

UNIVERSITY OF TECHNOLOGY, SYDNEY

**Affordance-Map : Learning Hidden
Human Context in 3D Scenes Through
Virtual Human Models**

by

Jayaweera Mudiyansele Lasitha Chandana Piyathilaka

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Engineering and IT
Centre of Autonomous Systems

February 2016

Declaration of 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.

Signature of Student:

Date:

UNIVERSITY OF TECHNOLOGY, SYDNEY

Abstract

Faculty of Engineering and IT
Centre of Autonomous Systems

Doctor of Philosophy

by Jayaweera Mudiyansele Lasitha Chandana Piyathilaka

Ability to learn human context in an environment could be one of the most desired fundamental abilities that a robot should possess when sharing workspaces with human co-workers. Arguably, a robot with appropriate human context awareness could lead to a better human robot interaction. This thesis addresses the problem of learning human context in indoor environments by only looking at geometrics features of the environment. The novelty of this concept is, it does not require to observe real humans to learn human context. Instead, it uses virtual human models and their relationships with the environment to map hidden human affordances in 3D scenes.

The problem of affordance mapping is formulated as a multi label classification problem with a binary classifier for each affordance type. The initial experiments proved that the SVM classifier is ideally suited for affordance mapping. However, SVM classifier recorded sub-optimum results when trained with imbalanced datasets. This imbalance occurs because in all 3D scenes in the dataset, the number of negative examples outnumbered positive examples by a great margin. As a solution to this, a number of SVM learners that are designed to tolerate class imbalance problem are tested for learning the affordance-map. These algorithms showed some tolerance to moderate class imbalances, but failed to perform well in some affordance types.

To mitigate these drawbacks, this thesis proposes the use of Structured SVM (S-SVM) optimized for F1-score. This approach defines the affordance-map building problems as a structured learning problem and outputs the most optimum affordance-map for a given set of features (3D-Images). In addition, S-SVM can be learned efficiently even on a large extremely imbalanced dataset. Further, experimental results of the S-SVM method outperformed previously used classifiers for mapping affordances.

Finally, this thesis presents two applications of the affordance-map. In the first application, affordance-map is used by a mobile robot to actively search for computer monitors in an office environment. The orientation and location information of humans models inferred by the affordance-map is used in this application to predict probable locations of computer monitors. The experimental results in a large office environment proved that the affordance-map concept simplifies the search strategy of the robot. In the second application, affordance-map is used for context aware path planning. In this application, human

context information of the affordance-map is used by a service robot to plan paths with minimal distractions to office workers.

Acknowledgements

I am sincerely in debt to my supervisor, Associate Professor Sarath Kodagoda for believing in me and for giving me enough flexibility to work on my lifelong passion. His continuous support and guidance made this difficult endeavor achievable.

I would also like to thank Professor Massimo Piccardi for the fruitful discussions we had together. He always had enough time for me in his busy schedule.

To all the academic and support staff at CAS, thanks for taking me in and making me part of the family. Never have I felt more at home than in these past four years.

I would also like to thank Australian Government and University of Technology Sydney for awarding me with International Postgraduate Research Scholarship (IPRS) and Australian Postgraduate Award (APA). Without these scholarships I certainly would have not studied for a PhD.

Finally, I am ever in debt to my family. Specially my parents who made great sacrifices to make me who am I today. A very special thanks to my wife Uthsu for continuously believing me in and encouraging me in this difficult journey. Thanks my lovely daughter Sethmi for making my life joyful and meaningful.

Contents

Certificate Of Original Authorship	i
Abstract	ii
Acknowledgements	v
List of Figures	x
List of Tables	xii
Abbreviations	xiii
Nomenclature	xiv
Glossary of Terms	xvi
1 Introduction	1
1.1 Human Context	1
1.2 Affordances	2
1.3 Motivation to Learn Hidden Human Context	2
1.4 Objectives and Problem Statement	3
1.5 Contributions	5
1.6 Thesis Outline	6
1.7 Publications	8
2 Related Work and Background	9
2.1 Introduction	9
2.2 Related Work	9
2.2.1 Learning Human Context	9
2.2.2 Human Context and Object Affordances	11
2.3 Types of Classifiers	15
2.3.1 Binary Classifier	15
2.3.2 Multi Class Classification	15
2.3.3 Multi Label Classification	16

2.4	Evaluation Measures	16
3	Basic Classifiers for Mapping Affordances	18
3.1	Introduction	18
3.2	Naive Bayes classifier	19
3.2.1	Probabilistic Approach	20
3.2.1.1	Classification Rule	22
3.2.1.2	Parameter Estimation	22
3.2.2	Gaussian naive Bayes	22
3.2.3	Flexible Naive Bayes	23
3.2.4	Support Vector Machine (SVM)	25
3.2.5	Learning and Inference Process	28
3.2.6	Voxelising 3D Scenes	28
3.2.7	Feature selection	29
3.2.7.1	Virtual Human Skeleton Model	29
3.2.7.2	Distance and Collision Features	31
3.2.7.3	Normal Features	32
3.3	Experimental Setup	33
3.3.1	Dataset	34
3.3.2	Ground Truth Labels	35
3.3.3	Balancing Class Imbalance	36
3.3.4	Parameter Selection	36
3.4	Results	37
3.4.1	Qualitative Analysis	37
3.4.1.1	Sitting-Relaxing Affordance	37
3.4.1.2	Sitting-Working Affordance	41
3.4.1.3	Standing-Working Affordance	43
3.4.2	Quantitative Analysis	45
3.5	Discussion and Limitations	46
4	Evaluation of SVM Learners for Mapping Affordances with Highly Imbalance Datasets	48
4.1	Introduction	48
4.2	Soft Margin Optimization with Class Imbalance	49
4.3	Imbalance Learning Methods for SVMs	52
4.3.1	Data pre-processing methods	53
4.3.1.1	Re-sampling methods	53
4.3.2	Algorithmic Modifications	55
4.3.2.1	zSVM Algorithm	56
4.3.2.2	Different Cost Model	57
4.4	Discussion and Limitations	59
5	Structured SVM for Learning Affordance Map	62
5.1	Introduction	62
5.2	Affordance Learning with SVM Optimized for Performance measures	64

5.2.1	Efficient Learning Algorithm	66
5.2.2	Maximization Step at Inference	67
5.2.3	Maximization Step at Learning	68
5.2.3.1	Loss Function Based on Contingency Table	69
5.2.4	Experiments	71
5.3	Training	72
5.4	Experimental Results	73
5.4.1	Qualitative Analysis	73
5.4.1.1	Sitting-Relaxing Affordance	73
5.4.1.2	Sitting-Working Affordance	75
5.4.1.3	Standing-Working Affordance	78
5.4.2	Quantitative Analysis	81
5.5	Discussions and Limitations	81
6	Active Visual Object Search using Affordance-map	84
6.1	Introduction	84
6.2	Human Context for Object Search	84
6.2.1	Affordance-map for Object search	86
6.2.1.1	Human-Object Relationship	87
6.2.1.2	Human-robot Relationship	88
6.2.2	Experiments- Active Object Search	90
6.3	Conclusions	93
7	Affordance-map for context aware path planning	94
7.1	Introduction	94
7.2	Human Context in Path Planning	95
7.3	Related Works	96
7.4	Affordance-Map for Context-Aware Path Planning	97
7.4.1	A* for Path Planning	98
7.4.2	Detecting Human Presence	100
7.4.3	Context-aware Path Planning Algorithm	101
7.5	Experiments	102
7.6	Conclusions	106
8	Conclusions	107
8.1	Summary of Contributions	108
8.1.1	Affordance-map for Learning Hidden Human Context	108
8.1.2	Learning Affordances with Extreme Class Imbalances	109
8.1.3	Affordance-map for Active Object Search	109
8.1.4	Affordance-map for Context-aware Path Planning	110
8.2	Discussion of Limitations	110
8.3	Future Work	111

Appendices

112

Bibliography

113

List of Figures

1.1	A 3D living room scene	2
1.2	The affordance-map of a living room	4
2.1	Predictions results for ‘sittable’ locations in an office space by the algorithm discussed in [1]	13
2.2	A Human model inferred by the algorithm discussed in [2]. It is highly influenced by the features of the monitor and its location is not supported by a chair.	14
3.1	The effect of using a single Gaussian versus kernel method to estimate the density of a continuous variable	24
3.2	Steps in learning and inference stages	28
3.3	Types of affordances and their associated skeleton models	30
3.4	Virtual human models sitting near a computer	31
3.5	Surface Normals calculated from point clouds	33
3.6	Sample rooms from the dataset	34
3.7	For sitting-relaxing affordance, class labels are identified by manually marking possible areas of hip and leg locations of the skeleton models	35
3.8	Test room for the Sitting-Relaxing Affordance. Best viewed in colour . . .	38
3.9	Classification results of the Sitting-Relaxing Affordance	38
3.10	Decision Function values that indicate the confidence level of the predictions	39
3.11	Predicted 3D skeleton Map of the room	40
3.12	Test room for the Sitting-Working Affordance. Best viewed in colour . . .	41
3.13	Classification results of the Sitting-Working Affordance	42
3.14	Decision Function values that indicate the confidence level of the predictions	42
3.15	Predicted 3D skeleton Map of the room	43
3.16	Test room for the Standing-Working Affordance. Best viewed in colour . .	43
3.17	Classification results of the Standing-Working Affordance	44
3.18	Decision Function values that indicate the confidence level of the predictions	45
3.19	Predicted 3D skeleton Map of the room	45
4.1	The effect of class imbalance on the SVM classifier. K-fold cross validation results with average F1-score vs different class imbalances	50
4.2	Prediction results for the Sitting-Relaxing Affordance with different positive :negative class imbalance ratios	51
4.3	SVM learning with focus re-sampling	55

4.4	SVM learning with zSVM method	57
4.5	SVM learning with different cost models. K-fold cross validation results . . .	59
5.1	The F1-score improves at each iteration of the S-SVM training process . . .	72
5.2	Test room for the Sitting-Relaxing Affordance. Best viewed in colour . . .	74
5.3	Classification results of the Sitting-Relaxing Affordance	75
5.4	Decision Function values that indicate the confidence level of the predictions	75
5.5	3D skeleton map for the Sitting-Relaxing affordance	76
5.6	Test room for the Sitting-Working Affordance. Best viewed in colour . . .	76
5.7	Classification results of the Sitting-Working Affordance	77
5.8	Decision Function values that indicate the confidence level of the predictions	77
5.9	3D skeleton map for the Sitting-Working affordance	78
5.10	Test room for the Standing-Working Affordance. Best viewed in colour . . .	79
5.11	Classification results of the Standing-Working Affordance	79
5.12	Decision Function values that indicate the confidence level of the predictions	80
5.13	3D skeleton map for the Standing-Working affordance	80
5.14	False positive prediction of the affordance type Sitting-Working	83
6.1	Affordance-Map models the relationship between virtual human skeleton model and the 3D environment. 1.Robot- Human relationship 2. Object -Human relationship 3. Probable robot's locations that give the best view of the object	87
6.2	Test room for active object search experiments	88
6.3	2D Affordance Likelihood Map	89
6.4	3D Skeleton Map	89
6.5	Positions of the Robot that give the best view of the object	90
6.6	Scaled view of the robot's position and orientation for object search	91
6.7	Path planning for active object search	91
6.8	Samples from object search and detection experiments. (1-4) True Positive, (5) True Negative, (6-7) False positive, (8) False Negative	92
6.9	Object detection results. The figure shows detected and undetected monitors	93
7.1	A robot with proper context-awareness will select a path along corridors rather than the shortest path which is through the working area	95
7.2	The learned affordance map for the office space. High affordance likelihoods can be observed near workbenches	99
7.3	Experiment 1. Context-aware path planner selects a longer path to minimize the distractions to humans work spaces.	103
7.4	Experiment 2. Affordance-map based 'Global Path with Minimal Distractions' (GPMD) selects the only path that always lies along corridors.	104
7.5	Experiment 3. Minimum distraction path along corridors is too long and shortest path goes through human works space.	105
7.6	Experiment 3. Minimum distraction path along corridors is too long. Hence robot tries to find a path through an empty working space.	105

List of Tables

2.1	Contingency table for Binary Classification	16
3.1	Quantitative Analysis of SVM and Naive Baye’s Classifiers	46
4.1	Summary of the class imbalances in the dataset	49
4.2	Imbalance Test Result Summary	60
5.1	Contingency table for Binary Classification	69
5.2	Comparison of performance measures in the K-fold cross validation test. Structured SVM and Other Methods.	82
6.1	Object detection results summary	92

Abbreviations

3D	Three Dimensional
2D	Two Dimensional
DCM	Different Cost Model
FNB	Flexible Naive Bayes
GPMD	Global Path with Minimum Distractions
HMM	Hidden Markov Model
IFTM	Infinite Factored Topic Model
KDE	Kernel Density Estimation
MRF	Markov Random Fields
PRM	Probabilistic Road Map
RGB-D	Red, Green, Blue and Depth
SVM	Support Vector Machine
S-SVM	Structured Support Vector Machine
SLAM	Simultaneous Localization and Mapping
SMOTE	Synthetic Minority Over Sampling Technique

Nomenclature

$f(\dots)$	A scalar valued function
$\mathbf{f}(\dots)$	A vector valued function
$P(x)$	Probability of a random variable x
$P(x z)$	Conditional probability of x , given evidence z
$P(x, z)$	Joint probability of x and z
$DT[\cdot]$	Distance transform
$[\cdot]^T$	Transpose of a vector or a matrix
$\ \cdot\ $	The magnitude of a vector
$\Phi(\cdot)$	Feature function
$E[\cdot]$	The expectation of a random variable
$K((\cdot), (\cdot))$	Kernel function

Specific Symbol Usage

A	Affordance-map
i_k	i^{th} Point cloud image
y_i	i^{th} Label
\mathbf{x}_n	Feature vector at i^{th} grid location
\mathbf{g}_i	i^{th} Grid location of the 3D map
C_k	k^{th} Class label
\mathbf{w}	The normal vector to the separating hyperplane of the SVM classifier
b	The offset of the hyperplane from the origin along the normal vector of the SVM classifier
ξ_i	i^{th} Slack variable

α_i	i^{th} Lagrange multiplier
$R_z(\theta_k)$	The rotational matrix about the z axis
H_l	Human skeleton model
\bar{y}_k	k^{th} Ground truth label
$\Delta(\bar{y}', \bar{y}_k)$	Loss function
μ_s	The mean of position and orientation of the robot in human skeleton co-ordinate system
Σ_s	The covariances of position and orientation of the robot in human skeleton co-ordinate system

Glossary of Terms

Affordance	All action possibilities latent in the environment, objectively measurable and independent of the individual's ability to recognize them.
Environment	A complex 3D unstructured place . Assumed to have some structural characteristics such as planar surfaces.
Features	Distinctly identifiable points in an image of a 3D environment.
Freespace	Areas in the environmental model or map that are known to be free of objects, obstacles and surfaces.
Iteration	A single step which is determined by optimisation.
Obstacle	An object within the environment which a robot can collide with.
Occlusion	Not visible from a viewpoint due to an obstruction.
Pose	The combination of position and orientation of an object in 3D space with respect to a reference coordinate frame.
Planning	The act of generating a path (and motion) course which the robot can then follow to get between two poses.
Surface	This is the face of an object in the environment.
Surface Normal	A 3D vector perpendicular to a surface.
Unstructured	Real-world environment that cannot be set up to facilitate experiment.
Viewpoint	A position in space and an orientation of a sensor that a corresponding robot pose can achieve.
Voxel	Volumetric Pixel which represents a 3D cube-like volume in Euclidean space.