A vision-based approach for navigation and control of multiple drones

Dr. Daniel Raviv and Dr. Naphtali Rishe

ndr@terrafly.com

Pre-proposal of Industrial-Academic Partnership under the auspices of National Science Foundation's Industry-University Cooperative Research Center for Advanced Knowledge Enablement

The problem

Drones can easily be controlled remotely or with pre-programmed paths to achieve certain navigation tasks. A more difficult problem is what can the drones sense and how should they locally and autonomously act to initiate a local path that avoid a-priori unknown obstacles or other moving drones.

We propose a “looming”-based solution that uses visual data.

When dealing with a moving camera-based autonomous navigation system, a huge amount of visual data is captured. For vision-based navigation tasks (such as obstacle avoidance, maintaining safe clearance, etc.), relevant information needs to be extracted from the visual data and used in real-time closed-loop perception-action control system. In order to accomplish safe visual navigation several questions need to be answered, including:

1. What is the relevant visual information to be extracted from a sequence of images?
2. How does one extract this information from a sequence of 2D images?
3. How to generate control commands to the drone based on the extracted information?

We propose answers to all three questions, i.e., generation of vision-based control signals for collision avoidance and maintenance of clearance using multiple images obtained from on-board video cameras, i.e., using visual information only.

The Idea

Usually, the process of navigating in a 3D environment involves a human operator. The operator acts in part as a sensory feedback in the perception-action closed loop control to ensure safe navigation in real time. It becomes a difficult problem to replace the human
operator by a vision-based system to achieve similar tasks for the following reasons: In outdoor navigation the environment is usually unknown and unstructured, and the same 3D scene may result in many different images due to changes in illumination conditions, relative distances, orientation of the camera, choice of fixation point, etc., as well as various camera parameters such as zoom, resolution, focus, etc.

There is a need for an approach to obtain relevant visual information about relative proximity in the presence of the above mentioned factors and employ it as a set of sensory feedback signals to accomplish the tasks of safe navigation.

We propose measurable visual cues such as visual looming (VL) and visual threat cue (VTC) that carry important information about the relative proximity. These cues do not need any 3D reconstruction process, which is usually computationally intense. In many cases these visual cues are sufficient for safe navigation. Clearly image processing is needed.

What we have done

Inspired by nature (e.g., school of fish or flock of birds do not show any signs of collisions [https://youtu.be/eakKfY5aHmY](https://youtu.be/eakKfY5aHmY)) and human-based behaviors (e.g., huge intersections in Saigon with many and different vehicles that do not collide while crossing the intersection [https://www.youtube.com/watch?v=4phFYiMGCIY](https://www.youtube.com/watch?v=4phFYiMGCIY)), we have researched the concepts of Visual Looming (VL) and Visual Threat Cue (VTC) as powerful visual cues for autonomous obstacle and collision avoidance. The cues, which are 3D extensions of the so called Time-To-Contact (TTC) are general (i.e., not ad-hoc) and are quantitatively measurable, using a single camera. They allow for local diversions from current paths before returning to the originally pre-planned path.

The methods are based on studying the relative temporal rate of expansion of surfaces and/or studying the texture and its temporal change near the camera fixation point. These visual cues or their equivalents have been used as a feedback signal of a closed loop system to implement obstacle avoidance using a six-degree-of-freedom simulator, as well as UNITY based simulations.

The VTC allows not only for collision avoidance but also for maintenance of clearance. The control schemes can be based on traditional control or fuzzy logic.

It is important to emphasize that both Visual Threat Cue (VTC) and the Visual Looming (VL) are measures that can be obtained directly from the raw data of gray level images, are independent of the 3D surface or its texture and need no optical flow information, 3D reconstruction, segmentation, feature tracking or preprocessing.
These motion cues are scale-independent, rotation independent and are measured in [the units of sec\(^{-1}\). There is no need to measure depth/range, since they are based on relative temporal change in images.

The VTC and VL correspond to visual fields surrounding a moving observer. The fields are time-based imaginary 3-D surfaces that move with the observer and can be segmented into multiple zones, e.g., safe, low-risk or dangerous zones that affect the degree of action to be taken.

Algorithms to extract the VTC and VL have been applied to several indoor as well as outdoor real images of objects textures.

We propose

1. A set of 3D simulations of multiple moving drones in different environments:

The drones will simultaneously change their velocities, i.e., directions and speeds, based on the different desired tasks at hand and local diversions from desired paths. The diversions from the paths will be based on Looming or Visual threat Cue values.

At first, the drones will have no dynamics (i.e., kinematics only). Later, they will be dynamically modeled to allow a better, closer to real life behaviors.

Drones will vary in size and dynamics

The simulated environments will mimic manmade structures, and later natural environments.

Simulations will be based on UNITY software.

2. A set of experiments using real drones and real cameras to test, evaluate, and modify the simulated algorithms or sensing mechanisms.

At first the experiment will be based on non-real time navigation, followed by real time behaviors.

The real experiment will also answer the three questions mentioned earlier, i.e.:

1. What is the relevant visual information to be extracted from a sequence of images?
2. How does one extract this information from a sequence of 2D images?
3. How to generate control commands to the drone based on the extracted information?
Appendix A: The Visual Threat Cue

The Visual Threat Cue (VTC) provides some measure for a relative change in range as well as for specific clearance between a 3D surface and a fixated observer in motion. This cue is independent of the 3D environment and needs no a-priori knowledge about it. It is time-based, rotation independent and does not need 3D reconstruction. This cue can be extracted directly from the raw gray level data of images and does not need optical flow information, segmentation, feature tracking and pre-processing. Mathematically the VTC (for $R > R_0$) is defined as follows:

$$VTC = R_0 \frac{d(R)/dt}{R(R - R_0)}$$

where $R$ is the range between the observer and a point on the 3D surface, $d(R)/dt$ is the differentiation of $R$ with respect to time, and $R_0$ is the desired minimum clearance and has the same units as $R$. Note that the units of the VTC are [time$^{-1}$]. The VTC has been shown to be independent of the rotational motion and can be measured without knowledge about $R$. The VTC corresponds to a visual field surrounding the moving observer, i.e., there are imaginary 3D surfaces attached to the observer that are moving with it, each of which corresponds to a value of the VTC. The points that lie on a relatively smaller surface correspond to a relatively larger value of VTC, indicating a relatively higher threat of collision. The VTC value on the minimum clearance sphere of $R_0$ centered at the location of the observer is the maximum which is infinity, indicating that the absolute distance between the observer and the camera is the minimum clearance. See Figures. Note that this field is not a sphere.

Figure A1: Sections of Equal VTC surfaces plotted for different velocity values
Appendix B: Visual Looming

Visual looming is related to an increasing projected size of an object on a viewer's retina as the relative distance between the viewer and the object decreases. It is an indication for threat that may be used to accomplish navigation tasks.

Visual looming is the time derivative of the relative distance (range) between the observer and the object divided by the relative distance itself, i.e.,

\[ L = -\frac{dR}{dt/dR} \]

where L is the looming value, and R in the range.

Visual looming can be calculated from the relative temporal change in the following attributes of a 2D image sequence: (i) image area, (ii) image brightness, (iii) texture density in the image, and (iv) image blur. It is shown that a closed-form unified expression can be adopted in all these methods. Experimental results illustrate how the measured values of looming are related to the actual values. Finally, looming is used in the sense of a threat of collision, along with visual fixation, to navigate in an unknown environment.
This visual cue is used as a feedback signal of a closed loop system to implement obstacle avoidance using a six-degree-of-freedom simulator.

Figure B1: Danger, High Risk and Safety Zones plotted relative to the moving vehicle (observer)
Figure B2: More Specific Danger Zone

References


K. Joarder and D. Raviv, “A New Method to Calculate Looming from Surface Texture and Visual Fixation,” Accepted, Transactions of the IEEE on Pattern Analysis and Machine Intelligence (Also NISTIR)
More on Looming

http://www.nist.gov/customcf/get_pdf.cfm?pub_id=820403

http://www.nist.gov/customcf/get_pdf.cfm?pub_id=820511