Scale Estimation for Monocular SLAM using Depth from Defocus

by

Tomoyuki Shiozaki

A thesis submitted in partial fulfilment of the
requirements for the degree of Master of Engineering (Research)

at the
Centre for Autonomous Systems
Faculty of Engineering and Information Technology
University of Technology Sydney

March 2018

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signed: _____

Date: _____

Scale Estimation for Monocular SLAM using Depth from Defocus

by

Tomoyuki Shiozaki

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Engineering (Research)

Abstract

An autonomous robot must map its environment and estimate its egomotion to perform effectively. Monocular simultaneous localization and mapping (SLAM) can generate maps of the robot's environment, except for the absolute scale. Alternatives based on stereo or RGB-D camera based SLAM systems can obtain the metric scale but have disadvantages in terms of the cost, size and power requirements. This thesis is focused on the development of an absolute metric scale monocular SLAM system for autonomous robots. A depth from defocus (DfD) technique that relies on image blur is used to estimate the metric scale. However, existing methods for DfD suffer from ambiguities caused by texture, motion blur, and the location of the focal plane. The novelty of this research is combining DfD with camera motion to resolve estimation errors caused by these ambiguities and compute a reliable measure of metric scale. Monocular SLAM algorithms are also prone to scale drift, where the scale gradually changes while mapping. It is demonstrated that integrating DfD into monocular SLAM eliminates scale drift and results in accurate metric scale maps.

Acknowledgements

This work would not have been accomplished without the support and encouragement of many others around me. In the following lines, I would like to take this opportunity to show my appreciation to those people who have helped me in the realization of this thesis.

Especially, my deepest gratitude and appreciation goes to my primary supervisor, Professor Gamini Dissanayake for accepting me as a Master student and for providing the opportunity to work on this topic. He has been supervising me with his immense patience, motivation, enthusiasm, and expertise. Thanks for the countless hours of thought-provoking discussions.

I am deeply grateful to my alternate supervisor Doctor Ravindra Ranasinghe for his continuous support which was indispensable for the accomplishment of this thesis.

I would like to express my greatest appreciation to Professor Tomonari Furukawa for giving me the initial idea of performing this study and for his helpful advice on this research.

I would like to offer my special thank to the funding received through Canon Inc. and Canon Australia Pty. Ltd. to undertake my Master's degree. I would never have been able to reach where I am today, without their support.

My gratitude also goes to Mr. John Hazelton for proofreading my thesis and providing me with insightful comments and suggestions.

I would like to thank Mr. Kuranage Asok Aravinda Perera and Mr. Clyde Webster for reviewing my thesis and providing valuable advice. I have had the support and encouragement of them.

I am appreciative to all my colleagues at Centre for Autonomous Systems (CAS), I find it to have been an exciting opportunity for me to work with such an intelligent and motivated group of people.

Finally, and most importantly, I would like to thank my wife Hiromi and my two little children, Shogo and Mizuho. Their support, encouragement, patience, and love were undeniably the bedrock upon which the past two years of my life in Sydney have been built.

Contents

List of Figures

List of Tables

Acronyms & Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
BRIEF	Binary Robust Independent Elementary Features
CAS	Centre for Autonomous Systems
CoC	Circle of Confusion
CPU	Central Processing Unit
DfD	Depth from Defocus
DfF	Depth from Focus
EKF	Extended Kalman Filter
FAST	Features from Accelerated Segment Test
IMU	Inertial Measurement Unit
KLT	Kanade-Lucas-Tomasi
PSF	Point Spread Function
PTAM	Parallel Tracking and Mapping
ORB	Oriented FAST and Rotated BRIEF
RGB-D	Red, Green, Blue, and Depth

RMSE Root mean square error

ROI Region of Interest

SfM Structure from Motion

SLAM Simultaneous Localization and Mapping

UKF Unscented Kalman Filter

UTS University of Technology Sydney

Nomenclature

	General Notations
A	Aperture diameter of a lens
A_m	Amplitude of a step function
$\square^*$	Unconstrained $\square$ in the two-step projection method
B	Unknown offset
$\boldsymbol{b}$	Motion blur vector
b_δ	Distance from the image plane to a virtual plane where the rays from an out-of-focus point converge
b_f	Distance from the lens center to the image plane along the optical axis
c	Diameter of the circle of confusion
$\boldsymbol{c}[\cdot]$	Constraint function for equality state constraints
d	Distance from the lens center to a point along the optical axis
$D(\cdot)$	Depth-Defocus function
d_f	Distance from the lens center to a focal plane along the optical axis
Δt	Length of time between discrete steps
e_{thl}, e_{thh}	Threshold values to select features with strong edges
$\boldsymbol{\epsilon}_k = [\epsilon_{\Lambda_k}\ \epsilon_{\lambda_k^i}\ \epsilon_{\sigma_k^i}]^T$	Process noise vector of EKF
$\boldsymbol{\eta}_k = [\eta_k^i]^T$	Observation noise vector of EKF
$\boldsymbol{F}$	State transition matrix
f	Focal length of a lens
$G(\cdot,\cdot)$	Gaussian-shaped point spread function

γ	Camera specific constant to approximate c with σ
$\boldsymbol{H}$	Observation matrix
$\boldsymbol{H_c}$	Constrained matrix
$\hat{\square}$	Prediction value of $\square$
$I(\cdot,\cdot)$	Blurred image
$I_f(\cdot,\cdot)$	Sharp image
I_x	Gradient magnitude along x_e axis of $I(x_e, y_e)$
I_y	Gradient magnitude along y_e axis of $I(x_e, y_e)$
$\square^i$	i-th feature point
$\square^{i,j}$	i-th feature point seen from j-th keyframe
k	Time (discrete step)
Λ	Scale factor which defines the relationship between the metric map and the estimated geometry
λ	Texture correction factor for DfD
Λ_{ini}	Initial guess of Λ
λ_{ini}	Initial guess of λ
m	Number of keyframes
N	Number of observed points
N_c	F-number of a lens
ν	Innovation vector
$\boldsymbol{P}$	State covariance matrix
$\boldsymbol{p} = [x\ y\ z]^T$	Non-scaled feature location in camera coordinates
$\boldsymbol{p_w} = [x_w\ y_w\ z_w]^T$	Non-scaled feature location in world coordinates
$\phi_\square$	Calibration parameters for Depth-Defocus function
$\square^+$	First-step constrained $\square$ in the two-step projection method
$\boldsymbol{Q}$	Covariance matrix of the process noise
R	Gradient ratio of input image and reblurred image
$\boldsymbol{R}$	Covariance matrix of the observation noise
r	Index to evaluate the constancy of λ
$\boldsymbol{R_c}$	Covariance matrix of the noise or the extent of constraint violations

r_{thl}, r_{thh}	Threshold values to select features with constant λ
$\boldsymbol{S}$	Innovation covariance matrix
$\boldsymbol{S_c}$	Constrained innovation covariance matrix
σ	Standard deviation of Gaussian-shaped PSF
σ_b	σ for motion blur
σ_m	Measured σ for defocus blur
σ_{mb}	Composite σ of σ_m and σ_b
σ_r	Standard deviation of reblurred Gaussian-shaped PSF
smg	Index to evaluate edge strength
t	Time (continuous)
θ	Edge direction angle
$u(\cdot)$	Step function
$\boldsymbol{u} = [f_u \ f_v]^T$	Optical flow vector
v	Image velocity
$\boldsymbol{W}$	Filter gain vector
$\boldsymbol{W_c}$	Constrained filter gain vector
x_e	X-axis on an image where the edge is placed at $x_e = 0$
$\boldsymbol{x}_k = [\Lambda_k \ \lambda_k^i \ \sigma_{m,k}^i]^T$	State vector of EKF
y_e	Y-axis on an image where the edge is placed at $y_e = 0$
z	Non-scaled distance from the lens center to a point along the optical axis
$\boldsymbol{Z}_k = [z_k^i]^T$	Observation vector of EKF
z_{th}	Threshold value to select features nearby camera
$\boldsymbol{\zeta}_k = [\zeta_k^i]^T$	Noise vector for the extent of constraint violations

Operations

$\square * \square$	Convolution operator
$\lvert \cdot \rvert$	Absolute value
$\sqrt{\cdot}$	Square root
$\exp(\cdot)$	Exponential
$\square^2$	Square of $\square$

$\bigtriangledown$	Gradient operator
$\|\| \cdot \|\|$	Norm
$\dot{\square}$	First derivative of $\square$
$E[\cdot]$	Expectation
$\Sigma(\cdot)$	Summation
argmin$[\cdot]$	Argument of the minimum

State Transitions

$\square_{k-1\|k-1}$	Previous state
$\square_{k\|k-1}$	Predicted current state
$\square_{k\|k}$	Updated current state

Transforms

$\square^T$	Transpose
$\square^{-1}$	Inverse
$\boldsymbol{r}$	Rotation matrix of keyframe pose
$\boldsymbol{t}$	Translation vector of keyframe pose
$\boldsymbol{t}_t$	Ground truth of $\boldsymbol{t}$
$\boldsymbol{T} = [\boldsymbol{r}\|\boldsymbol{t}]$	Transformation matrix of keyframe pose