - All edges of the tree satisfying the above inconsistency measure are most likely to be inter-cluster edges and are thereby removed resulting in a disjoint set of sub-trees each representing a separate cluster.
  - Each cluster consists of frames of same scene where the frames are likely to satisfy the epipolar constraint.
  - We then extract stereo image pairs from multiple such clusters as we intend to extract all possible safe-zones from different scenes

Digital Terrain Map Construction

- After obtaining several stereo pairs of a particular scene from one of the clusters, we first calibrate the camera to obtain the camera intrinsic parameters K

$$\mathbf{K} = \begin{bmatrix} \alpha_x & 0 & c_x \\ 0 & \alpha_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

- Based on Fundamental matrix E we can now generate 3D point cloud followed by Digital elevation Map (DEM).
  - **3D-Point Cloud Generation**: For the generation of the 3D point cloud we need to first estimate the camera pose between two images in terms of the rotation matrix $[R]_{3\times3}$ and the translation vector $[T]_{3\times1}$ that project a 3D world coordinate into a 2D image point.
  - **DEM Construction by 3D Delaunay Triangulation**: We now construct Digital Elevation Map (DEM) using Delaunay Triangulation (DT) over obtained 3D-point cloud.

Safe Landing Zone Detection

- An appropriate safe zone can be detected by measuring the slope and roughness of the landing site.
- While slope gives the measure of steepness, roughness indicates the possible presence of hazardous obstacles in the region.
- Once slope and roughness are determined we segment the “non-hazardous” regions.

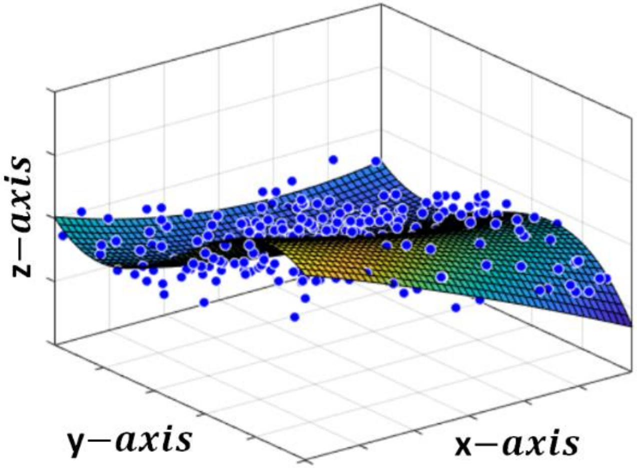


Fig 2. Smooth underlying surface of DEM obtained over 3D elevation points

- **Computation of Slope and Roughness Map**
- **Segmentation of Non-Hazardous Region**

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.

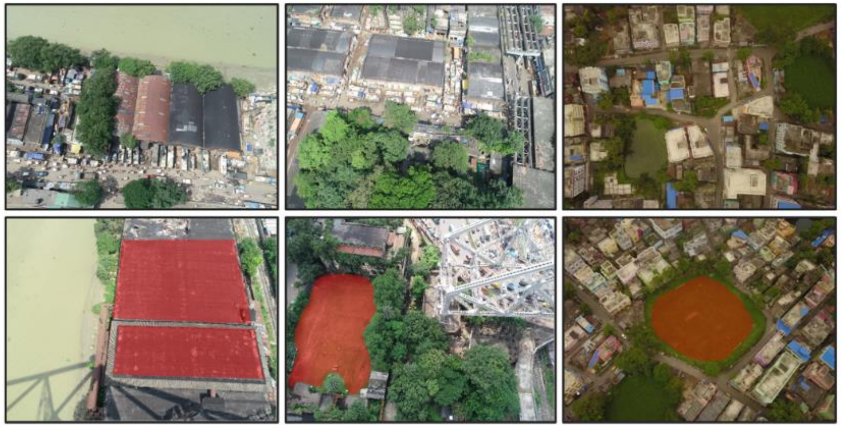


Fig 3. Various scenes of our dataset with and without safe-zones. Frames in first rows are without any safe landing zones while in second row the different safe zones are marked with red-mask

Experimental result (Quantitative)

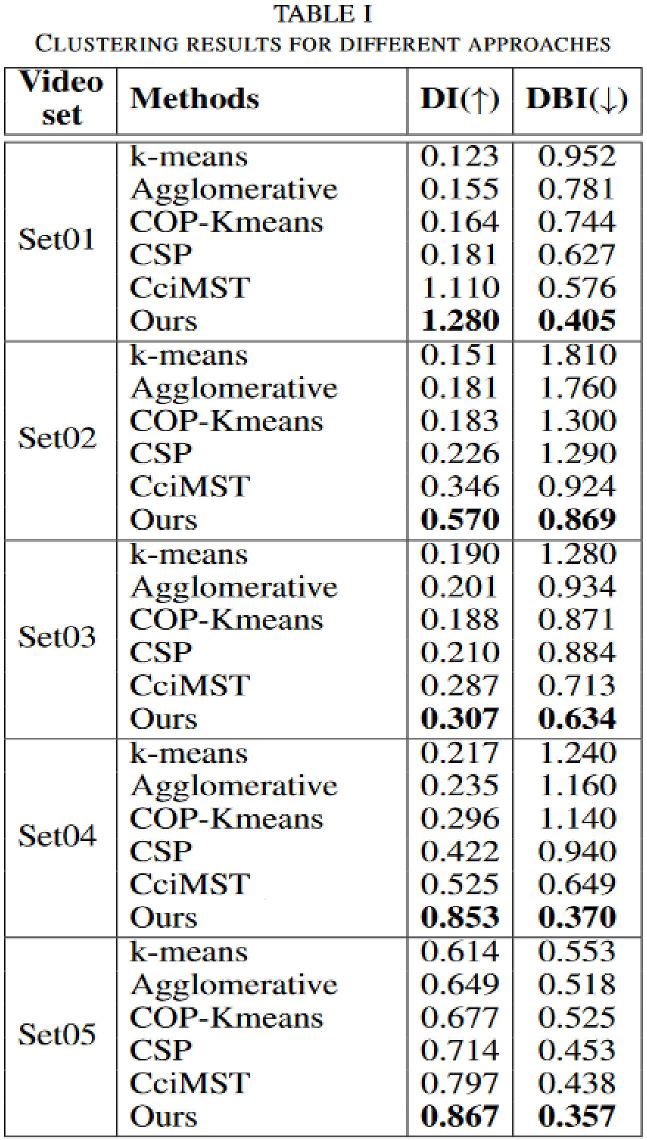


TABLE I
CLUSTERING RESULTS FOR DIFFERENT APPROACHES

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.

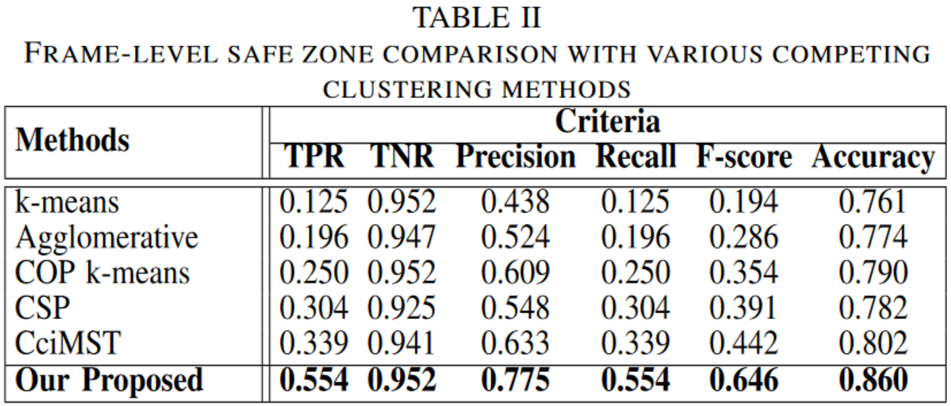
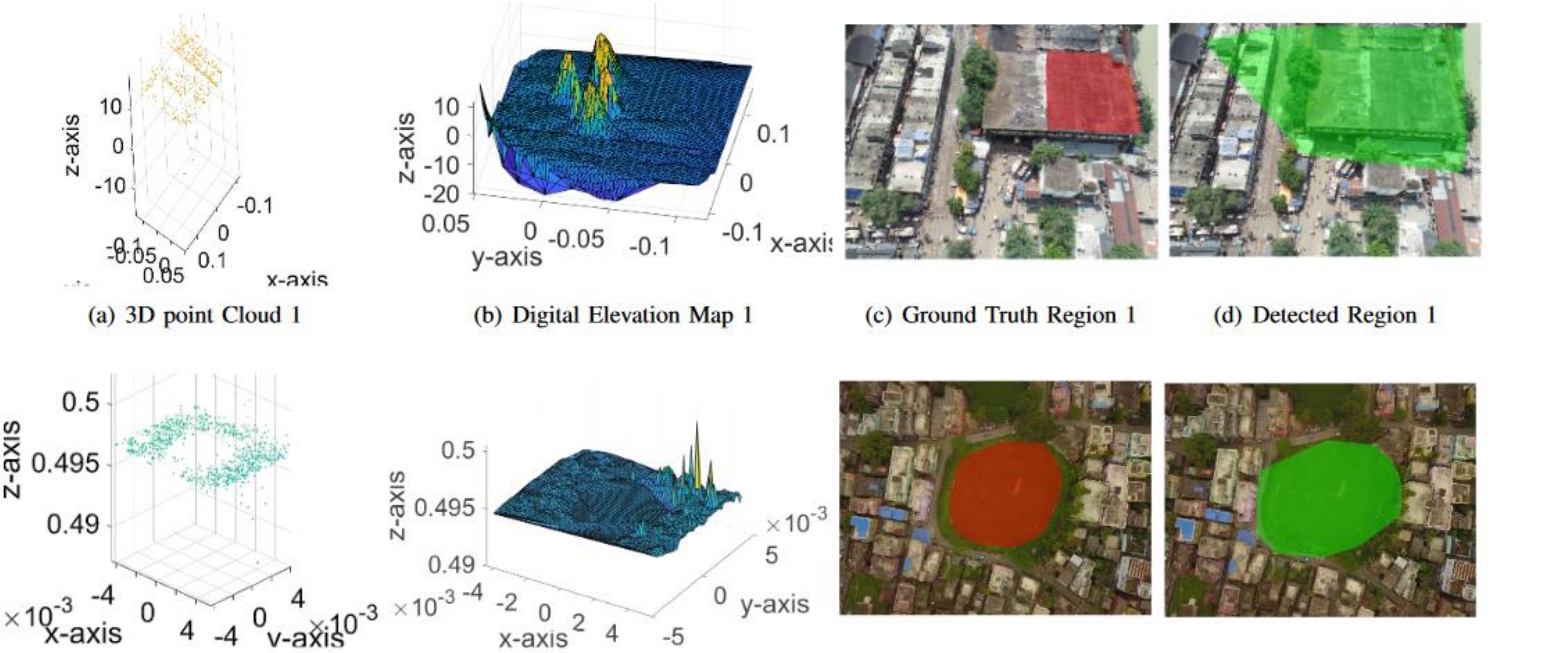


TABLE II
FRAME-LEVEL SAFE ZONE COMPARISON WITH VARIOUS COMPETING CLUSTERING METHODS

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**:



(a) 3D point Cloud 1 (b) Digital Elevation Map 1 (c) Ground Truth Region 1 (d) Detected Region 1

(e) 3D point Cloud 2 (f) Digital Elevation Map 2 (g) Ground Truth Region 2 (h) Detected Region 2

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)

Conclusions

- We have presented a solution to the safe zone detection for helicopter landing from videos captured by a drone in a completely unconstrained manner.

- An epipolar constraint driven MST based clustering algorithm is designed as a part of the solution.

- A new (AHL) dataset consisting of five sets of aerial videos has publicly been made available.

- Experiments clearly show the effectiveness of proposed formulation.

- In future, we will primarily focus on the pixel level safe zone detection by improving camera calibration.

Bibliography

- A. Johnson, J. Montgomery, and L. Matthies, “Vision guided landing of an autonomous helicopter in hazardous terrain,” in ICRA, 2005, pp. 3966–3971.

- S. Saripalli and G. S. Sukhatme, “Landing a helicopter on a moving target,” in ICRA, 2007, pp. 2030–2035.

- M. A. Kaljahi, P. Shivakumara, M. Y. I. Idris, M. H. Anisi, T. Lu, M. Blumenstein, and N. M. Noor, “An automatic zone detection system for safe landing of uavs,” Expert Syst Appl., vol. 122, pp. 319–333, 2019.

- J. Oliensis, “Exact two-image structure from motion,” IEEE Trans. PAMI, vol. 24, no. 12, pp. 1618–1633, 2002.

- O. Chum, T. Werner, and J. Matas, “Epipolar geometry estimation via ransac benefits from the oriented epipolar constraint,” in ICPR, vol. 1. IEEE, 2004, pp. 112–115.

- K. Wagstaff, C. Cardie, S. Rogers, S. Schrodl ¨ et al., “Constrained kmeans clustering with background knowledge,” in Icml, vol. 1, 2001, pp. 577–584.

For more information, please visit:
https://sites.google.com/site/ivprgroup/

Thank you!

Questions? .