Fortschritt-Berichte VDI

VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Marco Munderloh,
Langenhagen

Nr. 847

Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain

Der Fakultät für Elektrotechnik und Informatik

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor-Ingenieur

genehmigte

Dissertation

von

Dipl.-Ing. Marco Munderloh

geboren am 21.09.1977 in Wilhelmshaven.

2015

Referent:	Prof. Dr.-Ing. J. Ostermann
Korreferent:	Prof. Dr.-Ing. C. Heipke
Tag der Promotion:	24.03.2015

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Ing. Marco Munderloh,
Langenhagen

Nr. 847

Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

Munderloh, Marco
Detection of Moving Objects for Aerial Surveillance of Arbitrary Terrain
Fortschr.-Ber. VDI Reihe 10 Nr. 847. Düsseldorf: VDI Verlag 2016.
114 Seiten, 59 Bilder, 8 Tabellen.
ISBN 978-3-18-384710-5, ISSN 0178-9627,
€ 48,00/VDI-Mitgliederpreis € 43,20.

Keywords: Aerial Surveillance – Motion Detection – Motion Segmentation – Non-planar Motion Compensation

The detection of moving objects in aerial video sequences is a common application in safety and environmental monitoring. The challenge is the non-static camera, which is moving together with an aerial vehicle. To detect local changes due to movement of ground objects in such a scenario, the displacements of image pixels resulting from the motion of the camera need to be compensated. The most common method is to use a projective transformation and assume the observed scene to be planar. However, this is only valid for very high altitudes. It fails otherwise and results in falsely detected local motion. This work addresses the problem in two ways. After analyzing the error resulting from motion parallax, two detectors for moving objects in non-planar scenes are presented. One is based on a motion parallax model and one on a smooth optical flow approach. Following this, a motion compensation method for non-planar scenes is presented, allowing the use of image differences based methods for non-planar scenes.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter http://dnb.ddb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at http://dnb.ddb.de.

Als Manuskript gedruckt. Printed in Germany.
ISSN 0178-9627
ISBN 978-3-18-384710-5

Acknowledgments

The time working at the Institut für Informationsverarbeitung (TNT) of the Gottfried Wilhelm Leibniz Universität Hannover to earn the degree Doctor of Engineering has been intensive and highly instructive. It had a big influence on my life and career. Especially the wide-ranging research topics at the institute allowed me an insight into many areas of coding and computer vision, and by their connection this work was originated.

My special thanks go to Prof. Dr.-Ing. Jörn Ostermann as the supervisor of my thesis for his support, his guidance and ideas, and for always taking the time for discussions. I also like to thank Prof. Dr.-Ing. Christian Heipke for being the external examiner and for his detailed review and helpful comments and Prof. Dr.-Ing. Bodo Rosenhahn as the chair of the examination board for his suggestions and hints during writing.

Furthermore, I thank all my colleagues at the institute. Our discussions helped me focusing and finding the right path to go. In particular, I like to mention Thorsten Laude, Holger Meuel, Hendrick Hachmann, and Julia Schmidt for their support and encouragement and Stella Graßhof, who was always open-minded for questions, especially concerning math.

Finally, a huge thanks goes to my family: my parents Heidrun and Hans-Gerd, my brother Timo and his wife Nicole, and of course to my wife Claudia and my son Martin Justus, whose smile always brings sunlight into cloudy days. I am very grateful for your understanding not having me there at so many evenings and weekends during the finalization of this thesis.

Contents

1 Introduction **1**
1.1 Motion Estimation and Parameter Extraction 4
1.2 Problems of Current Ground Motion Models 5
1.3 Contributions 8
1.4 Outline 9

2 Basic Principles **10**
2.1 Scene Model 10
2.2 Camera Model 11
2.3 Epipolar Geometry 19
2.4 Essential and Fundamental Matrix 20
2.5 Projective Transformation and the Homography 21
2.6 Moving Object Detection by Background Subtraction 23
2.7 Motion Estimation from Image Sequences 25
2.8 Dense Optical Flow 31

3 Planar Landscape Error Model **34**
3.1 Aerial Surveillance Scene Model 34
3.2 Vertical Aerial Photo 37
3.3 Motion Parallax Displacement 42
3.4 Camera with a Tilt Angle 46
3.5 Arbitrary Camera Orientation 51
3.6 Height Restrictions based Outlier Detection 54
3.7 Detectability in Dependence of Speed and Direction of Motion 57

4 Multi-Planar Landscape Model based on Triangle Meshes **58**
4.1 Mesh Creation 59
4.2 Outlier Removal and Moving Object Detection 61
4.3 Mesh-based Motion Compensation 63
4.4 Accuracy Analysis of the Mesh-based Approach 65

5 Experiments **73**
5.1 Motion Vector Classification 73
5.2 Motion Compensation Performance 82
5.3 Moving Object Detection Performance 85

6 Summary and Conclusions **93**

Bibliography **97**

Abbreviations and Symbols

Abbreviations:

BMA	Block Matching Algorithm
CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide Semiconductor
CRF	Corner Response Function
FOV	Field Of View
GLONASS	GLObalnaja NAwigazionnaja Sputnikowaja Sistema
GPS	Global Positioning System
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
KLT	Kanade-Lucas-Tomasi feature tracker
MAV	Micro Air Vehicle
MPP	Motion Parallax Predictor classifier
PTZ	Pan, Tilt and Zoom camera system
RANSAC	RANdom SAmple Consensus
ROI	Regions of Interest
SIFT	Scale-Invariant Feature Transform
SURF	Speed-Up Robust Features
UAV	Unmanned Aerial Vehicle

Symbols:

α	angle of aperture of the camera
β, γ, θ	pan, roll, and tilt angle of the camera
λ	parameter of a line equation
λ_1,λ_2	Eigenvalues of $\mathtt{M}$
$\lambda_{\mathrm{m1}}, \lambda_{\mathrm{m2}}$	parameters of the triangle plane equation
$\mathtt{A}$	affine matrix of size 2×2
$b_k(\mathbf{n})$	binarized image intensity differences of the frame k
$\mathbf{c}(c_x,c_y)$	principal point offset
$\mathbf{C}(C_x,C_y,C_z)^\top$	position of the camera in world coordinates
$\mathbf{C}_k$	position of the camera in the frame k
$\Delta\mathbf{C}$	vector between two camera centers

$\Delta\mathbf{c}, \tilde{d}_0$	motion parallax of the ground plane in image plane coordinates
D_L	diameter of the camera lens
D_P	distance of scene point $\mathbf{P}$ to the triangle surface
d_f	minimum feature distance
$d_k(\mathbf{n})$	image intensity differences of the frame k
$\mathbf{d}(d_x,d_y)^\top$	displacement vector
$\mathbf{d}_i(d_{i,x},d_{i,y})^\top$	displacement of the ith feature
$\hat{\mathbf{d}}$	estimate of $\mathbf{d}$
$\mathbf{e}$	position of an epipole
$e_k(\mathbf{n})$	binarized image intensity differences of the frame k after erosion
$\mathtt{E}$	elementary matrix of size 3×3
f	focal length
$\mathbf{f}_{i,k}$	position of the ith feature in the frame k
$\mathtt{F}$	fundamental matrix of size 3×3
$\mathbf{g}_{k-1}$	holds the temporal derivatives of I
h	height of a scene point above the ground plane
$h_{11}...h_{33}$	the elements of $\mathtt{H}$
$\mathtt{H}$	homography matrix of size 3×3
$I(\mathbf{n})$	image intensity at the position $\mathbf{n}$
$I_k(\mathbf{n})$	image intensities of the frame k
I_x,I_y	partial derivatives of I
k	frame index
k_H	Harris weighting factor
Δk	number of frames between the source and destination frame used for feature tracking and motion compensation
$\mathtt{K}$	camera calibration matrix of size 3×3
$\mathbf{l}$	epipolar line
$\mathtt{M}$	Harris corner matrix
$\mathbf{M}_{0,1,2}$	mesh node positions of a triangle in world coordinates
$\mathbf{m}_{0,1,2}$	mesh node positions of a triangle on the image plane
n	number of features
N_x,N_y	amount of sensor elements in x- and y-direction
$\mathbf{n}(n_x,n_y)^\top$	point in image coordinates
$\bar{\mathbf{n}}, \bar{\mathbf{n}}_0$	normal vector and unit normal vector of a triangle surface
$\mathbf{N}$	position of the nadir in world coordinates
$\Delta\mathbf{n}_c, \Delta\mathbf{n}_m$	relief displacement and motion displacement in pel
$\mathbf{O}$	projection of the image plane origin onto the ground plane
$\mathbf{p}(x,y)^\top$	point on the image plane

$\mathbf{p}'(x',y')^\top$	displaced position of the point $\mathbf{p}$
$\tilde{\mathbf{p}}(x_d,y_d)^\top$	point on the image plane with lens distortions
$\mathbf{p}_k$	point on the image plane of camera $\mathbf{C}_k$
$\hat{\mathbf{p}}_k$	estimate of $\mathbf{p}_k$ through affine motion compensation
$\mathbf{p}_0,k$	projection of $\mathbf{P}'_0$ onto the image plane of camera $\mathbf{C}_k$
$\mathbf{p}_h,k$	projection of $\mathbf{P}_h$ onto the image plane of camera $\mathbf{C}_k$
$\mathbf{P}(X,Y,Z)$	point in world coordinates
$\tilde{\mathbf{P}}(X_c,Y_c,Z_c)^\top$	point in camera coordinates
$\mathbf{P}_0, \mathbf{P}'_h$	point on the ground plane in world coordinates
$\tilde{\mathbf{P}}_0, \tilde{\mathbf{P}}'_h$	point on the ground plane in camera coordinates
$\mathbf{P}_h$	point on an object width height h in world coordinates
$\tilde{\mathbf{P}}_h$	point on an object width height h in camera coordinates
$\Delta\mathbf{p}$	relief displacement of $\mathbf{p}$
$\Delta\mathbf{p}_m$	motion parallax of $\mathbf{p}$
$\Delta\mathbf{P}$	relief displacement projected to the ground plane
$\Delta\mathbf{P}_m$	motion parallax projected to the ground plane
$\mathbf{q}(q_1,q_2)^\top, q$	projective components of the homography
r, r_d	radii of $\mathbf{p}$ and $\tilde{\mathbf{p}}$ to the center of distortion
r_{fps}	frame rate
$r_{11}...r_{33}$	the elements of $\mathtt{R}$
$\mathtt{R} = \mathtt{R}_\theta\mathtt{R}_\gamma\mathtt{R}_\beta$	camera orientation matrix of size 3×3
$r_k(\mathbf{n})$	pixel-wise motion detection results of the frame k
Δr_c	relief displacement in radial direction
ΔR_c	relief displacement projected to the ground plane in radial direction
s_w, s_h	width and height of the camera sensor
T_1, T_2, T_3	thresholds of the cluster filter
T_b, T_r	binarization and erosion thresholds of the noise filter
T_d	distance threshold of the motion parallax outlier detector
$t_{i,k}$	set referencing the mesh nodes of the triangle i in the frame k
$\mathtt{T}_{i,k}$	matrix containing the nodes of the triangle i in the frame k
$\mathbf{t}$	translation vector component of a homography
$u,v,\mathbf{u},\mathbf{v}$	arbitrary feature indexes and positions
$\mathbf{v}_{\text{plane}} = (v_x,v_y)^\top$	velocity and flight direction of aircraft
$\mathbf{v}_{\text{thresh}}$	minimal object speed needed for detection
W	search window set of pixels
Δx_c	relief displacement in x direction
ΔX_c	relief displacement projected to the ground plane in x direction

Abstract

The detection and segmentation of moving objects in aerial video sequences is a common application in safety and environmental monitoring. The challenge hereby is the non-static camera which is attached to and moves together with an aerial vehicle. To be able to detect local changes due to movement of ground objects in such a scenario, the displacements of the image pixels resulting from the motion of the camera need to be compensated between the frames of the recorded sequence. For this purpose, the motion of the camera as well as the structure of the recorded scene needs to be known to compensate the global motion without errors. While the motion of the camera can be measured accurately enough by external sensors or can even be estimated from the video feed itself, the structure of the observed scene is commonly unknown. Therefore, easy to compute universal approximations of the scene structure are made instead. The most common method is to model the global motion of the pixels by a projective transformation using a homography, in which the observed scene is assumed to be planar. While this might be accurate enough for high altitudes, small focal lengths, and vertically downwards oriented cameras, the approximation fails for scenes with high buildings or low altitudes due to motion parallax effects. As a result, the global motion for large areas of the frame is estimated and compensated incorrectly, which leads to lots of falsely detected local motion for such scenarios. In this work, the problem is addressed in two ways: first, the approximation errors made by the projective transformation for scenes with high buildings, low altitudes, and tilted or sideways looking cameras are analyzed mathematically. From the resulting aberration equations, a predictor for the motion parallax of image pixels is created and used as an outlier and moving object detector. It is called motion parallax predictor classifier is this work and able to distinguish between global motion of the background, displacements resulting from the motion parallax of static objects such as buildings, and local motion of individual objects in the scene. In contrast to similar methods, e.g. the elementary or fundamental matrix conformance test, only a small range on the epipolar line is determined as a valid representation of the possible motion parallax of static scene objects. This allows the detection of local moving objects moving along the epipolar line, which is not possible with epipolar geometry alone. However, the predictor has restrictions when it comes to objects moving along the epipolar line in the direction of the motion parallax: only objects with a displacement larger than the motion parallax are detectable. Moreover, the intrinsic and extrinsic camera parameters must be known, which requires external sensors and calibrated cameras. For this reason, an additional detector based on the clustering of frame to frame displacements of feature points (cluster filter) is developed in this work. It uses similarity constraints to join

the estimated displacement into clusters of equal motion. This allows the detection of local moving objects without explicit knowledge of the flight altitude, camera parameters and motion, or the scene geometry. Compared to the homography and fundamental matrix based methods used as references, the cluster filter as well as the motion parallax predictor classifier were able to classify up to 100% of the moving objects correctly. Moreover, the negative predictive value is increased from 30 to over 90% at the same time. The second way of addressing the problem is the global motion compensation of image pixels for the use in an image differences based system. As the investigated scenario does not conform with the single planar model of the homography, a multi-planar approach using a mesh of locally adaptive triangle patches is presented. In contrast to the homography, the mesh is able to adapt to objects sticking out of the ground plane by using an individual affine mapping for each triangle, e.g. the wall of a building and the roof. This allows the compensation of the global motion nearly error free, leaving only the newly occurring background as a possible source of false detections. Compared to the single planar model, the multi-planar approach was able to reduce the amount of falsely classified pixels in the experiments by a factor of 4.

Keywords: aerial surveillance, motion detection, motion segmentation, non-planar motion compensation

Kurzfassung

Das Erkennen bewegter Objekte in Luftbildsequenzen ist eine häufige Aufgabe in der Luftüberwachung. Die Herausforderung liegt hierbei in der Unterscheidung der globalen Verschiebung der Pixel zwischen den Bildern, hervorgerufen durch die Bewegung der Kamera, und lokalen Bewegungen durch die zu erkennenden Objekte. Um diese trennen zu können, muss die globale Bewegung kompensiert werden. Hierfür muss sowohl die Bewegung der Kamera als auch die Geometrie der überwachten Szene bekannt sein. Während sich die Bewegung der Kamera mittels externer Sensoren oder aus der Bildsequenz selbst heraus ermitteln lässt, ist die überflogene Szene meist unbekannt und wird daher unter Verwendung eines einfach zu bestimmenden Modells approximiert. Das meist genutzte Modell ist hierbei die Homographie, welche die überflogene Szene durch einer Ebene annähert. Diese Approximation gilt allerdings nur für große Flughöhen, kleine Brennweiten und lotrechte Aufnahmen. Entspricht die Landschaft nicht diesem Modell, z.B. wegen hoher Gebäude, niedriger Flughöhe, etc., führen die unterschiedlichen Bewegungsparallaxen zwischen der Oberfläche und Gebäuden zu Fehldetektionen in großen Bereichen des Bildes. In dieser Arbeit wird das Problem auf zweierlei Arten angegangen. In der ersten Methode wird zunächst der Approximationsfehler des Homographiemodells mathematisch bestimmt. Aus den sich ergebenen Fehlergleichungen wird ein Prädiktor erstellt, der die Bewegungsparallaxen von statischen Objekten bis zu einer vorgegebenen Maximalhöhe voraussagt. Der in dieser Arbeit Bewegungsparallaxeklassifizierer genannte Detektor erlaubt die Unterscheidung zwischen einer Bild-zu-Bild Verschiebung aufgrund statischer Objekte wie Hintergrund oder Gebäuden und einer lokalen Verschiebung durch ein sich bewegendes Objekt anhand des Abstandes zu einem prädizierten Epipolarliniensegment. Im Gegensatz zu Verfahren, die auf der Elementar- oder der Fundamentalmatrix basieren, erlaubt dieser Ansatz auch die Detektion von Bewegungen entlang der Epipolarlinie. Allerdings werden Objekte, welche sich in Richtung der Bewegungsparallaxe bewegen, nur erkannt, wenn die Eigengeschwindigkeit ausreichend hoch ist. Außerdem müssen für dieses Verfahren die intrinsischen und extrinsischen Kameraparameter bekannt sein. Aus diesen Gründen wurde ein zweiter Detektor entwickelt, welcher auf dem Clustern von Bewegungsvektoren basiert und als Clusterfilter bezeichnet wird. Das Vektorfeld wird hierbei anhand von Ähnlichkeitsbedingungen in Bereiche gleicher Bewegung eingeteilt, wobei innerhalb der Bereiche eine sanfte Änderung der Bewegungsrichtung erlaubt wird. Hierdurch wird eine Erkennung von Objekten unabhängig von der Bewegungsrichtung und ohne zwingende Kenntniss der Flughöhe oder der Kameraparameter ermöglicht. Sowohl der Bewegungsparallaxeklassifizierer als auch das Clusterfilter erreichen dabei eine Erkennungsrate be-

wegter Objekte auch für nicht planare Sequenzen von bis zu 100%, bei gleichzeitig niedrigerer Fehlalarmrate als das Referenzverfahren. Die zweite Methode hat als Ziel die Verbesserung der globalen Bewegungskompensation, welche z.B. für Detektoren notwendig ist, die auf Bilddifferenzen arbeiten. Im vorgestellten Verfahren wird hierbei das Einzelebenenmodell der Homografie durch ein Multiebenenmodell auf Basis von stückweise planaren Dreiecksnetzen ersetzt, in dem jedes Dreieck eine individuelle Ebene darstellt. Hierdurch ist es möglich, auch Objekte abzubilden, die aus der Grundebene herausstehen, in dem z.B. die Bewegung der Wand oder des Daches eines Gebäudes individuell bewegungskompensiert wird. Im Gegensatz zum Referenzverfahren lässt sich hierdurch die Anzahl der fälschlicherweise als bewegt erkannten Pixel dramatisch reduzieren, so dass Fehldetektionen nur noch an neu auftauchendem Hintergrund auftreten. Im Experiment ließen sich die Fehldetektionen um den Faktor 4 reduzieren.

Schlagworte: Luftbildüberwachung, Bewegungserkennung, Bewegungssegmentierung, nicht-planare Bewegungskompensation