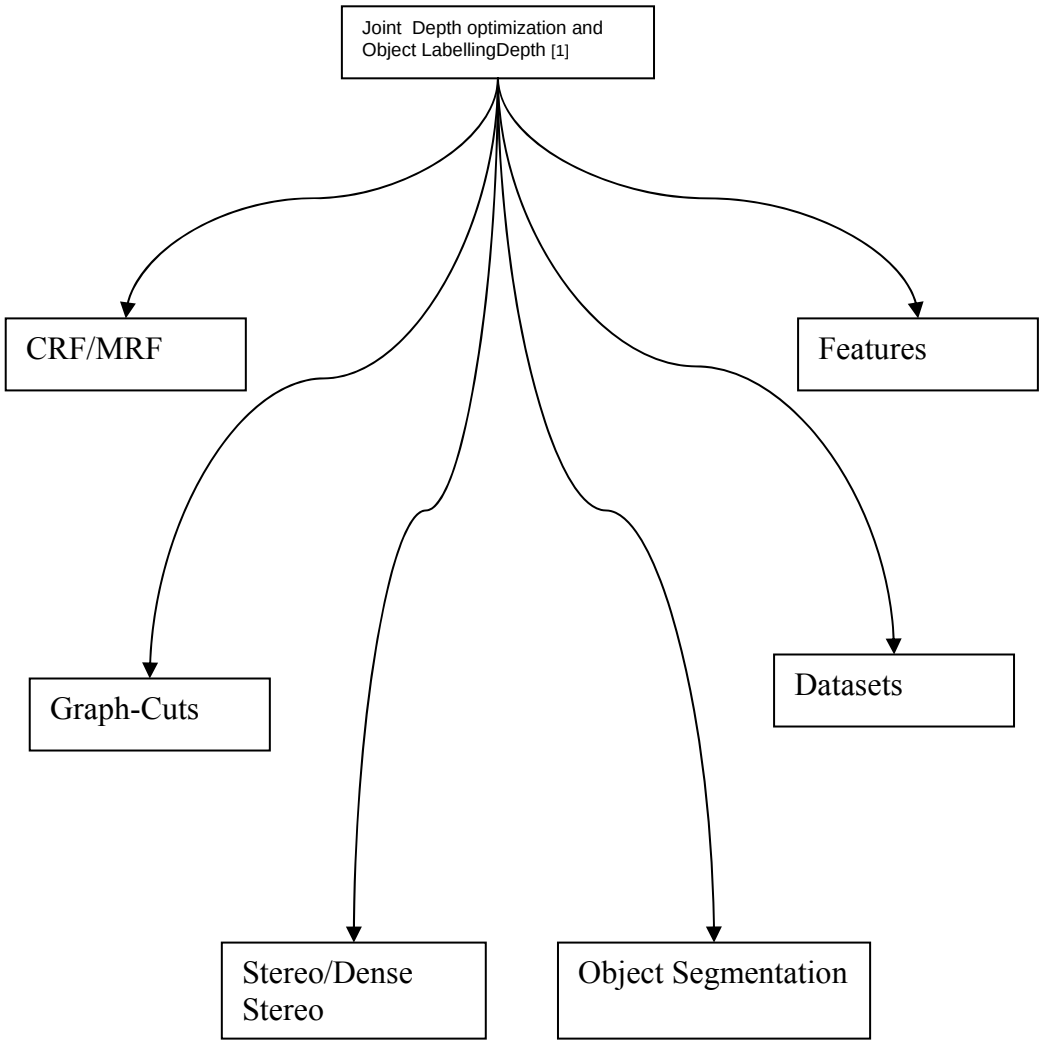
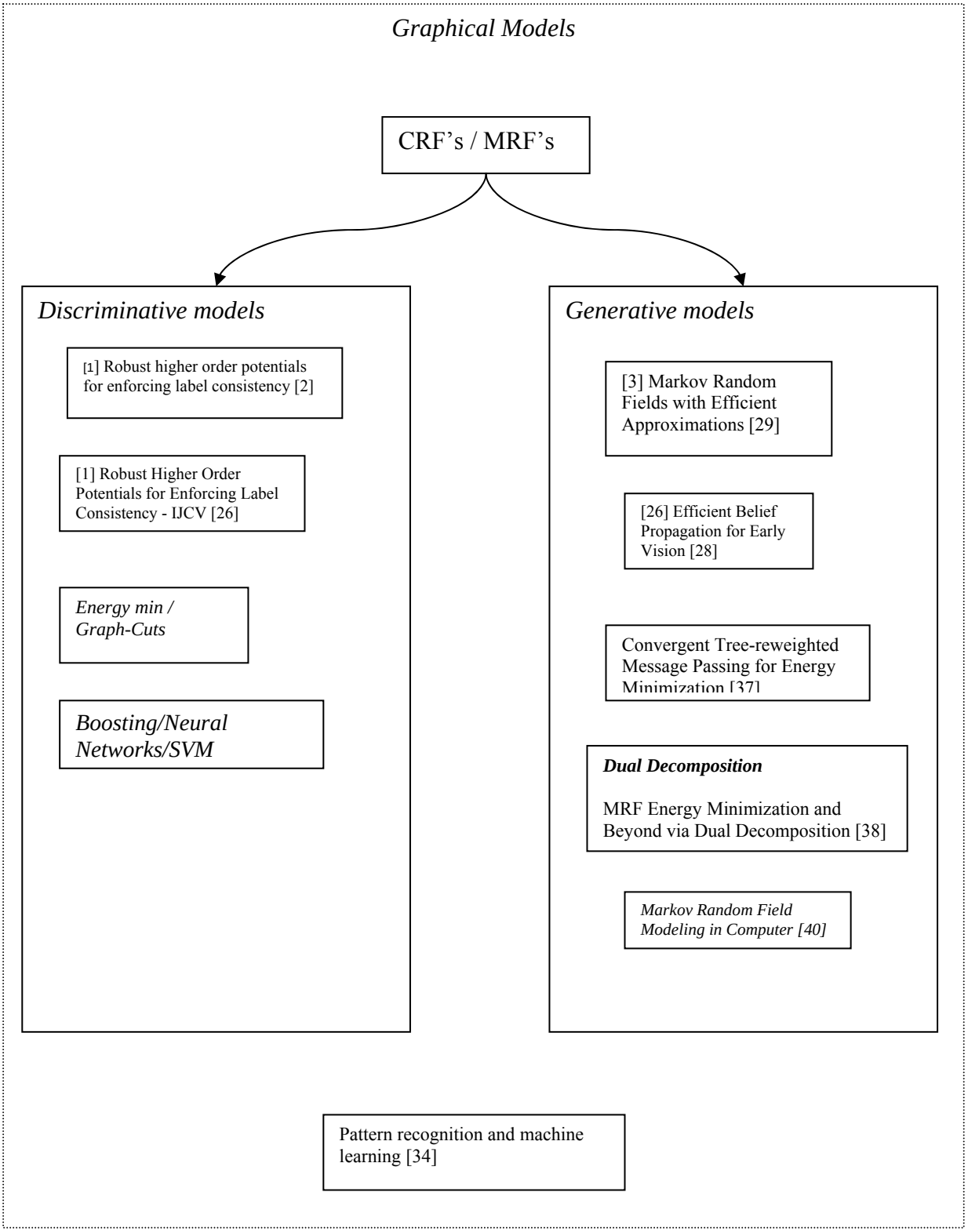
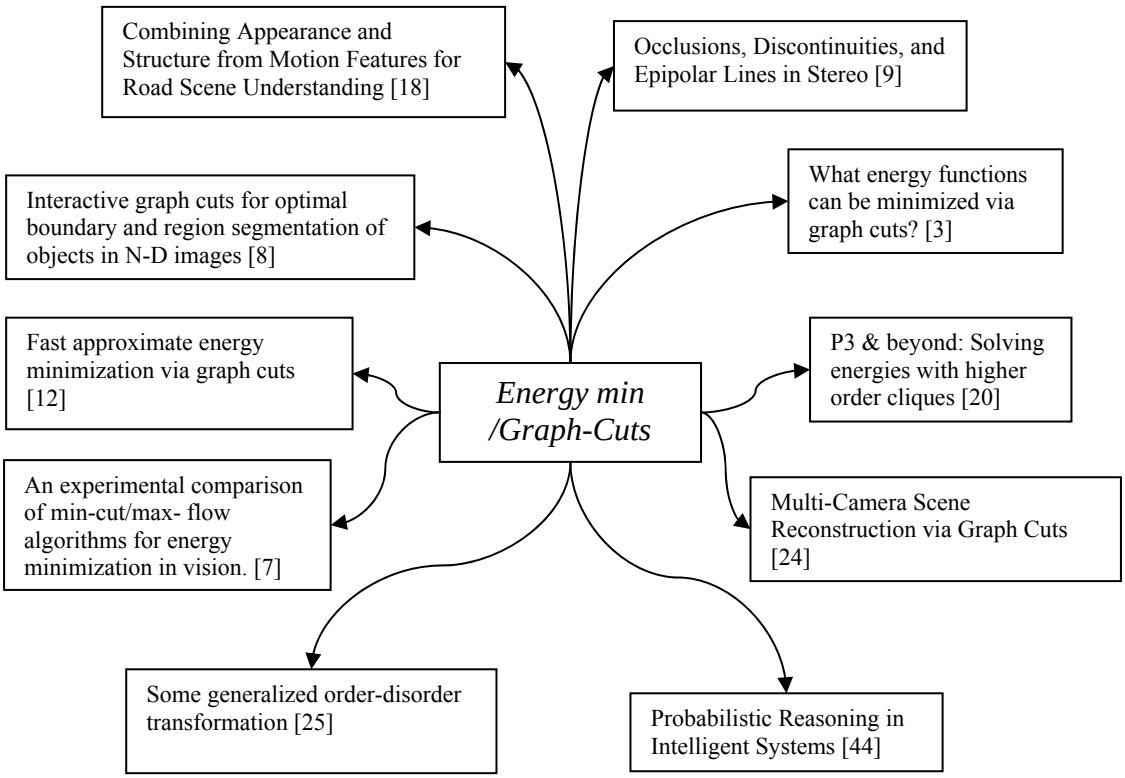
ICVSS Reading group Assignment
Sunando Sengupta

Topic - Energy Minimization in Computer Vision




Graphical Models
CRF's / MRF's
Discriminative models
[1] Robust higher order potentials for enforcing label consistency [2]
[1] Robust Higher Order Potentials for Enforcing Label Consistency - IJCV [26]
Energy min / Graph-Cuts
Boosting/Neural Networks/SVM
Generative models
[3] Markov Random Fields with Efficient Approximations [29]
[26] Efficient Belief Propagation for Early Vision [28]
Convergent Tree-reweighted Message Passing for Energy Minimization [37]
Dual Decomposition
MRF Energy Minimization and Beyond via Dual Decomposition [38]
Markov Random Field Modeling in Computer [40]
Pattern recognition and machine learning [34]



Max Flow/ Min Cut

Flows in Networks. [14]

Introduction to Algorithms[15]

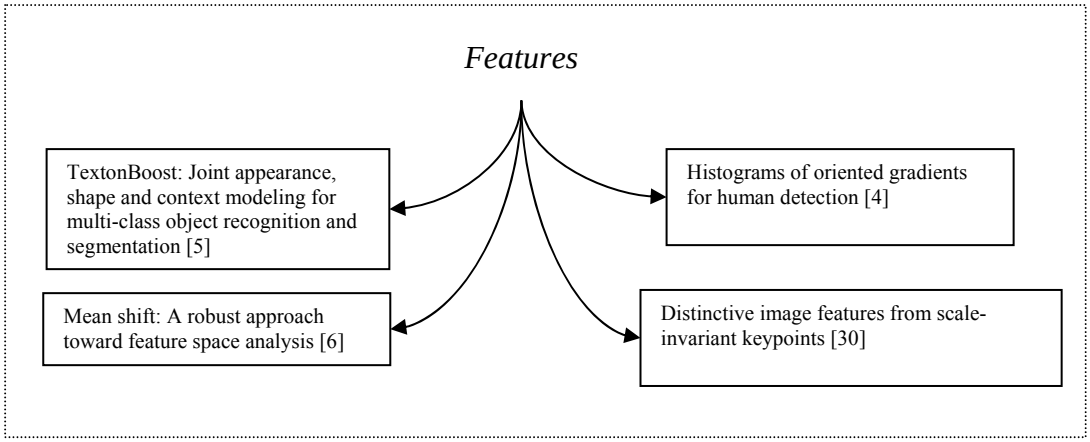
A Note on the Maximum Flow Through a Network [13]

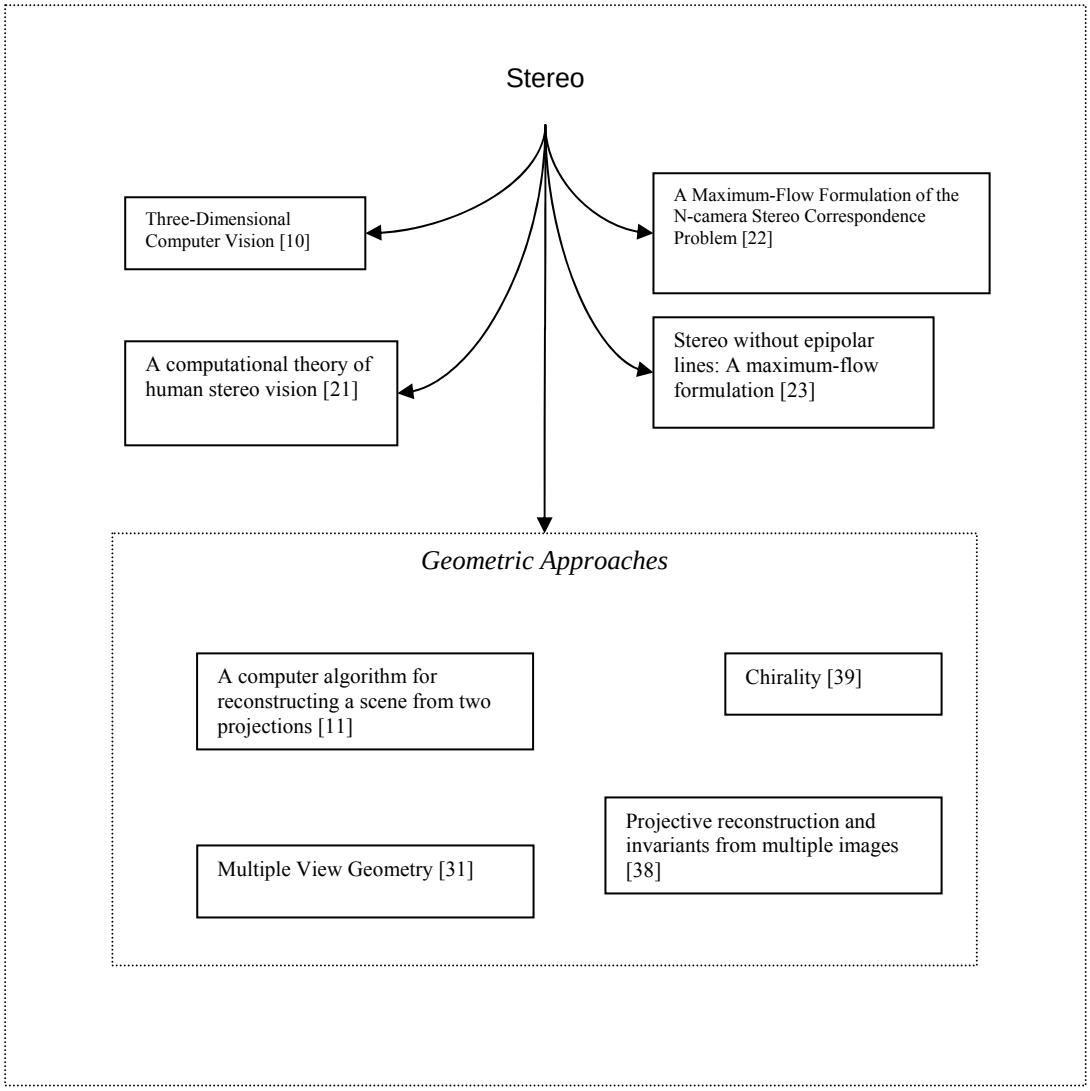
DataSets

Yotta [16]

Dynamic 3d scene analysis from a moving vehicle [17]

A high-definition ground truth database [19]


Features
TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation [5]
Histograms of oriented gradients for human detection [4]
Mean shift: A robust approach toward feature space analysis [6]
Distinctive image features from scale-invariant keypoints [30]


Stereo
Three-Dimensional Computer Vision [10]
A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem [22]
A computational theory of human stereo vision [21]
Stereo without epipolar lines: A maximum-flow formulation [23]
Geometric Approaches
A computer algorithm for reconstructing a scene from two projections [11]
Chirality [39]
Projective reconstruction and invariants from multiple images [38]
Multiple View Geometry [31]

The problems of dense stereo reconstruction, and object class segmentation, can both be formulated as a CRF based labelling problems in which every pixel in the image is assigned a label corresponding to either its disparity, or an object class such as road or building. While these two problems are mutually informative, no attempt has been made to jointly optimise their labelling. The work [1] provides a principled energy minimisation framework which unifies the two problems and demonstrates that, by resolving ambiguities in the data, joint optimisation of the two problems substantially improves performance on real world data sets. They evaluated the method, on the street view Leuven dataset [17].

The image labelling problem can be defined as follows: Given the image, we need to determine for each pixel what labels/class does it belong to. The labels can be object classes (leading to object segmentation) or might be disparity/depth or any meaningful representation. Consider a set of random variables $X = \{X_1, X_2, \ldots, X_n\}$ and a set of labels $L = \{l_1, l_2, \ldots, l_k\}$. The objective of a labelling problem defined over these random variables is to assign a label from the set L to each variable. Many computer vision tasks, such as image segmentation [16], stereo matching [22, 23,9], object recognition [18, 26,27], can be viewed as labelling problems. Typically, in such scenarios, the random variables correspond to pixels in an image, and the label set is defined according to the problem.

Random fields provide an elegant probabilistic framework to model labelling problems [26, 29, 34, 40]. They provide a neighbourhood relationship between variables, and incorporate not only (noisy) image measurements, but also a prior model over the labelling space in a principled manner. Let N represent the neighbourhood of the random field, which is defined by sets N_i, $\forall i \in \{1, 2, \ldots, n\}$. The set Ni denotes the set of all neighbours of the random variable X_i. Markov random fields (MRF) model the joint probability of the labelling x and the data y, denoted by Pr(x, y). It follows the *Markovian property* that the image pixel will take labels based on properties from its neighbours.

Conditional Random Fields (CRF) on the other hand uses the data observations to compute the pair-wise potentials. Two neighbouring pixels with different colour intensity are allowed to take different labels and thus the pair-wise potential was made to be dependent on the image data (observation). Kumar and Hebert [46] formalized the resulting probabilistic distribution as a conditional random field (crf) model in the context of computer vision problems.

The labelling problem can then be formulated as a discrete energy minimization problem (Gibbs energy formulation)

$$E(f) = E_{smooth}(f) + E_{data}(f)$$

$E_{data}(f) = \sum_{p \in P} D_p(f_p)$ The data term measuring how appropriate a label is given the pixel observation.

$E_{smooth}(f) = \sum_{p,q \in N} V_{p,q}(f_p, f_q)$ The smoothness term depends on the inter-pixel observations, should be discontinuity preserving across the object boundaries (often based on MRF).

The energy expression was augmented using higher order clique potentials, defined on the set of random variables derived from the pairwise interacting random variables [27, 45]. There the image is segmented [6] and some higher order cliques are defined over the super-pixels. The data term (also known as unary potential) derived from the image is obtained using features like textons [5] or disparity for stereo problems [3, 8, 42]. The pairwise energy terms are generally derived from image based on the gradient features.

The Gibbs energy minimization is in general NP Hard and depends on the properties of the energy function. Two such families of energy functions for which powerful algorithms exists are: (i) *Submodular* energy functions [3, 40] and (ii) Energy functions defined on tree structured [37, 38] mrf/crf. Submodular energy function minimization for certain random fields has been shown to be equivalent to a graph cut (specifically st-mincut) problem, which has several efficient polynomial time algorithms. Submodular energy function can be efficiently minimized using the Graph-cut based solution, while the multi-label problem can be converted to a binary problem and efficiently solved using st-mincut [13, 14] The graph-cut based solution has also been used to find approximate solutions for non-submodular energy functions.

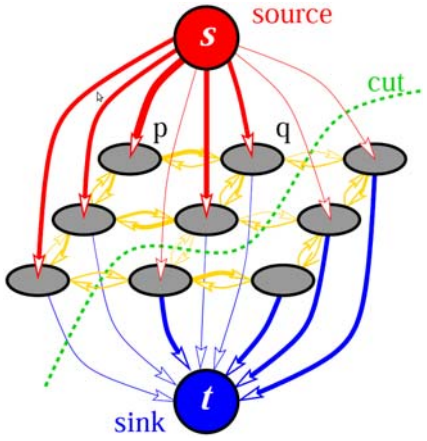


Fig: Graph-cut

Let us consider the binary image segmentation problem as an example. The nodes in the st-graph correspond to pixels in the image, and the terminals represent the two labels, say 0 and 1. The edge weights are set according to the energy function defined for the segmentation problem. The cost of the st-cut is equal to the energy of its associate labelling x, *i.e.* E(x). representaion of image as a graph has been suggested previously by Ishikawa[9] who did a stereo reconstruction over the pixel graph.

Moves making algorithms: Boykov[12] suggested efficient graph cut based α-expansion and αβ-swap algorithms for solving non-submodular problems. These algorithms belong to the class of move making algorithm, where an initial label is assigned to the image pixel and a set of moves are performed. Other set of moves are known as range moves [47] where the pixel/node is given a label from a range (used in disparity/stereo matching)

Another class of algorithms are the Message passing algorithms for MAP inference problem. For tree structures random field they are guaranteed to provide optimal solution (max-product belief propagation algorithm exactly minimizes energy functions defined over graphs with no loops) [37, 44]. The message passing algorithms for discrete MRF-based optimization in computer vision has been enhanced via the technique of dual-decomposition in [38].

The stereo/depth reconstruction problems have been attempted from the last eighties, where a pair of images (or more for n-view stereo) was used to get the depth for each feature points. Geometric techniques were employed to get the sparse point set reconstruction using the estimation of Fundamental/Essential matrix. More recently dense stereo reconstruction has been studied extensively and MRF based approaches [40] have been defined for them.

Similar to the the problem of object segmentation, dense stereo can be formulated as an image labelling problem where the image pixels take any of the predefined depth labels.
The dense stereo problems involve in estimating a 3D model of the scene by finding matching pixels in the images and converting their 2D positions into 3D depths [1].In the energy function we describe here, the set of vertices corresponds to pixels in the image, and the set of edges is given by 4-neighbourhood. The pairwise term is a Generalized Potts model, which encourages similar pixels to take the same label. This multi-label energy function can be minimized using the move making or message passing algorithms.

Road scene classification has been an important area for autonomous systems and interesting vision problem. Object labelling for scene reconstruction has been used using popup-methods [49] and through CRF modelling[18] on camvid database [19]. The CamVid [19] data set provides sparse SfM cues, which were used by several object class segmentation approaches to provide pixelwise labelling. The scene reconstruction in an urban area [17] involves identifying the objects such as road, car and sky and finds their 3D locations. Compared to typical stereo data sets [48] that are usually produced in controlled environments, stereo reconstruction on this real world data is noticeably more challenging due to large homogeneous regions and photo-consistency problems.

References

1. Lubor Ladicky, Paul Sturgess, Chris Russel, Yalin B, S. Sengupta, W. Clocksin, P.H.S. Torr. Joint Optimisation for Object Class Segmentation and Dense Stereo Reconstruction. Accepted at BMVC 2010.
2. P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. In CVPR, 2008.
3. V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts?. PAMI, 2004.
4. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005
5. J. Shotton, J. M. Winn, C. Rother, and A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In ECCV (1), pages 1–15, 2006.
6. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. PAMI, 2002
7. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. PAMI, 2004.
8. Y. Boykov and M. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In ICCV, 2001
9. Hiroshi Ishikawa , Davi Geiger. Occlusions, Discontinuities, and Epipolar Lines in Stereo. In ECCV 98
10. O. Faugeras *Three-Dimensional Computer Vision*. MIT Press. Cambridge, Mass., 1993
11. H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". *Nature* **293**
12. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;23(11):1222-1239.
13. P. ELIAS, A. FEINSTEININ, AND C. E. SHANNON. A Note on the Maximum Flow Through a Network.. IRE TRANXACTIONS ON INFORMATION THEORY 1956
14. L. Ford and D. Fulkerson. Flows in Networks. Princeton University Press, 1962

15. T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. McGrow-Hill, New York, 1990.
16. Yotta DCL. Yotta dcl case studies. http://www.yottadcl.com/surveys/case-studies/, April 2010
17. B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool. Dynamic 3d scene analysis from a moving vehicle. In CVPR, 2007.
18. Paul Sturgess, Karteek Alahari, L'ubor Ladicky, Philip H. S. Torr , Combining Appearance and Structure from Motion Features for Road Scene Understanding. BMVC 2009
19. G. J. Brostow, J. Fauqueur, and R. Cipolla. Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 30(2):88–97, 2009.
20. P. Kohli, M. P. Kumar, and P. H. S. Torr. P3 & beyond: Solving energies with higher order cliques. In CVPR, 2007
21. D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, 204:301–328, 1979
22. S. Roy and I. Cox. A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem In *Proc. Int. Conf. on Computer Vision, ICCV'98*, Bombay, India 1998.
23. Sebastien Roy. Stereo without epipolar lines: A maximum-flow formulation. International Journal of Computer Vision, 34(2/3):147–162, August 1999.
24. V. Kolmogorov and R. Zabih, "Multi-Camera Scene Reconstruction via Graph Cuts," Proc. European Conf. Computer Vision, vol. 3, pp. 82-96, 2002.
25. R. Potts. Some generalized order-disorder transformation. Proceedings of the Cambridge Philosophical Society, 48:106–109, 1952
26. Pushmeet Kohli, Lubor Ladicky, Philip H.S. Torr, Robust Higher Order Potentials for Enforcing Label Consistency, In Proceedings of the International Journal of Computer Vision, Volume 82, Issue 3, pages 302-324, 2009
27. P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation." *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
28. Pedro F. Felzenszwalb and Daniel P. Huttenlocher Efficient Belief Propagation for Early Vision International Journal of Computer Vision, Vol. 70, No. 1, October 2006
29. Y. Boykov, O. Veksler, and R. Zabih, "Markov Random Fields with Efficient Approximations," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 648-655, 1998
30. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91.110, 2004
31. R. Hartley and A. Zisserman. Multiple View Geometry, Cambridge University Press. 2002
32. Dimitri P. Bertsekas. NonLinear Programming. Athena Scientific, 2003
33. Edmund K. Burke, Graham Kendall. Search methodologies: introductory tutorials in optimization and decision support techniques. Springer
34. Christopher M. Bishop. Pattern recognition and machine learning.
35. Richard O. Duda, Peter E. Hart, David G. Stork. Pattern Classification, 2nd Edition , Wiley
36. Vladimir Kolmogorov. Convergent Tree-reweighted Message Passing for Energy Minimization. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), 28(10):1568-1583, October 2006.
37. N. Komodakis, N.Paragios and G. Tziritas. MRF Energy Minimization and Beyond via Dual Decomposition IEEE Transactions on Pattern Analysis and Machine Intelligence (in press).
38. R. I. Hartley, "Projective reconstruction and invariants from multiple images," PAMI, pp. 1036--1041, 1994
39. R. I. Hartley, "Chirality," IJCV, pp. 41--61, Jan. 1998
40. S. Z. Li. Markov Random Field Modeling in Computer Vision. Springer-Verlag, 2000.

41. P. Kohli. Minimizing Dynamic and Higher Order Energy Functions using Graph Cuts. PhD thesis, Oxford Brookes University, November 2007.
42. Vladimir Kolmogorov. Graph Based Algorithms for Scene Reconstruction from Two or More Views PhD thesis, Cornell University, September 2003.
43. G.H. Golub, C.F. Van loan, Matrix computations Johns Hopkins 1996
44. J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers 1988
45. Lubor Ladicky, Chris Russell, Pushmeet Kohli, and Philip H.S. Torr. Associative hierarchical crfs for object class image segmentation. In ICCV, 2009.
46. S. Kumar and M. Herbert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proc. IEEE Int'l Conf. Computer Vision*, 2003.
47. M. P. Kumar and P. Torr. Efficiently solving convex relaxations for map estimation. In ICML, 2008.
48. http://vision.middlebury.edu/stereo/
49. D. Hoiem, A. A. Efros, and M. Hebert. Automatic photo pop-up. In *ACM SIGGRAPH*, pages 577–584, 2005.