Dickmanns ED (1988). **Computer vision for flight vehicles**

Summary: The advantages of providing flight vehicles with a sense of vision are discussed. The use of integral spatio-temporal world models is considered to be a prerequisite for achieving real-time performance with the microprocessors currently available. Feature-based image sequence processing exploiting this approach has been tested in a real-time simulation loop, with a 10 Hz frame rate, for visual interpretation and control. Airspeed was the only quantity measured conventionally. Results for a fixed base simulation, using real image sequence processing hardware in the loop, are given for a final landing approach and flare at speeds around 130 kts.

1 Introduction

Man's millennia-old dream of flying like birds has become a reality over the last nine decades - at least to some degree. Today's great variety of airframes and propulsion systems permit a broad spectrum of vehicles but except for rather simple missions, sensor data integration and purposeful actions have to be done by man.

Piloting an aircraft is a rather demanding task and engineers have long been considering ways to alleviate this task by partial automation. In fact, automatic control experiments for flight vehicles date back further than successful flight because some aircraft pioneers considered the stabilization problem in six degrees of freedom to be too demanding to be handled by man [1]. With their flights in 1903, the Wright brothers demonstrated that man was able to control even basically unstable aircraft. The advent of the tail elevator, leading to a basically stable aircraft, meant that there was no real need for automatic control to enable man to fly.

However, when flying matured from pure adventure to professional activity, it was appreciated that automatic control could free the pilot from tiring monotonous control tasks. Instead, he could concentrate on mission planning and supervision, adjusting "autopilot"-parameters now and then and taking over control for the more complicated transitions between mission elements or for especially demanding mission elements.

Particularly during the last six decades, the sensing of single physical state variables has increasingly been transferred to more precise instruments. Certain control tasks, such as regulating the flight state in cruise conditions, have been taken over by autopilots relatively early; more complex mission elements, such as landing approach, only became autopilot functions about thirty years ago.

Man's ability to deal with unforeseen events, to evaluate new situations, develop new plans and action schemes and to physically manipulate devices on board will, in all probability, keep him in the on-board control loop for many missions for a long time to come.

There is, however, a need for high performance unmanned vehicles, especially in hazardous environments. Helicopters carrying TV-cameras will be useful for inspecting contaminated areas. Aerial intelligence by small drones has a wide field of military applications. Intelligent electronic devices with on-board decision capability and flexible response to a wide variety of situations are being sought.

Remote piloting requires a steady high bandwidth data link which is difficult to maintain, especially in a hostile environment. The human pilot, even if he is in the loop only by "tele-presence" via information- and control-telecommunication, introduces flexibility in reactions and a chance for real-time readjustment of the mission to the actual state unknown at the start, possibly including return and landing. It is his sense of vision which provides the human pilot with the most useful information, especially for interactions with the environment close to the Earth surface. For take-off and landing, a human pilot is able to extract almost all information he needs for control from the image flow on the retinas of his eyes. The task can even be performed with one eye (monocular vision) without much degradation in performance. A survey on the literature with respect to vision for aircraft landing approach may be found in [2].

For transcontinental or intercontinental flights, microwave-based (high frequency electromagnetic radiation, HF) navigation systems with all-weather performance capabilities have been developed and these may be sufficient even in the long term. The satellite based global positioning system (GPS) deployed in the US will yield global coverage and high navigational accuracy. Radar has been developed for detecting obstacles like other vehicles or storm cells and gives angular directions, radial distance and, by proper processing over time, the relative speed vector of other objects. Radar may by considered an active vision system for the translatory degrees of freedom of larger objects. In many applications today radar signals are displayed to the pilot by cathode ray tubes (CRT) as distance images and he has to interpret them to assess the situation and make correct decisions.

Computers are becoming powerful enough to take over image interpretation tasks and perform data fusion of data originating in different sensory modalities; i.e., combining them to an integrated representation of the environment and the egomotion state.

"Vision" in its proper sense, as used for living beings, has a somewhat more specific meaning - the sensed electromagnetic radiation field in the environment together with its temporal changes ("image flow") is used internally to reconstruct, during the interpretation process, a spatial and temporal (spatial velocities) representation of the environment and of the relative egomotion. It is essential to note that we humans perform cognition in a spatiotemporal unit, without separating space and time, even though the image flow on the retina is two-dimensional; the mapping by perspective projection eliminates direct depth information. In an even more restricted sense, the term "vision" is used in human culture for the ability of some individuals to preview developments or events and to ask for preparatory steps in order to be able to deal with these situations.

In today's technical systems, the sense of vision is at its most rudimentary. Two separate origins may be traced in aerospace engineering. In the first of these, the infrared seeker/goniometer in missiles, which measures the angular displacement between a point light source and the viewing direction, may be considered as very simple "eye-like" devices. More modern imaging systems permit resolution of areal intensity distributions but do not go beyond a two-dimensional interpretation. Imaging radar systems can add range and range rate information, thus yielding the full translatory state vector.

The second origin goes back to digital image processing. In the early sixties, mainframe computers were used to improve and regularize static images obtained from remote sensing by aircraft and spacecraft. Segmentation and classification tasks for areal elements were soon transferred to the computer. Comparing two images of the same area taken at different times makes it possible to detect changes. From this point, it seems a small step to try and identify objects -at least in a conceptual sense; in practice, this is a very big step which has certainly not been satisfactorily solved up to the present. Today, image processing and image sequence processing are very active areas of research. Many different types of computer systems from PC to super-mainframes find application and image sequence processing is one of the main driving forces for developing highly parallel computer architectures [3]. On the other hand, it is surprising what performance levels can be achieved by organic systems, with an amount of "wetware" (a term designating the biological nervous hardware in living beings), of the order of one cubic centimeter in birds: they can do formation flying, obstacle avoidance, partner and prey chasing, landmark navigation and precise landings.

When landing on the branch of a tree, they do not just perform a two-dimensional kinematic prediction in the image plane; they indulge in full three-dimensional dynamics. Approaching at an altitude lower than the branch they want to land on, they can, in a final pull-up maneuver, trade kinetic energy for altitude and arrive almost precisely above the branch with a vanishingly small, relative velocity. The final fine adjustment is done by active wing flapping while their claws already grip the branch in a well-coordinated way. In pursuit maneuvers, predatory birds like hawks watch the roll angle of the evading prey bird and initiate a similar egomotion component giving them a gain in the pursuit relative to pure translatory tracking. In both cases, knowledge of the bird's own capabilities and of cause and effect sequences in dynamic manoeuvers is used to generate egomotion behavior by means of proper aerodynamic control. Today's technical systems lack the ability

- to pick up the information needed and available in the electromagnetic radiation field,
- to reconstruct, from a temporal image flow, a sufficiently correct representation of the environment and of objects moving in translatory and rotatory degrees of freedom in space and
- to derive driven by a set of objectives and based on the situation as "seen" an action plan, which then is implemented with concurrent supervision by the situation assessment process described above.

Computing power and data processing schemes are now maturing to the point where complex behavioral patterns similar to those in animals may be implemented in technical systems in the near future. Highly parallel microprocessors in foreseeable hardware technology will be sufficiently powerful to attack moderately demanding tasks. There seems to be a lack of efficient data processing schemes unifying the well-proven recursive numerical estimation methods and the more modern so-called "artificial intelligence"-methods (AI).

A rather wide range of imaging sensors is available -black/white, color and low light level (LLL) television (in the mode proposed, TV would be better termed "machine"-vision, since it is no more "tele-", meaning distant), forward looking infrared (FLIR), microwave radiometry and imaging radar; combinations of these may provide powerful systems applicable over a wide range of situations.

Many fundamental research problems have yet to be solved. What models for object and state recognition are efficient? How can data fusion be achieved most economically? What are the best schemes for knowledge representation and for planning? How can all this be implemented on parallel computers? Work in these directions is under way at many institutions concerned with aerospace engineering and avionics, especially with respect to defense technology development. The DARPA Program on Strategic Computing with the "pilot's associate" as the demonstrator example may be mentioned as one major center of interest in the U. S. [3]. A helicopter application is discussed in [4].

An approach taken at the Universitaet der Bundeswehr Muenchen (University of the German Federal Armed Services in Munich, UniBwM) will be discussed in more detail later in the paper but a short survey of possible applications of machine vision in aircraft engineering is given first.

2 Possible applications

The possible uses of intelligent computer vision systems in flight vehicles may be broadly subdivided into three classes – support of aircraft crew members, independent monitoring systems, improvement of autopilot system performance and new capabilities, i.e. extensions of autopilot functions.

Support of aircrew members: Information derived from visual systems may be cross-checked and blended with other information from scalar sensors or from communication inputs by the artificial intelligence system; only the combined information, possibly enhanced by confidence measures, will normally be presented to the crew member. On request, more detailed data may be provided or options suggested for solving a problem. This would make the task of sensory data integration and situation assessment easier for the human operator. It requires, however, that the man/ machine interface should be similar to that between the individuals in a team cooperating to solve a certain task. Natural language interfaces to and from the computer being developed may soon be an essential item. Future enhanced graphics displays may be able to provide much better communication channels between the computer and a human operator than are available between two humans. This is mainly due to the highly parallel data acquisition in the human visual system which a human partner cannot utilize. Machine vision results may be blended with the crew member's natural view of the outside world by means of head-up displays.

Independent monitoring system: The functions performed may be similar to those discussed above, except for the possibly high level of interaction with the crew members which may, in critical situations, be reduced to warnings. In addition to the environmental state, the performance and the awareness state of the crew may also be monitored in order to give early warning should the situation require it.

Improvement of autopilot system performance: Depending on the type of imaging system installed and the environmental conditions given, several improvements are possible. The potential may be greatest if - as in the cases discussed above - the interpretation process analyzing the incoming data maintains a knowledge base, using an integrated spatio-temporal (four-dimensional) world model in which both the egomotion and the motion processes in the relevant outside world are modeled. If the mission goals are formulated within an appropriate framework and if behavioral models for different situations are available, very flexible automation schemes become possible. Some initial steps in this direction are described in [5, 6].

Landmark navigation will allow precise localization and navigation relative to some prominent features in the environment, such as shore lines, river beds, big cities (or other remarkable building complexes), mountains, large railway or roadway systems and combinations of them. This - the only orientation resource for the flight pioneers of the early days - would be very much upgraded, if fully weather-proof passive radiometry systems could be developed. Since this is unlikely in the near future, landmark navigation will be restricted to special applications and visual flight conditions. In tasks like rescue search or remote sensing, it may play a predominant role.

In addition to landmark navigation machine vision may make it possible for small unmanned aircraft to make their missions much more flexible and adaptable to the actual situation because of the three-dimensional object recognition and classification capability. The exchange of symbolic information with remotely piloted vehicles may permit drastic reductions in signal transmission with only a minor reduction in information flow; local autonomy may contribute considerably to effectiveness and to reducing sensitivity to interference.

A technical system with full four-dimensional vision capability should be able to land fully autonomously, like birds or piloted aircraft under VFR-conditions, i. e., with no ground support system whatsoever. Like pilots, these systems could even detect obstacles on the runway and react to them properly, e.g., touch down behind one or keep to one side of the runway. Since landing an aircraft is a well-defined, rather complex control task needing full four-dimensional capabilities for efficient solution, it had been selected as the demonstration task for our work in developing machine vision for aircraft. The objective is to manage with monocular vision, sometimes called "cyclopean vision", by systematically exploiting motion stereo.

An extension of the first origin, mentioned above, for a sense of vision in aircraft would be the capability to observe the motion of other vehicles relative to the aircraft's own motion in all six degrees of freedom. This could be used to introduce lead for trajectory control, e.g. by observing, in pursuit maneuvers, the roll angle of a vehicle which has banked in order to turn.

New aircraft capabilities possible with vision - in addition to obstacle recognition on runways, as mentioned above - could be an auto-taxiing system for following taxiways and for automatically keeping a safe distance when in convoy to and from the runway.

In formation flight, automatic relative positioning and fast reactions to atmospheric disturbances are easily implemented by machine vision. With respect to VTOL flight vehicles, hovering relative to ground locations or buildings becomes an important ability. In contaminated environments, after an accident in a nuclear power plant, for example, this may make it possible to avoid exposing human pilots to high radiation doses. By using proper imaging sensors, automatic systems for helicopter landings in dense fog may become possible.

In contrast to the autopilot and guidance functions based on the imaging sensors now available, a fully developed machine vision system must have at its core a spatiotemporal world model containing all the major relationships influencing the dynamic behavior of the vehicle.

The rest of the article describes one approach and the initial results achieved in simulated landing approaches of a business jet type aircraft with real vision hardware in the real time simulation loop at the Armed Services University in Munich.

3 The vision paradigm chosen

In contrast to the approach generally chosen in computer vision, in which an attempt is made to invert the three-dimensional into two-dimensional perspective mapping - a usually non-unique process for a pair of noise-corrupted images, the approach selected here tries to circumvent the inversion problem by establishing, internally in the interpretation process, a spatiotemporal world model which is sufficiently complete in terms of the task at hand. The geometric properties of the environment, a dynamical model of the egomotion and the laws of perspective projection are exploited in conjunction in order to introduce as many constraints as possible for the image sequence processing. It is initially assumed that time invariant objects only exist and that their shape is known; parameters may be adjusted during the recognition process. The projected shape in the image is first used to generate hypotheses as to what object is observed, at which aspect angles and at what distance. This, however, is not continued as new images come in; the egomotion state which leads to the image flow observed is estimated instead. This encompasses the entire three-dimensional state vector, including all the velocity components.

Knowing the present state and the dynamical model of the process, the state can be predicted for the next time at which measurements will be taken. Applying the laws of perspective projection in the unique forward direction to this predicted state, the position of features in the new image can be computed. These predicted feature positions are then compared with those actually measured. The differences, usually occurring due to incorrectness of the model or to disturbances, are not interpreted in image coordinates but are immediately transformed into the spatiotemporal world model in order to adjust the estimated state components by recursive least square fits. The internal four-dimensional model is therefore servo-driven in accordance with the incoming image sequence and the interpretation model chosen. Figure 1, lower part, shows this feedback interpretation loop based on prediction.

If a temporally extended matching of the model images and the measured images can be achieved, the process is considered as having been recognized. Note that it is always the last image only in the sequence which is used, thus eliminating the need for the costly storing and accessing of previous images. The spatial velocity components, in particular, are not derived by differencing between subsequent images -optical flow in the image sequence is not computed at all -but by the integration of prediction errors. This model-based method, well known from modern control theory [7], has good smoothing properties. It has been extended here to the case of feature-based motion recognition in a three-dimensional environment. In terms of modern philosophy [8], this may be considered to be an embryonic stage of a "world 2" for the interpretation process. The method can easily be generalized to deal with cases including other moving objects or an environment which varies with time - at least conceptually. The computational load, however, demands more parallel computing power than is available today. It seems probable that this situation will change very soon.

The approach described has been implemented in a specially developed multi-microcomputer image sequence processing system (BW) [9]. Two additional features have been found very useful for augmenting the performance level of the system:

- active control of the viewing direction by the interpretation process (permitting fixation on image feature combinations) and inertial stabilization for reducing motion blur;
- parallel use of two (coaxial) TV-cameras with different focal lengths for a wide viewing angle in the close-up range (corresponding to peripheral vision in humans) and for good resolution at larger distances (corresponding to foveal vision).

The resulting vision system architecture functionally resembles that of vertebrates although the hardware elements are completely different. Results reported here have been gathered by G. Eberl using one black and white CCD TV camera as the sensor [2].

.

. (technical part with figures)

.

7 Outlook

The feature-based four-dimensional dynamic scene analysis approach using integral spatiotemporal world models, including perspective projection, has shown promising results. Noting that the next generation of image sequence processing hardware B W [9] may increase processing power by two orders of magnitude in the next two years and that general parallel microprocessors will exhibit a similar increase in performance compared with the 1983 standard used here, real time motion by machine vision in well-structured civilized environments may become practically applicable in the near future.

In order to improve flexibility and robustness, active viewing direction control by the interpretation process in the computer and the integral use of visual and inertial feedback are being investigated. The internal four-dimensional world model comprising egomotion and perhaps other independently moving three-dimensional objects and their shapes is well suited for the fusion of measurement data coming from other imaging sensors (e.g., radar). It provides a natural basis for the man/ machine communication which is so important for establishing the confidence level necessary for introducing automation at this high degree of autonomy. Based on this type of very efficient state estimation and representation, "artificial intelligence" methods may be in a better position to deploy their full potential than they would be without the integral spatiotemporal world model.

8 Conclusions

By combining the dynamic models of modern control theory, shape representations of objects, their relative spatial orientation and the laws of perspective projection as part of the measurement model, a numerically efficient recursive scheme for real time image sequence processing has been obtained [6]. Its performance capability has been demonstrated in a task as complex as the fully autonomous final landing approach of an aircraft in six degrees of freedom in a real time "hardware in the loop" simulation [2] with moderate computer processing power.

In the near future, increases in computer power by several orders of magnitude compared to that used will make it possible to undertake flight tests with real aircraft under normal visual flight regime conditions, including small atmospheric perturbations. Autonomous visual autopilot functions may be beneficial for landmark navigation close to the Earth, including precise relative hovering for helicopters. With infrared or multispectral imaging sensors, day and night capabilities may be achieved. RPV's may become more autonomous, permitting reduced communication bandwidth; they may even achieve the capability of autonomous (bird like) landing without any ground support.

Literature

1. Maxim, H.S.: Artificial and natural flight. London: Whittacker 1908
2. Eberl, G.: Automatischer Landeanflug durch Rechnersehen. Dissertation UniBw Muenchen, LRT, 1987
3. Strategic computing. New-generation computing technology: A strategic plan for its development and application to critical problems in defense. DARPA AD-A141 982, 1982
4. Gilmore, J. F.; Semeco, A. C.: Knowledge-based approach toward developing an autonomous helicopter system. Opt. Eng. 25 (1986) 415-27
5. Dickmanns, E.D.: 4-d dynamic scene analysis with integral spatio-temporal models. Proc. 4th international symposium on robotics research (ISRR), Santa Cruz, 1987
6. Dickmanns, E. D.: 4-d-Szenenanalyse mit integralen raum-/ zeitlichen Modellen. 9. DAGM-Symposium Mustererkennung, Braunschweig 1987
7. Kalman, R. E.: A new approach to linear filtering and prediction problems. Proc. IFAC-Congr. 1960, Moscow, Vol. 1, 481-192
8. Popper, K. R.; Eccles, J. C.: The self and its brain - an argument for interactionism. Berlin, Heidelberg, New York: Springer 1977
9. Graefe, V.: Two multi-processor systems for low-level real-time vision. In: Brady, M.; Gerhardt, L. A.; Davidson, H. F. (eds.) Robotics and artificial intelligence. Berlin, Heidelberg, New York: Springer 1984
10. Dickmanns, E.D.; Eberl, G.: Automatischer Landeanflug durch maschinelles Sehen. Jahrbuch 1987 der DGLR, 294-300
11. Luenberger, D.G.: An introduction to observers. IEEE Trans. Autom. Control 16 (1971) 596-602
12. Kailath, Th.: Linear systems. Englewood Cliffs/NJ: Prentice Hall 1980

Received January 26, 1988