

TABLE OF CONTENTS

Signature Page		iii
Dedication		iv
Epigraph		v
Table of Contents		vi
List of Figures		x
List of Tables		xvi
Acknowledgements		xvii
Vita and Publications		xx
Abstract of the Dissertation		xxii
Chapter 1	Introduction	1
	1.1 Contributions of the thesis	5
	1.1.1 Semantic Image Representation	6
	1.1.2 Visual Recognition Systems	6
	1.1.3 Holistic Context Modeling	10
	1.2 Organization of the thesis	12
Chapter 2	Semantic Image Representation	13
	2.1 Preliminaries	14
	2.1.1 Notations	14
	2.1.2 Image Retrieval Systems	15
	2.1.3 Scene Classification Systems	16
	2.1.4 Image Representation	18
	2.2 Semantic Image Representation	20
	2.2.1 The Semantic Multinomial	22
	2.2.2 Robust estimation of SMNs	23
	2.2.3 SMNs as Posterior Probability Vector	24
	2.3 Computing the Semantic Multinomial	25
	2.4 Related Work	28
	2.5 Acknowledgments	29

Chapter 3	Image Retrieval: Query By Semantic Example	31
	3.1 Introduction	32
	3.2 Related Work	34
	3.3 Query by Semantic Example	36
	3.3.1 Query by Visual Example vs Semantic Retrieval	36
	3.3.2 Query by Semantic Example	38
	3.4 The Proposed Query by Semantic Example System	40
	3.4.1 Similarity Function	41
	3.5 Multiple Image Queries	42
	3.5.1 The Benefits of Query Fusion	43
	3.5.2 Query Combination	44
	3.6 Experimental Evaluation	46
	3.6.1 Evaluation Procedure	46
	3.6.2 Databases	46
	3.6.3 Model Tuning	48
	3.6.4 Performance Within the Semantic Space	51
	3.6.5 Multiple Image Queries	53
	3.6.6 Performance Outside the Semantic Space	54
	3.7 Acknowledgments	59
Chapter 4	Scene Classification with Semantic Representation	60
	4.1 Introduction	61
	4.2 Related Work	63
	4.3 Proposed Approach	64
	4.4 Experimental evaluation	66
	4.4.1 Datasets	66
	4.4.2 Experimental Protocol	66
	4.4.3 Results	67
	4.5 Acknowledgments	76
Chapter 5	Cross Modal Multimedia Retrieval	77
	5.1 Introduction	78
	5.2 Previous Work	80
	5.3 Fundamental Hypotheses	84
	5.3.1 The problem	84
	5.3.2 Multi-modal modeling	85
	5.3.3 The fundamental hypotheses	86
	5.4 Cross-modal Retrieval	88
	5.4.1 Correlation matching (CM)	89
	5.4.2 Semantic matching (SM)	92
	5.4.3 Semantic Correlation Matching (SCM)	95
	5.5 Experimental Setup	95
	5.5.1 Image and text representation	96

	5.6	Parameter selection	99
	5.7	Testing the fundamental hypotheses	104
	5.8	Acknowledgments	108
Chapter 6		Holistic Context Modeling	115
	6.1	Introduction	116
	6.2	Related Work on Context Modeling	118
	6.3	Semantics-based Models and Context Multinomials . . .	120
	6.3.1	Limitations of Semantic Representations	120
	6.3.2	From Semantics to Context	122
	6.3.3	Contextual Concept Models	123
	6.3.4	Contextual Space	125
	6.3.5	Data Augmentation	126
	6.4	Experimental Setup	127
	6.4.1	Datasets	127
	6.4.2	Appearance Features	128
	6.5	Results	129
	6.5.1	Designing the Semantic Space.	129
	6.5.2	Number of Mixture Components	130
	6.5.3	Choice of Appearance Features	132
	6.5.4	Some Examples	134
	6.5.5	Complexity	136
	6.6	Comparison with Previous Work	136
	6.6.1	Scene Classification	139
	6.6.2	Image Retrieval Performance	141
	6.7	Acknowledgments	143
Chapter 7		The Importance of Supervision	145
	7.1	Introduction	146
	7.2	Topic Models	148
	7.2.1	LDA model	148
	7.2.2	Class LDA (cLDA)	149
	7.2.3	Supervised LDA (sLDA)	150
	7.2.4	Geometric Interpretation	151
	7.3	The Importance of Supervision	152
	7.4	Limitations of Existing models	154
	7.4.1	Theoretical Analysis	154
	7.4.2	Experimental Analysis	157
	7.5	Topic supervision	158
	7.5.1	Topics supervision in LDA model	158
	7.5.2	Models and geometric interpretation	159
	7.5.3	Learning and inference with topic-supervision . .	159
	7.5.4	Experimental analysis	161

7.6	Acknowledgments	162
Chapter 8	Conclusions	164
Appendix A	Datasets.	169
A.1	Datasets	170
A.1.1	Natural Scene Categories (N8, N13, N15)	170
A.1.2	UIUC Sports Dataset (S8)	172
A.1.3	Corel Image Collection (C371, C50, C43, C15)	172
A.1.4	Flickr Images (F18)	177
A.1.5	TVGraz	178
A.1.6	Wikipedia	178
Appendix B	Generalized Expectation maximization (GEM)	180
Appendix C	Computation of Image-SMNs	182
Appendix D	Variational Approximation	184
Appendix E	Parameter Estimation in cLDA	186
Appendix F	Parameter Estimation in topic-supervised LDA models	188
F.1	Learning Topic Conditional Distributions	188
F.2	Learning Class Conditional Distribution	189
Appendix G	Implementation Details of the various systems	190
G.1	Image Representation	190
G.1.1	SIFT Features	190
G.1.2	DCT Features	191
G.1.3	Bag-of-Features	191
G.1.4	Bag-of-Words	192
G.1.5	Semantic Multinomial	192
G.2	Concept/Category Models	193
G.2.1	Appearance Based Models	193
G.2.2	Holistic Context Models	193
G.3	Topic Supervised LDA	194
Bibliography	195

LIST OF FIGURES

Figure 1.1:	Probability of a locomotive image belonging to a number of visual concept classes according to appearance based visual classifiers. Note that, while most of the concepts of largest probability are present in the image, the SMN assigns significant probability to “bridge” and “arch”. This is due to the presence of a geometric structure similar to that of “bridge” and “arch”, shown on the image close-up.	4
Figure 1.2:	An illustration of image representation on the <i>semantic space</i> . An image is represented as a <i>semantic multinomial</i> which is a weight vector obtained using an array of appearance based classifiers.	6
Figure 1.3:	An illustration of “semantic gap” — two images which are similar for humans as they depict the semantic concept of “beach”. However they have different low-level visual properties of color, shape, etc.	7
Figure 2.1:	The generative model underlying image formation at the appearance level. w represents a sample from a vocabulary of scene categories or semantic concepts, and an image \mathcal{I} is composed of N patches, \mathbf{x}_n , sampled independently from $P_{\mathbf{X} W}(\mathbf{x} w)$. Note that, throughout this work, we adopt the standard plate notation of [14] to represent graphical models.	16
Figure 2.2:	Learning the scene category (semantic concept) density from the set \mathcal{D}_w of all training images annotated with the w^{th} caption in $\mathcal{W}(\mathcal{L})$, using hierarchical estimation [21]	17
Figure 2.3:	Image representation in semantic space \mathcal{S} , with a semantic multinomial (SMN) distribution. The SMN is a vector of posterior concept probabilities which encodes the co-occurrence of various concepts in the image, based on visual appearance.	21
Figure 2.4:	SMN for the image shown on the top left computed using (top-right) (2.8), (bottom-left) (2.21) and (bottom-right) (2.23).	25
Figure 2.5:	Alternative generative models for image formation at the appearance level. (a) A concept is sampled per appearance feature vector rather than per image, from $P_{\mathbf{X} W}(\mathbf{x} w)$. (b) Explicit modeling of the contextual variable Π from which a single SMN is drawn per image.	26
Figure 3.1:	An image containing various concepts: ‘train’, ‘smoke’, ‘road’, ‘sky’, ‘railroad’, ‘sign’, ‘trees’, ‘mountain’, ‘shadows’, with variable degrees of presence.	37

Figure 3.2:	Semantic image retrieval. Top: Under QBSE the user provides a query image, probabilities are computed for all concepts, and the image represented by the concept probability distribution. Bottom: Under the traditional SR paradigm, the user specifies a short natural language description, and only a small number of concepts are assigned a non-zero posterior probability.	39
Figure 3.3:	SMN of the <i>train</i> query of 3.6 as a function of the ratio $\frac{L(\alpha-1)}{n}$ adopted for its regularization.	50
Figure 3.4:	Average precision-recall of single-query QBSE and QBVE, Left: Inside the semantic space (<i>Corel371</i>), Right: Outside the semantic space (<i>Flickr18</i>).	51
Figure 3.5:	MAP scores of QBSE and QBVE across the 50 classes of <i>Corel371</i> .	51
Figure 3.6:	Some examples where QBSE performs better than QBVE. The second row of every query shows the images retrieved by QBSE.	52
Figure 3.7:	MAP as a function of query cardinality for multiple image queries. Comparison of QBSE, with various combination strategies, and QBVE. Left: Inside the semantic space (<i>Corel371</i>), Right: Outside the semantic space (<i>Flickr18</i>).	54
Figure 3.8:	Effect of multiple image queries on the MAP score of various classes from <i>Corel371</i> . Left: Classes with highest MAP gains, Right: Classes with lowest MAP gains	55
Figure 3.9:	Best precision-recall curves achieved with QBSE and QBVE on <i>Corel371</i> . Left: Inside the semantic space (<i>Corel371</i>), also shown is the performance with meaningless semantic space. Right: Outside the semantic space (<i>Flickr18</i>).	55
Figure 3.10:	Examples of multiple-image QBSE queries. Two queries (for “Township” and “Helicopter”) are shown, each combining two examples. In each case, two top rows presents the single-image QBSE results, while the third presents the combined query. . .	56
Figure 3.11:	SMN of individual and combined queries from class ‘Township’ of 3.10. Left column shows the first query SMN, center the second and, right the combined query SMN.	57
Figure 3.12:	Performance of QBSE compared to QBVE, based on precision-scope curve for $N = 1$ to 100, Left: Inside the semantic space (<i>Corel371</i>), Right: Outside the semantic space (<i>Flickr18</i>). . . .	58
Figure 4.1:	The proposed scene classification architecture.	65
Figure 4.2:	Theme vectors from each of the scenes of fifteen scene categories.	68
Figure 4.2:	Theme vectors from each of the scenes of fifteen scene categories. (continued)	69
Figure 4.2:	Theme vectors from each of the scenes of fifteen scene categories. (continued)	70

Figure 4.3:	Confusion Table for our method using 100 training image and rest as test examples from each category of Natural15. The average performance is $72.2\% \pm 0.2$	71
Figure 4.4:	Some images from worst performing scene categories in Natural15. (\rightarrow) implies the scene category the image is classified into.	73
Figure 4.5:	Some images from the Corel50 dataset. (\rightarrow) implies the scene category the image is classified into. (a) and (b) show two examples of correctly classified images, (c) and (d) two reasonably misclassified images and (e) and (f) shows two examples of error.	74
Figure 4.6:	The theme vector for the image in Figure 4.5(a).	74
Figure 4.7:	Classification performance as a function of the semantic space dimensions. Also shown, is the growth of the variance of the semantic themes, scaled appropriately.	75
Figure 5.1:	Two examples of image-text pairs: (a) section from the Wikipedia article on the Birmingham campaign (“History” category), (b) part of a Cognitive Science class syllabus from the TVGraz dataset (“Brain” category).	83
Figure 5.2:	Each document (D_i) consists of an <i>image</i> (I_i) and accompanying <i>text</i> (T_i), <i>i.e.</i> , $D_i = (I_i, T_i)$, which are represented as vectors in feature spaces \mathfrak{R}^I and \mathfrak{R}^T , respectively. Documents establish a one-to-one mapping between points in \mathfrak{R}^I and \mathfrak{R}^T	85
Figure 5.3:	Correlation matching (CM) performs joint feature selection in the text and image spaces, projecting them onto two maximally correlated subspaces \mathcal{U}_T and \mathcal{U}_I	87
Figure 5.4:	Cross-modal retrieval using CM. Here, CM is used to find the images that best match a query text.	92
Figure 5.5:	Semantic matching (SM) maps text and images into a semantic space. For each modality, classifiers are used to obtain a semantic representation, <i>i.e.</i> , a weight vector over semantic concepts.	93
Figure 5.6:	Cross-modal retrieval using SM used to find the text that best matches a query image.	94
Figure 5.7:	MAP performance (cross-modal retrieval, validation set) of SCM using two image models: BOW (flat lines) and LDA, for (a) TVGraz and (b) Wikipedia.	101
Figure 5.8:	Cross-modal MAP for CM on TVGraz and Wikipedia (validation sets), as a function of (a) the number of image codewords, (b) the number of text LDA topics, and (c) the number of KCCA components (while keeping the other two parameters fixed at the values reported in 5.5).	104

Figure 5.9:	Confusion matrices on the test set, for both TVGraz (left) and Wikipedia (right). Rows refer to true categories, and columns to category predictions. The more confusion on Wikipedia motivates the lower retrieval performance.	105
Figure 5.10:	top) Precision recall curves, bottom) Precision at N curves for left) Text query, right) Image query for TVGraz	107
Figure 5.11:	top) Precision recall curves, bottom) Precision at N curves for left) Text query, right) Image query for Wikipedia	108
Figure 5.12:	Per-class MAP for the cross-modal retrieval tasks on TVGraz (left) and Wikipedia (right): text queries (top); image queries (middle); and average performance over both types of queries (bottom).	109
Figure 5.13:	Text query from Biology class of Wikipedia and the top 5 retrieved images retrieved using SCM. The query text, associated probability vector, and ground truth image are shown on the top; retrieved images are presented at the bottom.	110
Figure 5.14:	Text query from 'Warfare' class of Wikipedia and the top 5 retrieved images retrieved using SCM. The query text, associated probability vector, and ground truth image are shown on the top; retrieved images are presented at the bottom.	111
Figure 5.15:	Text query from 'Cactus' class of TVGraz and the top 5 retrieved images retrieved using SCM. The query text, associated probability vector, and ground truth image are shown on the top; retrieved images are presented at the bottom.	112
Figure 5.16:	Text query from 'Butterfly' class of TVGraz and the top 5 retrieved images retrieved using SCM. The query text, associated probability vector, and ground truth image are shown on the top; retrieved images are presented at the bottom.	113
Figure 5.17:	Image-to-text retrieval on TVGraz (first two columns) and Wikipedia (last two columns). Query images are shown on the top row. The four most relevant texts, represented by their ground truth images, are shown in the remaining columns.	114
Figure 6.1:	An image from the "street" class of the N15 dataset (See 6.4.1) along with its SMN. Also highlighted are the two notions of <i>co-occurrence</i> . <i>Ambiguity co-occurrences</i> on the right: image patches compatible with multiple unrelated classes. <i>Contextual co-occurrences</i> on the left: patches of multiple other classes related to "street".	121
Figure 6.2:	Learning the contextual model for the "street" concept, (6.1), on semantic space \mathcal{S} , from the set of all training images annotated with "street".	123

Figure 6.3:	3-component Dirichlet mixture learned for the concept “street”. Also shown, as “*”, are the SMNs associated with each image. The Dirichlet mixture assigns high probability to the concepts “street” and “store”.	125
Figure 6.4:	The Contextual multinomial (CMN) of an image as the vector of co-occurrence probabilities of contextually related concepts.	126
Figure 6.5:	(a) Classification accuracy as a function of the number of mixture components of the contextual class distributions, for both DCT and SIFT. (b) Dependence of appearance and contextual classification on the accuracy of the appearance modeling for SIFT-GRID features, (c) for DCT features. The performance of contextual classification remains fairly stable across the range of appearance models.	131
Figure 6.6:	Four cluster centers for the class “street” (top) and “forest” (bottom). Note that each class comprises different co-occurrence patterns.	132
Figure 6.7:	top) Two images from the “street” class of N15, and bottom) an image each from the “Ireland” and “Mayan ruins” CD of the Corel collection. Also shown with the images are the SMN and CMN vectors (middle and right column respectively). Notice that the CMN vectors are noise-free and capture the “gist” of the image.	135
Figure 6.8:	Class confusion matrix for classification on the N15 dataset. The average accuracy is 77.20%	141
Figure 6.9:	Precision-recall curves achieved with SMN, CMN, visual matching and chance level image retrieval.	142
Figure 6.10:	Retrieval results for four image queries shown on the left-most column. The first, second, and third row of every query show the five top matches using image matching, SMN, and CMN-based retrieval, respectively.	144
Figure 7.1:	Graphical models for (a) LDA and ts-LDA. (b) cLDA and ts-cLDA. (c) sLDA and ts-sLDA. All models use the standard plate notation [19], with parameters shown in rounded squares.	148
Figure 7.2:	Representation of cLDA and ts-cLDA on a three <i>word simplex</i> . Also shown are sample images from two classes: “o” from class-1 and “x” from class-2. a) cLDA model with two topics. The line segment depicts a one-dimensional <i>topic simplex</i> , whose vertices are topic-conditional word distributions. Each class defines a smooth distribution on the topic simplex, denoted by the contour lines. c) ts-cLDA model. Topic-conditional word distributions are learned with supervision which encapsulate the class attributes.	151

Figure 7.3:	left) Four groups of words with equal word histograms. right) Four groups of edge segments with the equal edge segment histograms. Note that each group can be derived from the others by a displacement of words or edge segments. (This figure is best viewed in color)	152
Figure 7.4:	Classification accuracy as function of the number of topics for sLDA and cLDA, using topics learned with and without class influence and codebooks of size 1024, on (a) N15, (b) N8 and (c) S8. Similar behavior was observed for codebooks of different sizes.	155
Figure 7.5:	Performance of ts-sLDA, ts-cLDA, sLDA, and cLDA as a function of codebook size on (a) N13, (b) N8 and (c) S8. For ts-sLDA and ts-cLDA the number of topics is equal to the number of classes. For sLDA and cLDA, results are presented for the number of topics of best performance.	160
Figure 7.6:	Some example images that were misclassified by cLDA, but correctly classified using ts-cLDA. The expected topic distributions for ts-cLDA and cLDA (using 13 topics) are shown in the middle and bottom rows respectively. For ts-cLDA, topic labels are same as the class labels and the high probability topics are indeed the ones which capture the semantic meaning of the image. For cLDA, the topic labels do not carry any clear semantic meaning.	161

LIST OF TABLES

Table 2.1: SMN Entropy.	26
Table 3.1: Retrieval and Query Database	47
Table 3.2: Effect of SMN regularization on the MAP score of QBSE	49
Table 3.3: Effect of the similarity function on the MAP score of QBSE	50
Table 3.4: MAP of QBVE and QBSE on all datasets considered.	58
Table 4.1: Classification Result for 15 scene categories.	75
Table 4.2: Classification Result for 13 scene category subset.	76
Table 5.1: Taxonomy of the proposed approaches to cross-modal retrieval.	88
Table 5.2: Cross-modal retrieval performance (MAP) on the validation set using different distance metrics for TVGraz. μ_p and μ_q are the sample averages for p and q , respectively.	97
Table 5.3: Cross-modal retrieval performance (MAP) on the validation set using different distance metrics for Wikipedia. μ_p and μ_q are the sample averages for p and q , respectively.	98
Table 5.4: MAP for CM hypothesis (validation sets).	102
Table 5.5: Best parameter settings for CM, SM and SCM, on both TVGraz and Wikipedia (validation sets).	103
Table 5.6: Cross-modal MAP on TVGraz and Wikipedia (test sets).	106
Table 6.1: Impact of inference model on classification accuracy.	129
Table 6.2: Impact of appearance space on classification accuracy.	133
Table 6.3: Classification Results on Natural Scene Categories.	137
Table 6.4: Classification Results on Natural Scene Categories.	138
Table 6.5: Classification Results on Natural Scene Categories.	138
Table 6.6: Classification Results on Corel Collection.	139
Table 7.1: Classification Results on Natural Scene Categories.	162
Table 7.2: Classification Results on Sports8 and Corel50.	163
Table A.1: Summary of the Natural Scene datasets.	171
Table A.2: Summary of the UIUC Sports dataset.	173
Table A.3: Summary of the C371 dataset.	173
Table A.4: Summary of the TVGraz dataset.	178
Table A.5: Summary of the Wikipedia dataset.	179

ACKNOWLEDGEMENTS

“It takes a village to raise a child” — an old African proverb. So it is with everything else in life. Although, I get the privilege of writing this thesis, it is certain that this thesis would not have been possible, had not the village helped raise it. In this small, yet important section, I take the opportunity to acknowledge many people, who helped shape this thesis. Foremost, my deepest gratitude to life itself, which provides eternal wonder and amazement, motivating me to always learn and know more.

Next, I would like to express my sincerest gratitude to my supervisor, Professor Nuno Vasconcelos. It won't be far from truth, if I said that I did not know the meaning of “research” and it's only through his mentoring and guidance that I slowly understood its true essence. From him I learned the benefits of striving for perfection, of hard work, of hanging on, of just doing it. His contributions to this thesis, definitely equal mine, if not exceeding it. I am also grateful for having an exceptional doctoral committee, and wish to thank Serge Belongie, Kenneth Kreutz-Delgado, David Kriegman, and Truong Nguyen for their valuable input, accessibility and informative courses.

I would like to thank all my colleagues at SVCL, Dr. Dashan Gao, Dr. Antoni Bert Chan, Dr. Hamed Masnadi-Shirazi, Dr. Sunhyoung Han, Vijay Mahadevan, Jose Maria Costa Pereira, Mandar Dixit, Mohammad Saberian, Kritika Muralidharan and Weixin Li for their support, friendship and assistance throughout the years and their collaboration on some of the experiments.

Although the six years, which went in the making of this thesis, were filled with its share of ups and downs, the most cherishable moments — the moments that would flash before my eyes — are the ones I shared with my friends. With the risk of offending the few, who were here with me all along, but to whom, my impaired memory prevents me from doing justice, I would like to thank Ankit Srivastava, Vijay Mahadevan, Himanshu Khatri, Nitin Gupta, Jose Maria Costa Pereira, Gaurav Dhiman, Mayank Kabra, Rathinakumar Appuswamy, Anshuman Gupta, Shikha Misra, Shibin Parameswaran, Adarsh Krishnamurthy, Sethuraman Sankaran, Karthik Sanji, Sravanthi K V, Arun Manohar, Bharath Kumar SV,

Aneesh Subramanian, Nikhil Karamchandani, Vikram Mavalankar, Kowsik Bodi and the many other from whom I again seek forgiveness. A special thanks to my rock climbing friends who showed me what true adventure meant. Also thanks to my friends who walked with me in the years gone by.

In the end, I reserve a special mention for my immediate and extended family. Since any amount of thanks would be incommensurate to their support, I just bow down and seek their blessings.

The text of Chapter 2, in part, is based on the material as it appears in: N. Rasiwasia, P. J. Moreno and N. Vasconcelos, ‘*Bridging the Semantic Gap: Query by Semantic Example*’, IEEE Transactions on Multimedia, 9(5), 923-938, August 2007, and N. Rasiwasia, P. J. Moreno and N. Vasconcelos, ‘*Query by Semantic Example*’, ACM International Conference on Image and Video Retrieval, LNCS 51-60, Phoenix, 2006. The dissertation author was a primary researcher and an author of the cited material.

The text of Chapter 3, in part, is based on the material as it appears in: N. Rasiwasia, P. J. Moreno and N. Vasconcelos, ‘*Bridging the Semantic Gap: Query by Semantic Example*’, IEEE Transactions on Multimedia, 9(5), 923-938, August 2007, N. Rasiwasia, P. J. Moreno and N. Vasconcelos, ‘*Query by Semantic Example*’, ACM International Conference on Image and Video Retrieval, LNCS 51-60, Phoenix, 2006, and N. Rasiwasia and N. Vasconcelos, ‘*A Systematic Study of the role of Context on Image Classification*’, IEEE Conference on Image Processing, 1720-1723, San Diego, Oct 2008. The dissertation author was a primary researcher and an author of the cited material.

The text of Chapter 4, in full, is based on the material as it appears in: N. Rasiwasia and N. Vasconcelos, ‘*Scene Classification with Low-dimensional Semantic Spaces and Weak Supervision*’, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-6, Anchorage, June 2008. The dissertation author was a primary researcher and an author of the cited material.

The text of Chapter 5, in part, is based on the material as it appears in: N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G.R.G. Lanckriet, R. Levy and N. Vasconcelos, ‘*A New Approach to Cross-Modal Multimedia Retrieval*’, ACM

Conference on Multimedia, 251-260, November 2010, and J. Costa Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G.R.G. Lanckriet, R. Levy and N. Vasconcelos, ‘*On the role of Correlation and Abstraction in Cross-Modal Multimedia Retrieval*’, submitted to Pattern Analysis and Machine Learning, Sept 2011. The dissertation author was a primary researcher and an author of the cited material. The author would like to thank J. Costa Pereira, E. Coviello and G. Doyle for their helpful comments and contributions to the project.

The text of Chapter 6, in full, is based on the material as it appears in: N. Rasiwasia and N. Vasconcelos, ‘*Holistic Context Models for Visual Recognition*’, Accepted to appear in IEEE Transactions on Pattern Analysis and Machine Intelligence, N. Rasiwasia and N. Vasconcelos, ‘*Holistic Context Modeling using Semantic Co-occurrences*’, IEEE Conference on Computer Vision and Pattern Recognition, Miami, June 2009, and N. Rasiwasia and N. Vasconcelos, ‘*Image Retrieval using Query by Contextual Example*’, ACM Conference on Multimedia Information Retrieval, pp. 164-171, Vancouver, Oct 2008. The dissertation author was a primary researcher and an author of the cited material.

The text of Chapter 7, in full, is based on the material as it appears in: N. Rasiwasia and N. Vasconcelos, ‘*Holistic Context Models for Visual Recognition*’, Accepted to appear in IEEE Transactions on Pattern Analysis and Machine Intelligence, and N. Rasiwasia and N. Vasconcelos, ‘*Generative Models for Image Classification*’, In preparation for IEEE Transactions on Pattern Analysis and Machine Intelligence. The dissertation author was a primary researcher and an author of the cited material.

VITA AND PUBLICATIONS

- 2001-2005 Bachelor of Technology,
Electrical Engineering,
Indian Institute of Technology, Kanpur, India
- 2005–2007 Master of Science
Electrical Engineering (Signal and Image Processing), Uni-
versity of California at San Diego
- 2005–2011 Research Assistant
Statistical and Visual Computing Laboratory
Department of Electrical and Computer Engineering
University of California at San Diego
- 2005-2011 Doctor of Philosophy
Electrical Engineering (Signal and Image Processing),
University of California at San Diego

Journals

N. Rasiwasia, P. J. Moreno and N. Vasconcelos, ‘*Bridging the Semantic Gap: Query by Semantic Example*’, IEEE Transactions on Multimedia, 9(5), 923-938, August 2007.

N. Rasiwasia and N. Vasconcelos, ‘*Holistic Context Models for Visual Recognition*’, Accepted to appear in IEEE Transactions on Pattern Analysis and Machine Intelligence.

J. Costa Pereira, E. Coviello, G. Doyle, **N. Rasiwasia**, G.R.G. Lanckriet, R. Levy and N. Vasconcelos, ‘*On the Role of Correlation and Abstraction in Cross-Modal Multimedia Retrieval*’, Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence.

N. Rasiwasia and N. Vasconcelos, ‘*Generative Models for Image Classification*’, In preparation for IEEE Transactions on Pattern Analysis and Machine Intelligence.

Conferences

A. Kannan, P. Talukdar, **N. Rasiwasia**, and Q. Ke, ‘*Improving Product Classification Using Images*’, To appear in IEEE International Conference on Data Mining, Vancouver, 2011

R. Kwitt, **N. Rasiwasia** and N. Vasconcelos, '*Learning Pit Pattern Concepts for Gastroenterological Training*', To appear in International Conference on Medical Image Computing and Computer Assisted Intervention, Toronto, September 2011 [**Oral**]

M. Dixit, **N. Rasiwasia** and N. Vasconcelos, '*Adapted Gaussian Models for Image Classification*', IEEE Conference on Computer Vision and Pattern Recognition, pp. 937-943, Colorado Springs, June 2011

N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G.R.G. Lanckriet, R. Levy and N. Vasconcelos, '*A New Approach to Cross-Modal Multimedia Retrieval*', ACM Conference on Multimedia, 251-260, November 2010 [**Oral**] [**Best student paper**]

N. Rasiwasia and N. Vasconcelos, '*Holistic Context Modeling using Semantic Co-occurrences*', IEEE Conference on Computer Vision and Pattern Recognition, Miami, June 2009

N. Rasiwasia and N. Vasconcelos, '*Image Retrieval using Query by Contextual Example*', ACM Conference on Multimedia Information Retrieval, pp. 164-171, Vancouver, Oct 2008

N. Rasiwasia and N. Vasconcelos, '*A Systematic Study of the role of Context on Image Classification*', IEEE Conference on Image Processing, 1720-1723, San Diego, Oct 2008 [**Oral**]

N. Rasiwasia and N. Vasconcelos, '*Scene Classification with Low-dimensional Semantic Spaces and Weak Supervision*', IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-6, Anchorage, June 2008

N. Rasiwasia and N. Vasconcelos, '*A study of Query by Semantic Example*', 3rd International Workshop on Semantic Learning and Applications in Multimedia, pp. 1-8, Anchorage, June 2008 [**Oral**]

N. Rasiwasia, P. J. Moreno and N. Vasconcelos, '*Query by Semantic Example*', ACM International Conference on Image and Video Retrieval, LNCSPhoenix, 2006 [**Oral**]