UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

Augmented reality over video stream
acquired from UAVs for operations support

Tesis Doctoral

Susana Ruano Sáinz
Licenciada en Matemáticas e Ingeniera en Informática
2018

DEPARTAMENTO DE SEÑALES,
SISTEMAS Y RADIOCOMUNICACIONES

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

**Augmented reality over video stream
acquired from UAVs for operations support**

Tesis Doctoral

Autora:

Susana Ruano Sáinz

Licenciada en Matemáticas e Ingeniera en Informática

por la Universidad Autónoma de Madrid

Directores:

Carlos Cuevas Rodríguez

Doctor Ingeniero de Telecomunicación

por la Universidad Politécnica de Madrid

Guillermo Gallego Bonet

Ph.D. in Electrical and Computer Engineering

por el Georgia Institute of Technology

TESIS DOCTORAL

**Augmented reality over video stream
acquired from UAVs for operations support**

Autora: Susana Ruano Sáinz

Directores: Carlos Cuevas Rodríguez y Guillermo Gallego Bonet.

Tribunal nombrado por el Sr. Rector Magnífico de la Universidad
Politécnica de Madrid, el día de de 2018.

Presidente: D.

Vocal: D.

Vocal: D.

Vocal: D.

Secretario: D.

Realizado el acto de defensa y lectura de la Tesis el día de
de 2018 en

Calificación:

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

Resumen

La realidad aumentada (RA) se ha convertido, gracias a los últimos avances tecnológicos, en una de las disciplinas con mayor crecimiento. El potencial de la RA propicia su estudio, no sólo para nuevos dispositivos específicos como gafas y cascos, sino para cualquier dispositivo dotado de una cámara. Con esta idea, Airbus impulsó un proyecto de innovación, Situational Awareness Virtual EnviRonment (SAVIER), para incorporar la RA a sus estaciones de tierra, mejorando así el flujo de vídeo capturado desde la cámara de sus vehículos aéreos no tripulados (UAV). Esta tesis se enmarca en ese proyecto y explora distintas estrategias para mejorar la conciencia de situación de los operadores de UAV durante el transcurso de una misión.

Inicialmente, la tesis aborda el geo-registro, que es una estrategia utilizada para localizar al UAV en zonas sin acceso a señal de posicionamiento global (GPS). Esto tiene interés porque conocer la posición del UAV es esencial para poder proporcionar información sobre los alrededores. Por ello, la tesis propone dos sistemas clave para el geo-registro, atendiendo a los diferentes datos de referencia que se utilicen. En primer lugar, un esquema de procesamiento estéreo multi-vista para construir un modelo de terreno denso a partir de imágenes del vídeo capturado por el UAV. Es útil cuando se necesita como referencia un modelo del terreno pero no está disponible, está desactualizado o tiene baja resolución. El método variacional propuesto impone continuidad no sólo a lo largo de la línea epipolar sino también transversalmente, en todo el dominio de la imagen. En segundo lugar, la tesis propone un método conjunto (geométrico y fotométrico) de registro de imágenes que puede lidiar con tipos de distorsión genéricos: deformaciones parametrizadas (como homografías) y transformaciones fotométricas no lineales. Es un método de registro basado en zonas de las imágenes, que permite operar en escenarios donde los métodos de geo-registro basados en puntos característicos no son fiables.

Por último, se considera el caso general, donde todas las medidas de los sensores

tienen suficiente precisión y la tesis se centra en mostrar elementos virtuales sobre el flujo de vídeo. Se desarrolla una herramienta de RA para mejorar la conciencia de situación de operadores de UAV durante misiones de inteligencia y vigilancia. El sistema de RA proporciona información sobre la ruta de vuelo y los objetivos; de esta forma el operador puede reducir el tiempo de búsqueda para encontrarlos incluso si están ocultos. La usabilidad de la herramienta propuesta se demostró con la adopción de estándares de la OTAN y fue plenamente integrada en el demostrador de SAVIER de Airbus, en Getafe, Madrid.

Abstract

Augmented reality (AR) has become, due to recent technology developments, a fast-growing discipline. The potential of AR supports its study not only for specific devices such as glasses or helmets, but for anything equipped with a camera. Following this idea, Airbus promoted an innovation project, Situational Awareness Virtual Environment (SAVIER), to incorporate AR in their ground control stations, thus allowing the enhancement of the video stream captured from Unmanned Aerial Vehicles (UAVs). This thesis is framed in that project and explores different approaches to improve the situational awareness of the UAV operators during a mission.

Initially, the thesis is focused on geo-registration, a strategy used for the localization of the UAV in GPS-denied environments. This is of interest because knowing the position of the UAV is essential to provide information about the surroundings. For this reason, we proposed two key systems for geo-registration with different reference data. First, a multi-view stereo processing pipeline for building a dense terrain model from images of the UAV video feed. This is helpful when a reference terrain model is needed for geo-registration but it is unavailable, outdated, or it has low resolution. The proposed variational method enforces continuity not only along epipolar lines but also across them, in the full image domain. Second, the thesis proposed a joint geometric and photometric image registration method that can deal with generic types of distortion: parametric warpings (such as homographies) and non-linear photometric transformations. It is built on top of area-based registration methods to be able to operate in scenarios where feature-based geo-registration methods are not reliable.

Finally, the general case was considered, in which every sensor measurement is known with enough accuracy and the thesis focused on displaying virtual elements over the video stream acquired by the UAV. An AR tool to improve the situational awareness of UAV operators during intelligence and surveillance missions was developed. The AR system provides information about the flying path and the targets, so

that the operator can reduce the time to find them even in the presence of occlusions. The usability of the proposed AR tool was proved by the adoption of NATO standards and it was fully integrated with the Airbus SAVIER demonstrator, in Getafe, Madrid.

Agradecimientos

En primer lugar quiero dar las gracias a Carlos y a Guillermo, mis tutores, por su tiempo, sus enseñanzas y su apoyo, desde Madrid y desde Suiza. Gracias a ellos he llegado hasta aquí. A Narciso, por dejarme formar parte del GTI. A Fernando, por la recta final. A Julián, a Francisco, a Nacho y a Luis.

Also, I want to thank to Prof. Yezzi to give me the opportunity to do my research stay in his lab. Thanks to Ping-Chang to be my "point of contact" there. Gracias de nuevo, Guillermo, por hacer que la estancia fuera posible, y por no dejar de ser mi tutor aunque cambiaras de país. Gracias al Consejo Social por financiarla.

Gracias a Airbus por apostar por un proyecto como SAVIER. Gracias a todo el equipo, especialmente, a Gemma y a todos los doctorandos con los compartí proyecto.

A Tomás, por el paralelismo GTI-SAVIER que compartimos. Por los muchos ratos buenos, por los agobios superados. Gracias, porque siempre he tenido un apoyo. A Carlos Roberto, porque consiguió que saliera a correr voluntariamente (lo más increíble que ha sucedido en estos años). A Ana, César y Dani. A todos con los que compartí momentos, café y chocolate en el GTI, al principio y al final de esta etapa.

A todos los que habéis estado en los momentos buenos y en los malos, escuchándome y apoyándome. Gracias a Elena por estar desde que tengo memoria, a Cris, porque se ha hecho imprescindible. A Susi, Vero, Iris, Maite, María, Javi y Katsu.

Gracias a Rafa por confiar en mi, ayudarme en todo y caminar a mi lado.

Gracias a mis padres que siempre, siempre, están ahí cuando los necesito y me apoyan en mis decisiones. A mi hermano, Pablo, porque siempre nos saca una sonrisa y tiene tiempo para escuchar las tonterías de su hermana. A mi tía, Marisa, que siempre se alegra por mi. A mis abuelos, que aunque ya no están aquí, me acuerdo de ellos y sé que estarían muy orgullosos.

Y, para terminar, a todos aquellos que hayáis contribuido a que este trabajo se haya podido realizar, mil gracias.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Objectives	3
1.3	Structure	5
1.4	Contributions	6
2	State of the Art	9
2.1	Introduction	9
2.2	Geo-registration	10
2.2.1	Geo-registration with 3D Reference Data	11
2.2.2	Geo-registration with 2D Reference Data	13
2.2.2.1	Direct Image Registration	15
2.2.3	Conclusions	18
2.3	Augmented Reality	18
2.3.1	Introduction	18
2.3.2	Definition	19
2.3.3	Available Technology	20
2.3.4	Applications	21
2.3.4.1	Education and Entertainment	23
2.3.4.2	Training and Maintenance	23
2.3.4.3	Medicine	24
2.3.4.4	Military and UAV	24
2.3.5	Techniques	25
2.3.5.1	Determining the Pose (Position and Orientation) of the Camera	25

2.3.5.2	AR for Situational Awareness	26
2.3.6	Conclusions	29
3	Geo-registration with 3D Reference Data	31
3.1	Motivation	31
3.2	Terrain Models	32
3.2.1	Building Terrain Models from Images	32
3.2.1.1	Multi-view Stereo Processing Overview	32
3.2.1.2	Dense Stereo Matching Via Variational Methods	35
3.2.1.3	Back-projection of Dense Disparity Maps and Surface Generation	40
3.2.2	Geo-registration with the Terrain Model	41
3.3	Experiments	41
3.4	Conclusion	45
4	Geo-registration with 2D Reference Data	47
4.1	Motivation	47
4.2	Image Registration	48
4.3	Image Registration Method	50
4.3.1	Geometric Registration Model. Matching Image Isophotes	50
4.3.2	Joint Geometric and Photometric Registration Model	53
4.3.3	Closed-Form Photometric Solution	53
4.3.4	Coarse-to-fine, Nested Iterative Optimization Approach	54
4.4	Experiments	55
4.4.1	Competing Joint Registration Methods	55
4.4.2	Experiments with Ground-truth Transformations	57
4.4.3	Recovering Non-linear Photometric Transformations	60
4.4.4	Joint Geometric and Photometric Registration “in the Wild”	64
4.4.5	Experiments on Aerial Imagery	66
4.5	Conclusion	69
5	Augmented Reality for Operation Support	71
5.1	Motivation	71
5.2	Functionalities of the Augmented Reality (AR) Tool	74
5.2.1	Awareness of the Flight Route	74

CONTENTS

5.2.2	Awareness of the Targets	75
5.2.2.1	Target Location and Viewpoint-based Classification .	76
5.2.2.2	Target Identification	76
5.2.2.3	Target Search	76
5.2.2.4	Target Detection	76
5.3	Structure of the AR Tool	77
5.3.1	Input Data Processing	77
5.3.1.1	Mission Planning Data	77
5.3.1.2	Mission Execution Data	79
5.3.2	AR Solution Module	83
5.3.2.1	Real World	83
5.3.2.2	Projection Model: Conversion between Coordinate Sys- tems	83
5.3.2.3	Virtual World	88
5.3.3	Augmented Video	91
5.4	Results	92
5.4.1	Results on Route Orientation	92
5.4.2	Results on Targets	94
5.4.2.1	Target Location and Viewpoint-based Classification .	94
5.4.2.2	Target Identification	95
5.4.2.3	Target Search	95
5.4.2.4	Target Detection	97
5.5	Conclusions	98
6	Conclusions	99
6.1	Conclusions	99
6.2	Future work	101
A	First Variation of the Stereo Disparity Functional	103
B	Functional Measuring the Area Mismatch between Contours	107
C	First Variation of the Registration Functional	109
	Bibliography	111

List of Figures

1.1	<i>AR experience.</i> A person is watching a 3D reconstruction of a football match over a table with the Microsoft Hololens. Courtesy of V-SENSE.	2
1.2	<i>Different types of UAVs.</i> On the left, a quadcopter be controlled from a hand-held device. On the right, a UAV that should be controlled from a GCS.	3
1.3	<i>Ground Control Station (GCS).</i> On the left, the operators controlling the UAV, inside the GCS. On the right, the field-deployable container housing the GCS.	4
2.1	Processing flow using geo-registration for aerial navigation [8].	10
2.2	The 3D geo-registration scheme forms a predicted image from the DEM, which is used as a reference to compare and register the actual image [10].	12
2.3	<i>2D geo-registration scheme.</i> Image courtesy of [26].	14
2.4	On the left, an area with large extensions of vegetation; on the right, a desert zone. Both are scenarios where direct methods are preferred.	15
2.5	Reality-virtuality continuum by Milgram in [58].	19
2.6	Magic Leap One™ Reveal, the AR device created by Magic Leap (2018).	20
2.7	AR examples in different fields.	22
2.8	An illustration of “Multi-source Information Fusion Augmented Reality Benefited Decision-making for Unmanned Aerial Vehicles” research paper [113].	28
3.1	Several frames of a video as acquired by an EO sensor of a UAV. Experiment 1: camera tilt is set to zero.	33
3.2	Sparse reconstruction of the viewed terrain in Experiment 1 (Sect. 3.3), including the trajectory (camera poses) of the UAV during the flight.	34

3.3	Epipolar geometric entities. On the left, the epipolar plane π , defined by a 3D point $\mathbf{X}$ and the camera centers, $\mathbf{C}_1$ and $\mathbf{C}_2$. On the right, the epipolar line ℓ given two different views, which consists of the positions where, given a point on a view, the corresponding point can lie in the other one.	35
3.4	Disparity parametrization. At every point $\mathbf{x}_1$ in the reference image, the disparity $\mathbf{d}$ with respect to the corresponding point $\mathbf{x}_2$ in another image can be parameterized by the tangential disparity $\lambda(\mathbf{x}_1)$ along the epipolar line ℓ , i.e., $\mathbf{d} \equiv \mathbf{d}(\lambda(\mathbf{x}_1))$	37
3.5	Variational disparity method. Predicted images by transferring intensities according to the correspondence given by the disparity map. Left: original image I_1 (outside the centered rectangle Ω , with some intensity scaling for visualization purposes) and predicted image $\hat{I}_1(\mathbf{x}_1) = I_2(\mathbf{x}_1 + \mathbf{d}(\lambda))$ (inside Ω). Center: Tangential disparity $\lambda(\mathbf{x}_1)$, pseudo-colored in grayscale expanding the range of λ , in this example $\lambda \in [-48.54, -41.11]$ pixels. Grid size (Ω): 513×257 pixels. Right: predicted image $\hat{I}_2(\mathbf{x}_2) = I_1(\mathbf{x}_2 - \mathbf{d}(\lambda))$ (matched region) and original image I_2 (outside).	39
3.6	Point cloud obtained by triangulation of densely matched points with the variational disparity method. Example with points from five disparity maps of size 513×257 , adding up to ≈ 0.66 million points. Left: zenithal view. Right: close-up view. Individual points are more distinguishable as they are closer to the selected viewpoint.	40
3.7	Experiment 1 (Downward-looking camera). Terrain elevation model in $43.079^\circ \leq \text{latitude} \leq 43.161^\circ \text{ N}$, $3.758^\circ \leq \text{longitude} \leq 3.802^\circ \text{ W}$, obtained by Algorithm 3.1: textured (left) and pseudo-colored (center), from blue (low) to red (high). Right: terrain elevation model (with shaded-relief details, obtained from Google Maps) of the area enclosing the region of interest (highlighted by a rectangle).	42
3.8	Experiment 1. A portion of the dense terrain elevation model and camera trajectory obtained from variational disparity method. Left: shaded model (geometry). Right: textured model (geometry and photometry).	42

3.9	Experiment 1. Dense terrain elevation model obtained from variational disparity method. Close-up point of view.	42
3.10	Experiment 2. Predicted images by transferring intensities according to the correspondence given by the disparity map (cf. Fig. 3.5). Left: original image I_1 (outside the rectangle Ω , of size 513×257 pixels) and predicted image $\hat{I}_1(\mathbf{x}_1) = I_2(\mathbf{x}_1 + \mathbf{d}(\lambda))$ (inside Ω). Center: Tangential disparity $\lambda(\mathbf{x}_1)$, pseudo-colored in grayscale expanding the range of λ , in this example $\lambda \in [-18.22, -8.21]$ pixels. Right: predicted image $\hat{I}_2(\mathbf{x}_2) = I_1(\mathbf{x}_2 - \mathbf{d}(\lambda))$ (matched region) and original image I_2 (outside).	44
3.11	Experiment 2 (Tilted camera). Terrain elevation model obtained by Algorithm 3.1: textured (left) and pseudo-colored (center), from blue (low) to red (high). Right: (low resolution) DEM of the surrounding area enclosing the region of interest (NASA Shuttle Radar Topographic Mission (SRTM) 90m DEM obtained from [140]), also pseudo-colored; cf. Fig. 3.7. Axes are latitude and longitude (in degrees); color legend (elevation), in meters.	44
4.1	High-level input/output diagram of the proposed method.	50
4.2	<i>Image comparison using level sets (isophotes)</i> . Two images to be registered I_1 and I_2 . For simplicity, each image is dominantly bimodal (black and white, except for few transition pixels), in the range $[0, 1]$. The intensity $\lambda = 0.5$ corresponds to the gray level and it is marked in red, for clarity. The dark pixels in the images coincide with the interior of the level set $S_\lambda^i, i \in \{1, 2\}$. Letting g be the identity transformation, then (c) shows the symmetric difference between the interior of S_λ^1 and the interior of the transformed level set $g^{-1}(S_\lambda^2)$, in the first image. In blue, the pixels with $(I_1(\mathbf{x}) < \lambda) \cap (I_2(\mathbf{x}) \geq \lambda)$, and in green, the pixels with the opposite case: $(I_1(\mathbf{x}) \geq \lambda) \cap (I_2(\mathbf{x}) < \lambda)$	51
4.3	A pair of aerial images related by a homography g and a non-linear photometric transformation f	52

4.4	<i>Experiment with ground-truth data.</i> Fig. 4.4a shows two images (Florida and Africa) that are used for the experiments with ground-truth. The images in Fig. 4.4a and the non-linear photometric transformations in Fig. 4.4b are used to generate multiple image pairs (I_1, I_2) with different geometric warps to test registration methods. Two sample generated images I_1 are shown in Fig. 4.4c, with a red square of 512×512 pixels indicating the part of the images used in the experiments.	56
4.5	<i>Accuracy Evaluation. Experiment with Florida images.</i> Sensitivity with respect to the amount of distortion σ_n used to generate the ground truth homography g . Fig. 4.5a shows the mean Euclidean distance between the four corners of the image domain transformed with ground truth g vs. with the estimated homography $\hat{g}$. Fig. 4.5b displays the distance between photometric transformations (4.8). Different registration methods are indicated by the color of the curves: in blue, the method minimizing the L^2 error with affine photometric model; in red, the proposed method.	58
4.6	<i>Experiment with Africa images.</i> Sensitivity with respect to the amount of distortion σ_n used to generate the ground truth homography g , analogously to Fig. 4.5. On the left the geometric error, on the right the photometric one. The red curves correspond to our method, and the blue ones to the competing one.	58
4.7	<i>Experiment to recover Non-linear Photometric Transformations.</i> The first two columns show pairs of input images to the algorithms (cf. Fig. 4.1). The remaining columns show the image absolute differences of two different methods, both using the same metric (L^1 norm) but different photometric models: affine vs. non-linear with closed-form (CF) solver. Results are displayed in false color using a color scale which goes from dark blue (minimum error) to red (maximum error) (see Section 4.4.3). Rows correspond to the Grand Canal , Memorial and Snowman sequences in [152]. Note that the range of the error scale is different for every row to better distinguish the errors, and in particular, the fourth row is in logarithmic scale.	61

4.8	<i>High Dynamic Range (HDR) example.</i> The photometric transformation f estimated by our method can be used to produce better exposed pictures.	63
4.9	<i>Experiment “in the wild” (without ground truth).</i> The five columns correspond to: sample image pairs from the dataset used (first two columns), results of joint geometric and photometric registration using the L^2 affine method (Section 4.4.2 and the proposed method (labeled as L^1 CF), and, finally, in the last column, the comparagram of the input images with the photometric functions recovered by both methods (L^2 Affine in green, and ours in red). Each row corresponds to a specific pair of images and their registration results, which are displayed using pseudo-color, as in Fig. 4.7.	65
4.10	Right column: sensed images. Left column: target images, with color notation: red (exact registration), blue (registration with histogram specification), green (registration with the proposed method).	68
5.1	Information available in the screens of the Ground Control Station (GCS) of Unmanned Aerial Vehicles (UAVs).	73
5.2	Principal modules of the AR tool: input data processing, AR solution and augmented video. The input data module encompasses the processing of data coming from the GCS and the UAV. In this module, four different colors are used: yellow, for the data used as input to the virtual world; red, for data corresponding to mission planning; green, for the metadata to create the projection model; blue, for the images of the real world. The AR solution module is responsible for achieving coherence of real and virtual worlds. Finally, the augmented video module manages the information displayed to the UAV operator. . . .	78
5.3	Generation scheme of MISP compliant file for Full Motion Video. Information sources are: (1) motion imagery and (2-4) metadata. The imagery is processed (5) and multiplexed together with the metadata in a unique metadata packet (6), then a time stamp is given to both of them to be synchronized (7), and then both are combined in the container (8) or sent using RTP (9).	80

5.4	Example of a metadata Key-Length-Value (KLV) packet. It is formed by a key (in green), the length of the whole packet (in purple), and a sequence of metadata. Each metadata is identified by a tag (in cyan), the length of the data (in magenta) and the information itself (in orange). Grid patterned colors have the same meaning as the solid colors [164].	81
5.5	Geographic coordinate systems used: ECEF and NED [167].	84
5.6	On the left, the heading angle of the platform in the plane $N-E$. In the middle, the pitch angle of the platform with respect to the plane $D-N$. On the right, the roll angle of the platform in the $D-E$ plane .	86
5.7	Azimuth of the sensor, given with respect to the platform reference frame.	87
5.8	Initialization process of the flying route. The top of each subfigure shows the UAV situation with respect to the <i>closest</i> waypoint (in red). The bottom of each subfigure shows the color coding of the legs and waypoints that will be presented to the operator. The <i>upcoming</i> waypoint and the currently flown leg are always displayed in green. . . .	88
5.9	Virtual target beacons, composed by a post with a cube at the top, a label and a semi-transparent sphere at the bottom. On the right, an icon has been included next to the cube.	90
5.10	Augmented video with highlighted route (waypoint and legs) and four targets. Same notation for virtual targets as in Figure 5.9. Same notation for waypoints and legs as in Figure 5.8: the camera is looking at the last visited waypoint (hence, it is colored in grey, as in Figure 5.8b).	91
5.11	Route orientation. Four different moments of the video stream augmented with the proposed AR tool during a mission.	93
5.12	Target identification. Results of the AR tool, displayed on the screens of the GCS. Blue circles surrounding the targets have been superimposed on both images to mark their true position.	94
5.13	Difference between the raw video stream (a) and the augmented with the proposed AR tool (b) for distinguishing buildings in reconnaissance missions. Blue circles surrounding the targets have been superimposed on both images to mark their true position.	96

LIST OF FIGURES

5.14	Augmented video with highlighted route (waypoint and legs) and two targets. Same notation for virtual targets as in Fig. 5.9.	96
5.15	Target search aid. A green arrow is displayed in the right part of the screen indicating that the target <i>tgt_2</i> will be found turning the joystick to the right.	97
5.16	On the left, a detected target reported to the GCS is presented in red with a yellow label. On the right, a target which has not been reported is displayed in blue with the label in red.	98

List of Tables

4.1	Competing methods classification.	55
4.2	Statistics of the registration errors (RMSE and MAE) on 40 images from the dataset [152].	66
4.3	Registration errors (RMSE and MAE) on 40 images pairs from the dataset [152] along with their corresponding identifier. The best results of each test are in bold. In addition, the results corresponding to images shown in Fig. 4.9, are marked in bold in in the first column, the order of appearance in the table is the same as in the figure.	67
5.1	UAS categories according to [1]	72
5.2	Example of Tag-Length-Value (TLV) packets contained in a Key-Length-Value (KLV) packet. The table shows: the TLV hexadecimal value (last column), the tag (first column) of the metadata (second column) and its value (third column), and the interpretation of the specific value (fourth column).	82

Nomenclature

AR	Augmented Reality
CCMI	Contextual Conditioned Mutual Information
CRD	Common Route Definition
DDS	Data Distribution System
DEM	Digital Elevation Model
DIC	Dual Inverse Compositional
DTED	Digital Terrain Elevation Data
ECC	Enhanced Correlation Coefficient
ECEF	Earth-Centered Earth-Fixed
EKF	Extended Kalman Filter
EL	Euler-Lagrange
EO	Electro-Optical
FLC	Forward-Looking Camera
FMG	Full Multigrid
FOV	Field Of View
GCS	Ground Control Station
GIS	Geographical Information System
GPS	Global Positioning System
GSD	Ground Sampling Distance
HALE	Medium-Altitude Long Endurance
HDR	High Dynamic Range
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
ISTAR	Intelligence, Surveillance, Target Acquisition, and Reconnaissance

KLV	Key-Length-Value
LOS	Line Of Sight
MALE	Medium-Altitude Long Endurance
MAV	Micro Aerial Vehicle
MI	Mutual Information
MISB	Motion Imagery Standards Board
MISP	Motion Imagery Standards Profile
MR	Mixed Reality
NATO	North Atlantic Treaty Organization
NED	North-East-Down
PDE	Partial Differential Equation
PMVS	Patch-based Multiview Stereo
RSNCC	Robust Selective Normalized Cross Correlation
RTP	Real-Time Transport Protocol
SAVIER	Situational Awareness Virtual EnviRonment
SfM	Structure from Motion
SLAM	Simultaneous Location And Mapping
SMPTE	Society of Motion Picture and Television Engineers
SSD	Sum-of-Squared-Differences
TLV	Tag-Length-Value
TS	Transport Stream
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
UL	Universal Label
VR	Virtual Reality
XML	Extensible Markup Language

Chapter 1

Introduction

1.1 Motivation

Augmented Reality (AR) techniques that have been studied for decades, but recently, they have gone beyond academia and promise to become commonplace. As a consequence of the improvements achieved on AR, leading companies such as Google and Apple have launched Software Development Kits (ARCore and ARKit, respectively) to incorporate these techniques in everyday life devices (e.g., smartphones and tablets). Additionally, other corporations, including high-tech giants like Microsoft and promising start-ups such as Magic Leap, have decided to create AR headsets. As an example of such devices, in Fig. 1.1, an AR experience with the HoloLens (developed by Microsoft) is shown. The new AR equipment improves the user's experience, thus opening new paradigms and increasing the possibilities of evolution in AR techniques. Thus, AR is not fully solved yet, and disciplines intrinsically related to AR (i.e., video, image processing and computer vision) are of interest in research laboratories driven by the needs of industry.

Not only AR devices, but also drones have fostered research on image-related disciplines, since they allow for new ways of capturing images. Drone is the common term used when referring to Unmanned Aerial Vehicles (UAVs). These vehicles allow to obtain aerial imagery and have become very popular recently thanks to the affordable price of the smaller versions, such as the quadcopter shown in Fig. 1.2a. According to the taxonomy in [1], in addition to the small platforms (micro- and mini-UAVs), there are other categories of UAVs with larger wingspan: the tactical,

Figure 1.1: *AR experience*. A person is watching a 3D reconstruction of a football match over a table with the Microsoft Hololens. Courtesy of V-SENSE.

the MALE (Medium-Altitude Long Endurance) and the HALE (Hing-Altitude Long Endurance). The UAVs that fall in those three categories are fix-wing models, as the one shown in Fig. 1.2b, and they are usually employed for Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR) missions. One of the main differences among all UAVs, which is related to the size, is the control system. Whereas small UAVs can be remotely piloted with a hand-held device, the big ones have to be controlled from a Ground Control Station (GCS), as shown in Fig. 1.3. The main reason why it is necessary to use the GCS for large UAVs is because they have a more complex command system as well as a more sophisticated payload equipment, which sometimes requires an extra operator.

Airbus Defense & Space in Getafe (Madrid, Spain), in particular the Unmanned Aircraft System (UAS) Ground Segment Department, is in charge of the GCS of the UAVs developed by Airbus. A few years ago, they started a project called Situational Awareness Virtual EnviRonment Open Innovation (SAVIER) with the aim of designing the GCS of the future. The objective of the project was to develop, in an integrated way, human-machine interfaces for UASs that would fulfill the necessities identified in their products. Particularly, the ones related to mission planning, decision support, stress management and situational awareness. Likewise, they tackled the problem from the perspective of global system security, training, air-to-air refueling and integration of UASs in the airspace. The project consists of twelve research lines and an integrated demonstrator. Each one of the research lines integrates in the

(a) Small drone by DJI.

(b) MALE UAV. Atlante ©Airbus

Figure 1.2: *Different types of UAVs.* On the left, a quadcopter be controlled from a hand-held device. On the right, a UAV that should be controlled from a GCS.

demonstrator a proof of concept under a common environment and mission.

This thesis arises as one of the research lines of the project SAVIER. In particular, it focuses on improving the situational awareness of the UAVs operators, i.e., facilitating the understanding of what is happening during a mission. Having an adequate situational awareness allows the operators to anticipate the next movements and ease decision making. At the moment, as it is shown in Fig. 1.3a, the operators need to focus on many screens to be aware of what is the current status of the UAV during the mission. They have to pay attention to the video stream from the on-board camera, displayed on one screen, while they are monitoring the mission planning in a different screen. At the same time, they have also to be aware of the targets detected by other sensors that, surely, will be displayed on another monitor. This situation turns decision making into a very stressful task. Therefore, to overcome that difficulty, the motivation of this thesis is to enrich the geo-registered video stream received from the UAV by overlaying synthetic spatial information of the mission and the targets. Having this essential information augmenting the video stream, as in AR scenarios, frees the operators from the burden of having to fuse the information displayed on different screens and improves their situational awareness of the mission status.

1.2 Objectives

The main goal of this thesis is to improve the situational awareness of the UAV operators during a mission to reduce their workload and stress. A procedure to achieve this

(a) UAV operators are seated at their positions in the Ground Control Station. US Air Force photo.

(b) Field-deployable Ground Control Station Container. By Seb5340 [CC BY-SA 4.0], from Wikimedia Commons

Figure 1.3: *Ground Control Station (GCS)*. On the left, the operators controlling the UAV, inside the GCS. On the right, the field-deployable container housing the GCS.

goal consists of fusing data coming from several sources in an integrated and natural way. The fusion of information from several on-board sensors and a-priori available information (e.g., satellite or Digital Elevation Models - DEMs) is essential to achieve visual coherence of the whole. In the context of aerial imagery AR, two different scenarios are considered regarding the on-board sensors availability and precision. In the first one, the Global Positioning System (GPS) and the Inertial Measurement Unit (IMU) signals are not available or not reliable. In this situation, the prior problem to address for rendering virtual content is having an accurate localization of the UAV by means of visual sensors and cartographic information (geo-registration). In the second scenario we assume those signals are accurate enough. The focus in this case is on displaying virtual content and information over the images received from the UAV, whose location is known (i.e., registered images).

Geo-registration was identified as the problem to address in the first part of the thesis toward a coherent processing of the information contained in multiple images. It is the process of assigning 3D world coordinates to the pixels of an image, and it is based on registering an image that lacks geographic information with respect to reference imagery that have been previously geo-tagged. In our research, geo-registration is of interest because it can be used to localize the UAV with respect to the referenced data. This offers the possibility of flying the UAV in GPS-denied environments, for example, due to the absence of the GPS signal or due to the presence of hostile

jamming signals that may be used to kidnap the UAV. A broad classification of geo-registration solutions can be done according to the dimension of the geo-referenced data: 3D (e.g., terrain elevation map), or 2D (e.g., satellite imagery).

First, we focused on the problem of using 3D reference data and we proposed a pipeline to build a 3D surface from a video sequence captured by the UAV, which is helpful when the available terrain model is out-of-date, does not have enough detail or, simply, is not available. Regarding geo-registration with 2D reference data, we focused on proposing a registration technique that considers the photometric differences between images. We built on top of area-based registration methods to be able to operate in scenarios where feature-based methods are not reliable.

In the second scenario of the thesis, it is assumed that the position and orientation of the UAV payload is known with sufficient accuracy. Under this assumption, we focused on AR problems, such as displaying virtual content over the video. As a result, a complete AR application that helps the operators to accomplish a specific mission was developed. The AR system followed the standards and was fully integrated in the SAVIER demonstrator. The main functionalities and benefits of the tool are oriented to improve the situational awareness of UAV operators. We have implemented four different functionalities: route orientation (which allows operators to check the route that the UAV follows while they are operating the camera), target search (to reduce the time of search when the camera is manually operated) and target location and identification (which help operators when the target is far away and possibly occluded by the terrain). Finally, the AR system provides confirmation when the GCS receives a target detection message.

1.3 Structure

The thesis is organized in six chapters, including this introduction. The content of each one is outlined below.

In Chapter 1 the motivation of the thesis is presented. After that, objectives of the work are explained. Then, the structure of the thesis and the main contributions are summarized.

In Chapter 2, the techniques of the state-of-the-art are analyzed in the two main scenarios (i.e., parts) of the thesis. The first part covers the scenario of the thesis where the on-board sensors are not reliable: geo-registration. The classification and

comparison of methods that use 3D or 2D reference data is done. Also, the specific techniques used for surface reconstruction and registration are reviewed. In the second part, it is considered that the sensor measurements are accurate enough, so a survey of AR information fusion techniques is pursued.

Chapter 3 presents the research carried out on 3D geo-registration. Mainly, we focus on the scenarios where the terrain model is missing or outdated, and therefore has to be built from images. We propose a geo-registration pipeline to build a terrain model from a video sequence captured by a UAV, which can be later used by another UAV flying over the same zone.

Chapter 4 is focused on geo-registration with 2D reference data. In this chapter we propose an area-based image registration technique that takes into account different illumination conditions. We present a rigorous analysis of the contour-based registration approach, which leads to an L^1 -norm energy functional. We show how our method, which can deal with non-linear photometric changes, performs better than the standard “gain and bias” model.

Chapter 5 presents the AR tool designed for the situational awareness improvement of UAV operators. The functionalities, methodology and the standards followed are explained, as well as the specific enhancements for the operators.

Finally, in Chapter 6 conclusions of the thesis are drawn and future research directions are suggested.

1.4 Contributions

Our main contributions are listed below in correspondence with the chapter where they are detailed.

- Chapter 3: A multi-view stereo processing pipeline for building a dense terrain model from images of the UAV video stream in case that a reference terrain model (DEM) is needed for geo-registration but it is out-of-date, unavailable or with low resolution. The proposed variational method enforces continuity not only along epipolar lines (as done by previous geo-registration techniques) but also across them, in the full image domain [2]. A video explaining this research can be found in [3].
- Chapter 4: A joint geometric and photometric image registration method that

can deal with generic types of distortion: parametric warps (such as homographies) and non-linear photometric transformations. We propose a closed-form photometric solution of the necessary optimality conditions, which allows to reduce the search space of the joint problem to that of the geometric parameters alone [4, 5].

- Chapter 5: An AR tool to improve the situational awareness of UAV operators during ISTAR missions with MALE UAVs. The AR system provides information about the flying route and the targets, so that the operator can reduce the time to find them even in the presence of occlusions. The usability of the proposed AR tool is assured by the adoption of North Atlantic Treaty Organization (NATO) standards for motion imagery and data formats [6]. The AR tool and videos of augmented targets and augmented route are available at [7].

Chapter 2

State of the Art

2.1 Introduction

Situational awareness using images, video or AR requires being able to “point” on the images to display virtual objects which in turn requires an understanding of the 3D geometry of the scene and the location of the user (UAV camera) with respect to the scene. Hence, multiple elements and algorithms are involved in the process of enabling situational awareness. We first start by reviewing the literature to identify the main challenges in the context of AR in aerial imagery, and we found that geo-registration is essential in order to have a global positioning of the UAV when the on-board sensors fail. This is a non-frequent situation but extremely sensitive, so we decided to explore the state-of-the-art techniques.

These techniques are described in Section 2.2. Then, according to the classification done with respect to the reference data, an overview of the techniques with 3D reference data is presented in Section 2.2.1. In addition, we include a review of geo-registration with 2D reference data in Section 2.2.2. Finally, in Section 2.2.3 some conclusions are drawn.

Assuming a proper operation of the UAV on-board sensors, with an accurate-enough knowledge of its position and imaging sensors, we focus on the challenge of enhancing the video acquired from the UAV to improve the situational awareness of the operators. Having additional information in the video feed can improve decision making, which is essential during a mission. For that reason, we focus on displaying virtual content on the images acquired by the UAV and we review AR techniques,

Figure 2.1: Processing flow using geo-registration for aerial navigation [8].

considering this approach. Therefore, Section 2.3 contains a review of the state-of-the-art in the topic of AR, and is organized as follows. First of all, a definition of the discipline and the AR systems is explained in Section 2.3.2. Then, the available technologies for their development as well as the interest of the industry in AR is presented in Section 2.3.3. Afterwards, in Section 2.3.4, the fields where the application of AR is more relevant are mentioned. Some examples of the latest developments are given in Section 2.3.5, and finally, a review of their constraints is presented in Section 2.3.6.

2.2 Geo-registration

In recent years, UAVs are being increasingly used in different domains, both for civil and military applications [9]. For these remotely piloted systems, electro-optical (EO) sensors are essential because they allow vehicle operability and enable the development of applications that build upon aerial imagery, such as surveillance, reconnaissance and remote sensing. One critical step towards processing aerial imagery is geo-registration [10], which involves the assignment of 3D world coordinates to the pixels of an image (depending on the author, this may be called geo-positioning [11]). Geo-registration is a well-known problem that requires precise measurements to achieve

accurate results.

Typically, these measurements will be given by on-board systems (e.g., GPS) or by previously geo-registered data. However, on-board positioning systems may not be reliable (due to the lack of accuracy or due to temporal inoperability), and so the availability of reference geo-registered data is paramount. Image-based geo-registration methods do not require GPS data to estimate the localization of the UAV, hence they can be used in situations where the GPS signal is not available or is being kidnapped. For example, in [8] aerial navigation based on geo-registration following the scheme shown in Fig. 2.1 is proposed. As it is illustrated, an aircraft flying over a terrain captures images which are used as input for the geo-registration software, which provides as output an updated navigation position. These methods can also be combined with GPS data; for example, GPS data can be used as a localization prior distribution or initialization of image-based methods that refine the estimated pose of the UAV.

A broad classification of geo-registration solutions can be done according to the dimension of the geo-referenced data: 2D if the system uses a collection of geo-referenced images [11, 12] (Section 2.2.2), or 3D if the system has an explicit terrain elevation map (e.g., DEM) (Section 2.2.1). In addition, in some solutions, geo-registration may be aided by inertial navigation systems (INS) and GPS [13].

2.2.1 Geo-registration with 3D Reference Data

Geo-registration of aerial imagery that uses 3D data as a reference is usually done with DEMs, which are raster representations of the real-world terrain surface [14], although some approaches use textured models [15]. The DEM provides the height information for a particular geographic position and can be easily converted to a mesh or point cloud that are suitable for 3D-3D or 3D-2D registration.

The former case, 3D-3D registration, is less frequent when dealing with aerial imagery because, commonly, the input data without geographic information is a single image or a video. Although there are methods such as Simultaneous Location and Mapping (SLAM) that can recover the 3D structure “on the fly” from a monocular camera, they are not the best option. As it is pointed out in [10], they can be useful to create the map, but when the objective is geo-registration, a method based on 3D-2D registration is simpler and more robust. Moreover, it only requires one image.

Figure 2.2: The 3D geo-registration scheme forms a predicted image from the DEM, which is used as a reference to compare and register the actual image [10].

The latter, 3D-2D registration, is more common when dealing with aerial platforms [10, 8, 16]. The approach followed in [10] consists on creating a predicted image from the elevation model, which will be used as a reference for the registration of the image captured by the aircraft, as illustrated in Fig. 2.2. They use Normalized Cross Correlation (NCC) because they assume that the predicted image preserves the scale and orientation of the actual image, so they do not need a more complicated algorithm for feature detection. However, they need high-resolution 3D models when dealing with uneven scenes to achieve good results. Later, they improved their method to allow for aircraft navigation in case of GPS spoofing attacks, but they also rely on the existence of DEMs with high resolution and an IMU data [8]. Furthermore, the navigation system proposed in [16] is capable of estimating the absolute 3D pose with inertial data and a monocular camera. They focused on the integration of the geo-registration system on the navigation algorithm in an efficient way more than improving the geo-registration itself.

Most recent works on geo-registration use 3D information [17, 18, 19, 20, 21]. In [17], a framework for 3D geo-registration by using Structure from Motion (SfM) techniques and georeferenced road maps, instead of independent georeferenced lidar scans, is proposed. In addition, in [18] an SfM and pose refinement method for aerial imagery are presented. They remark the usability for image geo-registration and explain how their method avoids 2D registration in the image space by using 3D

information from camera poses.

There are some geo-registration techniques that cannot use previously available 3D models because they lack terrain data. Therefore, they build those models from a collection of images using SfM techniques and a point cloud densification algorithm, such as Patch-based MultiView Stereo (PMVS) [19]. These geo-registration methods perform geo-registration using ground-to-aerial image matching. The quality of the results strongly depends on the quality of the dense multi-view stereo models obtained [20]. In [21], a low altitude UAV was used to perform an experiment of the 3D geometry of a mine, and they also considered the combination of SfM with PMVS.

In addition, other techniques may benefit from having a terrain 3D model, but it could be possible that the reference terrain data did not have enough resolution for their specific flying height or it was out-of-date. This situation is studied in [22], where the authors proposed to build a 3D dense surface. As it is done in the previously described methods that have to build the 3D model, the densification algorithm used after SfM sparse reconstruction restricts the search of dense correspondences to the (1D) epipolar lines.

In Chapter 3 we describe our proposed method, that is focused on including in the geo-registration pipeline the building of a 3D dense model, as the methods described above. All the methods previously described use a densification algorithm that are principled on the epipolar restriction, and use as a starting point a sparse reconstruction recovered with SfM methods. However, if the terrain model that should be built is a continuous surface, as it is presumably the case, the restriction of searching correspondences only along the epipolar line can be improved. In contrast, we propose a geo-registration algorithm that builds a dense terrain model that not only enforces the continuity (i.e., surface smoothness) along the epipolar lines, but also across them, in the full 2D domain of the image. In our approach, in contrast to previous methods that also use the epipolar geometry to formulate the problem, a regularization term was added to the energy functional to reinforce the continuity of the surface.

2.2.2 Geo-registration with 2D Reference Data

A literature review of geo-registration methods confirmed that techniques that use 2D reference data are extensively chosen [23, 24, 25]. Moreover, it is possible to distinguish a common processing scheme among the geo-registration procedures. In

Figure 2.3: *2D geo-registration scheme*. Image courtesy of [26].

general, the scheme, illustrated in Fig. 2.3, consists of a first stage where features are detected, a second stage where features are matched, and a final stage where the geometric transformation that fits the feature matches is estimated.

Navigation in a GPS-denied area is a problem that has been also addressed using 2D reference data. In [27], a navigation system is proposed, that combines visual and inertial data for state estimation. The absolute localization is obtained by registering the captured images with previously available satellite imagery. The goal is to localize the UAV in the reference image (i.e., the geo-referenced map) by finding the transformation between the sensed image and the reference image. They do not consider perspective deformation because they assume that in case of loss of GPS signal, the

Figure 2.4: On the left, an area with large extensions of vegetation; on the right, a desert zone. Both are scenarios where direct methods are preferred.

flight envelope is restricted. Then, they used as similarity measure NCC.

Registration methods based on feature detection and matching across images are considered the standard nowadays, but there are some situations in which these methods are not suitable. Fig. 2.4 shows some samples of scenarios where feature registration is not effective. The poor results are either due to the lack of sufficient feature points in the images or due to the deficit of correct feature matches. As a consequence, other registration approaches that directly match pixel intensities across images (called “direct methods” or “area-based” methods) are more appropriate in these situations. In this context, we developed novel direct registration methods. These methods can be used to improve the available geo-registration strategies that use 2D data as a reference.

2.2.2.1 Direct Image Registration

Classic direct registration methods assume brightness constancy in their formulation, which means that, to model the differences between the intensities at corresponding pixel locations, they only search for the appropriate geometric transformation [28]. However, that assumption is not valid in the presence of varying illumination conditions, when different sensors are used to capture the images, or even when different configurations are applied to the same capture device [29]. Therefore, direct registration techniques tend to overcome such situations by adapting traditional techniques to make them robust to illumination changes. Those techniques can be divided into three main groups depending on the similarity measure used: Correlation-like methods, Mutual-Information (MI) methods, and Sum-of-Squared-Differences (SSD) meth-

ods [30, 29].

Among correlation-like methods, the NCC is the most representative similarity measure, and it has been used as the basis for several registration methods [31, 32]. The main drawback of these techniques is the complexity of the similarity function and its sensitivity to noise and photometric changes [30, 32]. Robust similarity measures invariant to photometric distortions have been proposed, such as the enhanced correlation coefficient (ECC) [29] and the robust selective normalized cross-correlation (RSNCC) [33]. However, although correlation-based methods can deal with some degree of photometric variation, they do not recover the specific photometric transformation between the images.

The second group of direct methods builds upon the MI similarity measure, which is based on the statistical relationship between images. It quantifies the information that one image contains about the other, and consequently, the registration process consists of maximizing such a quantity [34]. It has been extensively used in registration of medical images, especially with images from different modalities [30, 35]. One limitation of MI-based methods is that they ignore the spatial and geometric transformations. Therefore, efforts have been carried out to incorporate spatial information in order to improve registration, such as the contextual conditioned mutual information (CCMI) [36], where MI is measured only on those parts of the images with similar structures. These techniques are also used for registering remote sensing images with different modalities (e.g., infrared images) [37].

The last group of direct registration methods is built upon the SSD similarity measure, which is a particular case of the L^p -norm of the difference of image intensities. Hence, we also consider in this category the methods based on the L^p -norm similarity measure. These methods rely on the brightness constancy assumption, i.e., corresponding pixels have the same value. Thus, to overcome the limitation of having different illumination across images, one must explicitly model photometric distortions along with the geometric ones, yielding a *joint registration* problem. Contrarily to previous methodologies, which aim at proposing similarity measures invariant to illumination changes, the methods in this category are actually able to recover the photometric transformation between the images.

Joint Registration Approaches

Photometric transformations are typically modeled by means of the camera response function, which is the mapping that relates irradiance to pixel (intensity) values. Determining the camera response function is equivalent to solving the comparametric equations [38]. Strategies to estimate the camera response function have been proposed in [39, 40], where a piecewise linear comparametric model was regressed along with geometric parameters. However, the experiments were limited to rotation and scaling (four parameters) between the images. Other approaches [41, 42, 43] have been proposed to deal with non-linear photometric transformations and affine geometric distortions by exploiting the algebraic structure of the joint problem. In these works, the problem is reduced to a sequence of two linear systems that are solved using least squares. Most joint methods model photometric distortions using the “gain and bias” model, consisting of an affine photometric transformation of the image intensities. For example, (i) a closed-form solution was proposed in [44], which was demonstrated on affine geometric transformations, (ii) a dual inverse compositional (DIC) algorithm was proposed in [45], and (iii) a total least-squares approach was developed in [46], which was evaluated on affine geometric transformations.

In Chapter 4 we describe our proposed method, that falls in the third category described in Section 2.2.2.1: methods using the L^p -norm as similarity measure. All the methods previously described use an L^2 -norm objective function, either as a starting point or as the final goal of the optimization algorithm. However, we propose to minimize the L^1 -norm of the difference of image intensities because it naturally arises as the result of formulating the problem in terms of matching isophote curves (i.e., contours) across images. Regarding the photometric model, previous methods use the “gain and bias” model because it is a simple two-parameter model. However, such model is very limited in the type of photometric distortions that it can represent. In contrast, our method is more flexible since it allows to model the broader class of monotonic photometric variations (linear and non-linear). There are methods that also consider non-linear photometric distortions, like we do, but their experiments are limited to affine geometric transformations. In contrast, we demonstrate our non-linear photometric technique on more complex geometric transformations, such as homographies.

2.2.3 Conclusions

Geo-registration is a classical problem whose goal is to achieve accurate results and therefore, precise knowledge of the UAV camera pose. Depending on the community, it may also be called geo-positioning, and it consists of the assignment of 3D world coordinates to the pixels of an image. For UASs, the information coming from the on-board positioning sensors may be inaccurate and, in such situations, geo-registration can provide valuable location information. Typically, due to the lack of reliability of the on-board sensors, the availability of previously geo-registered data is also important. The data commonly used as a reference are satellite images or DEMs, which are raster representations of the terrain surface. There are two main directions to follow in order to develop a geo-registration pipeline according to the previous classification: using 3D or 2D reference data. Therefore, it has been considered the geo-registration problem from two different points of view.

Firstly, in Chapter 3, considering geo-registration done with 3D data, we propose a geo-registration pipeline to build a dense terrain model. The terrain model, in contrast to previously cited methods, enforces the continuity of the surface in its full 2D domain. Secondly, considering that the geo-registration is done with 2D data, we focus on the image registration problem. We explore area-based methods because they can be used where feature-based approaches fail. The proposed registration method is explained in Chapter 4 and consists of a direct method that can deal with global illumination changes.

2.3 Augmented Reality

2.3.1 Introduction

AR is currently a relevant research area [47], as shown by the large amount of publications from different perspectives [47, 48, 49] and the interest in large technological companies (Microsoft, Facebook, Apple...) in developing their own products in this field. It is an interdisciplinary field of study which consists of many disciplines, such as signal processing, sensors, user interfaces, graphics, computer networks and computer vision. Their concepts are clearly applicable in a wide range of fields. AR can be useful in medicine [50, 51], education [52, 53], entertainment [54], and also in military applications [55] and robotics [56]. Furthermore, the growing interest both in indus-

try and among researchers is reflected in the organization of conferences such as the International Symposium of Mixed and Augmented Reality organized by IEEE, the Augmented World ExpoTM that embraces a conference and an expo, or SIGGRAPH, which this year (2018) features a new space devoted to Virtual Reality [57] (VR), AR and mixed reality (MR). The interest is also reflected in publications at top-level journals such as IEEE Computer Graphics and Applications, IEEE Biomedical Engineering or IEEE Industrial Informatics, where the recent developments of these techniques are discussed.

2.3.2 Definition

The term AR falls within the reality-virtuality continuum [58]. The continuum includes four categories and it extends from the completely real through the completely virtual environment as it is shown in Figure 2.5.

Figure 2.5: Reality-virtuality continuum by Milgram in [58].

The following classification is used: real environment, AR, augmented virtuality (AV) and virtual environment. On the one hand, in a real environment there is nothing modeled, on the other hand, in a virtual environment (placed in the opposite side) everything is modeled. AR provides a local virtuality concept which takes place in a real environment, whereas AV provides local reality in a virtual environment. Therefore, AR aims at enhancing reality with virtual elements. The term MR is a global concept that encompasses every term in the reality-virtuality continuum. It does not make a distinction on the amount of real and virtual elements, it only indicates that information from both worlds is merged. The term MR is useful for applications where the concept of locality is vague [59].

An AR system has to meet the following requirements [48, 60]:

- Combination of real and virtual elements in a real environment

Figure 2.6: Magic Leap One™ Reveal, the AR device created by Magic Leap (2018).

- Interactive in real time
- 3D registration between real and virtual elements

Three main points must be taken into account in these systems. First, the definition of AR is not restricted to the devices where the systems are developed. Second, AR is not only a concept related to sight, hence it also includes other senses such as hearing, touch, taste and smell [61, 62]. Third, an AR system can also be used to hide real elements by means of virtual ones.

2.3.3 Available Technology

Although AR systems can involve every sense, we focus on devices related to sight. There are three ways to present an AR system [49]: *i*) video see-through displays where virtual elements overlay on the top of live videos captured from cameras [63], *ii*) optical see-through devices in which only virtual information of the AR is provided through mirrors and transparent lenses [64], and *iii*) projection displays, where virtual elements are projected directly on the physical world [65].

It is possible to classify the devices that allow AR according to how users interact with them: head-worn displays [66], hand-held displays [67] and other ones that do not need to be worn [68].

It can be proved that AR is an attractive discipline for multiple companies. For example, the interest of Amazon in AR is proved by their recent patent of an AR device. Besides, ASUS is interested in making its own AR headset, which can be seen as an opportunity to make this kind of technology a widely available consumer product. Google is also interested in the development of AR apps; they have Project Tango [69], an experimental tablet equipped with depth sensors and a motion tracking camera which allows the development of AR apps.

Furthermore, Microsoft is also in the AR world with its device Hololens [70]. They describe it as a holographic computer that allows to integrate holograms with your world. It is composed by two depth sensors and a camera, and their presentations show how natural the integration of the virtual content is done. Nowadays, they are promoting developer programs to explore different applications for their devices. However, not all the AR is carried out by large companies, for example, Magic Leap [71] is a start-up that works on their own version of the mixed-reality computing paradigm. They have been building its technology in secret for years and they are not far from being ready to ship their product. They have notable investors, such as Google, but little is known about the project, although news and product demos are sometimes presented. The device is supposed to be a self-contained computer, small and comfortable to be used in public as the one shown in Fig. 2.6. It will be a competitor of Microsoft's Hololens.

Additionally, as a consequence of the interest on AR, leading companies such as Google and Apple have not only focused on building devices but in creating tools to incorporate AR in widespread devices such as smartphones and tablets. Proof of this are the SDKs ARCore (Google) and ARKit (Apple).

2.3.4 Applications

A review of recent research suggests that AR systems are of great and growing interest in various fields as it is illustrated in Fig. 2.7. In other ones, the exploration has not started yet because AR is an incipient discipline whose current development is result of recent technological advances in acquisition, processing and visualization

(a) A VR experience that simulates being an astronaut with Oculus Rift.

(b) AR system to facilitate root canals [72].

(c) DAQRI Smart glasses® used for training and maintenance.

(d) Map graphic overlaid providing AR capabilities after geo-registration of images taken from a UAV [10].

Figure 2.7: AR examples in different fields.

devices [47]. Because of this variety, multiple classifications can be established, for instance, a classification based on the fields where AR systems are frequently used: education, entertainment, training, maintenance, medicine and military. However, there are other ones, such as advertisement [73] or industrial prototyping [74].

2.3.4.1 Education and Entertainment

Educators are interested in analyzing the impact that AR systems have in learning. As a consequence, various surveys pointing out advantages and disadvantages have been held [75]. These systems seem to be better than traditional learning techniques that simply use books or other multimedia material [52]. Moreover, the student's enthusiasm to engage in AR experiences, generates an increment in motivation and an improvement in long-term learning [53]. Regarding entertainment, AR systems are used to improve game quality in terms of player satisfaction [76]. Apart from that, adaptation to different environments is being carried out, using mobile devices [54]. Having said that, these AR systems can be found not only in games, but also in shows at theme parks [65]. Besides, there is also a big market in VR games thanks to VR headset devices such as Oculus Rift or HTC Vive, in Fig. (2.7a). With them, the experience of the user is significantly improved. Furthermore, Oculus commitment to the improvement of user's experience with VR is proven with project Santa Cruz. They want to maintain the quality of the Rift but making it a standalone headset incorporating inside-out tracking.

2.3.4.2 Training and Maintenance

AR systems are very useful for teaching maintenance tasks of specific equipment in industry [77]. Thanks to AR systems, it is possible to highlight the components that must be used and to superimpose the instructions that should be followed [78]. This technology facilitates the training of specialized workers in maintenance operations, which is a key point in industry. In addition, it implies a cost reduction for the industry because workers spend less time learning [79]. In Fig. 2.7c, we can see an example of use of AR glasses by operatives in training and maintenance.

It is also worth pointing out the application of AR systems in learning and training high precision tasks, such as needle insertion [80]. However, AR systems find applications not only in this kind of tasks, but also in those required by a collaborative

environment [81], such as the coordination of rescue services during a crisis.

2.3.4.3 Medicine

One of the most important areas where AR systems are used is in medical applications. Over the past decades, the main surgery techniques have evolved, adopting a minimalistic invasive character [50, 51, 82]. One of the most widely used in robotic surgery and telesurgery is the Da Vinci robot, which motivates the development of improved AR systems [83, 84]. Recent research also suggests that these systems may be used to facilitate root canals [72]. Another great challenge is merging medical information coming from different sources to help the surgeon. As a proof of that, a development consisting in a complete visualization system to enable simultaneous visualization of X-ray images and 3D anatomical information in a common monitor in the operating room during angiography is proposed in [85]. The results show an improvement when showing real positions of aneurysms in X-ray images. It is also possible for AR systems to aid doctors augmenting preoperative images onto patients during the surgical processes [86].

2.3.4.4 Military and UAV

Military operations are one of the traditional applications of AR systems. Systems enabling soldiers the identification of friendly forces and targets in the battle field have been studied in [87]. In addition, computer vision algorithms incorporating AR are being elaborated for surveillance [88].

Besides that, cutting-edge technology such as unmanned vehicles, particularly UAV [89, 90], are highly suitable for military operations. Moreover, for these kind of systems, vision-based approaches are relevant because a wide range of UAV are equipped, at least, with an on-board camera [91]. Some research works are focused on enforcing vision-based auxiliary navigation methods with AR. As an example, an immersive user interface with overlying information to aid decision-making is implemented in [92]. Results achieved are promising but they are in an initial stage and have not been tested outdoors. Besides, geo-registered data allow for AR, such as shown in Fig. 2.7d.

2.3.5 Techniques

To develop an AR system, available devices have to be considered since they provide different types of information and enforce constraints on the system design [93]. In fact, current developments not only focus on a particular application area but also, in view of generic AR systems, on creating techniques to solve essential AR problems.

Some systems have an external tracking (e.g., Oculus Rift) while others track the user position based on a camera attached to the AR device, and therefore rely on SLAM methods to determine the location of the user. We are interested in this latter class of systems.

2.3.5.1 Determining the Pose (Position and Orientation) of the Camera

Several studies describe sensor fusion techniques [94] that allow to track the camera motion and determine its pose. Others, try to improve registration algorithms [95]. In addition, some investigations analyze computational constraints of some devices and try to fix them by using distributed systems so that dynamical models of the environment can be built in real time [96].

The popularity of smartphones boosts the interest in Location Based Systems (LBS) [97]. As a consequence, they are one of the most promising devices for AR, and many procedures are being developed to improve accuracy of both Global Positioning Systems (GPS) [98] and digital compass, which are frequently used to locate users outdoors [99].

There are also some developments where classical techniques for localization have been adapted. For instance, the ones based on matching visual landmarks with those of an off-line generated landmark database [100]. Among them, there are methods to improve the accuracy of these techniques by applying information coming from different sensors, such as, depth sensors (3D lidar). Likewise, the goal of some studies is to investigate improvements of standard techniques. So, in [101] improvements to Extended Kalman Filter (EKF) are proposed to deal with the localization problem.

For an AR system to show virtual objects in a real environment, it is first required that the system is aware of both the scene and the user's location with respect to it. Taking this into account, a 3D scene model is needed to allow an accurate tracking and 3D registration [95]. The level of detail in these models depends on the application and the distance from the user to the virtual objects [91]. Moreover, the majority

of these tracking algorithms work under specific conditions, and some of them need previously prepared environments [87, 102].

On the other hand, there are several approaches to determine the user position, which is a challenging task [94, 99]. Some of them allow knowing the user position with six degrees of freedom: three of them corresponding to his position in the world and the other ones to his orientation (*pitch*, *roll*, *yaw*) [49]. Although these techniques are needed in some applications, they are not of special interest to fulfill the goal of this thesis. This is so because we assume that the sensor measurements of the UAV pose are known with sufficient accuracy.

2.3.5.2 AR for Situational Awareness

AR can be also presented as part of computer graphics methodology. Systems with AR have numerous possibilities and they can be used to obtain information about the environment. This can be done by mixing virtual objects and annotations of the surroundings in the user's view, increasing the meaningful semantic context of the world. In [103], two hazards for the AR systems are presented: hardware to implement the systems and human factors issues. Their investigation is focused on the latter, the verification of usability to determine if the system is available. Furthermore, techniques to reference objects are really useful when pointing out an object in a shared environment. However, some specific problems due to surrounding conditions must be solved [104].

The work [105] focuses the attention in the support to the comprehension of spatial relationships between virtual objects and real world. Authors of this work discuss techniques such as X-ray visualization that are suitable for those AR systems to enhance the clarity of the relationship. X-ray visualization is also discussed in [106] for spatial understanding in outdoor AR, where three different types of ghosting combined with depth cues are studied: 1) alpha blending, 2) edge-based ghosting and 3) image-based ghosting. The results in such work show that the latter is the more effective. Apart from the X-ray visualization, there is another tendency in AR that tries to integrate the virtual objects in a realistic way, as it can be seen in [107]. It allows the presentation or inclusion of virtual elements in a real environment. What is more, some of the basic purposes of these strategies are to improve temporal coherence [108] and to achieve a truthful color reproduction [68]. However, for the purposes of this

thesis, photorealism is not a goal to achieve, because it does not try to include the virtual elements as if they were part of the natural environment but to make them distinguishable and easily noticeable to the user (UAV operator).

The outdoor environments are of special interest in the context of this thesis. In [109], the authors are also interested in outdoor AR applications to improve scene understanding. They identify the first-person view as a problem because it narrows the view of the environment, particularly in large areas, and present two techniques to overcome the problem: multi-view AR and variable perspective view. Additionally, wide area AR is chosen in [110] to conduct their experiments. The central concept of their work is to integrate GPS with localized mapping and tracking based on visual SLAM in real time for AR annotations (i.e., the process of adding virtual information to an object).

Furthermore, although annotations are a common purpose in a wide variety of AR applications, there is not a well-established taxonomy for them. In [111], a taxonomy to distinguish between two categories of annotations is proposed: direct and indirect. On the one hand, direct annotation consists of static information about a single real-world object. One example is the annotation used in maintenance where the semantic relevance is high because it provides information of a particular object. On the other hand, indirect annotation does not describe a particular object in the real world, but provides virtual information that makes sense in the environment. For example, in the case of navigation, virtual waypoints are used to route a user.

Besides this taxonomy, there are other research works related to annotations for outdoor AR. In [112], an enhanced technique for registering and tracking anchor points is presented where it is explained that browsing object annotations requires not only a first identification in the image, but also a tracking system to stay registered after the first detection.

Moreover, besides the sensed information by the UAV (such as a video stream), we have as input some additional geographic information. If we take into account that, research related to integrating this kind of information is paramount. As an example, in [114] an application that helps during evacuations of affected areas in nuclear accidents is presented. The application was capable of identifying the correct locations of users showing the escape routes. In [115], another example of the importance of integrating geographical information is given. This work remarks that one of the main challenges in this kind of integration is the establishment of a link between

Figure 2.8: An illustration of “Multi-source Information Fusion Augmented Reality Benefited Decision-making for Unmanned Aerial Vehicles” research paper [113].

the geographic information system (GIS) data and the real-world context (i.e., comprehensibility of the AR visualization and the geographic database). The goals to solve in such work are: (i) easy interpretation of the visualized information; (ii) understanding between the spatial location of virtual elements and the real world; and (iii) consistency in AR visualization when the GIS dataset is modified. However, it does not show results with aerial data, which is the type of information that is considered in this thesis.

The situational awareness problem has been addressed in the past [106, 111, 116]; some solutions proposed the enhancement of the acquired video stream with virtual elements to provide the operator with additional information. In this research direction, some techniques of AR have been used for outdoor elements annotation of the objects seen by the user [111]. A study of the optimal representation of occluded elements has been discussed in [106]. Finally, an AR system to improve the situational awareness and depth perception of UAV operators is proposed in [116], but with a focus on small aerial platforms.

Examples of how AR can be useful for UAV operations are given in [113, 117]. The AR system in [113] increases the safety during take-off and landing UAV operations under harsh conditions such as fog, or at night. However, the tests are not carried out for MALE UAVs, do not follow standards and are not oriented to reconnaissance missions. In [117], a vision-based navigation method for UAVs using AR is proposed. The idea is to develop a system that can be used to superimpose the virtual elements over the video captured by the UAV. However, they only show experiments detecting markers in a controlled environment.

2.3.6 Conclusions

AR systems are continuously evolving due to the birth of new devices where applications can be developed and also due to the improvement of already existing ones. Moreover, devices such as tablets and smartphones are opening a door for massive diffusion of AR applications. For these reasons, it may be predicted that the interest of AR in the areas mentioned in Section 2.3.4 will increase, and probably new applications will appear.

With this review of the state-of-the-art, it is shown that several challenges must be addressed when developing AR systems. As it was said in Section 2.3.5, these systems need to solve problems related to the understanding of the 3D scene and the localization of the user within it. Besides, one of the most important issues is to set the context conditions that determine where the system is going to be used. That is so because developing a system which is going to be used indoors is very different from developing an outdoors one, being the latter more complicated. Difficulty increases if the scene is not previously known.

Apart from that, taking into account the recent studies reviewed, there is a trend in techniques that improve the situational awareness of the users through the use of AR. The challenges of the AR systems include an easy interpretation of the visualized information and the optimal representation of occluded elements. Moreover, one remarkable thing is that when the AR systems are oriented to solve a problem of situational awareness, they are specifically designed for a task under very particular restrictions. As we have shown, the ones related to aerial platforms are focused on small platforms, or are oriented to take-off and landing operations. Indeed, they do not use standardized tools and protocols, so they have very limited real-world applicability.

In this thesis, we have developed an AR system for situational awareness in outdoor scenarios and for a MALE aerial platform, for the whole duration of the mission. Moreover, we have used standardized tools and protocols that facilitate its deployment in real-world applications. The system is described in Chapter 5.

Chapter 3

Geo-registration with 3D Reference Data

3.1 Motivation

The use of geo-registered video data in a UAV can improve the operator’s situational awareness by overlaying synthetic spatial information on the video received from the UAV, thus avoiding the burden of fusing related information in mission planning and mission execution displayed in separate screens [118]. In [119], a pipeline that allows geo-registration without a previous DEM, a low-resolution DEM or an outdated one is presented. Following this idea, the main purpose of our work is the generation of a dense surface from the video sequence to substitute the DEM. The generation of a 3D surface from multiple images when the extrinsic camera parameters are unknown is a central problem in computer vision known as SfM [120]. Traditional strategies provide a sparse 3D reconstruction which is not enough to replace a DEM, thus we consider dense reconstruction methods. Classical dense reconstruction methods operate by matching the intensities of corresponding epipolar lines (rectified scenario), which may produce noisy 3D reconstructions because epipolar lines are processed independently from one another. In contrast, our approach to obtain dense reconstructions is based on the estimation of dense and coherent disparity maps between image pairs of the video sequence using multi-resolution variational methods that enforce continuity of the maps in the full 2D image domain. This produces a dense and coherent surface elevation model that may be used as a reference model in order to avoid the use of

previous or outdated DEMs.

The rest of the chapter is organized as follows: Section 3.2 describes how to generate terrain models and mentions how to use such terrain surfaces to do geo-registration. The terrain generation method is tested and its results are analyzed in Section 3.3. Finally, conclusions and future work are drawn in Section 3.4.

3.2 Terrain Models

In this section an automated solution to the problem of motion imagery geo-registration is presented. The method is based on the scheme [119], which aims at reducing the need for an input DEM to geo-register images. Thus, it is intended to be used in case that such a DEM is not available (it may not exist or is out of date) or it is of low resolution. Instead, the terrain elevation model is built from aerial images acquired from a UAV.

The method consists of two main steps: (1) building or refinement of the terrain elevation model, and (2) geo-registration of images using the computed terrain elevation model as reference. This work focuses on using SfM techniques to implement the first step. The registration of new images is done according to the method in [10].

3.2.1 Building Terrain Models from Images

3.2.1.1 Multi-view Stereo Processing Overview

Let us show how to build accurate and robust terrain elevation models from aerial imagery. To this end, we present a stereo processing pipeline in Algorithm 3.1. The principles behind stereo processing pipelines and their building blocks are presented in [120, 121, 122].

Consider the scenario where a UAV is flying over a terrain of interest. The first step of the processing pipeline consists of acquiring images (see, for example, Fig. 3.1) using an EO sensor, looking downward or in the direction closer to a forward-looking camera (FLC). Next, point correspondences are established across images. To this end, repetitive and distinctive features are detected and matched in the images. There are multiple available detectors and descriptors [123] (SIFT [124], SURF [125], KAZE [126], etc.).

Figure 3.1: Several frames of a video as acquired by an EO sensor of a UAV. Experiment 1: camera tilt is set to zero.

Algorithm 3.1 Multi-view stereo processing pipeline to build dense terrain elevation models.

1. Acquire images from a UAV.
 2. Establish feature (e.g. point) correspondences across images.
 3. Estimate the (sparse) 3D scene structure and the relative camera poses from 2D image point correspondences using SfM techniques.
 4. Generate dense depth maps of the scene using a variational disparity method between image pairs.
 5. Combine multiple depth maps into a single surface by triangulation (i.e., back-projection) and averaging.
 6. Geo-reference the terrain elevation model using either known position and orientation of the UAV's camera or known landmarks.
-

The 2D point correspondences are the input of multi-view SfM [120] algorithms, which reconstruct both the 3D location of the object points as well as the pose of the cameras in the scene based on the observed point projections in the images. This is accomplished in an optimization framework with objective function given by the reprojection error between the predicted image points and the observed ones.

In most practical cases where there is access to the EO sensor, it is possible to assume that the internal camera parameters (focal length, principal point, etc.) are known. This constitutes valuable information because it significantly constrains the 3D reconstruction problem by reducing the number of unknowns and avoiding camera self-calibration [127, 128], a very sensitive step. This is our case, and so

Figure 3.2: Sparse reconstruction of the viewed terrain in Experiment 1 (Sect. 3.3), including the trajectory (camera poses) of the UAV during the flight.

we use SfM algorithms such as [129] or [130] that can exploit the calibrated camera constraint. First, two images with sufficient parallax are selected to initialize the reconstruction. The essential matrix [120] that encodes the relative Euclidean motion (rotation and translation) between both cameras is estimated as well as a set of 3D points that lie in front of the cameras. Then, the remaining images add new camera poses and object points to the existing reconstruction. Camera poses are estimated from object and image point correspondences (3D-to-2D matches) by solving the so-called resection [120] or Perspective-n-Point (PnP) problem. New object points are added to the reconstruction by triangulation [120] of matched image points of a previously added camera pose. A global optimization of the camera poses and object points is usually performed to improve the fit of the reconstruction, a step known as bundle adjustment [120]. Robustness of the algorithm is achieved in multiple stages by outlier rejection using RANSAC [131].

The output of the SfM step consists of a sparse 3D reconstruction of the scene (structure and camera poses), as shown in Fig. 3.2. Such a sparse reconstruction does not sample the objects in the scene (e.g., the terrain of interest) with enough density of points to provide accurate surface models. Nevertheless, it is a useful step for several reasons: it yields accurate information by processing small amounts of data (sparse features) in reasonable computational times, it allows the estimation of the camera poses in case they are unknown and it may be used as the starting point of a “densification” algorithm, as we show next.

Figure 3.3: Epipolar geometric entities. On the left, the epipolar plane π , defined by a 3D point $\mathbf{X}$ and the camera centers, $\mathbf{C}_1$ and $\mathbf{C}_2$. On the right, the epipolar line ℓ given two different views, which consists of the positions where, given a point on a view, the corresponding point can lie in the other one.

3.2.1.2 Dense Stereo Matching Via Variational Methods

Once the camera parameters are known, the stereo reconstruction problem is separable in two subproblems [132]: stereo matching (establishment of dense point correspondences between images) and depth recovery (back-projection of corresponding image points by triangulation, i.e., intersection of optical rays from the cameras). The first subproblem is significantly more difficult than the second one, and to solve it under the hypothesis of a Lambertian scene (where corresponding points in images have equal intensity levels) we use the variational disparity method between image pairs proposed by [133] and modified by [134] to reconstruct ocean waves. Ultimately, this variational method estimates a dense depth map of the scene with respect to both cameras indirectly via estimating the disparity between images. Such depth maps will be translated to terrain elevation displacements with respect to a reference plane.

Epipolar geometry

The epipolar geometry studies the geometric relation between two different views and it is usually motivated by the search of corresponding points in stereo matching (e.g., left and right cameras with centers $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively).

When given a 3D point, $\mathbf{X}$, as it is illustrated in Fig. 3.3a, a plane is determined by it and the camera centers, $\mathbf{C}_1$ and $\mathbf{C}_2$. This is the so-called, epipolar plane, π . Given a camera configuration, there is a pencil of epipolar planes defined by the baseline determined by the camera centers (i.e., the set of all epipolar planes through the line defined by the centers). Besides, it can be observed in the illustration that the projection of $\mathbf{X}$ on both, left and right image planes (points $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively), lies on the epipolar plane. So, when having a correspondence $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$, triangulation can be applied to recover the 3D point.

However, the main challenge in dealing with stereo images is the establishment of the appropriate correspondences between them. Given a 2D point on the left image plane, $\mathbf{x}_1$, it is impossible to determine, only with the information from that view, the exact 3D point, $\mathbf{X}$, that generates it. But, as illustrated in Fig. 3.3b, it can be said that the point $\mathbf{X}$ must lie on the line defined by $\mathbf{C}_1$ and $\mathbf{x}_1$. Therefore, the corresponding point $\mathbf{x}_2$ in the right image plane cannot be determined, but it can be assured that it is going to be on the projection of that line in the right image. That projection is the epipolar line, ℓ . The epipolar line is the intersection of the epipolar plane and the right image plane, and it allows to restrict the 2D search domain of $\mathbf{x}_2$ to a 1D space.

Disparity map

The disparity map is a vector field that establishes correspondences between two images. A point $\mathbf{x}_1$ in image I_1 is mapped to point

$$\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{d}$$

in image I_2 . The disparity $\mathbf{d}$ has two components: the displacements in both horizontal and vertical directions of the image. However, assuming there is no lens distortion (it has been corrected because we assumed calibrated cameras) and according to the epipolar constraint, the disparity can be parameterized by a single displacement: the signed distance along the epipolar line, also called the tangential disparity λ , thus obtaining $\mathbf{d}(\lambda)$ [133]. Specifically, the disparity at $\mathbf{x}_1$ can be decomposed in the orthonormal frame $\{\mathbf{T}, \mathbf{N}\}$ adapted to the epipolar line ℓ corresponding to $\mathbf{x}_1$:

$$\mathbf{d} = \lambda \mathbf{T} + \gamma \mathbf{N},$$

Figure 3.4: Disparity parametrization. At every point $\mathbf{x}_1$ in the reference image, the disparity $\mathbf{d}$ with respect to the corresponding point $\mathbf{x}_2$ in another image can be parameterized by the tangential disparity $\lambda(\mathbf{x}_1)$ along the epipolar line ℓ , i.e., $\mathbf{d} \equiv \mathbf{d}(\lambda(\mathbf{x}_1))$.

where the orthonormal frame is defined as

$$\mathbf{N} = \left(\frac{a}{\sqrt{a^2 + b^2}}, \frac{b}{\sqrt{a^2 + b^2}} \right) \quad \mathbf{T} = \left(\frac{-b}{\sqrt{a^2 + b^2}}, \frac{a}{\sqrt{a^2 + b^2}} \right),$$

γ is the signed distance from $\mathbf{x}_1$ to the closest point on its epipolar line ℓ , and so it solely depends on $\mathbf{x}_1$ and the camera parameters, but not on $\mathbf{x}_2$. It is given by the equation

$$\gamma = \frac{\mathbf{x}_1^\top F \mathbf{x}_1}{\sqrt{a^2 + b^2}}$$

where F is the fundamental matrix. Only λ depends on the location of $\mathbf{x}_2$ along ℓ , as illustrated in Fig. 3.4.

In this setting, the disparity map that solves the matching problem is obtained as the minimizer of the functional

$$E(\lambda; I_1, I_2) = E_{\text{data}}(\lambda; I_1, I_2) + \alpha E_{\text{smooth}}(\lambda), \quad (3.1)$$

which consists of a weighted sum of a data fidelity term and a smoothness prior ($\alpha > 0$). The data fidelity term measures the photometric consistency between stereo images caused by a candidate disparity map,

$$\begin{aligned}
 E_{\text{data}}(\lambda; I_1, I_2) &= \int_{\Omega} \frac{1}{2} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_1 + \mathbf{d}(\lambda(\mathbf{x}_1))))^2 d\mathbf{x}_1 = \\
 &= \int_{\Omega} \frac{1}{2} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_1 + \lambda(\mathbf{x}_1)\mathbf{T}(\mathbf{x}_1) + \gamma(\mathbf{x}_1)\mathbf{N}(\mathbf{x}_1)))^2 d\mathbf{x}_1,
 \end{aligned} \tag{3.2}$$

where $\mathbf{x}_1 \in \Omega$ lies in the reference image (origin for the disparity map) and $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2(\lambda)$ are corresponding points in images 1 and 2, respectively, with observed intensities in I_1 and I_2 . This data fidelity cost is not symmetric with respect to the role of each image, but it can be easily symmetrized. The regularizer,

$$E_{\text{smooth}}(\lambda) = \int_{\Omega} \frac{1}{2} \|\nabla \lambda(\mathbf{x}_1)\|^2 d\mathbf{x}_1, \tag{3.3}$$

$\nabla \lambda$ being the gradient of $\lambda(\mathbf{x}_1)$, enforces coherence (continuity) of the disparity map. The epipolar constraint allows to express (3.1) so that it depends on a single 2D function, $E(\lambda)$, rather than on two 2D functions (the disparity vector field $\vec{\mathbf{d}}$).

In other stereo approaches, e.g. those that match points in corresponding epipolar lines on rectified images, continuity of the solution is only enforced at most along epipolar lines, but not across them. In our approach, however, continuity is enforced in the full 2D domain of λ within the image and it does not require rectified images. In addition, the weight $\alpha > 0$ allows to control the amount of smoothness of the solution disparity map, which is directly transferred to the smoothness of the terrain model because the disparity map contains depth information of the scene with respect to the cameras.

The cost functional (3.1) is minimized by gradient descent according to its necessary optimality conditions (Euler-Lagrange (EL) equations). These yield a non-linear elliptic partial differential equation (PDE) in the tangential disparity λ ,

$$\alpha \Delta \lambda + (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} = 0, \tag{3.4}$$

with homogeneous Neumann boundary conditions, where the symbol Δ stands for the Laplacian operator. Details are given in Appendix (A).

The PDE is discretized on a rectangular 2D grid using finite differences and is

Figure 3.5: Variational disparity method. Predicted images by transferring intensities according to the correspondence given by the disparity map. Left: original image I_1 (outside the centered rectangle Ω , with some intensity scaling for visualization purposes) and predicted image $\hat{I}_1(\mathbf{x}_1) = I_2(\mathbf{x}_1 + \mathbf{d}(\lambda))$ (inside Ω). Center: Tangential disparity $\lambda(\mathbf{x}_1)$, pseudo-colored in grayscale expanding the range of λ , in this example $\lambda \in [-48.54, -41.11]$ pixels. Grid size (Ω): 513×257 pixels. Right: predicted image $\hat{I}_2(\mathbf{x}_2) = I_1(\mathbf{x}_2 - \mathbf{d}(\lambda))$ (matched region) and original image I_2 (outside).

numerically solved using iterative multigrid methods [135], which are among the most efficient numerical tools for boundary value problems. The Neumann boundary conditions correspond, numerically, to extending the image replicating the pixels in the boundary.

In the same spirit of the hierarchical motion estimation framework [119], multigrid methods are well founded multi-resolution tools. This increases robustness and significantly reduces the computational cost of establishing dense matches. Specifically, the full multigrid (FMG) method combines multi-resolution with a coarse-to-fine initialization, which is a fast and sensible approach to improve convergence of the iterative method and avoid local minima of the cost functional.

The numerical solver can be initialized in several ways. One option consists of using available software such as PMVS [136] to provide a dense point cloud representing the terrain model and project it to compute an initial estimate of λ . This is, however, overkill: λ does not need to be initialized by such a detailed map since it is a slowly varying signal (in our use cases, see Fig. 3.5, center). Instead, an interpolation of the disparities given by the sparse reconstruction often suffices, specially if it is used to initialize the solver at the coarsest level of FMG. A third initialization option (used in [133]) consists of using a block-based correlation matching algorithm.

Figure 3.5 shows the output of the variational disparity method overlaid on the original images (with some intensity scaling for visualization purposes). It can be observed, e.g., at the boundary of the highlighted matched region, that the predicted images via the estimated disparity map are a good fit to the original images.

Figure 3.6: Point cloud obtained by triangulation of densely matched points with the variational disparity method. Example with points from five disparity maps of size 513×257 , adding up to ≈ 0.66 million points. Left: zenithal view. Right: close-up view. Individual points are more distinguishable as they are closer to the selected viewpoint.

3.2.1.3 Back-projection of Dense Disparity Maps and Surface Generation

Once the variational disparity method has densely matched points between two images (Sect. 3.2.1.2), we recover the depth of the corresponding terrain points with respect to the cameras by intersecting the optical rays through the image points (i.e., triangulation [120]), according to the camera parameters obtained in the sparse reconstruction. This operation gives a dense cloud of 3D points for every image pair densely matched. Figure 3.6 shows the combined point cloud from a subset of the image pairs matched in a video sequence. Clearly, overlapping between point clouds is prone to exist and they must be combined to produce a consistent point cloud and/or meshed terrain model.

Next, 3D points are expressed with respect to a plane through the scene [137]. If this plane coincides with a horizontal plane, whose normal direction may be obtained by means of an IMU in the UAV, then this operation produces elevation information so that 3D points are expressed as $z_i = (x_i, y_i)^\top$. Interpolation methods are used to fit a surface $z^k(x, y)$ through the points corresponding to each (k -th) disparity map. Then, surfaces are merged by using statistical estimators (e.g., mean, median) on the functions z^k over a common grid domain. Figure 3.7 shows the merged surface from all disparity maps in a video sequence.

An alternative approach consists of using Poisson surface reconstruction [138] or similar algorithms to fit a surface model (e.g., polygonal mesh) to the combined point cloud from all disparity maps, without referring the points to a plane. The fitted

surface model needs to be constrained to be in the form of a graph $z(x, y)$ with respect to the horizontal plane so that it represents an elevation terrain model (e.g., DEM).

Finally, the resulting terrain model is geo-referenced (step 6 in Algorithm 3.1). This can be accomplished using additional information from the UAV (IMU, GPS, etc.) or from the scene (e.g., known geographic coordinates of landmarks). This is just a Euclidean change of coordinates, possibly including a scaling, between reference frames: the geographic world and the world in which the 3D reconstruction (cameras, terrain points) is given.

3.2.2 Geo-registration with the Terrain Model

Once the terrain elevation model has been formed, it can be used as a reference to find the camera pose (location and orientation) of new images of the scene acquired, for example, by another UAV flying over the same terrain. According to the method in [10], the surface model, illumination conditions and candidate camera parameters of the new images are combined in a computer graphics pipeline to generate predicted images. Both, predicted and observed images are compared in an optimization framework that updates the camera parameters to achieve an optimal fit, i.e., to geo-register the new image with respect to the reference surface. A detailed discussion of this step can be found in [10, 118].

3.3 Experiments

To validate the proposed method, we test it on aerial video sequences obtained with Google Earth [139]. Two experiments are carried out: in the first one, the UAV camera is looking downward (with zero tilt and heading), whereas in the second one the camera is closer to the FLC setting, with a tilt of 45 degrees. The terrain under study corresponds to a mountainous area in the north of Spain. The videos consist of images of size 840×377 pixels, acquired at a frame rate of 20 Hz. The UAV flies at an altitude of 4.66 km, simulating a MALE UAV (from 3 km to 10 km approximately) covering an area of 10×5 km in approximately half a minute. Figure 3.1 shows several images of the input videos to the terrain modeling pipeline.

Next, the images are fed to Algorithm 3.1 described in Sect. 3.2.1.1. Figure 3.2

Figure 3.7: Experiment 1 (Downward-looking camera). Terrain elevation model in $43.079^\circ \leq \text{latitude} \leq 43.161^\circ \text{ N}$, $3.758^\circ \leq \text{longitude} \leq 3.802^\circ \text{ W}$, obtained by Algorithm 3.1: textured (left) and pseudo-colored (center), from blue (low) to red (high). Right: terrain elevation model (with shaded-relief details, obtained from Google Maps) of the area enclosing the region of interest (highlighted by a rectangle).

Figure 3.8: Experiment 1. A portion of the dense terrain elevation model and camera trajectory obtained from variational disparity method. Left: shaded model (geometry). Right: textured model (geometry and photometry).

Figure 3.9: Experiment 1. Dense terrain elevation model obtained from variational disparity method. Close-up point of view.

shows the result of the sparse reconstruction stage (step 3). Although the density of terrain points is not sufficient to accurately model the surface, this step provides precise locations of the cameras with respect to the terrain, which is used in the next step of Algorithm 3.1.

Figures 3.7 to 3.9 show dense terrain 3D reconstruction results obtained with the variational disparity method, after step 5 of Algorithm 3.1. The domain Ω matched in the reference image(s) was discretized on a grid of 513×257 pixels, and a 6-level multigrid solver [135] with 200 iterations, 2 V-cycles per iteration and one pre- and post-relaxation sweeps per level was used. The weight $\alpha = 4000$ was empirically chosen to provide a sufficiently smooth surface. Figure 3.8 shows on the left, the terrain elevation model (geometric information only) and on the right, the same model with its corresponding texture (geometry + photometry). The figure also displays the camera position in which the images were collected during the UAV flight. As it is expected, the variational method shows a remarkable performance in terms of the high density of sampling points in the surface model compared to the sparse reconstruction obtained by SfM methods, an effect that is even more perceptible in Figure 3.9, which shows a close-up of the reconstruction from a low altitude point of view.

Hence, these high-resolution models are appropriate for image geo-registration. Figure 3.7 shows a comparison of the obtained terrain elevation model with a high-resolution terrain model of the surrounding area. Each input image pixel corresponds to a real-world spacing of approximately 4.6 meters. The generated terrain model (Figure 3.7, center) is defined on a grid of 1000×2480 points, with a ground sampling distance (GSD) of 3.7 meters, thus covering an area of 9.17×3.7 km. Indeed, for fixed focal length acquisition conditions, distance between terrain samples depends on the UAV flight altitude. The generated terrain model will be most useful as a reference in the geo-registration phase for UAVs flying at altitudes around the acquisition altitude (4.66 km) and above. This is so because the accuracy of the model is limited by its resolution (GSD) and UAVs flying at lower altitudes may require higher resolution models.

Experiment with tilted camera. Figures 3.10 and 3.11 show the results of the second experiment (tilted camera). The UAV flies at the same altitude as Experiment 1 and covers approximately the same area in half a minute. The video acquisition settings and the processing steps are also the same. The reconstruction is more

Figure 3.10: Experiment 2. Predicted images by transferring intensities according to the correspondence given by the disparity map (cf. Fig. 3.5). Left: original image I_1 (outside the rectangle Ω , of size 513×257 pixels) and predicted image $\hat{I}_1(\mathbf{x}_1) = I_2(\mathbf{x}_1 + \mathbf{d}(\lambda))$ (inside Ω). Center: Tangential disparity $\lambda(\mathbf{x}_1)$, pseudo-colored in grayscale expanding the range of λ , in this example $\lambda \in [-18.22, -8.21]$ pixels. Right: predicted image $\hat{I}_2(\mathbf{x}_2) = I_1(\mathbf{x}_2 - \mathbf{d}(\lambda))$ (matched region) and original image I_2 (outside).

Figure 3.11: Experiment 2 (Tilted camera). Terrain elevation model obtained by Algorithm 3.1: textured (left) and pseudo-colored (center), from blue (low) to red (high). Right: (low resolution) DEM of the surrounding area enclosing the region of interest (NASA Shuttle Radar Topographic Mission (SRTM) 90m DEM obtained from [140]), also pseudo-colored; cf. Fig. 3.7. Axes are latitude and longitude (in degrees); color legend (elevation), in meters.

challenging than the previous one because the tilt angle implies that terrain points that are closer to the camera will be more accurately estimated than those further away. However, some of the latter will be better estimated as the UAV moves since they will be closer with respect to a later camera position. For this reason, as shown in Fig. 3.10 (left), we choose the location of the disparity grid (Ω) in the region of the image where the closer points to the camera project, since such points will be triangulated with less uncertainty.

In addition, observe that in this setup a rectangle in the image plane maps to a trapezoidal shape in the scene (in a quasi flat world), and therefore a uniform sampling in the image will give a non-uniform sampling of the terrain according to the perspective transformation. The front-to-parallel setup in Experiment 1 is a better strategy for sampling the terrain in a more uniform manner.

In Fig. 3.11 (center), the GSD of the obtained terrain is approximately 4.5 m, larger than in Experiment 1 due to the deviation from the front-to-parallel configuration: an image pixel covers more terrain area than before. Fine details can be better resolved than in the 90 m DEM of Fig. 3.11 (right), specifically this is more evident in the textured model (Fig. 3.11, left), e.g., in the road at the southernmost region.

3.4 Conclusion

We have presented a stereo processing pipeline for building a dense terrain model from images of the UAV video feed in case that a reference DEM is needed for geo-registration but it is unavailable or not present. The proposed method uses variational methods and multi-resolution to generate a dense disparity map, providing surface continuity not only along epipolar lines but also across them, and yielding a dense and coherent surface model.

Experiments have been carried out simulating a MALE UAV with two different camera orientations, showing that terrain recovery is possible with both a downward-looking camera and a tilted one, the first configuration being better than the second one from the point of view of a uniform sampling of the terrain. The surface models obtained with the proposed method are better than the sparse ones achieved with SfM techniques, providing the capability to create a high-resolution DEM. A video explaining this research can be found in [3].

Chapter 4

Geo-registration with 2D Reference Data

4.1 Motivation

Considering state-of-the-art solutions to the geo-registration problem, a broad classification was done in Section 2.2 according to the dimension of the geo-referenced images: 3D or 2D. The previous chapter discussed methods related to 3D geo-referenced data, whereas this one takes into consideration geo-registration with 2D geo-referenced data. Generally, when using 2D geo-referenced data, it is necessary to adjust and align the input images with a standard reference image. Following this idea, we aim at improving the adjustment and alignment of image pairs. Therefore, the parameters of the geometric transformation that relate the input image space to the reference image space need to be recovered. This is an optimization problem on the transformation parameters that will allow us to geo-register the images in the video. The aim of this work is to improve the estimation of the geo-registration parameters in practical scenarios with image intensity variations by modeling global illumination changes (in terms of image histogram transformations). This implies a joint optimization of both the registration parameters and the intensity transformation that maps the histograms of the images.

4.2 Image Registration

Image registration is the process of aligning two or more images regardless of the viewpoint, the sensor and the acquisition time. It is a classic problem in image processing that still remains an important research topic [141, 142]. Registration is highly relevant in various disciplines, among them: medical imaging [143], remote sensing [144], and augmented reality [145]. There is no single technique that outperforms the others, the choice of the appropriate one depends on several factors such as the application scenario and the characteristics of the images.

Registration methods can be broadly classified into two groups: feature-based and area-based methods [30]. Feature-based methods rely on the extraction and association of a set of salient features prior to their registration [146]. In contrast, area-based methods (also known as direct methods [147]) do not need this intermediate processing step and they use (directly) the intensity of the pixels in the image [33]. Feature-based methods are useful in highly-textured images where it is easy to find distinctive and repetitive features due to large intensity changes. However, their performance degrades as texture decreases and good features become more difficult to extract. Recently, there has been a resurgence of direct methods, specially using image alignment for ego-motion estimation [148, 149, 150].

In most cases, the images to be registered are not acquired under the same illumination conditions, with the same camera parameters or even with the same sensor. This implies that not only geometric distortions are present but also photometric ones. Those photometric distortions can be resolved by either (i) performing some pre-processing on the images (e.g., histogram transformation) and then using a simple comparison function [151], or (ii) using a comparison function (e.g., zero-mean normalized cross-correlation) [27, 152] or feature descriptor [146] that is more robust to photometric changes. Although these techniques provide some invariance to illumination changes, they do not provide information about the actual photometric distortion between the registered images. The latter can only be achieved by explicitly modeling the photometric distortion during the geometric alignment process, i.e., by considering the problem of jointly estimating *both* types of distortion between the images: geometric and photometric.

Joint frameworks allow to solve connected problems (see, e.g., [153]) at the expense of increased complexity. In the problem at hand, mutual feedback between geometric

and photometric solvers may yield better image registrations. The complexity issue is usually tackled by devising simplifying strategies or assumptions. For example, [39] simplifies the solution of the joint geometric and photometric registration problem by transforming the non-linear problem into a linear one, using an approximate piecewise linear photometric model comprising six segments.

In this chapter, we tackle the joint geometric and photometric registration problem by proposing a nested iterative optimization method that exploits the structure of the problem to reduce the search space of the solution to that of the geometric distortion alone. At each iteration, for a given geometric transformation, the optimal photometric distortion is obtained via a closed-form solver. The photometric distortion considered is given by a global (spatially-invariant) intensity transformation, and therefore, is not limited to the usual “gain and bias” (affine) model. The closed-form photometric solver that we propose is principled since it is based on the analysis of the necessary optimality condition. Additionally, as opposed to previous methods, we do not rely on the L^2 -norm error minimization but on the L^1 -norm error, which is known to be more robust to outliers [30]. We show that this error criterion naturally arises when images are compared by means of the areas enclosed by corresponding isophotes of the images. Our contributions are the following:

- A problem formulation based on matching isophote curves across images that leads to the L^1 -norm error functional.
- A closed-form photometric solution of the necessary optimality condition, which allows to reduce the search space of the joint problem to that of the geometric parameters alone.
- A joint geometric and photometric image registration method that can deal with generic types of distortion: parametric warpings (such as homographies) and non-linear photometric transformations.

The chapter is organized as follows: Section 4.3 describes the proposed methodology to solve the joint estimation problem; in Section 4.4 the experimental results are presented and, finally, conclusions are drawn in Section 4.5.

Figure 4.1: High-level input/output diagram of the proposed method.

4.3 Image Registration Method

A review of the related work on direct image registration was given in Section 2.2.2.1. We first motivate our proposed registration minimization criterion from a geometric point of view, in terms of matching the isophotes (level sets) of the two images being registered. We establish a link between the symmetric difference of level sets of two images and the L^1 -norm of the difference image (Section 4.3.1). Then, we extend the criterion to allow for non-linear, spatially-invariant photometric distortions between the images (Section 4.3.2). We derive the necessary optimality condition of the proposed error functional and provide a closed-form solution to estimate the photometric distortion (Section 4.3.3). Finally, we describe our approach to jointly optimize for the geometric and photometric transformations relating both images (Section 4.3.4).

As shown in Fig. 4.1, our method consists of a registration algorithm that takes two input images I_1, I_2 , and returns the geometric and photometric transformations between them, the registration error and $\hat{I}_1$, which is the prediction of I_1 using the estimated transformations on I_2 .

4.3.1 Geometric Registration Model. Matching Image Isophotes

Let $I_i : \Omega_i \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, $i \in \{1, 2\}$ be two grayscale images to be registered. Let $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be a geometric transformation between image domains. As it is standard, assume that g is described by a finite number of parameters, i.e., it is a parametric geometric transformation. We define the prediction of I_1 by $\hat{I}_1 \doteq I_2 \circ g$.

Let λ be an intensity level in the range of I_1 or I_2 , with an associated level set S_λ , which consists of the subset of Ω_i (Ω_1 or Ω_2) given by the points with intensity λ (i.e., an isophote), i.e.,

$$S_\lambda^i = \{\mathbf{x} \in \Omega_i \mid I_i(\mathbf{x}) = \lambda\}, \quad i \in \{1, 2\},$$

Figure 4.2: *Image comparison using level sets (isophotes)*. Two images to be registered I_1 and I_2 . For simplicity, each image is dominantly bimodal (black and white, except for few transition pixels), in the range $[0, 1]$. The intensity $\lambda = 0.5$ corresponds to the gray level and it is marked in red, for clarity. The dark pixels in the images coincide with the interior of the level set S_λ^i , $i \in \{1, 2\}$. Letting g be the identity transformation, then (c) shows the symmetric difference between the interior of S_λ^1 and the interior of the transformed level set $g^{-1}(S_\lambda^2)$, in the first image. In blue, the pixels with $(I_1(\mathbf{x}) < \lambda) \cap (I_2(\mathbf{x}) \geq \lambda)$, and in green, the pixels with the opposite case: $(I_1(\mathbf{x}) \geq \lambda) \cap (I_2(\mathbf{x}) < \lambda)$.

as depicted in red in Figs. 4.2a and 4.2b. For continuous images, the level sets are given by curves (except in the regions of constant intensity).

The registration problem consists of finding the transformation g that best matches the level sets S_λ^1 and S_λ^2 of both images, i.e., that best warps the level sets of one image onto the other one. For now, assume that corresponding level sets have the same intensity λ ; this will be relaxed in Section 4.3.2. In our approach, the dissimilarity between two corresponding level set curves is measured using the area enclosed by them. We propose, for each intensity level λ , the following error functional measure:

$$E(\lambda, g) \doteq \int_{I_1(\mathbf{x}) < \lambda} H(\hat{I}_1(\mathbf{x}) - \lambda) d\mathbf{x} + \int_{I_1(\mathbf{x}) \geq \lambda} H(\lambda - \hat{I}_1(\mathbf{x})) d\mathbf{x}, \quad (4.1)$$

where H is the Heaviside function. The first term in (4.1) corresponds to the points in Ω_1 , $\mathbf{x} \in \Omega_1$, with an intensity less than λ in I_1 but greater than λ in the prediction, $\hat{I}_1$ (i.e., I_2 after warping by g). Those points correspond to the blue pixels in Fig. 4.2c. The second term in (4.1) accounts for the opposite case: $\mathbf{x} \in \Omega_1$ with an intensity in I_1 greater than λ and an intensity in the prediction $\hat{I}_1$ less than λ (green pixels in

Figure 4.3: A pair of aerial images related by a homography g and a non-linear photometric transformation f .

Fig. 4.2c). Hence, (4.1) is the area of the symmetric difference between two sets: the interiors of S_λ^1 and $g^{-1}(S_\lambda^2)$, that is, the area mismatch between both level sets in Ω_1 .

Our approach to align the images consists of minimizing the area of the mismatch (4.1). To remove the dependency with λ , we sum (4.1) for each level set, producing the accumulated error

$$E(g) \doteq \int_{\lambda_{\min}}^{\lambda_{\max}} E(\lambda, g) d\lambda, \quad (4.2)$$

which, after some simplifications (see Appendix B), becomes

$$E(g) = \|I_1 - \hat{I}_1\|_{L^1(\Omega_1)} \doteq \int_{\Omega_1} |I_1(\mathbf{x}) - \hat{I}_1(\mathbf{x})| d\mathbf{x}. \quad (4.3)$$

Therefore, we arrived at the *following interpretation*: the L^1 norm of the difference between I_1 and $\hat{I}_1$ is the sum (for all intensities λ in the range of I_1 and I_2) of the areas of the symmetric difference between the interior of the level set S_λ^1 and the interior of the warped level set $g^{-1}(S_\lambda^2)$.

The functional (4.3) is not symmetric with respect to the roles of both images, but it can be symmetrized by adding the L^1 norm of the difference between I_2 and its prediction in Ω_2 , $\hat{I}_2$. For simplicity, we consider unsymmetrized functionals like (4.3), with I_1 as the reference image and I_2 as the image to be registered by generating a prediction $\hat{I}_1$ (cf. Fig. 4.1).

4.3.2 Joint Geometric and Photometric Registration Model

We relax the condition that corresponding isophotes have the same intensity value, which is sensible in case of photometric distortion between the images, as illustrated in Fig. 4.3. We do so by using a monotonic, spatially invariant transformation of the intensity, $f(\mu)$ (with $f'(\mu) \geq 0$), to map the intensity levels of I_2 , μ , to those of I_1 , λ , i.e., $\lambda = f(\mu)$. The function $f(\mu)$ can be monotonically decreasing, but in that case, it will imply an inversion of gray levels (i.e., a negative image); so we do not consider that type of transformations. The functional (4.1) can be augmented with f , leading to an accumulated error functional (4.2) with the same formal expression as (4.3), but using the prediction¹ $\hat{I}_1 \doteq f \circ I_2 \circ g$:

$$E(g, f) \doteq \|I_1 - \hat{I}_1\|_{L^1(\Omega_1)}. \quad (4.4)$$

This functional explicitly accounts for both types of distortion relating the images: geometric (g) and photometric (f). Our goal is the optimization of (4.4) with respect to both g and f , i.e., the joint optimization of the variables. In the next section, we analyze the optimality condition with respect to the photometric distortion and propose a closed-form solution.

4.3.3 Closed-Form Photometric Solution

The necessary optimality condition of the functional (4.4) with respect to f (assuming, for the moment, that g is fixed) is given by

$$\frac{\delta E}{\delta f}(\mu) = 0, \quad (4.5)$$

where the functional derivative $\delta E/\delta f$ is obtained from the first variation of (4.4) (see Appendix C):

$$\frac{\delta E}{\delta f}(\mu) = \int_{\Omega_1} \text{sgn}(f(\mu) - I_1(\mathbf{x})) \delta(\mu - (I_2 \circ g)(\mathbf{x})) d\mathbf{x}, \quad (4.6)$$

where $\text{sgn}(\cdot)$ is the sign function and $\delta(\cdot)$ is the Dirac delta.

We propose a direct solver for (4.5) by numerically solving the necessary optimality

¹One may interpret (4.3) as the particularization of (4.4) in case f is the identity.

condition using (4.6). Let us define $\tilde{I}_2 \doteq I_2 \circ g$ and its corresponding level set $\tilde{I}_2^{-1}(\mu) \doteq \{\mathbf{x} \in \Omega_1 \mid \tilde{I}_2(\mathbf{x}) = \mu\}$. For each value of μ in the range of I_2 , we replace the Dirac delta in (4.6) with a smooth approximation. This selects the pixels $\mathbf{x}$ around the level curve $\tilde{I}_2^{-1}(\mu)$; then, the value of $f(\mu)$ that solves (4.5) is given by the median of $I_1(\mathbf{x})$ at these pixels. Observe that the optimal transformation f is different from that one given by, e.g., histogram specification [154] between images I_1 and $\tilde{I}_2$. Histogram specification does not depend on the spatial distribution of the intensities, whereas the solution of the optimality condition (4.5) does.

Remark Using the co-area formula, (4.6) can be rewritten as a contour integral over the warped image $\tilde{I}_2$,

$$\int_{\tilde{I}_2^{-1}(\mu)} \frac{\text{sgn}(f(\mu) - I_1(\mathbf{x}(s)))}{\|\nabla \tilde{I}_2(\mathbf{x}(s))\|} ds = 0, \quad (4.7)$$

where s is the arc-length. This provides the interpretation that the optimal value $f(\mu)$ is almost the median of I_1 along the level curve $\tilde{I}_2^{-1}(\mu)$, except for the weight $\|\nabla \tilde{I}_2(\mathbf{x}(s))\|^{-1}$. The closer the term $\|\nabla \tilde{I}_2(\mathbf{x}(s))\|$ is to being constant, the closer $f(\mu)$ is to the median of I_1 along the level curve.

4.3.4 Coarse-to-fine, Nested Iterative Optimization Approach

The closed-form photometric solver in Section 4.3.3 allows to simplify the complexity of the joint registration problem since the optimal f is efficiently computed for any given g . Hence, the search space of the solution (g, f) is reduced to that of the geometric transformation g alone. To jointly optimize the functional (4.4) with respect to g and f , we design a nested strategy that exploits this reduced search space: an iterative quasi-Newton method [155] searches for the parameters of the geometric transformation g (i.e., the “outer loop”) while, for each candidate g , the closed-form solution of (4.5) provides the optimal intensity transformation f (i.e., the “inner-loop”). The method is combined with a multiresolution, coarse-to-fine framework, where the reduced-search method is applied to refine the geometric parameters at each resolution level.

	Photometric model	
	Affine	Closed-form
L^1 norm	-	Proposed method
L^2 norm	Aguiar [44] Bartoli [45]	-

Table 4.1: Competing methods classification.

4.4 Experiments

The proposed joint registration method has been tested on different sets of experiments. In Section 4.4.2, we demonstrate the capabilities of the proposed technique to handle non-linear photometric transformations and non-affine geometric transformations, outperforming the state-of-the-art methods summarized in Section 4.4.1. Then, Section 4.4.3 shows the benefits of considering non-linear photometric transformations over the simplistic “gain and bias” model, and we show how the non-linear transformation recovered by our method can be used in High Dynamic Range (HDR) imaging applications. Section 4.4.4 shows the performance of our method to register images “in the wild”, without ground truth, in a publicly available dataset [152] comprising multiple scenarios. Finally, Section 4.4.5 describes some experiments on aerial imagery.

4.4.1 Competing Joint Registration Methods

For comparison of the proposed method in the experiments, we have analyzed the most related and publicly available state-of-the-art joint, direct registration methods [44, 45]. The first one, proposed by Aguiar, is an iterative, multi-resolution method that can handle parametric geometric distortions, but was validated only on affine warpings [44]. The second strategy, a dual inverse compositional method proposed by Bartoli, is a gradient-based method designed for geometric and photometric transformations that commute, and can handle non-affine warpings [45].

(a) Florida and Africa images (I_2).

(b) Non-linear photometric distortions (f).

(c) Sample distorted images (I_1).

Figure 4.4: *Experiment with ground-truth data.* Fig. 4.4a shows two images (Florida and Africa) that are used for the experiments with ground-truth. The images in Fig. 4.4a and the non-linear photometric transformations in Fig. 4.4b are used to generate multiple image pairs (I_1, I_2) with different geometric warps to test registration methods. Two sample generated images I_1 are shown in Fig. 4.4c, with a red square of 512×512 pixels indicating the part of the images used in the experiments.

As mentioned in Section 2.2.2.1, both methods aim at minimizing the L^2 -norm of image differences and tackle photometric distortions using the “gain and bias” (affine) model, $f(\mu) = a\mu + b$, where a is the gain and b is the bias. Therefore, both methods should, in principle, recover the same minimizer (i.e., registration parameters) provided that they are initialized sufficiently closed to it. In practice, this is not the case in the code released by [44, 45]. The code implementing Aguiar’s method has two limitations: it only accounts for affine warpings and it only estimates the “gain” of the photometric model. Thus, it does not minimize what is proposed in the paper [44]. Additionally, Bartoli’s method prioritizes efficiency over accuracy, and often does not converge to the true minimizer of the L^2 objective function.

Because we want to focus on comparing the true minimizers of the approaches (the most accurate registrations) rather than the results of particular implementations suffering from limitations or trade-offs, we have implemented (using the minimization strategy in Section 4.3.4) an algorithm to minimize the L^2 -norm with affine photometric model. This provides a better comparison between the objective function aimed at by [44, 45] and the proposed one (4.4): the differences in the upcoming results are exclusively due to the error criterion and the photometric model used. For clarity, Table 4.1 depicts a classification of the mentioned methods in terms of the objective function and photometric model used.

4.4.2 Experiments with Ground-truth Transformations

In this section, we show the capabilities of our method to handle non-linear photometric variations and parametric geometric distortions. As parametric motion model, we use homographies, described by 8 degrees of freedom (DOF), since they are more challenging than the usual affine motion model (6-DOF). As photometric distortion, we use a monotonic, non-linear function (e.g., Fig. 4.4b). Thus, in this section we use known (ground truth) transformations to generate pairs of images to be registered, and we compare the ability of the selected methods to recover such transformations.

More specifically, for simplicity, let one of the images in Fig. 4.4a be image I_2 , and let the photometric transformation f be the function in Fig. 4.4b below the selected I_2 . Next, we generate an homography g by displacing the four corners of I_2 randomly by certain amount and computing the transformation between the original and the displaced corners. It is a well-known fact of projective geometry that four

Figure 4.5: *Accuracy Evaluation. Experiment with Florida images.* Sensitivity with respect to the amount of distortion σ_n used to generate the ground truth homography g . Fig. 4.5a shows the mean Euclidean distance between the four corners of the image domain transformed with ground truth g vs. with the estimated homography $\hat{g}$. Fig. 4.5b displays the distance between photometric transformations (4.8). Different registration methods are indicated by the color of the curves: in blue, the method minimizing the L^2 error with affine photometric model; in red, the proposed method.

Figure 4.6: *Experiment with Africa images.* Sensitivity with respect to the amount of distortion σ_n used to generate the ground truth homography g , analogously to Fig. 4.5. On the left the geometric error, on the right the photometric one. The red curves correspond to our method, and the blue ones to the competing one.

non-colinear points in the plane uniquely determine a homography [156]. Each corner $\{(x_i, y_i)\}_{i=1}^4$ is displaced by adding zero-mean Gaussian noise of standard deviation σ_n (coordinate-wise). Both transformations f and g are applied to I_2 to generate the test image I_1 (Fig. 4.4c). Then, the central part of the image pair (I_1, I_2) (marked

in red in Fig. 4.4) is passed to the different compared methods as the image pair to be registered. Cropping is done to avoid using non-valid pixels that arise when applying the geometric transformation, since it is not realistic when comparing two independently acquired images. The experiment is repeated for different random homographies and values of σ_n . As σ_n increases, the homographies present larger spatial distortion, and, therefore, registration becomes more difficult. Figs. 4.5 and 4.6 show the results of this experiments using, respectively, the **Florida** and **Africa** images from Fig. 4.4a along with the photometric transformation below them depicted in Fig. 4.4b. The resulting error curves are averaged over $N_r = 100$ trials for each selected σ_n (and discarding the worst 10% trials to remove outliers).

The accuracy in recovering the geometric transformation as a function of the increasing amount of distortion (σ_n) is shown on Figs. 4.5a and 4.6a. The accuracy is measured using the mean Euclidean distance between the points in the domain transformed with (i) the ground-truth homography g and (ii) the recovered one $\hat{g}$ by the compared registration methods. The two curves in the plot correspond to the different registration methods: the proposed method (in red) and the competing one (Section 4.4.1) (in blue). The homography recovered with the proposed technique has a remarkable performance compared to previous methods. In the experiment with **Florida**, in Fig. 4.5a, a mean deviation of almost zero pixels for $\sigma_n \lesssim 50$ pixels has been obtained. This is not possible with the competing joint registration method, which can only recover the geometric transformation with a similar error if the distortion is smaller ($\sigma_n \lesssim 30$ pixels). Moreover, the performance in the second experiment with **Africa**, shown in Fig. 4.6a, is even better. Our method can recover the geometric transformation for $\sigma_n \lesssim 40$ pixels whereas the other method, only recovers the transformation for a small σ_n and the error rapidly increases.

The accuracy in recovering the photometric transformation is presented in Figs. 4.5b and 4.6b, also as a function of σ_n . We report the distance between the ground truth f and the estimated one $\hat{f}$ by any of the registration methods, using a weighted root-mean-square (RMS) criterion,

$$D(f, \hat{f}) = \sqrt{\int w(\mu) (f(\mu) - \hat{f}(\mu))^2 d\mu}, \quad (4.8)$$

where $w(\mu)$ is the statistical frequency in image I_2 with value μ , i.e., weighting the

error according to the relative frequency of occurrence of the intensity values μ . The distance $D(f, \hat{f})$ has the same units as f , i.e., intensity levels.

As shown in Figs. 4.5b and 4.6b, our method recovers the photometric parameters with a very small error (less than one intensity level on a scale of 256 levels) even for large amounts of the geometric distortion (e.g., $\sigma_n \lesssim 50$ pixels in the experiment with **Florida** images and $\sigma_n \lesssim 40$ pixels in the case of the **Africa** images). Additionally, Figs. 4.5b and 4.6b show that the “gain and bias” photometric model has limited approximation capacity, since, even for small values of σ_n , the photometric error is considerably larger than the one of the proposed method. Finally, if we observe the geometric and photometric accuracy for each experiment separately, we can see that in both experiments (and in both methods) the value of σ_n from which the photometric error increases coincides with the value at which the geometric error increases too, thus showing a coupling between geometric and photometric errors.

4.4.3 Recovering Non-linear Photometric Transformations

In this section we show that the assumption of an affine model for the photometric variations between images (used in [44, 45]) is not enough in some situations. To this end, we test the registration methods on pairs of images acquired with a static camera. As will be shown, even providing with the exact geometric alignment at initialization, the methods based on the affine photometric model fail to recover the joint registration. Moreover, it can happen that a geometrically misaligned solution is obtained in favor of an overall smaller joint registration error.

We use three pairs of images (1st and 2nd columns of Fig. 4.7) from the benchmark dataset in [152], which consists of different sequences that are frequently used for HDR imaging. The first pair of images (1st and 2nd rows of Fig. 4.7) are part of the **Grand Canal** sequence. The second pair (3rd and 4th rows of Fig. 4.7) correspond to the **Memorial** sequence, and the last pair of images (5th and 6th rows of Fig. 4.7) are part of the **Snowman** sequence. Both, the **Memorial** and the **Snowman** sequences are considered in [152] as the most challenging scenarios.

Figure 4.7: *Experiment to recover Non-linear Photometric Transformations.* The first two columns show pairs of input images to the algorithms (cf. Fig. 4.1). The remaining columns show the image absolute differences of two different methods, both using the same metric (L^1 norm) but different photometric models: affine vs. non-linear with closed-form (CF) solver. Results are displayed in false color using a color scale which goes from dark blue (minimum error) to red (maximum error) (see Section 4.4.3). Rows correspond to the Grand Canal, Memorial and Snowman sequences in [152]. Note that the range of the error scale is different for every row to better distinguish the errors, and in particular, the fourth row is in logarithmic scale.

To show the effect of using the affine vs. the proposed photometric model, we factor out the effect of having a different metric, and so, in this section we compare our method (4.4) against the one that minimizes the same criterion (L^1 -norm) but uses an affine photometric model². The registration results of these two methods are compared in the third and fourth columns of Fig. 4.7. The error of the absolute differences between the images is shown in a common color scale for each pair of images tested (each row), from blue (small error) to red (high error).

As observed in Fig. 4.7, the method with the affine photometric model cannot register the pair of **Grand Canal** images: it achieves a good result in terms of geometric alignment but fails at recovering the photometric transformation between the images (red and yellow areas in the first and second rows). In contrast, our method (as observed in the first row) does a better job at registering both images spatially and photometrically. It can be observed that, due to moving objects in the scene (e.g., the water ripples), registration is not perfect: errors from the moving objects leak onto the static ones (i.e., buildings), since the algorithm finds a smaller overall error in this way. Nevertheless, registration is still better than in the affine case. When the role of the reference and the capture images is exchanged (in the second row), we observe that in some parts of the images (in light blue) the intensity values cannot be recovered. This is so because the photometric function cannot produce different output intensity values for the same input (by definition of function). Therefore, if the sky has the same intensity in the captured image but different ones in the reference, that part of the image cannot be recovered.

In the **Memorial** images (third and fourth rows of Fig. 4.7), it is shown that the method with the affine photometric model cannot recover the photometric transformation more accurately than the proposed method. Besides, in the fourth row, we can see that the geometric transformation with the affine photometric model leads to a large geometric registration error. Registration is visualized using a logarithmic scale to emphasize the errors, and this lets us see how the dome is transformed from a circle into an ellipse, and the windows are misaligned (in light blue). In contrast, our method correctly recovers the alignment and produces a very good approximation of the photometric registration.

²The same, multiresolution, nested strategy of Section 4.3.4 is used, but with the closed-form solver for f replaced by a solver for the gain and bias of the affine model.

Figure 4.8: *High Dynamic Range (HDR) example*. The photometric transformation f estimated by our method can be used to produce better exposed pictures.

Finally, in the **Snowman** images (fifth and sixth rows of Fig. 4.7), the method with the photometric affine model fails at recovering photometric distortions. Nevertheless, the proposed method obtains an almost perfect joint registration in the fifth row, and in the sixth one it behaves in the same way as on the second row of **Grand Canal** images.

These experiments show that the proposed method can handle photometric changes, such as those due to exposure variation. The affine photometric model commonly used in other methods [44, 45] is not sufficient to properly represent such intensity changes and obtain good registrations. In contrast, using a photometric model that accounts for non-linear intensity variations, such as the one we propose, allows to obtain better, almost perfect, image registration results.

HDR Imaging

Moreover, the recovered photometric transformation allows to fuse the spatially aligned images to produce an HDR image, as shown in Fig. 4.8. Here, images I_1 and I_2 in Fig. 4.7 are fused using the estimated transformation f to establish correspondences between their intensities. As can be seen, I_1 presents a good image exposure everywhere except on the sky pixels, which are over-exposed. In contrast, I_2 presents the opposite situation: some parts are under-exposed whereas the sky is correctly acquired. Fig. 4.8b combines the well-exposed parts from each input image to produce

an HDR image where details are preserved in both dark and bright areas (cf. Fig. 4.8b to Fig. 4.8a).

The above example highlights the advantages of jointly modeling geometric and photometric transformations between images: in addition to producing a better image alignment, it recovers the information that enables HDR imaging. Registration methods based on designing similarity measures invariant to photometric changes do not produce HDR images in a single step. Instead, they require two: first, image alignment, and then, photometric alignment (i.e., resolving the invariance) to be able to merge intensities across images.

4.4.4 Joint Geometric and Photometric Registration “in the Wild”

In this section, we show experiments “in the wild”, without knowledge about ground-truth transformations or even without guarantee that the geometric and/or photometric distortions are among the type of transformations considered by our model. We compare our registration method against the state-of-the-art on 50 image pairs from sequences of the public dataset *Real Images* [152] that have been acquired with a NIKON D300 and a CANON EOS-1Ds Mark II. The image pairs include various scenarios, among them there are images with small/large camera motions, static/dynamic scenes, and small/large illumination changes. A sample of five different image pairs used in our experiments is shown in the first two columns of Fig. 4.9. The following column corresponds to the results obtained with the method than minimizes the L^2 -norm of the image difference and uses an affine photometric model (Section 4.4.1). The next one shows the results obtained with the proposed method. The color difference images displayed on these two columns are in the same style as those in Fig. 4.7. Finally, the last column shows the comparagrams [38] of the input images with the photometric transformation recovered with the proposed method (in red) and with the other method (in green).

Table 4.2 reports the accuracy of the above-mentioned registration methods, averaged over all forty tested image pairs in the previously described set. The accuracy of the registration methods is computed using the overlapping pixels of images I_1 and $\hat{I}_1$ by means of the Mean Absolute Error (MAE) and the RMSE, as proposed in [46], because the motion parameters and the photometric transformation between the im-

Figure 4.9: *Experiment “in the wild” (without ground truth)*. The five columns correspond to: sample image pairs from the dataset used (first two columns), results of joint geometric and photometric registration using the L^2 affine method (Section 4.4.2 and the proposed method (labeled as L^1 CF), and, finally, in the last column, the comparagram of the input images with the photometric functions recovered by both methods (L^2 Affine in green, and ours in red). Each row corresponds to a specific pair of images and their registration results, which are displayed using pseudo-color, as in Fig. 4.7.

ages are unknown. As observed from Table 4.2, the proposed method, which models the photometric transformation with a non-linear function (closed-form solver), outperforms, on the average values over the entire dataset, the methods that minimize the L^2 -norm with affine photometric model ([44, 45]) in both quality metrics (RMSE and MAE). The individual results obtained in each experiment are shown in table 4.3.

		L^2 -Affine	L^1 -CF (Ours)
RMSE	mean	20.33	12.12
	std	9.39	6.82
	max	46.28	28.65
	min	6.4	1.36
MAE	mean	14.9	5.18
	std	8.09	2.84
	max	38.39	12.8
	min	3.72	0.65

Table 4.2: Statistics of the registration errors (RMSE and MAE) on 40 images from the dataset [152].

In both tables, the best results on each row are highlighted in bold.

The results shown in Fig. 4.9 corroborate that the proposed method achieves better results than the L^2 -norm minimization of the difference with affine photometric model [44, 45]. Particularly, this can be observed in the comparagrams (last column). The comparagrams are represented in grayscale, with darker values indicating pairs of intensities that are more frequent. As it can be seen, the non-linear photometric model (closed-form solver, in red) produces a better fit to the dark areas than the affine model (in green).

In summary, these experiments show that modeling the photometric changes between the images as a monotonic, non-linear function is beneficial for image registration, not only to recover the appropriate geometric alignment of the images but also to be able to obtain a predicted image $\hat{I}_1$ as similar as possible to the reference one I_1 , which may be used for HDR rendering or other image merging purposes. In addition, we have shown that the proposed method is useful not only from a theoretical point of view and in controlled experiments but also in real scenarios (images with unknown geometric and photometric transformations), as those represented by the Real Images dataset [152].

4.4.5 Experiments on Aerial Imagery

We have also evaluated the proposed method on multiple pairs of aerial images containing critical situations (e.g., poor textures or large illumination changes). All of them have been acquired using the same procedure as in Chapter 3 to simulate a

Identifier I_1/I_2	RMSE		MAE	
	L^2 -Aff.	L^1 -CF (Ours)	L^2 -Aff.	L^1 -CF (Ours)
Orson&Snow_1_0400.JPG /8000.JPG	46.28	9.59	38.39	6.20
Orson&Snow_1_1000.JPG /4000.JPG	25.31	5.67	21.45	3.73
Orson&Snow_1_4000.JPG /8000.JPG	8.99	2.21	7.14	1.60
SnowValley_1_0640.JPG /2500.JPG	17.8	4.20	15.1	2.68
SnowM_1_3200.JPG/8000.JPG	12.38	5.49	9.97	3.73
SnowValley_1_2500.JPG/8000.JPG	17.52	5.20	14.97	3.74
Family_1_0100.JPG/0200.JPG	13.04	10.72	9.01	4.73
Family_1_1600.JPG /4000.JPG	20.95	17.13	12.45	6.18
DSC_0002_1_05.jpg/0003_1_05.jpg	15.25	12.85	9.22	4.57
DSC_0003_1_05.jpg/0004_1_05.jpg	18.11	14.36	11.63	4.36
DSC_0004_1_05.jpg/0005_1_05.jpg	19.14	16.26	11.97	6.00
DSC_0006_1_05.jpg/0007_1_05.jpg	17.5	15.17	13.25	7.96
DSC_0001_1_06.jpg/0002_1_06.jpg	24.4	24.52	11.12	8.43
DSC_0004_1_06.jpg/0005_1_06.jpg	31.9	28.65	19.05	11.44
DSC_0006_1_06.jpg/0007_1_06.jpg	26.72	25.20	17.25	11.87
YC_01_0250.tif/01_0160.tif	14.02	16.18	11.46	12.8
wYC_1_100_11.tif/1_160_11.tif	7.98	8.05	3.72	2.70
YC_1_40_12.tif/1_60_12.tif	12.19	11.62	6.23	3.38
YC_01_0640_13.tif/1000_13.tif	6.93	1.36	5.74	0.65
YC_1_0100_13.tif/0160_13.tif	23.38	21.88	12.65	7.36
YC_1_0160_17.tif/0250_17.tif	41.7	24.68	30.77	8.15
YC_1_0100_18.tif/0160_18.tif	11.09	8.12	7.82	2.16
YC_1_0160_18.tif/0250_18.tif	18.54	16.7	11.28	4.94
YC_1_0160_20.tif/250_20.tif	12.99	9.12	9.21	3.18
YC_1_250_20.tif/0400_20.tif	13.18	11.73	9.31	4.52
YC_1_2500_20.tif/4000_20.tif	6.4	5.42	3.98	2.36
DSC_1_0640_21.JPG /2500_21.JPG	25.26	5.59	21.65	2.44
DSC_1_2500_21.JPG /5000_21.JPG	14.77	3.52	12.24	2.07
DSC_1_0640_21.JPG /5000_21.JPG	36.11	6.18	29.88	3.00
DSC_05_1_23.JPG /06_1_23.JPG	19.7	14.85	13.7	5.25
DSC_06_1_23.JPG /07_1_23.JPG	16.56	9.47	12.44	3.80
DSC_05_1_23.JPG /07_1_23.JPG	26.9	11.97	21.48	4.88
DSC_0048_1_24.JPG /0049_1_24.JPG	13.66	8.35	10.46	5.33
DSC_0047_1_24.JPG /0049_1_24.JPG	25.77	13.14	19.72	7.13
DSC_04_1_25.JPG /05_1_25.JPG	23.86	21.83	14.77	8.94
DSC_05_1_26.JPG /06_1_26.JPG	19.88	16.05	13.66	6.93
DSC_04_1_26.JPG /06_1_26.JPG	22.77	15.44	16.51	5.83
DSC_06_1_27.JPG /07_1_27.JPG	17.40	6.42	15.70	2.61
DSC_04_1_27.JPG /06_1_27.JPG	24.64	6.81	21.85	2.96
DSC_04_1_27.JPG /07_1_27.JPG	42.11	12.97	37.64	6.62

Table 4.3: Registration errors (RMSE and MAE) on 40 images pairs from the dataset [152] along with their corresponding identifier. The best results of each test are in bold. In addition, the results corresponding to images shown in Fig. 4.9, are marked in bold in in the first column, the order of appearance in the table is the same as in the figure.

Figure 4.10: Right column: sensed images. Left column: target images, with color notation: red (exact registration), blue (registration with histogram specification), green (registration with the proposed method).

UAV flying at altitudes between 6 and 10 km and using a downward-looking camera. For simplification, for testing purposes, the geometric warping applied to the images is limited to a rotation. In the performed experiments we have tested two types of illumination changes: (1) slight intensity changes due to the use of different camera sensor parameters (gain, shutter speed, etc.) and (2) natural illumination changes (e.g., day or night). The former is simulated by applying an arbitrary (e.g., linear) transformation and the latter is done by changing the acquisition settings.

Figure 4.10 illustrates some representative results comparing the estimations obtained by the coarse-to-fine nested optimization strategy using either histogram specification or the proposed method. The first row (Figure 4.10(a)) shows the results of applying the registration method to an image of an urban area that is rich in texture. It can be observed that both methods provide an accurate registration in this situation. Figure 4.10(b) depicts two images of a desert area, poor in texture. This example shows that the choice of the proposed method is more reliable than using histogram specification in these situations because an accurate registration is obtained whereas it is not with the later. Finally, in Figure 4.10(c), a scenario of two images with natural illumination changes is shown. In this case, the proposed method is also the only one that achieves an accurate registration.

4.5 Conclusion

In this chapter we proposed a joint geometric and photometric registration method that can deal with non-linear photometric changes as well as with parametric geometric distortions. We presented a formulation of the problem based on matching the isophote curves of the images that leads to an L^1 -norm error functional. Furthermore, we proposed a multiresolution, nested iterative method to simplify the solution of the joint registration problem by searching only over the space of geometric transformations, by devising a closed-form solver for the photometric distortion based on variational principles.

Additionally, we have compared the proposed method with other joint, direct registration methods publicly available. We showed that our method outperforms state-of-the-art methods in presence of non-linear photometric transformations and complex geometric distortions such as homographies. Moreover, we showed that the widespread “gain and bias” photometric model used in other methods falls short in some scenarios, whereas more complex photometric models, such as the ones we consider, produce better registration results.

Chapter 5

Augmented Reality for Operation Support

5.1 Motivation

The use of UAVs has become increasingly popular in recent years, especially because the majority of them are equipped with at least an EO sensor. Consequently, the use of these platforms has been considered for applications that traditionally use aerial imagery, such as surveillance [157] and remote sensing [158]. Additionally, UAVs have enabled the use of aerial images to improve tasks in which they were not previously used, such as structure inspection [159].

There are several categories of UASs, as described in Table 5.1: micro UAV, mini UAV, Tactical, MALE and HALE. The research work involving the use of UAVs is oriented to solve problems that are directly related to the type of used platforms. This is because the quality and accuracy of the sensors, as well as the mission specifications, are different for each one. There are studies for the autonomous navigation of micro aerial vehicles (MAV) [160], while others focus on target detection with platforms that can reach higher altitudes and can fly during longer periods of time [161].

MAVs can be operated by people without previous experience piloting aircrafts. On the contrary, MALE UAVs are usually operated by pilots of traditional manned aerial platforms that have adapted their expertise to control the aerial systems from a GCS. These MALE UAVs are usually operated in automatic or semi-automatic mode and are used in ISTAR missions. Therefore, the operator is simultaneously

UAS category	Description
Micro	Smallest available UAS systems. The UAV is small enough to be held in the palm of the hand, usually weighing less than a kilogram.
Mini	UAS systems designed to be launched and operated by a single person. The UAV is small enough to be lifted by a human.
Tactical	UAVs of several hundred kilograms, with an endurance of several hours and an operating radius of 200 km or less, based on line-of-sight (LOS) data links. They typically require more than one person to be deployed and operated.
MALE	Medium-Altitude Long Endurance UAVs with endurance of about 24 hours and long-range capability, generally used for operational reconnaissance.
HALE	High-Altitude Long Endurance UAVs with endurance of a day or more and long-range capability, generally used for strategic reconnaissance.

Table 5.1: UAS categories according to [1]

supervising the flight and manually operating the payload to fulfill the specific mission requirements.

The GCS has at least two different screens. One of them shows the mission plan, as it is represented in Figure 5.1a, and the other screen shows the video stream captured by the drone, as depicted in Figure 5.1b. In the example, the operator must accomplish a task as fast as possible: to determine the region of the images captured by the UAV (Figure 5.1b) where the target of interest is placed (illustrated in Figure 5.1c). However, such a task can be tedious and sometimes leads to misinterpretations when the targets are occluded. To this end, we propose to use AR to help the UAV operators.

We focus on AR solutions that merge virtual content over the video to increase the meaningful semantic context of the world by adding annotations. As a starting

(a) Screen where the mission planning is shown during the operation. In a standard GCS, the route (in green), the waypoints (in red), the targets (in blue) and the UAV (aircraft icon) position are shown.

(b) Raw video stream from the UAV, without any aid.

(c) Augmented video: video stream from the UAV camera, augmented with the proposed Augmented Reality (AR) tool.

Figure 5.1: Information available in the screens of the Ground Control Station (GCS) of Unmanned Aerial Vehicles (UAVs).

point, we have analyzed what kind of information is going to be available for our AR system. As a result of the analysis, besides the sensed information by the UAV (such as a video stream), we have as input some additional geographic information, the mission planning and also information about the targets.

As we have pointed in Section 2.3.5.1 determining the position and pose of the camera is necessary. But in the context of this AR solution, we assume a proper operation of the UAV positioning and imaging sensors to focus on displaying virtual content on the images acquired by the UAV. In addition, after the review done on Section 2.3.5.2 discarding the photorealism as a goal of the AR system, we detected, as summarized in Section 2.3.6, that research done on situational awareness is very dependent on the context. Previous work is focused on small aerial platforms, or in

different tasks to accomplish such a take-off or landing, and they do not use standardized protocols and tools.

The purpose of this chapter is to improve the situational awareness of UAV operators by providing a tool with AR capabilities valid for MALE UAVs in reconnaissance missions, and to make the tool available to the public on a website [7]. Our system avoids the burden of fusing information that comes from different sources and is displayed on separate screens. Additionally, the tool complies with NATO standards, which increases its usability: the application can be executed in different GCSs that follow the same standards. The contribution of this work is shown by the development of different AR capabilities centered on:

- **Route:** Enhancement of the video stream with the UAV flight route.
- **Targets:** Identification and location of targets (as illustrated in Figure 5.1c), viewpoint-based classification of targets according to occlusion with respect to geographical information, target search and target detection confirmation.

The chapter is organized as follows. Section 5.2 describes the main functionalities of the AR tool. Section 5.3 describes the structure of the proposed AR system. Input data processing is explained in Section 5.3.1. Then, the AR solution is presented in Section 5.3.2 and the augmented video that is shown to the operators is detailed in Section 5.3.3. Finally, results and conclusions are presented in Sections 5.4 and 5.5, respectively.

5.2 Functionalities of the Augmented Reality (AR) Tool

The functionalities of the AR solution are oriented to improve the situational awareness of the UAV operators when they have to accomplish a specific mission. The proposed tool provides different functionalities centered on improving the awareness around two main elements: the route and the targets.

5.2.1 Awareness of the Flight Route

MALE UAVs are usually equipped with a gimbal camera sensor payload. The sensor can operate in automatic or manual mode. When the manual mode is activated, the

operator can steer the UAV camera with a joystick and point it in any direction. This freedom of movement gives flexibility to explore the world because the sensor is not limited to the flight or the downward-looking directions, but it can also lead to disorientation of the operator. After managing the payload for a while, it may be difficult for the operator to distinguish if the camera is aligned with the flight direction. The proposed AR solution overcomes that problem by superimposing the flying route on the video stream.

The visual assistance given to the operators allows their situational awareness without having to establish correspondences between the 2D map mission information and the video stream, which are shown on different screens. Operators become aware of the camera orientation, not in a global frame but with respect to the flying route at a glance. This can benefit the world exploration because they can infer the remaining time to visit a waypoint and what the next movements of the UAV will be. Besides, having the route superimposed in the video stream allows the operators to control whether the route is being followed as planned in advance or not, i.e., there is an anomaly.

5.2.2 Awareness of the Targets

During a mission, the UAV operator could be in charge of identifying some strategic positions. The strategic positions are usually a list of targets that are known in advance and should be monitored. Sometimes, it can be difficult for the operators to distinguish the exact position of these targets in the video stream if they are carrying out manual inspection operations. Besides, some targets can be detected by additional sensors (e.g., radar). If the targets are far from the UAV, the operators may not distinguish them easily unless they use the camera zoom (if available). This situation, i.e., looking for a target with a close-up view, reduces the situational awareness of the operators.

The proposed AR tool includes the following functionalities that give assistance to the operators to easily determine where the targets are in the video stream: target location and viewpoint-based classification, target identification, target search and target detection.

5.2.2.1 Target Location and Viewpoint-based Classification

The proposed AR tool overcomes this problem by superimposing virtual beacons on the acquired video stream. The tool informs the operators about the visibility of the targets with respect to the terrain. This reduces the impact of using the camera zoom because the operators can distinguish if the target is going to be visible when zooming in or if it is occluded by the terrain (e.g., by a mountain) and, consequently, they should wait to reach a more appropriate viewpoint along the UAV trajectory.

5.2.2.2 Target Identification

When additional information (e.g., specific names for labels or the type of target) is known, it is included in the virtual beacons by adding the label and the corresponding icon. The size of the labels that identify the target remains constant no matter the distance from the camera to the target. Labels are visible from far away distances so that operators can easily distinguish among them. Together with the label, the icon is added to each target according to standard MIL-STD-2525. The operators, familiar with this symbology, can easily identify the type of target in addition to the information given in the label. The size of the icons remains constant despite of the camera movement and is configured in accordance to the size of the label.

5.2.2.3 Target Search

Sometimes, manual search of a target is challenging because *(i)* the target should be found as soon as possible and *(ii)* the on-board camera can cover a wide area. A system to assist the operator with this task, that helps to reduce the time of search of a specific target, has been designed. It consists of an arrow pointing toward the direction that should be followed to find the target.

5.2.2.4 Target Detection

The GCS centralizes a wide variety of information that comes from different sources. Particularly, it can receive information of target detection with different systems, such as radar. The AR tool incorporates a communication channel to be aware of these target detection messages so that the operator is able to confirm the detection of a

target while operating the camera. When a target is reported as detected to the GCS its color is changed to inform the operator about this fact.

5.3 Structure of the AR Tool

The AR system is structured in three main modules, as it is shown in blue in Figure 5.2. The first part of the system is that of input data processing; this module encompasses the processing of data coming from the UAV (video and metadata) and the information available in the GCS, part of which is available before the mission begins. The AR solution module comprises the information of both real and virtual worlds as well as the establishment of the relationship between them to achieve visual coherence of the whole. Finally, the augmented video module is where the semantic meaning of the virtual elements is explained.

5.3.1 Input Data Processing

The input data of the AR tool is provided by the GCS. Two types of data are distinguished according to their availability: (i) mission planning data and (ii) mission execution data.

5.3.1.1 Mission Planning Data

Mission planning data are available before the flight is carried out. During mission planning, the route and the flying parameters are settled to allow the autonomous flight of the UAV. The proposed AR tool uses the route path and the digital terrain model from the Geographical Information System (GIS).

The route information is obtained from the mission plan, exported in an Extensible Markup Language (XML) file using the Common Route Definition (CRD) that is defined according to the standard interfaces of the UAV Control System for NATO UAV interoperability. The latitude and the longitude of each waypoint are obtained from the XML file. The XML file also encapsulates the information about the order in which the waypoints will be visited and the flight altitude when the UAV visits them.

Figure 5.2: Principal modules of the AR tool: input data processing, AR solution and augmented video. The input data module encompasses the processing of data coming from the GCS and the UAV. In this module, four different colors are used: yellow, for the data used as input to the virtual world; red, for data corresponding to mission planning; green, for the metadata to create the projection model; blue, for the images of the real world. The AR solution module is responsible for achieving coherence of real and virtual worlds. Finally, the augmented video module manages the information displayed to the UAV operator.

The terrain data available in the GCS is the Digital Terrain Elevation Data (DTED). This data consists of a grid of square pixels that contain height information as well as a header with geographic information used to geo-reference the data. The DTED used in the performed experiments is DTED-Level 2, which has a GSD of 30 meters.

5.3.1.2 Mission Execution Data

Mission execution data provides the information that the GCS receives from the UAV payload (i.e., camera, GPS, IMU) after the mission starts. During the flight, the GCS collects information through the datalink established with the UAV.

The proposed AR tool uses the payload data provided by a Motion Imagery Standards Board (MISB) stream file constructed from technologies approved by the MISB and referenced in the NATO Motion Imagery Standard STANAG 4609 [162]. Any Motion Imagery Standards Profile (MISP) compliant file for Full Motion Video must have three components [163]: (i) motion imagery; (ii) metadata; and (iii) a media container. The motion imagery is the essence of the file. It corresponds to the imagery obtained from the EO capture device. The imagery can be in compressed or uncompressed format. The metadata contains the information coming from other sensors (e.g., GPS, IMU) and is encapsulated in Key-Length-Value (KLV) form. Finally, the media container carries the motion imagery and the metadata in two possible ways: using MPEG-2 transport stream (TS) or using a Real-Time Transport Protocol (RTP). The former is the one used in the proposed AR tool.

Fig. 5.3 illustrates the file composition process. This file consists of motion imagery (1) compressed or not (5) and metadata (6) with absolute timestamps applied to both motion imagery and metadata (7), then they are combined into a unified MPEG-2 TS container (8).

The metadata is usually collected by the mission computer, although in some occasions more information, such as an operator command, can be included. In the end, the source is not relevant until the information is interpreted. For this reason, the most important thing is to know how the metadata is encoded. All the MISB metadata is encoded following the Society of Motion Picture and Television Engineers

Figure 5.3: Generation scheme of MISP compliant file for Full Motion Video. Information sources are: (1) motion imagery and (2-4) metadata. The imagery is processed (5) and multiplexed together with the metadata in a unique metadata packet (6), then a time stamp is given to both of them to be synchronized (7), and then both are combined in the container (8) or sent using RTP (9).

(SMPTE) KLV with the specifications given in MISB STD 0902.1 [164]. This document specifies the Minimum Metadata Set to enable the functionality required for the Situational Awareness Product for Intelligence, Surveillance and Reconnaissance (ISR) missions.

Each metadata package associated to an image has the same scheme, which is illustrated in Figure 5.4. The set starts with the 16-bytes Universal Label (UL) key (in green), it is followed by the length of the KLV packet (in purple) and a sequence of Tag-Length-Value (TLV) encoded data. The TLV consists of a byte with the metadata tag (in cyan), the length of the metadata package (in magenta), and the value itself depending on the tag data type (in orange). All the metadata as well as the bytes, are represented using big-endian encoding (with the most significant bit first). The information relative to the interpretation of each TLV is given in the document MISB Standard 0601.2 [165]. The tag of the TLV is a unique identification of the type

06	0E	2B	34	02	0B	01	01	0E	01	03	01	01	00	00	00
81	AE	02	08	00	04	60	50	58	4E	01	80	03	0A	4D	69
73	73	69	6F	6E	20	31	32	05	02	71	C2	06	02	FD	3D
07	02	08	B8	0A	08	50	72	65	64	61	74	6F	72	0B	07
45	4F	20	4E	6F	73	65	0C	0E	47	65	6F	64	65	74	69
63	20	57	47	53	38	34	0D	04	55	95	B6	6D	0E	04	5B
53	60	C4	0F	02	C2	21	10	02	CD	9C	11	02	D9	17	12
04	72	4A	0A	20	13	04	87	F8	4B	86	14	04	00	00	00
00	15	04	03	83	09	26	16	02	12	81	17	04	F1	01	A2
29	18	04	14	BC	08	2B	19	02	34	F3	30	1C	01	01	01
02	01	07	03	05	2F	2F	55	53	41	0C	01	07	0D	06	00
55	00	53	00	41	16	02	04	01	41	01	02	01	02	29	72

Figure 5.4: Example of a metadata Key-Length-Value (KLV) packet. It is formed by a key (in green), the length of the whole packet (in purple), and a sequence of metadata. Each metadata is identified by a tag (in cyan), the length of the data (in magenta) and the information itself (in orange). Grid patterned colors have the same meaning as the solid colors [164].

of encapsulated metadata, and the interpretation is reported in the first two columns of Table 5.2. The rest of the columns correspond to the values and interpretation of an example of KLV packet.

The KLV packets contain:

- information that remains constant throughout a mission: mission identification (tag 3), the type of image sensor (tag 11) and the coordinate system (tag 12).
- information that changes in time: the UNIX Time Stamp (tag 2), the camera sensor, the platform position and orientation.

This information is used to obtain the intrinsic and extrinsic camera parameters that are needed to create the projection matrix. The intrinsic camera parameters are obtained from the Horizontal and Vertical Field of Views (FOVs). They correspond to the metadata with tags 16 and 17, respectively.

Concerning extrinsic parameters, the position of the sensor in the world and the pose are needed. The former is given by the Sensor Latitude (tag 13), Sensor Longitude (tag 14) and the Sensor True Altitude (tag 15), and the latter is calculated from

Tag	Name	Value	Interpretation	TLV Hex Bytes
2	UNIX Time Stamp	1,231,798,102,000,000 ms	Mon Jan 12 2009 22:08:22 (UTC)	02 08 00 04 60 50 58 4E 01 80
3	Mission ID	Mission 12	Mission 12	03 0A 4D 69 73 73 69 6F 6E 20 31 32
5	Platform Heading Angle	0x71C2	159.9744 Degrees	05 02 71 C2
6	Platform Pitch Angle	0xFD3D	−0.4315251 Degrees	06 02 FD 3D
7	Platform Roll Angle	0x08B8	3.405814 Degrees	07 02 08 B8
11	Image Source Sensor	EO Nose	EO Nose	0B 07 45 4F 20 4E 6F 73 65
12	Image Coordinate System	Geodetic WGS84	Geodetic WGS84	0C 0E 47 65 6F 64 65 74 69 63 20 57 47 53 38 34
13	Sensor Latitude	0x5595B66D	60.17682296 Degrees	0D 04 55 95 B6 6D
14	Sensor Longitude	0x5B5360C4	128.42675904 Degrees	0E 04 5B 53 60 C4
15	Sensor True Altitude	0xC221	14190.72 Meters	0F 02 C2 21
16	Sensor Horizontal FoV	0xCD9C	144.5713 Degrees	10 02 CD 9C
17	Sensor Vertical FoV	0xD917	152.6436 Degrees	11 02 D9 17
18	Sensor Rel. Azimuth Angle	0x724A0A20	160.71921147 Degrees	12 04 72 4A 0A 20
19	Sensor Rel. Elevation Angle	0x87F84B86	−168.79232483 Degrees	13 04 87 F8 4B 86
20	Sensor Rel. Roll Angle	0x00000000	0.0 Degrees	14 04 00 00 00 00

Table 5.2: Example of Tag-Length-Value (TLV) packets contained in a Key-Length-Value (KLV) packet. The table shows: the TLV hexadecimal value (last column), the tag (first column) of the metadata (second column) and its value (third column), and the interpretation of the specific value (fourth column).

the Platform Heading Angle (tag 5), Platform Pitch Angle (tag 6), Platform Roll Angle (tag 7), Sensor Relative Azimuth Angle (tag 18), Sensor Relative Elevation Angle (tag 19) and Sensor Relative Roll Angle (tag 20).

The video stream with the motion imagery data should be parsed, decoded, and decompressed. We have used the Fast Forward MPEG (FFMPEG) libraries to read a User Datagram Protocol (UDP) stream, demultiplex the MPEG-2 TS into video and metadata, and perform video decoding. However, it is not possible to decode the KLV metadata format with them, so we have implemented a decoder to collect the information associated with each frame.

In addition to the video stream, the AR tool is also listening to the target report that the GCS receives under the Data Distribution System (DDS) protocol following the NATO STANAG 4586 [166].

5.3.2 AR Solution Module

The AR solution module encloses the procedures to achieve visual coherence of the whole scene, so that virtual elements can be included in the video stream in a natural way. Three essential elements of an AR system are distinguished: the real world information, the virtual world, and the relation between the real and virtual worlds.

5.3.2.1 Real World

Real-world information is obtained from the UAV in real time through the GCS encapsulated in the MPEG-2 TS, as it is explained in Section 5.3.1.2. Motion imagery that is captured by the gimbal camera of the drone is the canvas where the virtual elements will be displayed. In addition, the information corresponding to the acquisition of the images and the UAV position and orientation is needed to build the appropriate projection model.

5.3.2.2 Projection Model: Conversion between Coordinate Systems

The inclusion of virtual elements in a scene is done by rendering the projection of such elements on the images: virtual elements are defined in the same reference system as the scene and are then rendered on the image using a perspective projection model.

Let us consider that the scene is given in a Cartesian world reference system called Earth-Centered Earth-Fixed frame (ECEF) (Figure 5.5a). In this system, the

Figure 5.5: Geographic coordinate systems used: ECEF and NED [167].

Earth's center corresponds to the origin of the ECEF frame, the x -axis points to the intersection of the prime meridian with the Equator (point at $(0^\circ$ latitude, 0° longitude)); the y -axis points to $(0^\circ$ latitude, 90° longitude), and the z -axis points to 90° latitude along the Earth axis of rotation.

The perspective projection is given by the pinhole camera model [168], which may include lens distortion parameters. In this work, we consider that the lens distortion is negligible so that the projection model is solely described by a 3×4 projection matrix $\mathbf{P}$. In practice, the optics of the camera may be calibrated using an algorithm such as [169], so that lens distortion can be considered compensated. In homogeneous coordinates [168], a world point $\mathbf{X} = (X, Y, Z, 1)^\top$ projects onto the image point $\mathbf{x} = (x, y, 1)^\top$ according to $\mathbf{x} \sim \mathbf{P}\mathbf{X}$, where $\sim$ means an equality up to a non-zero scale factor. The projection matrix $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ consists of the intrinsic ($\mathbf{K}$) and the extrinsic camera parameters ($\mathbf{R}, \mathbf{t}$).

The intrinsic parameter matrix

$$\mathbf{K} = \begin{pmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.1)$$

comprises the focal lengths in horizontal and vertical directions (f_x and f_y , respectively) and the principal point (x_0, y_0) , assumed to be at the center of the im-

age. For an image of size $w \times h$ pixels (width $\times$ height), the principal point is at $(x_0, y_0) = (w/2, h/2)$. The focal lengths, f_x and f_y , may be calculated from the horizontal and vertical FOVs (FoV_x and FoV_y) and the image size as follows: $f_x = (w/2)/\tan(FoV_x/2)$ and $f_y = (h/2)/\tan(FoV_y/2)$.

The extrinsic camera parameters (translation $\mathbf{t}$ and rotation $\mathbf{R}$) provide the position and orientation (i.e., the pose) of the camera in the world, but they are not as straightforward to set as the intrinsic parameters. We assume that the metadata used to build the camera orientation and position matrix are error free. However, when working with real data, the provided extrinsic parameters will be subject to error, which could lead to virtual objects not being projected on the desired image location. Accurate pose estimation is a different problem from the one tackled here. The proposed AR tool assumes that the pose provided in the metadata is accurate enough for a correct projection of virtual elements.

The camera pose in the world reference system is obtained by interpreting the metadata, which is given in a different coordinate system: World Geodetic System 1984 (tag 12 in Table 5.2). This standard states that a world location is specified by its latitude, longitude and height with respect to an oblate spheroid that models the shape of the Earth. Thus, the position of the camera in the scene coordinate system is obtained by converting the sensor geodetic coordinates given in the metadata (latitude α , longitude β , and true altitude γ —tags 13–15 in Table 5.2) to the ECEF reference system:

$$\begin{aligned} X &= (N + \gamma) \cos(\alpha) \cos(\beta), \\ Y &= (N + \gamma) \cos(\alpha) \sin(\beta), \\ Z &= (N(1 - ec^2) + \gamma) \sin(\alpha), \end{aligned} \tag{5.2}$$

where ec and N are the eccentricity and prime vertical radius of curvature of the spheroid, respectively. Such parameters are given by

$$ec = \frac{a^2 - b^2}{a^2}, \quad N = \frac{a}{\sqrt{1 - ec^2 \sin^2(\alpha)}}, \tag{5.3}$$

where $a = 6378137$ and $b = 6356752.3142$ are the lengths, in meters, of the semi-major and semi-minor axes of the spheroid.

The orientation of the camera is obtained by concatenating the rotations that

Figure 5.6: On the left, the heading angle of the platform in the plane $N-E$. In the middle, the pitch angle of the platform with respect to the plane $D-N$. On the right, the roll angle of the platform in the $D-E$ plane

define (i) the orientation of the UAV's local navigation frame (NED) with respect to the world (ECEF) frame; (ii) the orientation of the UAV with respect to its NED frame; and (iii) the orientation of the camera with respect to the UAV. These three changes of coordinates are specified next.

The local geographic reference system used by the UAV is defined by the gravity direction and its perpendicular plane (i.e., plane parallel to the ground). It is called the North-East-Down (NED) system and is shown in Figure 5.5b. As it can be observed, a plane tangent to the Earth spheroid contains the North and East directions, and the Down (i.e., gravity) is normal to this plane. The center of this local geographic reference system coincides with the position of the sensor in the world. As illustrated in Figure 5.5b, the NED frame is constructed in two steps, by applying the longitude (β) and latitude (α) rotations to the $N_0E_0D_0$ frame defined by the vectors

$$\mathbf{n}_0 = (0, 0, 1)^\top, \quad \mathbf{e}_0 = (0, 1, 0)^\top, \quad \mathbf{d}_0 = (-1, 0, 0)^\top. \quad (5.4)$$

Recall that a frame is rotated by multiplying each of its basis vectors by the rotation matrix. Let the matrix describing the rotation of a point by an angle θ around axis $\mathbf{n}$ (in a right-handed way) be $\mathbf{R}_{\mathbf{n}}(\theta)$, which is given by Rodrigues' formula ([168], p. 585):

$$\mathbf{R}_{\mathbf{n}}(\theta) = (1 - \cos(\theta))\mathbf{n}\mathbf{n}^\top + \cos(\theta)\mathbf{I}_3 + \sin(\theta)[\mathbf{n}]_\times, \quad (5.5)$$

where $\mathbf{n}$ has unit norm, $\mathbf{I}_3$ is the 3×3 identity matrix, and $[\mathbf{n}]_\times$ is the cross-product matrix associated with $\mathbf{n}$. Hence, the NED frame $(\mathbf{n}, \mathbf{e}, \mathbf{d})$ is obtained from the

Figure 5.7: Azimuth of the sensor, given with respect to the platform reference frame.

$N_0E_0D_0$ frame (5.4) by concatenating two rotations (Figure 5.5b): $(\mathbf{n}_1, \mathbf{e}_1, \mathbf{d}_1) = \mathbf{R}_{\mathbf{n}_0}(\beta)(\mathbf{n}_0, \mathbf{e}_0, \mathbf{d}_0)$, and $(\mathbf{n}, \mathbf{e}, \mathbf{d}) = \mathbf{R}_{-\mathbf{e}_1}(\alpha)(\mathbf{n}_1, \mathbf{e}_1, \mathbf{d}_1)$.

Next, the rotation of the UAV with respect to the NED system is performed (Tags 5–7 in Table 5.2). The rotation angles of the platform are illustrated in Figure 5.6. First, the heading is rotated around $\mathbf{d}$, $(\mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0) = \mathbf{R}_{\mathbf{d}}(\text{heading})(\mathbf{n}, \mathbf{e}, \mathbf{d})$; then, the pitch is rotated around the rotated $\mathbf{e}$, $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1) = \mathbf{R}_{\mathbf{y}_0}(\text{pitch})(\mathbf{x}_0, \mathbf{y}_0, \mathbf{z}_0)$, and, lastly, the roll angle is rotated around the latest rotated $\mathbf{n}$ (i.e., $\mathbf{x}$), yielding $(\mathbf{x}_2, \mathbf{y}_2, \mathbf{z}_2) = \mathbf{R}_{\mathbf{x}_1}(\text{roll})(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1)$.

Finally, the rotation of the sensor with respect to the platform, as it can be seen in Fig. 5.7, is applied (Tags 18–19 in Table 5.2). A rotation of the sensor relative azimuth angle, $(\mathbf{x}_a, \mathbf{y}_a, \mathbf{z}_a) = \mathbf{R}_{\mathbf{z}_2}(\text{azimuth})(\mathbf{x}_2, \mathbf{y}_2, \mathbf{z}_2)$, is followed by a rotation of the relative elevation angle, $(\mathbf{x}_e, \mathbf{y}_e, \mathbf{z}_e) = \mathbf{R}_{\mathbf{y}_a}(\text{elevation})(\mathbf{x}_a, \mathbf{y}_a, \mathbf{z}_a)$.

In summary, the camera rotation matrix is $\mathbf{R} = (\mathbf{x}_e, \mathbf{y}_e, \mathbf{z}_e)$, and the camera translation is $\mathbf{t} = -\mathbf{R}\mathbf{C}$, where the optical center of the camera ($\mathbf{C}$) is the position of the camera in ECEF coordinates (5.2).

Figure 5.8: Initialization process of the flying route. The top of each subfigure shows the UAV situation with respect to the *closest* waypoint (in red). The bottom of each subfigure shows the color coding of the legs and waypoints that will be presented to the operator. The *upcoming* waypoint and the currently flown leg are always displayed in green.

5.3.2.3 Virtual World

The proposed AR tool has been developed with OpenSceneGraph [170] as a 3D rendering engine for the virtual world. This is an open source 3D graphics engine written in C++ that acts as an object-oriented wrapper for OpenGL. The virtual world is defined according to the WGS84 Earth model but in the ECEF reference system. Therefore, to place virtual elements in correspondence with their real-world positions, the latitudes, longitudes and altitudes are converted to the ECEF reference system [171].

Flying Route

The flying route is composed of two different entities: the waypoints and the legs (i.e., the part of the route between two waypoints). The positions of the waypoints are defined in the mission planning step, obtained from the XML file, and then transformed from WGS84 to the ECEF reference system. Each waypoint is represented by a semi-transparent sphere and its correspondent leg (a semi-transparent cylinder that

extends from one waypoint to another). The route can be interpreted as a directed graph $G = (W, L)$ where $W = \{w_0, w_1, \dots, w_i, \dots, w_{n-1}\}$ denotes the set of waypoints and $L = \{l_{0,1}, l_{1,2}, \dots, l_{i-1,i}, \dots, l_{n-2,n-1}\}$ denotes the directed legs (i.e., $l_{i-1,i}$ is the directed leg from w_{i-1} to w_i). The colors of the waypoints and legs change according to the UAV position during flight to give additional information to the operators. Three different colors are used: green, gray and white. They correspond to three different states of the waypoints and legs, respectively: *upcoming*, *visited*, and *not visited*.

Two different situations are considered in the algorithm that controls the waypoint state: (i) initialization and (ii) standard use case. The initialization step is not crucial if the video stream is received before take-off, but it becomes essential if the datalink is lost and the connection is re-established at any time during the flight. The route information is also taken into consideration to distinguish the connectivity among waypoints. Although a more complex color code could have been chosen, the one selected is a trade-off between giving enough relevant information to the operators and not overloading them.

Figure 5.8 illustrates the initialization process. The first step in this stage is determining the closest waypoint (highlighted with a dash red circle), w_{closest} , to the current position of the UAV (in blue). Once this is known, the next step is distinguishing whether the UAV is getting closer or getting further away from w_{closest} . If the UAV is getting closer, as it is shown in Figure 5.8a, w_{closest} is the *upcoming* one, so w_{closest} and the leg $l_{\text{closest}-1, \text{closest}}$ are rendered in green. Otherwise, if the UAV is getting further away, as it is illustrated in Figure 5.8b, w_{closest} is set as *visited* (in gray) and the next one, $w_{\text{closest}+1}$, is set as the *upcoming* one, so that the leg shown in green is $l_{\text{closest}, \text{closest}+1}$. Finally, when the *upcoming* waypoint is identified (i.e., the green one is selected), the previous ones in the route are settled as *visited* (in gray) and the others as *non-visited* (in white). During the standard case of use, a waypoint changes from *upcoming* to *visited* when the UAV is in a sphere of a predefined radius centered at the waypoint.

Targets

The targets are represented, as it is illustrated in Figure 5.9, using four different virtual entities: a cylinder, a cube, a label and a semi-transparent sphere. It can also include a fifth element, an icon, as it is shown on the right part of the figure. The

Figure 5.9: Virtual target beacons, composed by a post with a cube at the top, a label and a semi-transparent sphere at the bottom. On the right, an icon has been included next to the cube.

cylinder represents a post that starts at the target position and ends at the same latitude and longitude but with a higher altitude. At the top of the post, a cube is added to highlight where the targets are. Both the cylinder and the cube are colored in blue. Additionally, next to the cube, a label that contains information of the specific target is shown in red. Besides, it can be included an icon to allow the operator not only to distinguish where the targets are, but also to identify them. Finally, a semi-transparent red sphere is displayed at the bottom of the post. This sphere encloses the part of the image where the target should appear, and it is used to represent a region of uncertainty around the target.

The positions of the targets are obtained from an XML file and then transformed from the WGS84 to the ECEF reference system. The use of an XML provides the possibility to exchange information between different sensor types regardless of the specific format used in each one of them.

To achieve a correct visualization of the targets and therefore improve the situational awareness of the operators, it is paramount to take into account occlusions. To this end, the proposed AR tool incorporates the terrain information into the virtual world. The terrain is built from the DTED-Level 2 information of the area that is going to be flown over during the mission, which is established in the mission planning stage. Occlusions are computed on the fly using the 3D rendering engine. The

Figure 5.10: Augmented video with highlighted route (waypoint and legs) and four targets. Same notation for virtual targets as in Figure 5.9. Same notation for waypoints and legs as in Figure 5.8: the camera is looking at the last visited waypoint (hence, it is colored in grey, as in Figure 5.8b).

terrain information is stored in a polygon mesh defined with pointers to a vertex list. In particular, we have chosen Wavefront Object (OBJ) format [172], a useful standard for representing polygonal data in the ASCII form that is widely used in computer graphics. This format is chosen because it is not limited to a specific terrain model and it provides the possibility to add terrain models built using different methods such as computer vision techniques, like those in [2]. In addition to the terrain, the use of the OBJ format allows the incorporation of other modeled elements such as buildings, which can improve the situational awareness when the mission requires target identification in an urban environment.

5.3.3 Augmented Video

The result of fusing the real-world images and the virtual world elements with the correct projection model, as explained in Section 5.3.2, is an augmented video stream, as illustrated in Figure 5.10. The UAV operators reduce their workload by visualizing the information contained in the video: they can distinguish the route that the UAV is going to follow and the targets that must be monitored. The route has three different

colors: green, for the upcoming waypoint, gray for the visited ones, and white for the non-visited. The operators can easily infer from the colors the direction of flight. Indeed, they can also infer when a waypoint is going to be reached because the closer you get to the waypoint, the bigger it is displayed on the screen.

The targets can be identified even if the UAV is flying far away from them. The operators can obtain information about the target positions quickly by observing the video. Then, they can identify them with the labels.

5.4 Results

The proposed AR tool for improving drone operations has been tested in a GCS demonstrator at Airbus facilities in Getafe, Madrid, Spain. The input data of the application is a synthetic video and metadata stream following the NATO standard 4609 [162] transmitted through UDP protocol. An XML file containing mission planning information, following the CRD standards, is also provided. Additionally, digital terrain information and a list of targets that should be identified by the operator are given.

The AR tool has been tested during a mission that takes place in the south of Spain. The objective of the mission was the identification of several targets that were reported to the operator. The targets chosen for the test were buildings, and the operator had to check if the targets were actually present in the indicated locations. This assignment was framed in a reconnaissance procedure. The UAV followed a route that was predefined according to several restrictions (e.g., non-flying zones) during mission planning. The UAV was flown with an automated control system and the operator was responsible for the supervision of the flight, the alerts and the payload. The operator can control the camera sensor manually with a joystick. Several tests were carried out with different operators and some representative moments of the mission are discussed below. In addition, supplementary material in [7] include augmented videos generated with the AR tool.

5.4.1 Results on Route Orientation

Figure 5.11 shows four different frames of the video stream, augmented with the route during the flight. In Figure 5.11a, the next waypoint of the route as well as the leg

that is being currently followed are shown in green. The operator can easily infer from the image that there is enough time to inspect the environment and the UAV is then going to turn left. Therefore, it is important to explore the zone to the right of the UAV.

(a) Currently flown leg and next waypoint (in green).

(b) Same view as Figure 5.11a, after a short time.

(c) Next waypoint: the UAV starts turning left.

(d) The leg has changed the color after the UAV has turned to the left.

Figure 5.11: Route orientation. Four different moments of the video stream augmented with the proposed AR tool during a mission.

In Figure 5.11b, the same waypoint is shown after approximately half a minute. It can be seen that the waypoint is bigger than before, thus indicating that it is going to be reached shortly. Figure 5.11c shows the waypoint when the UAV is turning left. It can be seen that the UAV is turning left because of the white leg in the image.

Finally, Figure 5.11d shows, in green, the leg that is being currently followed. This leg corresponds to the one shown in white in Figure 5.11c. This illustrates that, as the UAV progresses, the color information changes accordingly to the situation, thus improving the situational awareness of the operator.

(a) Screen displaying the map with the mission plan: in green, the route followed by the UAV; in blue, the targets and two different UAV positions during the flight; in red, the labels.

(b) Augmented video (AR tool) from position A. Occluded targets are easy to locate using the virtual beacons.

(c) Augmented video from position B. Targets are not occluded. Beacons highlight location and uncertainty of the targets.

Figure 5.12: Target identification. Results of the AR tool, displayed on the screens of the GCS. Blue circles surrounding the targets have been superimposed on both images to mark their true position.

5.4.2 Results on Targets

5.4.2.1 Target Location and Viewpoint-based Classification

During the mission, the operator is in charge of identifying different targets, which correspond to buildings. Figure 5.12 illustrates how some of such targets are seen from different positions along the UAV route. In Figure 5.12a, the map with the route (in green) is shown. In the middle of the map, two targets labeled with “3” and “4” are depicted in red. The operator should find these targets with the camera.

However, in this example, the targets are placed over a mountainous terrain. Thus, an appropriate point of view should be found. The results from two different UAV positions are shown in Figure 5.12b,c. The targets seen from “position A” are displayed

on Figure 5.12b: the operator can infer from the image that the targets are occluded by the terrain because the post beacons do not end in semi-transparent red spheres; they end with the shape of the mountain. With this information, the operator is aware of the fact that the camera zoom is not going to improve the visibility, hence another point of view is needed. In contrast, the augmented video observed from “position B”, displayed in Figure 5.12c, allows the operator to note that such a position is appropriate to distinguish the targets since they are not occluded. In such a case, the semi-transparent red spheres are shown at the end of the posts, which indicate that the targets are visible from that point of view. Therefore, target inspection can be carried out from that perspective.

Finally, Figure 5.13 illustrates the difference between having the raw video stream and the augmented one. The images in this figure show how difficult it is for the operator to know if there are visible targets in the video stream. As it can be seen, the proposed tool significantly benefits the target search. The benefits of the tool have been validated by the GCS experts of Airbus. Their comments have been considered in order to improve the tool because they are aware of the operator needs. For further validation, the AR tool could be tested by UAV operators under some experimental cases in order to provide objective time search measurements.

5.4.2.2 Target Identification

Along the mission there are two other targets to identify. This time, instead of just being numerated as before, there is information of the type of target so, as it is illustrated in Fig. 5.14, next to the cube, a label that contains information of the specific target is shown along with an icon that depends on the target type. The sizes of the label and the icon remain constant during the mission. This allows the operator not only to distinguish where the targets are, but also to identify them.

5.4.2.3 Target Search

The AR tool incorporates a module which indicates to the operator where to point the camera to find a specific target if it is not in the field of view (i.e., there is no virtual information about it over the video stream). The module will show a green arrow pointing to one of the following directions: up, down, left and right. Those directions show the operator how to move the joystick that control the camera to

Figure 5.13: Difference between the raw video stream (a) and the augmented with the proposed AR tool (b) for distinguishing buildings in reconnaissance missions. Blue circles surrounding the targets have been superimposed on both images to mark their true position.

Figure 5.14: Augmented video with highlighted route (waypoint and legs) and two targets. Same notation for virtual targets as in Fig. 5.9.

Figure 5.15: Target search aid. A green arrow is displayed in the right part of the screen indicating that the target *tgt_2* will be found turning the joystick to the right.

find the target. The arrow is shown accompanied by the label of the searched target. Fig. 5.15 illustrates a moment of the mission when the operator is searching for target labeled with *tgt_2*, a green arrow is shown in the right part of the screen pointing to the right. With that visual aid, the operator knows that the target will be found by moving the joystick to the right, thus reducing the uncertainty and the time of search.

5.4.2.4 Target Detection

Regarding target detection, the AR tool adds a component to reveal which ones are included in the target detection report received by the GCS. The report, that is continuously received by the GCS, is processed in the solution and if the target is being augmented in the video, then the color of the beacon (post and cube) and the label is changed. An example of the two different color configurations is illustrated in Fig. 5.16. On the one hand, the target on the left has been reported as detected, so the beacon is shown in red, and the label, in yellow. On the other hand, the target on the right has not been reported and is presented with the beacon in blue and the target in red.

Figure 5.16: On the left, a detected target reported to the GCS is presented in red with a yellow label. On the right, a target which has not been reported is displayed in blue with the label in red.

5.5 Conclusions

An AR tool to improve the situational awareness of UAV operators during ISTAR missions with MALE UAVs has been presented. The tool is available on-line with test data in a public website [7]. The AR system provides information about the flying path, letting the operator know the direction of flight and the next waypoints to visit. Additionally, the targets are highlighted, allowing the operator to easily identify them. Moreover, the presence of occlusions is taken into account so that the operator can reduce the time to find them and prevent the use of the camera zoom when it is not necessary. Besides, an aid for searching a specific target and visual confirmation of target detection is included. The usability of the proposed AR tool is assured by the adoption of NATO standards for motion imagery, KLV for metadata and CRD for mission plans. Therefore, the tool is valid for any GCS that follows the same standards.

The performance of the AR tool has been tested in an Airbus GCS demonstrator, where it has been shown how the enhancement of the video stream with virtual elements avoids the burden of fusing information displayed in separate screens and improves the situational awareness of the UAV operators.

Chapter 6

Conclusions

6.1 Conclusions

In this thesis we have proposed solutions to overcome the challenges of providing the UAV operators with the appropriate information for improving their situational awareness during a mission. We have explored them from different perspectives: first, considering that the sensor measurements of the UAV are not reliable, and then under the assumption of precise sensor-location performance.

The first scenario examined exceptional situations, such as being under a GPS-jamming attack or flying over a GPS denied area. Those situations are not only exceptional but critical when the objective is to provide an appropriate situational awareness to the operators. The state-of-the-art review (Section 2.2) points out that geo-registration is the chosen strategy to deal with such situations. Geo-registration contributes to determine the UAV/vehicle localization in world coordinates. For UASs, a classification of the strategies was done according to the reference data used. Therefore, we addressed the problem of 3D-2D geo-registration when 3D data was used, and 2D-2D, when the geo-reference data was an image.

We found limitations in the geo-registration approaches when the 3D models were not available or deprecated because, in the literature, the multiview stereo methods that produced accurate dense 3D models were based on the principles of patch-based multiview stereo along corresponding epipolar lines (1D primitives on the image plane). To overcome the shortcomings of multiview stereo methods for geo-registration we proposed, in Chapter 3, a geo-registration pipeline which used a different densi-

fication algorithm. The multiview stereo method proposed used variational methods and multi-resolution to generate a dense disparity map, providing surface smoothness not only along epipolar lines but also across them, thus producing a dense and coherent surface model in the full 2D image domain. We accomplished two experiments simulating two different camera configurations for a MALE UAV. We concluded that the downward-looking setup was the best one because it produced a more uniform terrain sampling. The surface models obtained with the proposed method significantly improved the sparse reconstruction achieved with feature-based (i.e., sparse) SfM techniques. We provided the capability of creating a high-resolution DEM.

When considering 2D-2D geo-registration we detected that the majority of existing registration methods are feature-based. Therefore, they rely on finding enough feature matches on both images: the geo-referenced one and the acquired by the UAV. However, we detected some scenarios where feature-based approaches failed. Hence, we explored area-based registration methods to provide solutions in such challenging scenarios. Specifically, we wanted to also model changes of illumination, thus we proposed a direct image registration method in Chapter 4. The registration method was formulated in a generic framework, which allowed to extend its applicability to other fields where the registration is needed, besides aerial imaging.

We proposed a joint geometric and photometric registration method that can deal with parametric geometric changes and non-linear photometric distortions between the images. We presented a complete and principled problem formulation that puts in correspondence the isophote contours of the images to be registered. The formulation led to an L^1 -norm error functional. Moreover, to reduce the complexity of the joint registration problem, we described a closed-form solver for the photometric transformation based on the analysis of the necessary optimality condition of the proposed functional. Additionally, we proved in the experiments that our method outperformed state-of-the-art methods in the presence of complex parametric geometric distortion (e.g., homographies) and in presence of non-linear photometric transformation. Furthermore, we revealed that in some scenarios the assumption of the “gain and bias” photometric model commonly used in the literature is not appropriate, and more complex photometric models are needed, such as the ones we considered, to produce better registration results.

When considering the second scenario of the thesis, we examined the augmented reality (AR) systems because AR constitutes a powerful technique to improve the

situational awareness. As it was reviewed in Section 2.3, the understanding of the 3D scene and the localization of the user within it are challenging. However, we have assumed in this second scenario that the pose with respect to the scene is known with enough accuracy (e.g., provided by a visual-inertial state estimation system with geo-referenced data), so that we could focus on exploring techniques aiming at an easy interpretation of the acquired, and possibly augmented, information. Each AR system is designed for a particular task and, when dealing with aerial platforms, they usually take into account small UAVs and do not use standardized protocols. Therefore, in this thesis, we developed an AR system for situational awareness in outdoor scenarios for MALE aerial platforms in ISTAR missions, and we used standardized protocols, thus facilitating the deployment in real-world ground control stations (GCS). The AR tool was presented in Chapter 5 and it provided information about the path planning stage and the targets over the video stream acquired from the UAV during a mission, in real time. The usability was achieved by the adoption of NATO standards, KLV metadata format and CRD for mission planning, making it valid for any GCS that followed the same specifications. Moreover, the AR tool was tested in an Airbus CGS demonstrator, where the improvement of the situational awareness of UAV operators was confirmed by the GCS expert team.

6.2 Future work

During the development of the thesis the research work has been driven to solve relevant problems within the geo-registration strategies and augmented reality systems. However, there are some limitations of the work that can be improved following the guidelines listed below.

- Regarding the geo-registration done with 3D reference data, we proposed a multiview stereo method to densify the surface which enforces the smoothness of the surface. This is assumable when dealing with a terrain but it can be relaxed by testing with other regularizers that allow for the existence of discontinuities. In addition, we have not taken advantage of the temporal coherence of the images that are given to the SfM algorithm to find the sparse reconstruction. This could be of interest to simplify the search of feature correspondences. Besides, it will be desirable to test complete the geo-registration pipeline proposed by

using the generated 3D model for geo-registering a real UAV flight over the area. State-of-the-art visual odometry and SLAM methods, such as ORB-SLAM [173], SVO [174], REMODE [175], etc. could be adapted to also use geo-referenced data and provide the skeleton onto which dense terrain models could be built. The resulting pipeline could make use of accelerated hardware (GPU) to provide dense terrain and city models in real time, to be geo-referenced and used to update long-term DEMs.

- Regarding the geo-registration done between 2D images, we have proposed a general method of registration between any pair of images. We have searched for a method that contemplates global photometric changes. The method could be extended to consider local (spatially varying) illumination changes. In addition, the implementation and test of the method was focused on the accuracy of the registration but, a study of its viability for a geo-registration system for on-line localization should be explored by looking for a more efficient implementation. Besides, as it was the case in the 3D reference geo-registration method, we have limited the research to images, so a clear improvement would be the use of the temporal coherence, for example, the photometric changes recovered for one frame could be used as initialization for the next frame.
- Regarding the AR tool created for the improvement of the situational awareness of UAV operators, we have tested the proposed approach in a GCS demonstrator, but it would be adequate to do tests with real video acquired by a UAV. In addition, a module that not only considers fixed, but also moving targets would be the next step. A tracking system for the targets could be used as well as additional information of the areas where the likelihood of finding a position of interest are higher. Additionally, it would be interesting to relax the assumption of having precise sensor measurements and include a module to refine the UAV pose, such as GPS-visual-inertial pose estimation system. Moreover, this system can be integrated with a geo-registration module to allow AR even in GPS-denied areas.

Appendix A

First Variation of the Stereo Disparity Functional

We want to minimize the energy functional (3.1), so we make the functional depend on t and we compute its derivative:

$$\frac{dE}{dt} = \frac{d}{dt} (E_{\text{data}} + \alpha E_{\text{smooth}}) = \frac{dE_{\text{data}}}{dt} + \alpha \frac{dE_{\text{smooth}}}{dt} \quad (\text{A.1})$$

Data fidelity term

The derivative of the data fidelity term (3.2) is

$$\begin{aligned} \frac{dE_{\text{data}}}{dt} &= \frac{d}{dt} \int_{\Omega} \frac{1}{2} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2))^2 d\mathbf{x}_1 \\ &= \int_{\Omega} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{d}{dt} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) d\mathbf{x}_1 \\ &= \int_{\Omega} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{\partial}{\partial \lambda} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{d\lambda}{dt} d\mathbf{x}_1 \\ &= \int_{\Omega} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) (-1) \frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} \frac{d\lambda}{dt} d\mathbf{x}_1. \end{aligned} \quad (\text{A.2})$$

Regularizer

We also differentiate the regularizer (3.3),

$$\begin{aligned}
 \frac{dE_{\text{smooth}}}{dt} &= \frac{d}{dt} \int_{\Omega} \frac{1}{2} \|\nabla \lambda(\mathbf{x}_1)\|^2 d\mathbf{x}_1 \\
 &= \int_{\Omega} \frac{1}{2} \frac{d}{dt} ((\nabla \lambda)^\top (\nabla \lambda)) d\mathbf{x}_1 \\
 &= \int_{\Omega} (\nabla \lambda)^\top \frac{d}{dt} (\nabla \lambda) d\mathbf{x}_1
 \end{aligned} \tag{A.3}$$

Assuming that λ is smooth enough ($\lambda \in C^2(\mathbb{R})$), by Schwarz's theorem we have the equality of mixed partial derivatives, so we can swap the order of the partial derivatives

$$\frac{dE_{\text{smooth}}}{dt} = \int_{\Omega} (\nabla \lambda)^\top \nabla \left(\frac{d\lambda}{dt} \right) d\mathbf{x}_1 \tag{A.4}$$

Before continuing, let us recall multivariable integration by parts. Given two functions $u(x^{(1)}, \dots, x^{(m)})$ and $v(x^{(1)}, \dots, x^{(m)})$ for which the following integrals are meaningful, integration by parts is defined as follows:

$$\int_{\Omega} \frac{\partial u}{\partial x^{(i)}} v d\Omega = - \int_{\Omega} u \frac{\partial v}{\partial x^{(i)}} d\Omega + \int_{\partial\Omega} u v \mathbf{N}^{(i)} d\partial\Omega, \tag{A.5}$$

where $\mathbf{N}^{(i)}$ is the i -th component of the outward unit normal to Ω , at the considered point, and $\partial\Omega$ is boundary of the fixed integration domain Ω .

So, let us do the following substitution

$$u \equiv \frac{d\lambda}{dt}, \quad v \equiv \frac{\partial \lambda}{\partial \mathbf{x}_1^{(i)}}, \quad x_i \equiv \mathbf{x}_1^{(i)} \quad i \in \{1, 2\}, \tag{A.6}$$

in the integration by parts formula (A.5):

$$\int_{\Omega} \frac{\partial}{\partial \mathbf{x}_1^{(i)}} \left(\frac{d\lambda}{dt} \right) \frac{\partial \lambda}{\partial \mathbf{x}_1^{(i)}} d\mathbf{x}_1 = - \int_{\Omega} \frac{d\lambda}{dt} \frac{\partial^2 \lambda}{\partial^2 \mathbf{x}_1^{(i)}} d\mathbf{x}_1 + \int_{\partial\Omega} \frac{d\lambda}{dt} \frac{\partial \lambda}{\partial \mathbf{x}_1^{(i)}} \mathbf{N}^{(i)} d\partial\Omega, \tag{A.7}$$

And combining both results ($i = 1$ and $i = 2$), if we substitute (A.7) in (A.4) we can move the derivatives from $\frac{d\lambda}{dt}$ to the other terms:

$$\begin{aligned}
\frac{dE_{\text{smooth}}}{dt} &= \underbrace{\int_{\Omega} (\nabla \lambda)^{\top} \nabla \left(\frac{d\lambda}{dt} \right) d\mathbf{x}_1}_{\text{area term}} \\
&= \underbrace{- \int_{\Omega} (\Delta \lambda) \frac{d\lambda}{dt} d\mathbf{x}_1}_{\text{area term}} + \underbrace{\int_{\partial\Omega} \frac{d\lambda}{dt} ((\nabla \lambda)^{\top} \mathbf{N}) d\partial\Omega}_{\text{boundary term}}
\end{aligned} \tag{A.8}$$

Recalling (A.1) and substituting the derivatives of both, the data fidelity term and the regularizer ((A.2) and (A.8)), we obtain

$$\begin{aligned}
\frac{dE}{dt} &= \int_{\Omega} (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) (-1) \frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} \frac{d\lambda}{dt} d\mathbf{x}_1 \\
&\quad - \alpha \int_{\Omega} (\Delta \lambda) \frac{d\lambda}{dt} d\mathbf{x}_1 + \alpha \int_{\partial\Omega} \frac{d\lambda}{dt} ((\nabla \lambda)^{\top} \mathbf{N}) d\partial\Omega \\
&= \underbrace{\int_{\Omega} (-1) \left(\alpha (\Delta \lambda) + (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} \right) \frac{d\lambda}{dt} d\mathbf{x}_1}_{\text{area term}} \\
&\quad + \underbrace{\alpha \int_{\partial\Omega} ((\nabla \lambda)^{\top} \mathbf{N}) \frac{d\lambda}{dt} d\partial\Omega}_{\text{boundary term}}.
\end{aligned} \tag{A.9}$$

Decreasing the energy

To decrease the energy we need $\frac{dE}{dt} \leq 0$. The question is: how to evolve $\lambda(\mathbf{x}_1, t)$ in time in order to decrease the functional $E(\lambda)$? The answer is choosing as the evolution rule of $\lambda(\mathbf{x}_1, t)$ in time

$$\frac{d\lambda}{dt} = \mu \left(\alpha \Delta \lambda + (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} \right) \tag{A.10}$$

being μ a small, non-negative scalar. Besides, to eliminate the boundary term in (A.9), we choose $(\nabla \lambda)^{\top} \mathbf{N} = 0$, that is, homogeneous Neumann boundary conditions, since $D_{\mathbf{N}} \lambda \equiv \frac{\partial \lambda}{\partial \mathbf{N}} \equiv \nabla \lambda \cdot \mathbf{N} = 0$.

Then, if we substitute the choices in (A.9) we get

$$\frac{dE}{dt} = -\mu \int_{\Omega} \left(\alpha \Delta \lambda + (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)) \frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} \right)^2 d\mathbf{x}_1 \leq 0. \quad (\text{A.11})$$

The steady state of (A.10) is (3.4), and so, we solve the necessary optimality condition given by the “elliptic” PDE by running the “parabolic” (A.10) until convergence (vanishing temporal derivative of the solution, λ).

Appendix B

Functional Measuring the Area Mismatch between Contours

Equation (4.1) can be rewritten as

$$\begin{aligned} E(\lambda, g) = & \int_{\Omega_1} H(\lambda - I_1(\mathbf{x})) H(\hat{I}_1(\mathbf{x}) - \lambda) d\mathbf{x} \\ & + \int_{\Omega_1} H(I_1(\mathbf{x}) - \lambda) H(\lambda - \hat{I}_1(\mathbf{x})) d\mathbf{x}, \end{aligned} \quad (\text{B.1})$$

so that, the functional in equation (4.2), after changing the order of integration, becomes

$$\begin{aligned} E(g) = & \int_{\Omega_1} \left(\int_{\lambda_{\min}}^{\lambda_{\max}} H(\lambda - I_1(\mathbf{x})) H(\hat{I}_1(\mathbf{x}) - \lambda) d\lambda \right. \\ & \left. + \int_{\lambda_{\min}}^{\lambda_{\max}} H(I_1(\mathbf{x}) - \lambda) H(\lambda - \hat{I}_1(\mathbf{x})) d\lambda \right) d\mathbf{x}. \end{aligned} \quad (\text{B.2})$$

Let us analyze the inner integrals in (B.2). For each point $\mathbf{x}$, the first term is non-vanishing if $I_1(\mathbf{x}) < \lambda < \hat{I}_1(\mathbf{x})$, whereas the second term is non-vanishing if $\hat{I}_1(\mathbf{x}) < \lambda < I_1(\mathbf{x})$. If the ranges of I_1 and I_2 are in the interval $[\lambda_{\min}, \lambda_{\max}]$, then both

terms simplify to

$$\begin{aligned}
 & \int_{\lambda_{\min}}^{\lambda_{\max}} H(\lambda - I_1(\mathbf{x})) H(\hat{I}_1(\mathbf{x}) - \lambda) d\lambda \\
 &= H(\hat{I}_1(\mathbf{x}) - I_1(\mathbf{x})) \int_{I_1(\mathbf{x})}^{\hat{I}_1(\mathbf{x})} d\lambda \\
 &= (\hat{I}_1(\mathbf{x}) - I_1(\mathbf{x})) H(\hat{I}_1(\mathbf{x}) - I_1(\mathbf{x})),
 \end{aligned} \tag{B.3}$$

$$\begin{aligned}
 & \int_{\lambda_{\min}}^{\lambda_{\max}} H(I_1(\mathbf{x}) - \lambda) H(\lambda - \hat{I}_1(\mathbf{x})) d\lambda \\
 &= (I_1(\mathbf{x}) - \hat{I}_1(\mathbf{x})) H(I_1(\mathbf{x}) - \hat{I}_1(\mathbf{x})),
 \end{aligned} \tag{B.4}$$

which are non-negative. Since, for each point $\mathbf{x}$, the above integrals cover mutually-exclusive cases, their sum is $(\text{B.3}) + (\text{B.4}) = |I_1(\mathbf{x}) - \hat{I}_1(\mathbf{x})| \geq 0$, and the total error (B.2) becomes (4.3).

Appendix C

First Variation of the Registration Functional

The functional (4.4) can be written in the form (4.2) using (B.1) with the new definition of $\hat{I}_1$ that includes f . Carrying out the change of variable $\lambda = f(\mu)$, $d\lambda = f'(\mu)d\mu$, then (4.4) becomes

$$E = \int_{\mu_{\min}}^{\mu_{\max}} \left(\int_{\Omega_1} H(f(\mu) - I_1(\mathbf{x})) H(\hat{I}_1(\mathbf{x}) - f(\mu)) d\mathbf{x} + \int_{\Omega_1} H(I_1(\mathbf{x}) - f(\mu)) H(f(\mu) - \hat{I}_1(\mathbf{x})) d\mathbf{x} \right) f'(\mu) d\mu.$$

Since $H(\hat{I}_1(\mathbf{x}) - f(\mu)) = H(f((I_2 \circ g)(\mathbf{x})) - f(\mu))$ and f is monotone, we have $H(f((I_2 \circ g)(\mathbf{x})) - f(\mu)) = H((I_2 \circ g)(\mathbf{x}) - \mu)$. So, letting

$$L(f, f', \mu) \doteq m(f, \mu) f'(\mu), \tag{C.1}$$

with

$$m(f, \mu) \doteq \int_{\Omega_1} H(f(\mu) - I_1(\mathbf{x})) H((I_2 \circ g)(\mathbf{x}) - \mu) d\mathbf{x} + \int_{\Omega_1} H(I_1(\mathbf{x}) - f(\mu)) H(\mu - (I_2 \circ g)(\mathbf{x})) d\mathbf{x},$$

we have

$$E = \int_{\mu_{\min}}^{\mu_{\max}} L(f, f', \mu) d\mu. \tag{C.2}$$

APPENDIX C. FIRST VARIATION OF THE REGISTRATION FUNCTIONAL

Having expressed the functional in standard form (C.2), its derivative is the left hand side of the Euler-Lagrange equation,

$$\frac{\delta E}{\delta f} = \frac{\partial L}{\partial f} - \frac{d}{d\mu} \left(\frac{\partial L}{\partial f'} \right). \quad (\text{C.3})$$

Deriving L in (C.1) yields $\frac{\partial L}{\partial f} = \frac{\partial m}{\partial f} f'$ and

$$-\frac{d}{d\mu} \left(\frac{\partial L}{\partial f'} \right) = -\frac{d}{d\mu} (m(f, \mu)) = - \left(\frac{\partial m}{\partial f} f' + \frac{\partial m}{\partial \mu} \right),$$

which allows to rewrite (C.3) as $\frac{\delta E}{\delta f} = -\frac{\partial m}{\partial \mu}$, where

$$\frac{\partial m}{\partial \mu} = \int_{\Omega_1} \text{sgn} (I_1(\mathbf{x}) - f(\mu)) \delta(\mu - (I_2 \circ g)(\mathbf{x})) d\mathbf{x},$$

thus producing the desired expression (4.6).

Bibliography

- [1] E. Atkins, A. Ollero, and A. Tsourdos, *Unmanned Aircraft Systems*. Wiley, 2017.
- [2] S. Ruano, G. Gallego, C. Cuevas, and N. García, “Aerial video georegistration using terrain models from dense and coherent stereo matching,” in *SPIE Int. Conf. Motion Imagery for ISR and Situational Awareness II*, vol. 9089, 2014, pp. 90 890V–1–10.
- [3] Aerial Video Geo-registration Using Terrain Models From Dense And Coherent Stereo Matching, <https://www.youtube.com/watch?v=k9wtS7JZ07U>.
- [4] S. Ruano, G. Gallego, A. Yezzi, C. Cuevas, and N. García, “Robust image registration with global intensity transformation,” in *2015 Int. Symp. on Consum. Electron. (ISCE)*, June 2015, pp. 1–2.
- [5] S. Ruano, G. Gallego, C. Cuevas, A. Yezzi, and N. García, “Joint geometric and photometric image registration based on contour dissimilarity. (under review),” *IEEE Trans. Image Process.*, 2018.
- [6] S. Ruano, C. Cuevas, G. Gallego, and N. García, “Augmented reality tool for the situational awareness improvement of uav operators,” *Sensors*, vol. 17, no. 2, p. 297, 2017.
- [7] Augmented Reality Tool for the Situational Awareness Improvement of UAV Operators, http://www.gti.ssr.upm.es/data/AR_SituationalAwareness.html.
- [8] M. D. Pritt and K. J. LaTourette, “Aircraft navigation by means of image registration,” in *Applied Imagery Pattern Recognition Workshop (AIPR): Sensing for Control and Augmentation, 2013 IEEE*. IEEE, 2013, pp. 1–6.

- [9] J. Zhang, W. Liu, and Y. Wu, “Novel technique for vision-based uav navigation,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2731–2741, 2011.
- [10] M. D. Pritt and K. J. LaTourette, “Automated georegistration of motion imagery,” in *2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 2011, pp. 1–6.
- [11] P. Doucette, J. Antonisse, A. Braun, M. Lenihan, and M. Brennan, “Image georegistration methods: A framework for application guidelines,” in *2013 IEEE Applied Imagery Pattern Recognition Workshop (AIPR): Sensing for Control and Augmentation*. IEEE, 2013, pp. 1–14.
- [12] T. McKay and H. Hirsch, “A fast, accurate, cross-modality image geo-registration and target/object detection algorithm,” in *Geospatial InfoFusion III*, vol. 8747. International Society for Optics and Photonics, 2013, p. 874708.
- [13] H. Eugster and S. Nebiker, “Geo-registration of video sequences captured from mini uavs: Approaches and accuracy assessment,” in *Mobile Mapping Technologies Symposium MMT*, vol. 1000, 2007, pp. 1–8.
- [14] T.-X. Yue, Z.-P. Du, D.-J. Song, and Y. Gong, “A new method of surface modeling and its application to dem construction,” *Geomorphology*, vol. 91, no. 1-2, pp. 161–172, 2007.
- [15] M. Shah and R. Kumar, *Video Registration, Chapter 8*, ser. The International Series in Video Computing. Springer US, 2003.
- [16] H.-P. Chiu, A. Das, P. Miller, S. Samarasekera, and R. Kumar, “Precise vision-aided aerial navigation,” in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 688–695.
- [17] L. Ding, A. Elliethy, and G. Sharma, “3d georegistration of wide area motion imagery by combining sfm and chamfer alignment of vehicle detections to vector roadmaps,” in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1487–1491.

BIBLIOGRAPHY

- [18] H. AliAkbarpour, K. Palaniappan, and G. Seetharaman, “Parallax-tolerant aerial image georegistration and efficient camera pose refinement - without piecewise homographies,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4618–4637, 2017.
- [19] K. Ni, N. Armstrong-Crews, and S. Sawyer, “Geo-registering 3d point clouds to 2d maps with scan matching and the hough transform,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1864–1868.
- [20] Q. Shan, C. Wu, B. Curless, Y. Furukawa, C. Hernandez, and S. M. Seitz, “Accurate geo-registration by ground-to-aerial image matching,” in *3D Vision (3DV), 2014 2nd International Conference on*, vol. 1. IEEE, 2014, pp. 525–532.
- [21] Q. Wang, L. Wu, S. Chen, D. Shu, Z. Xu, F. Li, and R. Wang, “Accuracy evaluation of 3d geometry from low-attitude uav collections a case at zijin mine,” *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, no. 4, p. 297, 2014.
- [22] K. J. LaTourette and M. D. Pritt, “Dense 3d reconstruction for video stabilization and georegistration,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 6737–6740.
- [23] B. Fan, Y. Du, L. Zhu, and Y. Tang, “The registration of uav down-looking aerial images to satellite images with image entropy and edges,” in *Intelligent Robotics and Applications*. Springer, 2010, pp. 609–617.
- [24] E. Molina and Z. Zhu, “Persistent aerial video registration and fast multi-view mosaicing,” 2014.
- [25] G. Conte and P. Doherty, “An integrated uav navigation system based on aerial image matching,” in *Aerospace Conference, 2008 IEEE*. IEEE, 2008, pp. 1–10.
- [26] B. Seema, H. Kumar, and V. Naidu, “Geo-registration of aerial images using ransac algorithm,” 2014.
- [27] G. Conte and P. Doherty, “A visual navigation system for UAS based on geo-referenced imagery,” *Linköping University, Sweden*, 2011.

- [28] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science and Business Media, 2010.
- [29] G. D. Evangelidis and E. Z. Psarakis, “Parametric image alignment using enhanced correlation coefficient maximization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1858–1865, 2008.
- [30] B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [31] P. Cheng and C.-H. Menq, “Real-time continuous image registration enabling ultraprecise 2-D motion tracking,” *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2081–2090, 2013.
- [32] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, “A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, 2014.
- [33] X. Shen, L. Xu, Q. Zhang, and J. Jia, “Multi-modal and multi-spectral registration for natural images,” in *Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2014, pp. 309–324.
- [34] J. P. Pluim, J. A. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: a survey,” *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, 2003.
- [35] A. Sotiras, C. Davatzikos, and N. Paragios, “Deformable medical image registration: A survey,” *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, 2013.
- [36] H. Rivaz, Z. Karimaghhaloo, V. S. Fonov, and D. L. Collins, “Nonrigid registration of ultrasound and mri using contextual conditioned mutual information,” *IEEE Trans. Med. Imag.*, vol. 33, no. 3, pp. 708–725, 2014.
- [37] J. Liang, X. Liu, K. Huang, X. Li, D. Wang, and X. Wang, “Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 603–615, 2014.

- [38] S. Mann, “Comparametric equations with practical applications in quantitative image processing,” *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1389–1406, 2000.
- [39] F. M. Candocia, “Jointly registering images in domain and range by piecewise linear comparametric analysis,” *IEEE Trans. Image Process.*, vol. 12, no. 4, pp. 409–419, 2003.
- [40] F. M. Candocia and D. A. Mandarino, “A semiparametric model for accurate camera response function modeling and exposure estimation from comparametric data,” *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1138–1150, 2005.
- [41] R. M. Frenkel and J. M. Francos, “Registration of geometric deformations in the presence of varying illumination,” in *IEEE Int. Conf. Image Process. (ICIP)*, vol. 3, 2007, pp. 489–489.
- [42] S. Z. Kovalsky, G. Cohen, and J. M. Francos, “Registration of joint geometric and radiometric image deformations in the presence of noise,” in *Statistical Signal Processing, 2007. SSP’07. IEEE/SP 14th Workshop on*, 2007, pp. 561–565.
- [43] S. Z. Kovalsky, G. Cohen, R. Hagege, and J. M. Francos, “Decoupled linear estimation of affine geometric deformations and nonlinear intensity transformations of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 940–946, 2010.
- [44] P. M. Aguiar, “Unsupervised simultaneous registration and exposure correction,” in *IEEE Int. Conf. Image Process. (ICIP)*, 2006, pp. 361–364.
- [45] A. Bartoli, “Groupwise geometric and photometric direct image registration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2098–2108, 2008.
- [46] H. Q. Luong, B. Goossens, A. Pižurica, and W. Philips, “Joint photometric and geometric image registration in the total least square sense,” *Pattern Recog. Lett.*, vol. 32, no. 15, pp. 2061–2067, 2011.
- [47] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, “Augmented reality technologies, systems and applications,” *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 341–377, 2011.

- [48] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.
- [49] D. Van Krevelen and R. Poelman, “A survey of augmented reality technologies, applications and limitations,” *International Journal of Virtual Reality*, vol. 9, no. 2, 2010.
- [50] C. Linte, J. White, R. Eagleson, G. Guiraudon, and T. Peters, “Virtual and augmented medical imaging environments: Enabling technology for minimally invasive cardiac interventional guidance,” *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 25–47, 2010.
- [51] C. Basdogan, M. Sedef, M. Harders, and S. Wesarg, “Vr-based simulators for training in minimally invasive surgery,” *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 54–66, 2007.
- [52] M. Sugimoto, “A mobile mixed-reality environment for children’s storytelling using a handheld projector and a robot,” *IEEE Transactions on Learning Technologies*, vol. 4, no. 3, pp. 249–260, 2011.
- [53] S. Matsutomo, T. Miyauchi, S. Noguchi, and H. Yamashita, “Real-time visualization system of magnetic field utilizing augmented reality technology for education,” *IEEE transactions on magnetics*, vol. 48, no. 2, pp. 531–534, 2012.
- [54] W. Broll, I. Lindt, I. Herbst, J. Ohlenburg, A.-K. Braun, and R. Wetzel, “Toward next-gen mobile ar games,” *IEEE Computer Graphics and Applications*, vol. 28, no. 4, 2008.
- [55] D. Roberts, S. Snarski, T. Sherrill, A. Menozzi, B. Clipp, and P. Russler, “Soldier-worn augmented reality system for tactical icon visualization,” in *SPIE Defense, Security, and Sensing*, vol. 8383. International Society for Optics and Photonics, 2012, pp. 838 305–838 305.
- [56] K. Chintamani, A. Cao, R. D. Ellis, and A. K. Pandya, “Improved telemanipulator navigation during display-control misalignments using augmented reality cues,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 29–39, 2010.

BIBLIOGRAPHY

- [57] S. LaValle, *Virtual Reality*. Cambridge University Press, 2017.
- [58] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE Transactions on Information Systems*, vol. 77, no. 12, 1994.
- [59] A. J. Fairchild, S. P. Champion, A. S. García, R. Wolff, T. Fernando, and D. J. Roberts, “A mixed reality telepresence system for collaborative space operation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 814–827, 2017.
- [60] R. T. Azuma, “A survey of augmented reality,” *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.
- [61] M. Harders, G. Bianchi, B. Knoerlein, and G. Szekely, “Calibration, registration, and synchronization for high precision augmented reality haptics,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 1, pp. 138–149, 2009.
- [62] S. Jeon, S. Choi, and M. Harders, “Rendering virtual tumors in real tissue mock-ups using haptic augmented reality,” *Haptics, IEEE Transactions on*, vol. 5, no. 1, pp. 77–84, 2012.
- [63] G. Klein and D. Murray, “Simulating low-cost cameras for augmented reality compositing,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 3, pp. 369–380, 2010.
- [64] M. Figl, C. Ede, J. Hummel, F. Wanschitz, R. Ewers, H. Bergmann, and W. Birkfellner, “A fully automated calibration method for an optical see-through head-mounted operating microscope with variable zoom and focus,” *Medical Imaging, IEEE Transactions on*, vol. 24, no. 11, pp. 1492–1499, 2005.
- [65] M. Mine, J. van Baar, A. Grundhofer, D. Rose, and B. Yang, “Projection-based augmented reality in disney theme parks,” *Computer*, vol. 45, no. 7, pp. 32–40, 2012.
- [66] K. Kiyokawa, “Trends and vision of head mounted display in augmented reality,” in *International Symposium on Ubiquitous Virtual Reality*. IEEE, 2012, pp. 14–17.

- [67] S. Kurkovsky, R. Koshy, V. Novak, and P. Szul, “Current issues in handheld augmented reality,” in *2012 International Conference on Communications and Information Technology (ICCIT)*. IEEE, 2012, pp. 68–72.
- [68] C. Menk and R. Koch, “Truthful color reproduction in spatial augmented reality applications,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 2, pp. 236–248, 2013.
- [69] “Project Tango,” <https://www.google.com/atap/project-tango/>.
- [70] “Microsoft Hololens,” <https://www.microsoft.com/microsoft-hololens/en-us>.
- [71] “Magic Leap,” <http://www.magicleap.com/>.
- [72] D. Bruellmann, H. Tjaden, U. Schwanecke, and P. Barth, “An optimized video system for augmented reality in endodontics: a feasibility study,” *Clinical oral investigations*, vol. 17, no. 2, pp. 441–448, 2013.
- [73] V. Valjus, S. Jarvinen, and J. Peltola, “Web-based augmented reality video streaming for marketing,” in *2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 2012, pp. 331–336.
- [74] M. A. Oikawa, I. de Souza Almeida, T. Taketomi, G. Yamamoto, J. Miyazaki, and H. Kato, “Augmented prototyping of 3d rigid curved surfaces,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 307–308.
- [75] I. Radu, “Why should my students use ar? a comparative review of the educational impacts of augmented-reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 313–314.
- [76] G. Yannakakis and J. Hallam, “Real-time adaptation of augmented-reality games for optimizing player satisfaction,” in *2008 IEEE Symposium On Computational Intelligence and Games*. IEEE, 2008, pp. 103–110.
- [77] S. Henderson and S. Feiner, “Exploring the benefits of augmented reality documentation for maintenance and repair,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 10, pp. 1355–1368, 2011.

- [78] B. Besbes, S. Collette, M. Tamaazousti, S. Bourgeois, and V. Gay-Bellile, “An interactive augmented reality system: A prototype for industrial maintenance training applications,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 269–270.
- [79] M. Hincapie, A. Caponio, H. Rios, and E. Mendivil, “An introduction to augmented reality with applications in aeronautical maintenance,” in *2011 13th International Conference on Transparent Optical Networks (ICTON)*. IEEE, 2011, pp. 1–4.
- [80] C. T. Yeo, T. Ungi, U. Paweena, A. Lasso, R. C. McGraw, G. Fichtinger *et al.*, “The effect of augmented reality training on percutaneous needle placement in spinal facet joint injections,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 7, pp. 2031–2037, 2011.
- [81] S. Nilsson, B. J. Johansson, and A. Jonsson, “Cross-organizational collaboration supported by augmented reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 10, pp. 1380–1392, 2011.
- [82] M. Nakamoto, K. Nakada, Y. Sato, K. Konishi, M. Hashizume, and S. Tamura, “Intraoperative magnetic tracker calibration using a magneto-optic hybrid tracker for 3-d ultrasound-based navigation in laparoscopic surgery,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 2, pp. 255–270, 2008.
- [83] T. Xia, S. Leonard, A. Deguet, L. Whitcomb, and P. Kazanzides, “Augmented reality environment with virtual fixtures for robotic telemanipulation in space,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2012, pp. 5059–5064.
- [84] N. Padoy and G. Hager, “3d thread tracking for robotic assistance in tele-surgery,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2011, pp. 2102–2107.
- [85] J. Wang, P. Fallavollita, L. Wang, M. Kreiser, and N. Navab, “Augmented reality during angiography: integration of a virtual mirror for improved 2d/3d visualization,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 257–264.

- [86] T. Blum, R. Stauder, E. Euler, and N. Navab, “Superman-like x-ray vision: Towards brain-computer interfaces for medical augmented reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 271–272.
- [87] R. Azuma, I. Neely, H., M. Daily, and J. Leonard, “Performance analysis of an outdoor augmented reality tracking system that relies upon a few mobile beacons,” in *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2006, pp. 101–104.
- [88] H. Oh, S. Kim, H.-S. Shin, A. Tsourdos, and B. A. White, “Behaviour recognition of ground vehicle using airborne monitoring of unmanned aerial vehicles,” *International Journal of Systems Science*, vol. 45, no. 12, pp. 2499–2514, 2014.
- [89] R. Schneiderman, “Unmanned drones are flying high in the military/aerospace sector [special reports],” *IEEE Signal Processing Magazine*, vol. 29, no. 1, pp. 8–11, 2012.
- [90] M. Suresh and D. Ghose, “Uav grouping and coordination tactics for ground attack missions,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 673–692, 2012.
- [91] M. E. Angelopoulou and C.-S. Bouganis, “Feature selection with geometric constraints for vision-based unmanned aerial vehicle navigation,” in *2011 18th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 2357–2360.
- [92] H. Wu, Z. Cai, and Y. Wang, “Vison-based auxiliary navigation method using augmented reality for unmanned aerial vehicles,” in *2012 10th IEEE International Conference on Industrial Informatics (INDIN)*. IEEE, 2012, pp. 520–525.
- [93] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr, “Tablet versus phone: Depth perception in handheld augmented reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 187–196.
- [94] T. Oskiper, S. Samarasekera, and R. Kumar, “Multi-sensor navigation algorithm using monocular camera, imu and gps for large scale augmented reality,” in *2012*

BIBLIOGRAPHY

- IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 71–80.
- [95] Y. Taguchi, Y.-D. Jian, S. Ramalingam, and Y. Feng, “Slam using both points and planes for hand-held 3d sensors,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 321–322.
- [96] W. Yii, W. H. Li, and T. Drummond, “Distributed visual processing for augmented reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 41–48.
- [97] H. Hile and G. Borriello, “Positioning and orientation in indoor environments using camera phones,” *IEEE Computer Graphics and Applications*, vol. 28, no. 4, 2008.
- [98] A. Cheok and L. Yue, “A novel light-sensor-based information transmission system for indoor positioning and navigation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 1, pp. 290–299, 2011.
- [99] J. Park, D. Lee, and J. Park, “Digital map based pose improvement for outdoor augmented reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 309–310.
- [100] Z. Zhu, H.-P. Chiu, T. Oskiper, S. Ali, R. Hadsell, S. Samarasekera, and R. Kumar, “High-precision localization using visual landmarks fused with range data,” in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 81–88.
- [101] A. Nemra and N. Aouf, “Robust ins/gps sensor fusion for uav localization using sdre nonlinear filtering,” *IEEE Sensors Journal*, vol. 10, no. 4, pp. 789–798, 2010.
- [102] J. Ventura and T. Höllerer, “Wide-area scene mapping for mobile visual tracking,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 3–12.
- [103] M. Livingston *et al.*, “Evaluating human factors in augmented reality systems,” *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 6–9, 2005.

- [104] O. Oda and S. Feiner, “3d referencing techniques for physical objects in shared augmented reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 207–215.
- [105] D. Kalkofen, E. Mendez, and D. Schmalstieg, “Comprehensible visualization for augmented reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 2, pp. 193–204, 2009.
- [106] S. Zollmann, R. Grasset, G. Reitmayr, and T. Langlotz, “Image-based x-ray visualization techniques for spatial understanding in outdoor augmented reality,” in *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. ACM, 2014, pp. 194–203.
- [107] C. Sandor, M. Fuchs, A. Cassinelli, H. Li, R. Newcombe, G. Yamamoto, and S. Feiner, “Breaking the barriers to true augmented reality,” *arXiv preprint arXiv:1512.05471*, 2015.
- [108] J. Chen, G. Turk, and B. MacIntyre, “A non-photorealistic rendering framework with temporal coherence for augmented reality,” in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 151–160.
- [109] E. Veas, R. Grasset, E. Kruijff, and D. Schmalstieg, “Extended overview techniques for outdoor augmented reality,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 4, pp. 565–572, 2012.
- [110] A. P. Gee, M. Webb, J. Escamilla-Ambrosio, W. Mayol-Cuevas, and A. Calway, “A topometric system for wide area augmented reality,” *Computers & Graphics*, vol. 35, no. 4, pp. 854–868, 2011.
- [111] J. Wither, S. DiVerdi, and T. Höllerer, “Annotation in outdoor augmented reality,” *Computers & Graphics*, vol. 33, no. 6, pp. 679–689, 2009.
- [112] T. Langlotz, C. Degendorfer, A. Mulloni, G. Schall, G. Reitmayr, and D. Schmalstieg, “Robust detection and tracking of annotations for outdoor augmented reality browsing,” *Computers & graphics*, vol. 35, no. 4, pp. 831–840, 2011.

BIBLIOGRAPHY

- [113] Z. Cai, M. Chen, and L. Yang, “Multi-source information fusion augmented reality benefited decision-making for unmanned aerial vehicles: A effective way for accurate operation,” in *2011 6th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2011, pp. 174–178.
- [114] M.-K. Tsai, Y.-C. Lee, C.-H. Lu, M.-H. Chen, T.-Y. Chou, and N.-J. Yau, “Integrating geographical information and augmented reality techniques for mobile escape guidelines on nuclear accident sites,” *Journal of environmental radioactivity*, vol. 109, pp. 36–44, 2012.
- [115] S. Zollmann, G. Schall, S. Junghanns, and G. Reitmayr, “Comprehensible and interactive visualizations of gis data in augmented reality,” in *Advances in Visual Computing*. Springer, 2012, pp. 675–685.
- [116] S. Zollmann, C. Hoppe, T. Langlotz, and G. Reitmayr, “Flyar: Augmented reality supported micro aerial vehicle navigation,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 4, pp. 560–568, 2014.
- [117] H. Wu, Z. Cai, and Y. Wang, “Vison-based auxiliary navigation method using augmented reality for unmanned aerial vehicles,” in *IEEE 10th International Conference on Industrial Informatics*. IEEE, 2012, pp. 520–525.
- [118] M. D. Pritt and K. J. LaTourette, “Georegistration of motion imagery with error propagation,” in *Full Motion Video (FMV) Workflows and Technologies for Intelligence, Surveillance, and Reconnaissance (ISR) and Situational Awareness*, vol. 8386. International Society for Optics and Photonics, 2012, p. 838606.
- [119] K. J. LaTourette and M. D. Pritt, “Dense 3d reconstruction for video stabilization and georegistration,” in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2012, pp. 6737–6740.
- [120] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
- [121] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3D Vision*. Springer Verlag, 2003.
- [122] O. Faugeras, Q.-T. Luong, and T. Papadopoulos, *The Geometry of Multiple Images*. The MIT Press, 2001.

- [123] T. Tuytelaars, K. Mikolajczyk *et al.*, “Local invariant feature detectors: a survey,” *Foundations and trends® in computer graphics and vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [124] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [125] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [126] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “Kaze features,” in *European Conference on Computer Vision*. Springer, 2012, pp. 214–227.
- [127] J. I. Ronda, A. Valdés, and F. Jaureguizar, “Camera autocalibration and horopter curves,” *International Journal of Computer Vision*, vol. 57, no. 3, pp. 219–232, 2004.
- [128] J. I. Ronda, A. Valdés, and G. Gallego, “Line geometry and camera autocalibration,” *Journal of Mathematical Imaging and Vision*, vol. 32, no. 2, pp. 193–214, 2008.
- [129] N. Snavely, S. M. Seitz, and R. Szeliski, “Modeling the world from internet photo collections,” *International journal of computer vision*, vol. 80, no. 2, pp. 189–210, 2008.
- [130] N. Herrero, J.-L. Landabaso, G. Gallego, and J.-C. Pujol-Alcolado, “In-loop feature tracking for structure and motion with out-of-core optimization,” in *2010 17th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2010, pp. 2937–2940.
- [131] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [132] E. Trucco and A. Verri, *Introductory techniques for 3-D computer vision*. Prentice Hall Englewood Cliffs, 1998, vol. 201.

BIBLIOGRAPHY

- [133] L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert, “Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach,” *Journal of Visual Communication and Image Representation*, vol. 13, no. 1-2, pp. 3–21, 2002.
- [134] G. Gallego, A. Yezzi, F. Fedele, and A. Benetazzo, “Two variational stereo methods for space-time measurements of ocean waves,” in *ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers, 2013, pp. V005T06A041–V005T06A041.
- [135] W. L. Briggs, S. F. McCormick *et al.*, *A multigrid tutorial, Second Edition*. Siam, 2000, vol. 72.
- [136] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.
- [137] G. Gallego, “Variational image processing algorithms for the stereoscopic space-time reconstruction of water waves,” Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, USA, 2011, directors: Yezzi, A. and Fedele, F.
- [138] M. Kazhdan, M. Bolitho, and H. Hoppe, “Poisson surface reconstruction,” in *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, ser. SGP ’06. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2006, pp. 61–70.
- [139] Google Earth, <http://www.google.com/earth/>.
- [140] The CGIAR Consortium for Spatial Information, <http://srtm.csi.cgiar.org/>.
- [141] J. Woo, M. Stone, and J. L. Prince, “Multimodal registration via mutual information incorporating geometric and spatial context,” *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 757–769, 2015.
- [142] M. A. Viergever, J. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. Pluim, “A survey of medical image registration—under review,” *Medical image analysis*, vol. 33, pp. 140–144, 2016.

- [143] Z. Li, D. Mahapatra, J. A. Tielbeek, J. Stoker, L. J. van Vliet, and F. M. Vos, “Image registration based on autocorrelation of local structure,” *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 63–75, 2016.
- [144] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, “Robust feature matching for remote sensing image registration via locally linear transforming,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, 2015.
- [145] N. Martinel, C. Micheloni, and G. L. Foresti, “Robust painting recognition and registration for mobile augmented reality,” *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1022–1025, 2013.
- [146] Y. Zhu, S. Cheng, V. Stanković, and L. Stanković, “Image registration using BP-SIFT,” *J. of Visual Communication and Image Representation*, vol. 24, no. 4, pp. 448–457, 2013.
- [147] M. Irani and P. Anandan, “About direct methods,” in *Int. Workshop on Vision Algorithms*. Springer, 1999, pp. 267–277.
- [148] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, “DTAM: Dense tracking and mapping in real-time,” in *Int. Conf. Comput. Vis. (ICCV)*, 2011, pp. 2320–2327.
- [149] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-scale direct monocular SLAM,” in *Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 834–849.
- [150] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 611–625, 2018.
- [151] H. Tuo, L. Zhang, and Y. Liu, “Multisensor aerial image registration using direct histogram specification,” in *IEEE Int. Conf. Networking, Sensing and Control*, vol. 2, 2004, pp. 807–812.
- [152] S. Wu, Z. Li, J. Zheng, and Z. Zhu, “Exposure-robust alignment of differently exposed images,” *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 885–889, 2014.
- [153] A. Yezzi, L. Zöllei, and T. Kapur, “A variational framework for integrating segmentation and registration through active contours,” *Medical Image Analysis*, vol. 7, no. 2, pp. 171–185, 2003.

BIBLIOGRAPHY

- [154] R. Gonzalez and R. Woods, *Digital Image Processing*. Prentice Hall, 2008.
- [155] J. E. Dennis, Jr and J. J. Moré, “Quasi-newton methods, motivation and theory,” *SIAM review*, vol. 19, no. 1, pp. 46–89, 1977.
- [156] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-d vision: from images to geometric models*. Springer Science & Business Media, 2012, vol. 26.
- [157] J. J. Acevedo, B. C. Arrue, I. Maza, and A. Ollero, “Cooperative large area surveillance with a team of aerial mobile robots for long endurance missions,” *Journal of Intelligent & Robotic Systems*, vol. 70, no. 1-4, pp. 329–345, 2013.
- [158] I. Colomina and P. Molina, “Unmanned aerial systems for photogrammetry and remote sensing: A review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 92, pp. 79–97, 2014.
- [159] K. Máthé and L. Buşoniu, “Vision and control for uavs: A survey of general methods and of inexpensive platforms for infrastructure inspection,” *Sensors*, vol. 15, no. 7, pp. 14 887–14 916, 2015.
- [160] K. Schauwecker and A. Zell, “On-board dual-stereo-vision for the navigation of an autonomous mav,” *Journal of Intelligent & Robotic Systems*, vol. 74, no. 1-2, pp. 1–16, 2014.
- [161] T. Oliveira, A. P. Aguiar, and P. Encarnação, “Moving path following for unmanned aerial vehicles with applications to single and multiple target tracking problems,” *IEEE Transactions on Robotics*, vol. 32, no. 5, pp. 1062–1078, 2016.
- [162] *STANAG 4609 Ed.3, NATO Motion Imagery, 13 Oct 2009*, http://www.gwg.nga.mil/misb/docs/nato_docs/STANAG_4609_Ed3.pdf.
- [163] *MISB TRM 0909.3, Constructing a MISP Compliant File/Stream, 19 Apr 2011*.
- [164] *MISB STD 0902.1 Motion Imagery Sensor Minimum Metadata Set, Jun 2010*.
- [165] *MISB Standard 0601.2, UAS Datalink Local Metadata Set*.
- [166] *STANAG 4586, Standard interfaces of UAV control system (UCS) for NATO UAV interoperability*.

- [167] D. Koks, “Using rotations to build aerospace coordinate systems,” DTIC Document, Tech. Rep., 2008.
- [168] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [169] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [170] “OpenSceneGraph,” <http://www.openscenegraph.com/>.
- [171] “Department of defense world geodetic system 1984, its definition and relationships with local geodetic systems, nima tr8350.2,” , Tech. Rep.
- [172] “WavefrontOBJ,” <http://www.fileformat.info/format/wavefrontobj/egff.htm>.
- [173] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “Orb-slam: a versatile and accurate monocular slam system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [174] C. Forster, M. Pizzoli, and D. Scaramuzza, “Svo: Fast semi-direct monocular visual odometry,” in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 15–22.
- [175] M. Pizzoli, C. Forster, and D. Scaramuzza, “Remode: Probabilistic, monocular dense reconstruction in real time,” in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2609–2616.