

# Segmentation of Interstitial Lung Disease Patterns in HRCT Images

Jatindra Kumar Dash<sup>a</sup>, Vaddepalli Madhavi<sup>a</sup>, Sudipta Mukhopadhyay<sup>a</sup>, Niranjan Khandelwal<sup>b</sup>  
and Prafulla Kumar<sup>c</sup>

<sup>a</sup>Department of Electronics & Electrical Communication Engineering,  
Indian Institute of Technology Kharagpur, India;

<sup>b</sup>Department of Radio-diagnosis & Imaging,  
Post Graduate Institute of Medical Education & Research, Chandigarh, India

<sup>c</sup>Department of Electronics & Information Technology,  
New Delhi, India

## ABSTRACT

Automated segmentation of pathological bearing region is the first step towards the development of lung CAD. Most of the work reported in the literature related to automated analysis of lung tissue aims towards classification of fixed sized block into one of the classes. This block level classification of lung tissues in the image never results in accurate or smooth boundaries between different regions. In this work, effort is taken to investigate the performance of three automated image segmentation algorithms those results in smooth boundaries among lung tissue patterns commonly encountered in HRCT images of the thorax. A public database that consists of HRCT images taken from patients affected with Interstitial Lung Diseases (ILDs) is used for the evaluation. The algorithms considered are Markov Random Field (MRF), Gaussian Mixture Model (GMM) and Mean Shift (MS). 2-fold cross validation approach is followed for the selection of the best parameter value for individual algorithm as well as to evaluate the performance of all the algorithms. Mean shift algorithm is observed as the best performer in terms of Jaccard Index, Modified Hausdorff Distance, accuracy, Dice Similarity Coefficient and execution speed.

**Keywords:** Lung Diseases, segmentation, Gaussian Mixture Model, Hidden Markov Random Field, Mean Shift

## 1. INTRODUCTION

Segmentation of the lung tissue patterns into several regions is one of the most important steps for the computer aided detection of abnormal region in High Resolution Computed Tomography (HRCT) images of the thorax. Most of the work reported in the literature<sup>1-7</sup> related to lung tissue pattern classification aim to classify the fixed sized (either square or circular) region of interest (ROIs). These types of block level classification never result in accurate or smooth boundary of the pathological bearing region (PBR). So many boundary pixels are misclassified to one among the unintended neighboring regions. In practice, the spread of the disease is not symmetric but arbitrary. In addition at times, these predefined size ROIs are not representative of those diseases patterns. For example, an ROI of size  $32 \times 32$  may represent an ILDs pattern called consolidation or emphysema well but may not represent patterns like fibrosis or micronodule. Therefore, the entire spread of the disease patterns needs to be considered to represent the pathology for computer-aided detection or analysis of the disease. It is also found that for most of the content-based image retrieval (CBIR) system for lung images reported in the literature<sup>8-11</sup> delineation of these ROIs is a manual process which not only involves huge human interventions, but also leaves chances of human error. Automated unsupervised clustering techniques could result in different regions with smooth boundaries between them which is helpful for representing the pathology better for both of these applications (classification and retrieval). Automated segmentation also substantially reduces the amount of human intervention for the image annotation task during the database creation for a CBIR system. The aim of this work is to evaluate the performance of three such algorithms in the context of automated clustering of lung tissues affected with Interstitial Lung Diseases (ILDs) into different regions of interest.

Jatindra Kumar Dash (jatinkdash@gmail.com), Sudipta Mukhopadhyay (smukho@gmail.com)

Medical Imaging 2015: Computer-Aided Diagnosis, edited by Lubomir M. Hadjiiski, Georgia D. Tourassi,  
Proc. of SPIE Vol. 9414, 94142R · © 2015 SPIE · CCC code: 1605-7422/15/\$18 · doi: 10.1117/12.2079072

## 2. EXPERIMENTAL SETUP

The publicly available database<sup>12</sup> is used to evaluate the performance of three segmentation algorithms. This database contains 110 ILD cases. All the images in the database are in DICOM format and are obtained with a slice thickness of 1 mm, and inter-slice distance of 10 mm. The images in the database are annotated with 2-D arbitrary region of interests (AROI) representing 17 image-findings related to ILDs by two experienced radiologists with 15 and 20 years of experience at University Hospitals of Geneva (HUG). The diagnosis of each case are confirmed either by pathology (biopsy, bronchoalveolar washing) or a laboratory/specific test. Among these, six most frequently seen lung tissue patterns<sup>13</sup> such as consolidation, emphysema, fibrosis, ground glass, micronodule and normal are considered for this study. From each lung tissue pattern, 25 AROIs are collected, and the final database contains a total of 150 images each with a single arbitrary region of interest (AROI). Three popular algorithms selected for this study are: Gaussian Mixture Model with Expectation Maximisation (GMM-EM)<sup>14-16</sup>, Hidden Markov Random Field with Expectation Maximisation (HMRF-EM)<sup>17</sup> and Mean Shift Clustering.<sup>18</sup>

Both the GMM and the HMRF are parametric approaches. In GMM, the intensity distribution of the image is modeled as linear combinations of fixed number Gaussian. In Markov Random Field (MRF) the intensity distribution of the image is modeled as a random field which posses the Markovian property. It expresses the dependency of the pixel intensity on its neighboring pixels in terms of Markov parameters. The segmentation using both GMM-EM and HMRF-EM algorithms are iterative procedures where each iteration consists of two distinct steps such as (i) determination of appropriate labels for each pixel by Maximum-a-Posteriori (MAP) Estimation, and (ii) estimation of model parameters by an Expectation Maximisation (EM) algorithm. Both of the methods (GMM-EM and HMRF-EM) need the parameter number of clusters (' $K$ ') to be initialised. Mean Shift (MS) is a data clustering algorithm. Mean Shift considers each pixel as data point and searches local maximal density pixels. Then it groups all the pixels defined by each local maximal density pixel into different clusters. There are mainly three parameters associated with MS. They are the spatial bandwidth  $Hs$ , the range bandwidth  $Hr$ , and the minimum segment area  $S$ . For our implementation, we consider  $Hs = Hr$  and  $S = 30$  pixels. All three segmentation methods are implemented using MATLAB 2011b on a PC with Intel(R) Core(TM)2 Due 2.80 GHZ CPU, 8GB RAM and Microsoft Windows 7 Professional(64-bit) Operating System. Both the algorithm HMRF-EM and GMM-EM are implemented using Matlab toolbox. Implementation of Georgescu and Christoudias's EIDSON system<sup>19</sup> is used for mean shift segmentation. For each of the algorithms, the segmentation is performed considering only the pixel intensity value as a feature. As the objective is to cluster lung tissues to different regions with similar properties, only the lung field is given as input to each algorithm. The lung field is extracted using the lung mask provided with the database.<sup>12</sup> For each of the algorithm, the output cluster which has maximum overlap with the ROI drawn by the radiologists is considered as the segmented output.

## 3. RESULTS AND DISCUSSION

A 2-fold cross validation approach is followed for the selection of optimum parameter and evaluation of each algorithm. In each fold, the optimum parameter value for the algorithm is selected using the training data and performance evaluation is done using the test data. As the Jaccard Index (JI) and Modified Hausdorff Distance (MHD) are the most popular measures for the evaluation of segmentation algorithm, these two are considered for the selection of optimum parameter value for each of the algorithm. In each fold, the parameters such as number of clusters ' $K$ ' (used in case of HMRF-EM and GMM-EM) and the spatial bandwidth  $Hs$  used in case of mean shift algorithm are varied ( $K=1$  to 25 with a step size of 1 and  $Hs=2$  to 64 with a step size of 2). The performance indices, such as JI and MHD are plotted against the respective controlling parameter. Fig. 1a through Fig. 1c show the variation in the JI and MHD (in number of pixels) in each of the 2-folds as well as the average of them with respect to the corresponding parameter value of each algorithm. The parameter value that results in maximum JI is chosen for each algorithm. It is observed that at the same operating point each of the algorithms results in near minimum value of MHD. The average accuracy and Dice Similarity Coefficient obtained using 2-fold cross validation by each of the methods are reported in Table 1. The execution time for each of the algorithm is reported for comparison. MS algorithm is found to have the highest JI (0.4133) and lowest MHD (19.02) among all the methods under consideration. The highest value of accuracy, Jaccard Index,

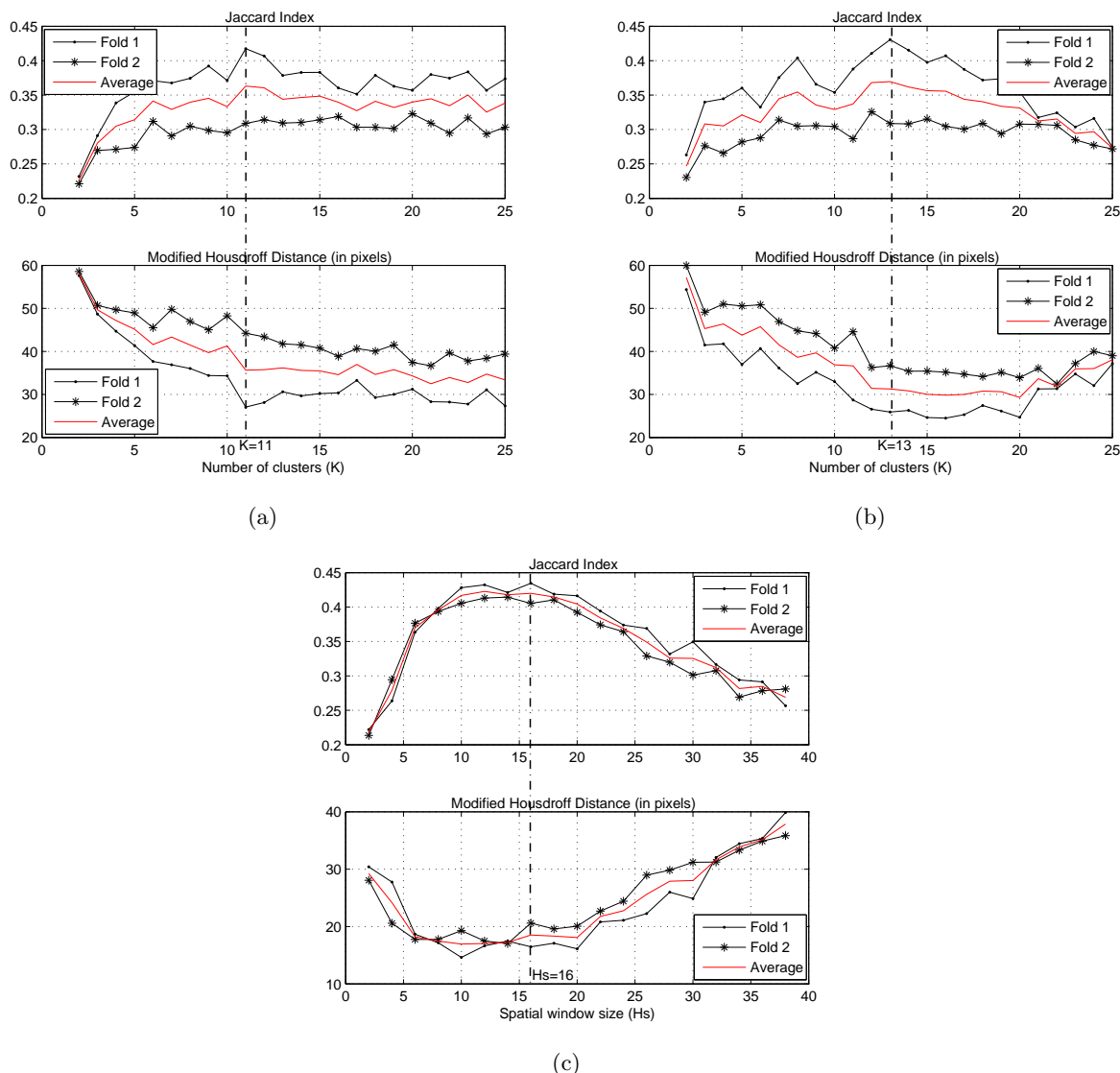


Figure 1: Average performance vs. parameter (a) GMM-EM (b) HMRf-EM (c) Mean Shift

Table 1: Average performance and its standard deviation at optimum parameter setting

Metric	GMM-EM	HMRf-EM	MS
Jaccard Index	$0.3328 \pm 0.03$	$0.3596 \pm 0.03$	$0.4133 \pm 0.06$
MHD	$37.7195 \pm 6.46$	$31.6199 \pm 7.27$	$19.0215 \pm 7.09$
Accuracy	$0.7215 \pm 0.09$	$0.7893 \pm 0.10$	$0.8036 \pm 0.08$
DSC	$0.4558 \pm 0.03$	$0.4964 \pm 0.04$	$0.5555 \pm 0.07$
Execution Time (Sec.)	18.98	171	11.71

DSC, lowest execution time and MHD along with low standard deviation make mean shift most suitable for the segmentation of ILDs patterns. Fig. 1 shows that the MS is more consistent in its result for both the data sets in 2-fold cross validation approach compared to other algorithms. The superiority of MS over other two methods (GMM-EM and HMRf-EM) is also supported by the qualitative results shown for all the lung tissue patterns in Fig. 2. The results are obtained for the corresponding optimal parameter of individual algorithm.

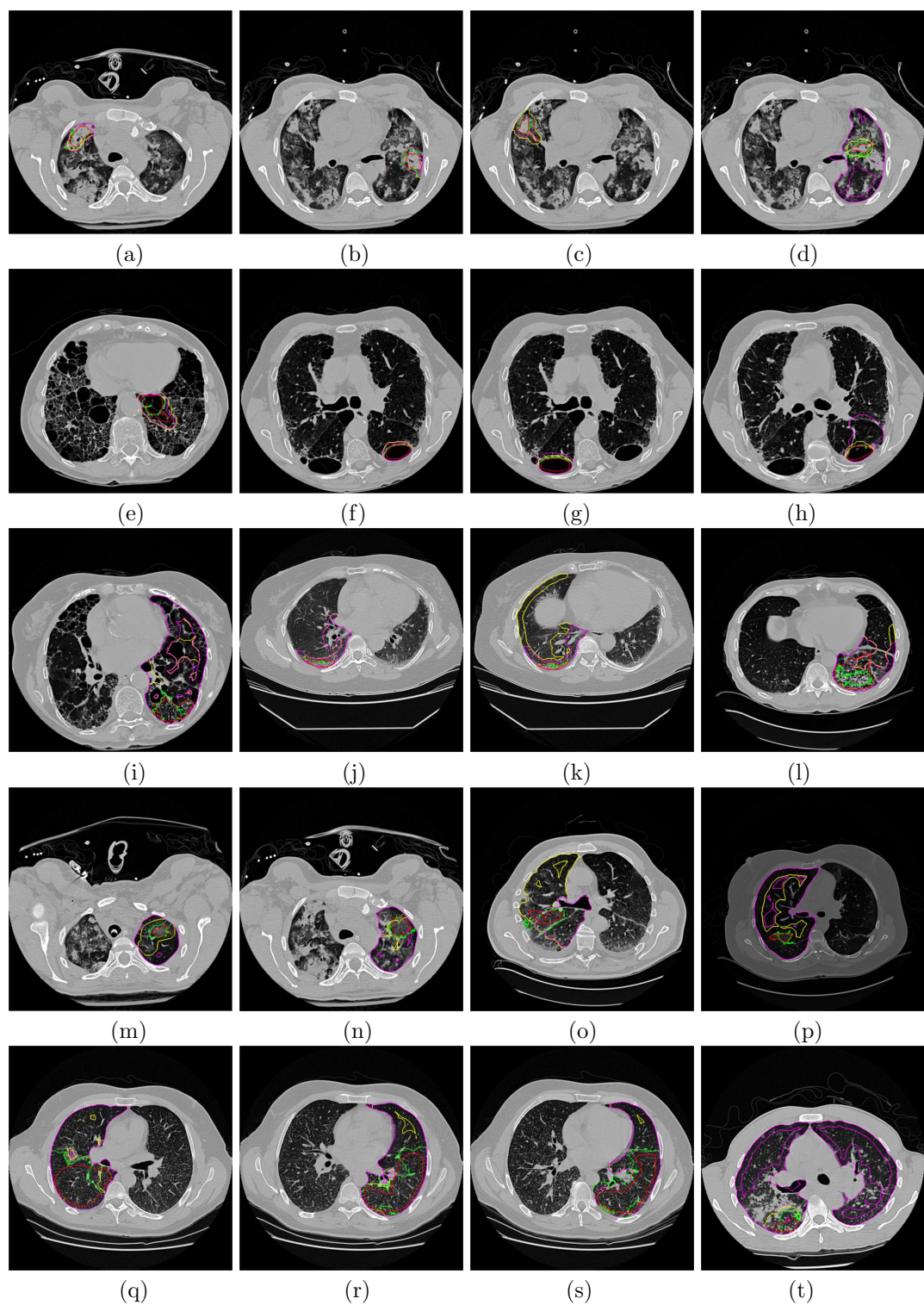


Figure 2: Qualitative segmentation result for GMM-EM (Magenta), HMRF-EM (Yellow) and MS (Green) for five abnormal lung tissues such as Consolidation (a-d), Emphysema (e-h), Fibrosis (i-l), Ground glass (m-p), Micronodule (q-t). Readers are requested to see the soft copy for the interpretation of colour use.

## 4. CONCLUSION

Automated segmentation of pathological lung tissue is the first step towards the development of lung CAD. This article evaluates the performance of three popular image segmentation algorithm (viz. GMM-EM, HMRf-EM and Mean Shift) in the context of automated segmentation or clustering of lung tissues affected with Interstitial Lung Diseases on HRCT lung images. The performance evaluation is done through a set of experiments on 2D HRCT lung images collected from publicly available database.<sup>12</sup> Optimum parameters for each of the said algorithm are empirically observed using 2-fold cross validation approach. The segmentation result obtained using each method for its optimum selected parameter are compared. The Mean Shift is found to be the best and consistent performer among all methods considered for the application.

## 5. ACKNOWLEDGMENT

This work has been supported by Ministry of Communications and Information Technology, Department of Electronics and Information Technology, Govt. of India, Grant number 1(3)2009-ME&TMD and 1(2)2013-ME&TMD/ESDA. The authors are thankful to Dr. Naveen Kalra, Dr. Mandeep Kumar Garg, Dr. Nidhi Prabhakar and Mr. Amitabh Sinha, Department of Radio-diagnosis & Imaging, Post Graduate Institute of Medical Education & Research, Chandigarh, India for providing valuable advice for our research. Thanks to Indian Institute of Technology Kharagpur for supporting the institute scholars with scholarship and providing the research facilities. The authors would like to acknowledge all the members of Project Review and Steering Group for their support, encouragement and valuable suggestions. Finally, thanks to all the members of Image Processing and Computer Vision Laboratory, Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, India for their timely help and support.

## REFERENCES

1. S. C. Park, J. Tan, X. Wang, D. Lederman, J. K. Leader, S. H. Kim, and B. Zheng, "Computer-aided detection of early interstitial lung diseases using low-dose ct images," *Physics in medicine and biology* **56**(4), p. 1139, 2011.
2. L. Sorensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *Medical Imaging, IEEE Transactions on* **29**(2), pp. 559–569, 2010.
3. K. T. Vo and A. Sowmya, "Multiple kernel learning for classification of diffuse lung disease using hrct lung images," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 3085–3088, IEEE, 2010.
4. K. T. Vo and A. Sowmya, "Directional multi-scale modeling of high-resolution computed tomography (hrct) lung images for diffuse lung disease classification," in *Computer Analysis of Images and Patterns*, pp. 663–671, Springer, 2009.
5. L. Sørensen, S. B. Shaker, and M. De Bruijne, "Texture classification in lung ct using local binary patterns," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 934–941, Springer, 2008.
6. M. Prasad and A. Sowmya, "Multi-level classification of emphysema in hrct lung images using delegated classifiers," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 59–66, Springer, 2008.
7. K. G. Kim, J. M. Goo, J. H. Kim, H. J. Lee, B. G. Min, K. T. Bae, and J.-G. Im, "Computer-aided diagnosis of localized ground-glass opacity in the lung at ct: Initial experience 1," *Radiology* **237**(2), pp. 657–661, 2005.
8. C.-R. Shyu, C. Pavlopoulou, A. C. Kak, C. E. Brodley, and L. S. Broderick, "Using human perceptual categories for content-based retrieval from a medical image database," *Computer Vision and Image Understanding* **88**(3), pp. 119–151, 2002.
9. C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "Assert: a physician-in-the-loop content-based retrieval system for hrct image databases," *Computer Vision and Image Understanding* **75**(1), pp. 111–132, 1999.

10. J. G. Dy, C. E. Brodley, A. Kak, L. S. Broderick, and A. M. Aisen, "Unsupervised feature selection applied to content-based retrieval of lung images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **25**(3), pp. 373–378, 2003.
11. S. D. MacArthur, C. E. Brodley, A. C. Kak, and L. S. Broderick, "Interactive content-based image retrieval using relevance feedback," *Computer Vision and Image Understanding* **88**(2), pp. 55–75, 2002.
12. A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," *Computerized Medical Imaging and Graphics* **36**(3), pp. 227–238, 2012.
13. W. R. Webb, N. L. Müller, D. P. Naidich, *et al.*, *High-resolution CT of the lung*, vol. 1, Lippincott Williams & Wilkins Philadelphia, 2001.
14. D. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics* , pp. 659–663, 2009.
15. Z. Fu and L. Wang, "Color image segmentation using gaussian mixture model and em algorithm," in *Multimedia and Signal Processing*, pp. 61–66, Springer, 2012.
16. R. Farnoosh and B. Zarpak, "Image segmentation using gaussian mixture model," *IUST International Journal of Engineering Science* **19**(1), pp. 29–32, 2008.
17. Q. Wang, "Hmrf-em-image: implementation of the hidden markov random field model and its expectation-maximization algorithm," *arXiv preprint arXiv:1207.3510* , 2012.
18. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(5), pp. 603–619, 2002.
19. B. Georgescu and C. Christoudias, "The edge detection and image segmentation (edison) system," *Robust Image Understanding Laboratory, Rutgers University. Code available at <http://web.archive.org/web/20060617020918/www.caip.rutgers.edu/riul/research/code/EDISON/index.html>* , 2003.