Evaluating Influence of Nonlinear Disturbances on Image Registration Based on Virtual Forces

Claudius Stern\textsuperscript{a}\textsuperscript{*}, Lisa Kleinjohann\textsuperscript{b}

\textsuperscript{a}Opsolution GmbH, Wilhelmshöher Allee 273a, 34131 Kassel, Germany
\textsuperscript{b}C-LAB University of Paderborn, Fürstenallee 11, 33102 Paderborn, Germany

\textsuperscript{*Corresponding Author: Claudius.Stern@gmx.net}

Abstract

In case of a disaster it is of utmost importance to obtain an overview of the affected area in order to coordinate rescue operations. Recent research and development has led to affordable unmanned aerial vehicles (UAVs) and enabled (semi-)autonomous surveillance and mapping of a disaster areas by UAVs. Mapping an area from many images incorporates image registration. Commonly used registration techniques expect static images to be registered. However, while mapping a disaster area, it is very likely that objects like vehicles or persons are moving on the ground, thus leading to dynamic scenes.

We formerly introduced a fast and robust approach to register large amounts of successively arriving images, e.g., from UAVs. The approach only uses rigid-body transformations, but the usage of virtual forces enables the registration to tolerate small perspective distortions. This paper investigates the performance of our approach when applied to dynamic scenes, which are represented by nonlinear disturbances within image correspondences. We show that our approach not only tolerates small perspective distortions but also is able to register dynamic scenes.

Keywords: image registration, virtual forces, UAV.

1. Introduction

Once a disaster occurs, it is of foremost interest for the first responders to get an overview of the affected area. This enables the rescue forces to enter the area and reach casualties more effectively. At present, such an overview often is gathered using, e.g., a manned helicopter. A human observer watches the scene and informs the responding forces of possible casualty positions.

Over the last decade, much research has been done in the area of unmanned aerial vehicles (UAVs). Especially, with ongoing development, UAVs have become affordable and thus a viable alternative to the use of manned aerial vehicles.

Nevertheless, often manual control of the flight or a manual inspection of the images taken is necessary. Recent advances enable swarms of UAVs to autonomously explore a previously unknown area, as shown by Rasche et al. [9].

The manual inspection of the often shaky live video streams delivered by several UAVs is a time consuming task. Hence, rescue forces should be supported in this process in order to reduce their workload. One approach is to register the successively received images into an overview map, which is provided to the rescue forces. In previous work, we presented a novel approach for image registration using virtual forces [10]. This paper concentrates on the evaluation of the quality regarding the registration of dynamic scenes, where object movement on the ground takes place. In contrast, other registration approaches for rescue scenarios are usually evaluated in static environments only.

For the evaluation of our approach in dynamic scenes we introduce a simulation environment, which induces errors like different types of ground movements would.

The remainder of this paper is organized as follows. In Section 2 related work in the field of autonomous mapping is presented and other approaches using UAVs in disaster scenarios. Image registration based on virtual forces is compared to other approaches. In Section 3 our image registration approach based on virtual forces is presented and in Section 4, the resulting registration quality is discussed. The generation of synthetic data is explained and the method how nonlinear disturbances are induced. In Section 5 the presented results are validated using images from real flight.
experiments. Finally, Section 6 concludes this paper and gives a future perspective regarding the use of virtual forces for image registration.

2. Related work

Generating overview maps of an area captured by a down-looking camera has been studied for more than two decades. An early work was, for instance, presented by Wiesel [14] in 1985. With continuing progress, many different approaches have been developed for various application scenarios.

Verhoeven et al. [13] present an approach to automatically generate a 3D terrain visualization precisely mapping archaeological sites. They use a camera mounted on a heli-kite—a moored balloon with a stabilizing kite—to take bird’s-eye images. Afterwards, the images are analyzed and converted into a textured 3D point cloud. Together with manually selected ground control points they achieve a precision of less than 13 cm positional deviation and less than 30 cm deviation in height. Like most image registration approaches, Verhoeven et al. need to capture the whole terrain in advance, i.e., for a successful registration, all images are needed beforehand. Besides that, the computation time of approximately one day is much too long in face of a rescue scenario.

Mapping an area with a down-looking camera does not necessarily involve an aerial vehicle. Many approaches in this area have been developed for underwater applications. One example is the work of Elibol et al. [5], where an autonomous underwater vehicle is utilized to map large areas of the seabed. Small amounts of images are processed directly for navigational purposes. Nevertheless, for generating the resulting map all images are necessary.

In a rescue scenario, these methods are not applicable. Since such scenarios are characterized by their time critical nature, overview maps have to be generated nearly instantaneously.

Most commonly used image registration approaches assume a static scene, even more if they have their origins in image stitching and therefore expect the images to be taken at almost the same time.

However, in contrast to archaeological sites, disaster areas are very dynamical and even major changes within a short period of time are possible. Especially, the scene has to be captured multiple times while the mission is in progress. Assuming a large disaster area, this can only be done successively.

Caballero et al. [2] present an approach to build an overview map using a small UAV. They focus on the use of image registration to estimate the position of the UAV itself, but by exploiting the positional calculations, a map can be constructed, too. The authors interlink the images’ positions by stochastic functions, which incorporate the position and an error estimation. That enables a correction of the map, if a previously mapped area is captured again.

In comparison to the approach of Caballero et al., we are also capable of dynamically changing the map. Like in their approach, we use additional information of former images to correct the map. In contrast to their approach, we do not need to store connection information for each predecessor of an image, which is very memory consuming.

3. Image registration with virtual forces

Using virtual forces for image registration has been introduced in [10]. Here, a short recapitulation is given.

Images are regarded as virtual masses connected by virtual forces. The masses are regarded to be two-dimensional. They are movable, rotatable and stretchable by means of aspect-ratio-keeping scaling. All operations are controlled by the system of forces attached to the masses. The dimensions of the masses are related to the image’s size.

Feature detection results in a set of key points which have positions on the image. The key point positions are used as force application points. After feature matching, corresponding key points are connected by virtual forces. A perfect registration is reached, if all forces have a length of zero. Therefore, the goal of the presented registration algorithm is to minimize the overall length of all virtual forces.

To compute the forces’ effect, a resulting force application point has to be calculated. Let \( F \) be the set of all virtual forces \( f_i \in F \). For each mass, a center of gravity (CoG) is calculated out of the force application points. Let \( P \) and \( Q \) be sets of corresponding points, such that \( \overrightarrow{q_i} = \overrightarrow{p_i} + f_i \) and thus, let \( n = |P| = |Q| \). Then, e.g.,

\[
\overrightarrow{CoG_P} = \frac{1}{|P|} \sum_{\overrightarrow{p_i} \in P} \overrightarrow{p_i}
\]

is the calculated force application point for mass \( m_P \). The CoG point for mass \( m_Q \) is calculated accordingly. The translation vector between those two masses can be calculated using the mean of all individual forces, which is the difference of the respective CoG points as stated below.
The CoG points are used not only to calculate the translation vectors but also as parameters of heuristic functions to estimate rotation and scale factor, as formerly described in [11].

\[
\frac{1}{n} \sum \vec{p} = \frac{1}{n} \sum \vec{q} - \vec{p} = \frac{1}{n} \sum \vec{q} - \frac{1}{n} \sum \vec{p},
\]

4. Quality of registration

The quality of a registration process is usually measured using the remaining distances between corresponding key points [15]. As shown in previous work [12], rather good registration results could be generated using the force based approach.

In addition to the real-world data, generated synthetic data is used to evaluate the algorithm's capabilities, especially large amounts of objects to be aligned.

Another point for using synthetic data, is the possibility to generate data, which incorporates only one specific disturbance type at a time. Hence, the approach can be evaluated for its robustness against different types of disturbances.

One significant error type is caused by the camera movement itself. There are two assumptions made, which are not met exactly: the area to be captured is flat and the camera's view axis is orthogonally aligned to the surface. If one of these assumptions is not met, the captured image becomes perspective distorted. The induced errors can be corrected using linear equations (the homography).

Regarding dynamic scenes, another error source for registration is the movement within the scene. This gets even worse when both the camera and objects move. In the scenario focused in this paper, camera-equipped UAVs take images of the ground where objects—rescue forces, victims, etc.—move. Regarding image registration, valid matching pairs of key points are displaced from one image to another. If the movement is fast, the displacement is large and can be filtered out. If the movement is small, the displacement is also. Nevertheless, a non-predictable, non-linear error is induced, if that key point pair is used for registration.

These error types are integrated in our simulation environment—each can be separately controlled—to evaluate the robustness of the force based image registration.

4.1 Generating synthetic data

Using synthetic data has another advantage: the possibility to check against a ground truth. In the case of this paper, the ground truth consists of camera-view areas on a common plane. The camera is held almost orthogonal with respect to the plane, i.e., it has a random deviation angle in the range of ±5 degrees. A camera looking skewed to the plane will have a perspectively distorted rectangle as view area.

To simulate a UAV's flight, overlapping camera views are generated, which are connected with perfectly matching key points. After creating the ground truth, every single image is passed as a rectangle to the registration algorithm.

Figure 1 depicts an exaggerated illustration of perspectively distorted rectangles—generated as ground truth—and the respective ``camera images'' passed to the registration algorithm.

The both camera views in the ground truth are overlapping. Within the overlapping area the images are connected by generated key points. In Figure 1 this is depicted by four key points, forming a square. The extra line is introduced for better visibility of the effect. In the ground truth, the key points mark the corners of a perfect square. To deliver rectangular images to the registration algorithm, the camera views are “undistorted” in a backwards manner from perspective to rectangular. With the outer bounds changed, the inlying key point positions are also changed. The figure illustrates how the square gets distorted as a result. Thus, without using perspective transformations, a perfect registration is not possible.

Nevertheless, a registration is possible, even without an accumulating error. Figure 2 shows the error evaluation of 10 independent runs, with 1000 virtual masses.
Each mass was connected with its predecessor, using 30 virtual forces. Even though perspective distortion of the camera views has been used, the errors do not increase over time and stay well below one pixel.

To take account of the circumstances in a rescue scenario, ground movements have to be considered, too. A moving object in an image is a displacement of a part of the image over time, i.e., if key points are located in this area, they are also likely to move from one image to its successor.

Regardless of how small those movements are, an error is introduced when they are considered for registration. The next section describes the method of how to induce those errors and discusses the effects on our force-based image registration approach.

4.2 Nonlinear disturbances

In Figure 3, two different types of nonlinear disturbances are depicted. Figure 3 (a) shows randomly distributed individual movements. In a rescue scenario, these movements could be, e.g., victims in panic. However, fleeing people might choose a similar direction. So, Figure 3 (b) depicts a group of key points moving in the same direction. Thus, they are displaced by the same vector.

The randomly chosen displacement vectors have a length of up to 3% of the largest image dimension, e.g., up to 19.2 pixels, assuming a VGA-resolution camera. They are chosen either individually for key points or once per connected mass pair. In both cases, the disturbance is applied to randomly selected key points until the user-given threshold of disturbance percentage has been reached.

4.3 Results

According to the two types of disturbances (induced by individual or group movements), we conducted tests for each type with increasing percentage of disturbance. The tests consisted of 10 independent runs with 100 virtual masses, each of them connected by 50 virtual forces to its respective predecessor. The resulting displacement errors are depicted in Figure 4. The diagram lines represent the average displacement error of 10 independent runs with the according amount of induced disturbance.

In case of random movements, only a slight increase of displacement errors can be observed. With increasing percentage of induced disturbance, the magnitude of displacement errors only slightly increases.

As can be seen easily by comparing both diagrams, group-wise movements cause a higher magnitude of displacement errors. However, even when 40% of the key points are moving, the errors do not increase significantly.

Nevertheless, the most important fact is the constant trend of the error development over time. Despite of the significant amount of disturbed key point positions, the introduced errors are tolerated and do not lead to an
Besides the registration quality, the even more important factor in a rescue scenario is the computation time needed to build the overview map. It has been previously shown, e.g., by Pesti et al. [8], that building an orthographic view is possible even without using expensive equipment. The drawback of commonly used approaches is the massive computational effort, which has to be spent to generate a rather exact map. However, an overview map dedicated to be used in a rescue scenario does not necessarily have to be overly exact. Usually—of course depending on the actual operation—errors in meter-scale are tolerable. Hence, a fast approach is necessary, even if the quality slightly suffers.

Figure 5 shows the calculation times for ten independent runs of force-based image registration. The mean calculation time was 10.5ms with slight increase over time. It is comparable with the homography estimation in classical image registration, e.g., in a similar situation, RANSAC will need approximately 50 ms per image pair [4].

5. Evaluation with real images

In the course of the funded research project SOGRO (immediate rescue in a large-scale accident with mass casualties), several flight experiments were conducted, regarding the use of UAVs to generate an overview of a disaster site. In Figure 6, a comparison of two mappings are shown. The source images were taken at Germany's largest rescue exercise ever since the foundation of the Federal Republic of Germany [7]. All the images were taken approximately at the same height and from an electronically stabilized camera platform. As the overflown area was practically flat, the images should contain almost rectangular areas of the ground. Nevertheless, much ground movement took place during the exercise, e.g., rescue forces, vehicles, made-up victims, etc. In this particular map excerpt, movements take place on one side of the individual images only. This leads to an imbalance of nonlinear disturbances, making image registration even harder. The upper map was built using RANSAC [6] to estimate homographies between each pair of images. The lower map was built using virtual forces for registration. For better visibility of the effects, a grid has been overlaid the first and last image.

As easily can be seen, the last image of this sequence was heavily distorted when concatenated homographies were used. The problems caused by successively registered images, through concatenating homographies, were already described by Brown and Lowe [1]. Often, this problem is circumvented by using bundle adjustment, which implies the necessity to provide all images at one time.

Our approach is to maintain the connections between
the images flexible, to be able to correct the alignment later. We use our force system to maintain the images’ positional parameters. Newly added images contribute to the force system of our map and the masses (images) move until an equilibrium is reached again. Furthermore, by interconnecting images not only with their direct predecessors but also with other images that are overlapped, the map gets enhanced over time. The remaining small displacement error at the right side of the map has been caused by a systematic distortion due to an uncalibrated camera. Nevertheless, a complete map of the area could be built.

The partly similar approach by Caballero et al. [2] uses stochastic functions to maintain the positional parameters between images. In contrast to this approach, we do not need to consider every predecessor of an image but only the ones with actually overlapping areas. Hence, our approach has a better memory performance. Nevertheless, merging these two approaches could be of future interest.

6. Conclusion and outlook

In this paper, we presented an evaluation of our formerly introduced approach regarding its capability to be used to register dynamic scenes, e.g., in a rescue scenario. Besides the motion caused by a UAV itself, images of such scenes incorporate additional movements from one image to its predecessor caused by, e.g., fleeing people, moving vehicles, etc. Those unsystematical motions cause additional nonlinear errors within corresponding key point positions through displacements not related to the UAV’s motion.

Earlier, we showed that our image registration approach is robust against small perspective distortions, despite the use of rigid-body transformations. In this paper, we showed that our approach even is robust against additional nonlinear distortions within the correspondences. The results achieved
using images from real UAV flights over a rescue-exercise scenario validate the results of our simulated tests.

Future work aims at integrating full homographies into our approach like Caballero et al. [3] did, but in an iterative manner. This should preserve the advantage of lower memory consumption using virtual forces for image registration and should improve the quality. Another aim is to integrate a fast 3D processing, e.g., to make use of a previously known 3D terrain model.

Acknowledgment

Parts of this work were funded by the German Federal Ministry of Education and Research in the course of the research project SOGRO (immediate rescue in a large-scale accident with mass casualties), funding codes 13N10162 to 13N1016.

References


