REMOTE CONTROLLED MODEL AIRCRAFT AS SENSOR PLATFORM FOR URBAN MODELLING

Wolfgang von Hansen **Ulrich Thönnessen**
FGAN-FOM Research Institute of Optronics and Pattern Recognition
Gutleuthausstr. 1, 76275 Ettlingen, Germany
`wvhansen@fom.fgan.de`

ABSTRACT

Several types of commercial sensor platforms exist for data acquisition in urban terrain. But when it comes down to purchasing data for a particular purpose, a mismatch between level of detail, age of information and price of data emerges. High resolution images that have an age of less than one week are not available at any reasonable price. Such data will be important to generate up to date city models with a high level of detail for various applications. Current research begins to focus on the computation of sensor trajectories from video sequences in real time which will lead to fast “structure from motion” and finally to the fully automatic generation of city models in very short time. In order to acquire suitable input data for such a processing chain, we have chosen to equip a small remote controlled model aircraft with a video camera. This paper will give a detailed overview of the system, report some experiences we had with a first setup, outline the processing chain and present preliminary results we obtained from the first video sequences.

1 INTRODUCTION

Several types of commercial sensor platforms exist for data acquisition in urban terrain. Typically these are satellites, airplanes and more recently also vehicles that gather data either at a more or less regular time interval or on demand. But when it comes down to purchasing data for a particular purpose, a mismatch between level of detail, age of information and price of data emerges. Optimizing one or two of these properties inevitably makes the remaining worse. High resolution images with a pixel size of approx. 10 cm, covering a district of a town, that show both roofs and facades of buildings, and that have an age of less than one week are not available at any reasonable price. Such data will be important to generate up to date city models with a high level of detail for various applications such as decision making for city planning or augmented reality systems. Current research begins to focus on the computation of sensor trajectories from video sequences in real time. (Nistér, 2004) This will lead to fast “structure from motion” and finally to the fully automatic generation of city models in very short time. In order to acquire suitable input data for such a processing chain, we have chosen to equip a small remote controlled model aircraft with a video camera. This paper will give a detailed overview of the system, report some experiences we had with a first setup, outline the processing chain and present preliminary results we obtained from the first video sequences.

2 SENSOR PLATFORM

2.1 The first try: Wingo Porter

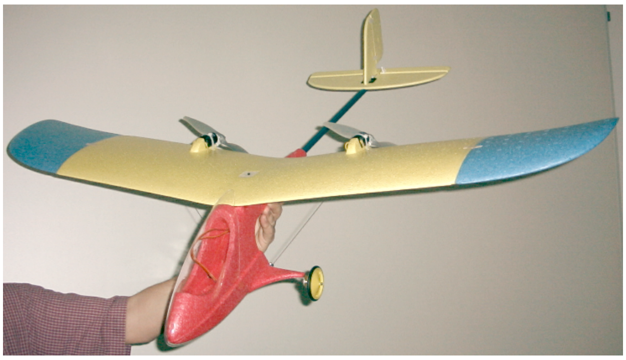

Figure 1: Styrofoam model Wingo Porter.

The focus for our first model had been based on two major points, payload capacity and size. Rough calculations yielded an anticipated payload of not more than 500 g. The smallest model aircraft that could carry this much according to its technical specifications (Tab. 1) was the “Wingo Porter” (Fig. 1). This is a styrofoam aircraft with two engines, an open loading bay for the camera and other equipment and a wingspan of 1.3 m.

Even though the design seems perfect at a first glance, it soon turned out that the Wingo Porter had not been the best choice:

- Having two airscrews mounted on the wings instead of one on the nose leads to a slow response to steering commands because the steering gear is not inside the main airflow. This makes it difficult to operate the aircraft and avoid obstacles like e. g. trees when there is some wind.
- Tests with dummy weights revealed that the engines were to weak even for safe flights even with a lower payload of about 0.3 kg. Stronger engines would be a solution but can not be mounted because this leads to unwanted additional stress to the wings.
- Styrofoam wears out under strain. The wings became noticably more flexible after a few flights so that this material is not fit for long term use.

Figure 2: Model of Piper J-3.

Table 1: Specifications of Wingo Porter

Wingspan	1.3 m
Length	1.2 m
Basic weight	1.2 kg
Payload	0.5 kg
Propulsion	170 W electric
Flight duration	7–10 min
Material	styrofoam

Table 2: Specifications of Piper J-3

Wingspan	2.1 m
Length	1.3 m
Basic weight	1.8 kg
All up weight	3–4 kg
Propulsion	700 W electric
Flight duration	15 min
Materials	wood / textile covering

- Styrofoam breaks easily on harsh landings. Even though it can be glued quite quickly, it remains instable.
- During a test flight the cockpit broke apart in mid-air because the wing struttings are attached to the landing gear, leading to strain during a flight maneuver.

The risk of a crash landing that might damage the camera was too high so that we stopped further testing of this model with the full load. In the future, we will use the Wingo Porter as a secondary platform to test single components of the sensor equipment.

2.2 Second chance: Piper J-3

After the disappointment with the Wingo Porter we now switch to a larger but more robust model aircraft. The new aircraft will be the model of a Piper J-3 made of wood and textile covering and with a wingspan of 2 m. (Fig. 2) For the technical specifications see Tab. 2. The sensor equipment can be mounted inside the hull and therefore is less vulnerable to damage in the case of a crash landing. We anticipate this model to be strong enough for a payload of 1–2 kg because it is frequently used to tow model sailplanes into the air. First test flights will commence in Spring 2005.

3 SENSOR EQUIPMENT

Since model aircrafts are not built to carry any significant payload only small and lightweight devices can be used. The sensor equipment that we have chosen will be described in this section along with comments on our first experiences. According to the planned usage, the emphasis is on video acquisition and storage. In all cases, the power supply for each device should be seperated from that of the model aircraft in order to avoid interference with the high frequencies from the remote control.

3.1 Tiny camcorder for onboard recording

The technically most simple solution to acquire and store a video stream is to use a camcorder that contains the CCD sensor as well as a storage device and a power supply. This simplifies mounting of the camera because only an installation frame is necessary; there is no need for any wiring for data transmission or power.

The consumer market offers a wide range of camcorders, but most of them weigh at least 0.4 kg. Minimum weight and size are mainly due to the type of recording device which often is a tape drive. Finally, we made a decision for the Panasonic SV-AV 100 (Fig. 3), a tiny video camera that records on SD-RAM and weighs only 185 g but nevertheless features the full video resolution of 704x576 @ 25 fps (Tab. 3).

The basic idea is to record the video onboard instead of transmitting it via a downlink. Advantages are no loss in data quality due to digital-analog-digital conversion or transmission errors / dropouts. This camera seemed optimal because it is very compact and has both power supply and recording media in the same case.

There are a number of points that can be made about camcorders in general and the SV-AV 100 in particular:

- All camcorders at the lower end of the price range feature only interlaced CCD readout. Progressive scan, i. e. non-interlaced video, is only available for expesive semi-professional cameras. Since a small model aircraft is a rather unstable platform, one has to deal

Figure 3: Camera Panasonic SV-AV 100. Size comparison to a hole punch for A4 paper.

Table 3: Specifications of Panasonic SV-AV 100

Size	33×90×65 mm^3
Weight	185 g
Sensor	$1/6''$ CCD
Optics	2.3–23 mm
Video format	704×576 pixels, 25 fps, MPEG-2
Memory	512 MB (10 min)

Figure 4: Diversity receiver.

Table 4: Specifications of TV camera module

Image sensor	$1/4''$ CCD
Effective pixels	512×582
Resolution	380 TV lines
Signal system	PAL
Lens	f=3.6 mm, F=2.0
FOV	68°
Size	26×22×31 mm^3
Weight	30 g

with the interlace effects in the video in an appropriate way.

- The SV-AV 100 uses MPEG-2 compression to store video streams at the highest possible resolution.[1] The image quality differs noticably from videos recorded on DV-AVI due to the strong compression.
- The maximum recording time of the SV-AV 100 is limited by the memory size of the SD-RAM card. It is shipped with 512 MB which allows for 10 min, whereas larger SD-RAM cards of 1 GB are already on the market.
- The small size and simple form of the camera made it easy to build a simple mounting frame.

Currently, there are not many competitors to the SV-AV 100 at the market, so that a real comparison with other products is not possible. The handling is very fine, but the image quality could be improved.

3.2 Video downlink for external recording

As an alternative we also have built up a video transmission downlink which allows to visualize the images of the on-board camera on a monitor and/or recording of data from camera modules with no such capabilities.

3.2.1 Transmitter and diversity receiver The video link consists of a small PLL transmitter with a transmission power of 10 mW and a transmission frequency of 2.4 GHz (Fig. 5). In order to enhance transmission quality, an additional booster with a transmission power of 200 mW is used. According to the manual, this brings the range from 300 m without the booster up to 1300 m. Transmitter and booster weigh without power supply a total of 13 g.

The second part of the video link is a diversity receiver (Fig. 4) which is designed to reduce the disturbance of video caused by reflections of the signal on walls. The diversity box receives all four incoming signals so that the receiver can evaluate which signal is most suitable and automatically switch to that antenna.

First tests show an excellent image quality over this video link which most probably is due to the diversity receiver. Only very few glitches are contained in a video stream under normal operating circumstances. The image quality is similar to that of the camcorder and might even be better, if a high quality sensor module and a good video recorder would be used.

3.2.2 TV camera module The first camera module is a mini board camera with a color CCD (Tab. 4). Its image quality is comparable to that of the SV-AV 100 camcorder even though it has a lower nominal resolution. One disadvantage is that it only has got an analog composite signal, so that either a digital recorder or a framegrabber must be used for input to the computer. On the other hand, a video monitor can be directly attached to the receiver. The main purpose of this camera is to provide the pilots sight during the flight to adjust altitude or course while the camcorder records the main data for evaluation. But as the image quality is surprisingly good, it might as well serve as the only sensor for data capture.

3.2.3 IR camera module The second option is a small thermal imager which allows to create city models with

---

[1] MPEG-4 is used for lower resolutions.

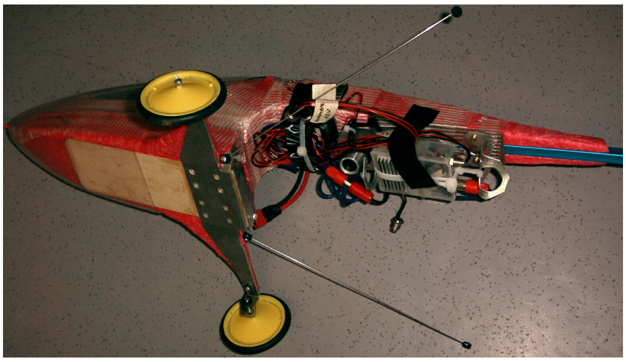

Figure 5: TV camera module and transmitter mounted on Wingo Porter. The transmitter is contained in the transperent box at the rear end with the camera lens located directly to the front of it.

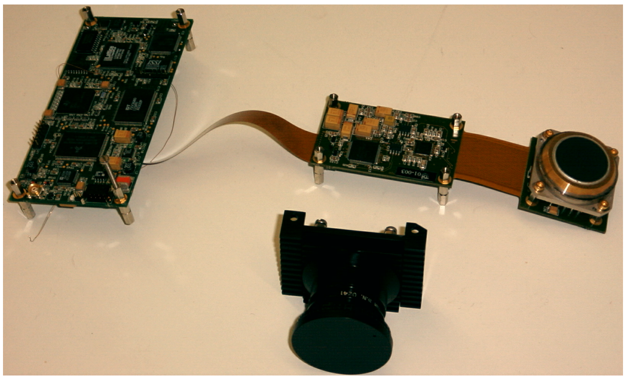

Figure 6: Small IR-Camera with sensor head and additional circuit boards.

thermal signatures (Fig. 6). These could e. g. be used to demonstrate heat loss from badly insulated buildings or for urban activity monitoring. Since the thermal imager does not come with onboard recording capabilities, it must be used with the video downlink. One of the most important features of this particular uncooled IR camera is the very low integration time of only 4 ms. Longer integration times lead to motion blurring due to platform movement that lowers the image quality drastically.

When used along with the SV-AV 100, the camcorder's zoom can be set according to the focal length of the thermal imager so that both fields of view match and data fusion is possible. However it has not yet been tried to fit both cameras into the small aircraft at the same time.

Table 5: Specifications of the IR camera module

Image sensor	a-Si focal plane array
Pixels	320×240
Pitch	45 μm
Thermal resolution NETD	$<$ 120 mK @ 300 K
Spectral range	8–14 μm
FOV	41°×31°
Weight	250 g

3.3 GPS receiver

One of the main goals of the video processing is to recover the sensor trajectory from the video stream alone. The advantage would be that no additional cross-referencing of navigation sensor and camera is necessary. On the other hand, absolute coordinates are required for georeferencing. For this purpose we have tested the small GPS receiver Garmin Geko 301 which has size and shape of a cellular phone, weighs 85 g and is originally made for outdoor trips, having the advantage of ruggedness and water resistance. Besides standard GPS functionality and the ability to record track points every two seconds it also features a magnetic compass and a barometric altimeter.

First tests indicate that the position sampling every two seconds is somewhat coarse, but the shape of the ground path can be recognized and seems plausible. The height readings – both with and without the altimeter enabled – suffer from drift effects and are unusable. A different problem is the synchronization of the GPS and the video camera because there exists no common time signal. A solution could be to record the GPS receivers display of the clock but there still would be an unknown time offset left.

As mentioned above, the acquisition of the flight path has only a low priority. If future results show that a precise external measurement of the sensor trajectory would be advantageous, a better GPS system based on high quality modules is necessary.

3.4 Planned equipment for the future

In the future, we plan to mount the camera on a remote controlled tilt platform so that the viewing direction can be adjusted during the flight.

4 PROCESSING CHAIN

In the preceding sections both the platform as well as the sensor equipment has been presented. The task to be solved is first to reconstruct the sensor trajectory and some 3D points from such sequences and then to model the urban environment. In this section, the general workflow will be given along with some results for the first processing steps. (Thoennessen et al., 2004)

The generation of a geometric model from a video sequence requires the knowledge about the pose of the cameras as well as their calibration parameters. If these are not known, such as for a model aircraft without navigation sensors and stock camera modules, they have to be computed from given point assignments. Such a task – simultaneous computation of inner and outer camera parameters when no initial values are known – is commonly referred to as auto or self calibration (Hartley and Zisserman, 2004). An approach to self calibration and creating both model and texture from only one data source is outlined.

It is well known among photogrammetrists and in the computer vision community, that it is possible to retrieve structure from motion. Several images taken from different

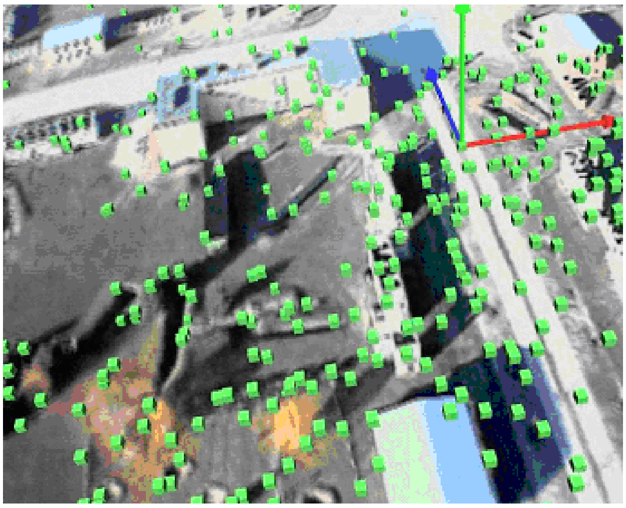

Figure 7: Points generated from interest operator.

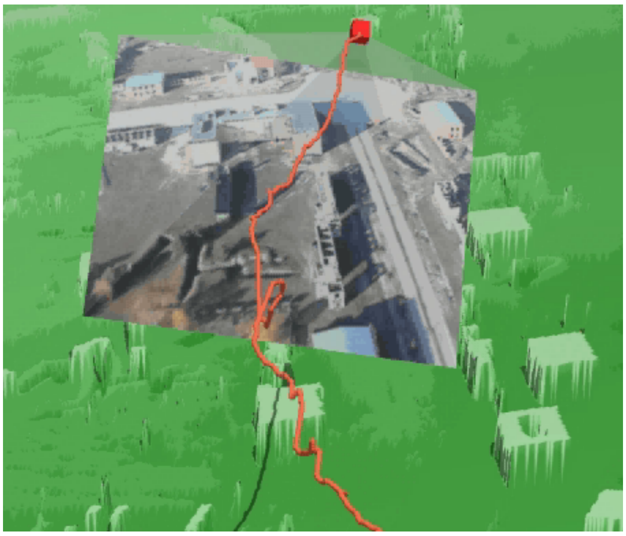

Figure 8: Flight trajectory recovered from video sequence.

viewpoints or the video stream of a moving camera provide enough information to reconstruct both the sensor pose and trajectory along with calibration parameters for the camera, and the 3D-scene viewed by the camera. In (Hartley and Zisserman, 2004) many aspects are covered in detail so that only a brief overview will be given here.

Suppose an object point is imaged by one camera so that the coordinates of its image are known. If a second camera takes an image of the same scene, what is known about the location of that particular object point in this image? It turns out that its position is restricted to lie on a straight line – namely the image of the viewing ray of the first camera to the object point. This line is called the epipolar line and its parameters for any point are defined by the relative pose of the two cameras and their inner parameters (e. g. the focal length) which describe the image formation inside the camera. Every known pair of corresponding points thus yields one constraint. A total of at least seven corresponding points between both images are exploited to compute the fundamental matrix which expresses their mathematical relation.

To generate the full sensor trajectory for a long image sequence the processing chain can be divided into three parts: Point tracking, initial projective reconstruction and complete reconstruction. The first part is to detect suitable image features and track their position through the sequence. The main reason is that in a typical video sequence the camera shift in space is only small from one frame to the next, but in order to retrieve 3D-information, different object movement due to different depths must be visible in the images. On the other hand, since neighboring images do not change much it is easy to follow one object point through the sequence. Initial track points are generated using a point interest operator like e. g. the Förstner operator. Tracking of such points through the sequence is accomplished by point matching between image frames where the cross correlation coefficient of the region surrounding the points serves as similarity measure. As an additional constraint for point displacements it can be exploited that two neighboring images are linked by a planar projective transform. Point tracking is the crucial part of the algorithm because any error introduced here could lead to a wrong result later on. Therefore robust schemes like e. g. RANSAC must be used for outlier detection.

Once all point tracks are completed, an initial reconstruction can be carried out. This consists of the creation of a coordinate frame for two cameras and the computation of the coordinates of some 3D points in that frame. Two images are selected such that they are sufficiently apart to form a proper stereo base, but still are connected by at least seven points so that the fundamental matrix can be computed. The two camera projection matrices can be recovered from the fundamental matrix – but not uniquely. The first camera can be chosen arbitrarily and for the second camera there are still four degrees of freedom left. What can not be determined from the images alone are absolute location and orientation of the two cameras and their calibration. The whole coordinate frame defined in this way differs from a metric coordinate frame by a projective transform. However, it already is possible to compute 3D-coordinates of the object points in the projective coordinate frame by triangulation of corresponding image points.

The two remaining tasks are the calibration of the cameras which also yields the transform from the projective to a metric reference frame and the inclusion of all other images into the model. With the introduction of constraints on the so far unconstrained inner parameters – e. g. focal length is constant for all images – it is possible calibrate the cameras. This has been done using the approach of the absolute quadric; a virtual object which is located on the plane at infinity. Its projection into the images is linked to the calibration parameters of the cameras. Using constraints, the absolute quadric can be recovered, where an appropriate parametrization directly results in both camera calibration and the transform to a metric reference frame.

Using the already known object points and corresponding image points, the camera pose can be estimated for other

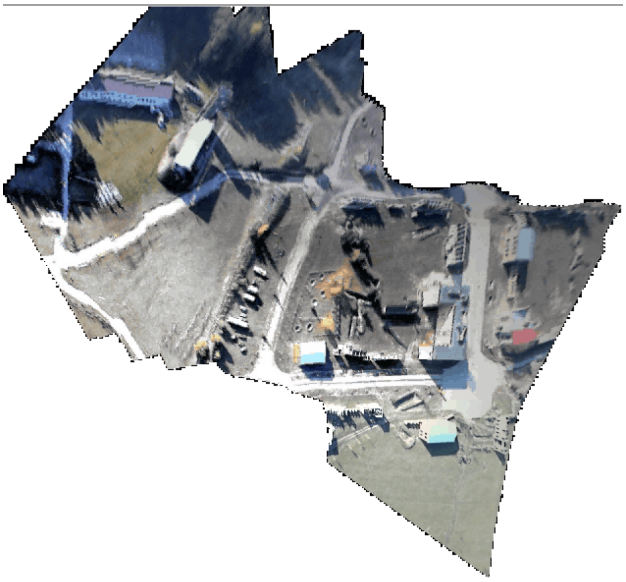

Figure 9: Mosaic created by projection of image sequence onto the ground.

images through resection in space. With the additional images there are more corresponding pairs of image points so that their 3D-coordinates can be found via triangulation. Repeating these two steps it is possible to cover the complete video sequence. With known camera poses and parameters, detailed 3D-structure can be generated through a dense stereo matching. Texture information is readily available as the complete viewing geometry is known.

The advantage of this approach is that camera calibration is not necessary to generate a result and that any simple camera can be used. The cameras we intend to use have a low geometric quality compared to true photogrammetric devices because of weight restrictions. Nevertheless it is anticipated that in combination with real time processing, rapid generation of 3D-models is possible with such a simple sensor system.

5 RESULTS

We already dispose of video sequences taken by a different system. The basic question was whether or not self calibration would work on such imagery at all. Therefore we have used the commercial program MatchMover for some test runs. In most cases the program came to plausible results, but some instability could be noted. If the constraints on the internal camera parameters were too loose, the solution would not converge. This is probably is due to the restricted movement of the sensor platform which leaves some of the camera parameters correlated. (Sturm, 2002)

Fig. 7 shows one frame of such a sequence with marked interest points. The reconstructed sensor trajectory is shown in Fig. 8 along with a projection of the frame onto the ground. If all frames of the whole flight path are mapped onto a surface, an image mosaic as shown in Fig. 9 can be produced. An application could be rapid terrain mapping e. g. for planning of rescue operations after a natural disaster.

These are only preliminary results but they already show the capabilities of a rather inexpensive airborne platform coupled with automatic image processing techniques.

6 CONCLUSIONS

We have presented a simple and inexpensive sensor platform that we intend to use for urban modelling purposes. One of the experiences in the beginning was that a simple model aircraft might seem appropriate from the technical specifications but actually it is far better to choose a slightly oversized model made of robust materials.

Regarding data acquisition, there currently exist two techniques: Onboard recording and transmission via a video downlink. The diversity receiver seemed to be the key to an excellent quality of the transmitted videos. However, it can not be said without more in depth tests which system will be better in the long run. For special cameras like the IR camera, a video link definitely is an alternative because the transmitter weighs much less than an additional video recorder on board the aircraft.

Probably the most unique feature of our system might be the IR camera that allows to collect thermal information both at day and night time. No tests have been conducted so far, but we anticipate that this extra information could be very valuable for urban monitoring purposes.

Finally, an overview of the processing chain along with some preliminary results has been shown. Once the system is running, the main work will focus here with the emphasis on building reconstruction from the image sequences.

ACKNOWLEDGEMENT

We would like to acknowledge the help from our colleague Markus Hebel, who created nice animations and the mosaic from the MatchMover output.

REFERENCES

Hartley, R. and Zisserman, A., 2004. Multiple View Geometry in Computer Vision. 2. edn, Cambridge University Press.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 26, pp. 756–770.

Sturm, P., 2002. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. In: Image and Vision Computing, Vol. 20, pp. 415–426.

Thoennessen, U., Gross, H. and von Hansen, W., 2004. Enhanced situational awareness in mout scenarios. In: 6th Joint International Military Sensing Symposium, Dresden, Germany.