# A Hierarchical Random Field Model For Image Segmentation And Cataloging

**Ms. Nidhi Soni[1], Prof. Sachin Malviya[2]**
[1, 2]Dept of CSE
[1, 2] SBITM Betul, MP, India

***Abstract-*** *The partitioning also called as segmenting and cataloging (labeling) images is a crucial problem in Computer Vision. The efficient and most appropriate hierarchical Conditional Random Field model deal with the problem of labeling images by identification of object. Labeling a new image, is a supervise learning where select the cluster and use the associated CRF model to label the selected image. In this research a given test image is firstly use the Conditional Random Field model to obtain initial labels then find the cluster of the image. Finally, inference model is used to relabel the image by the CRF model associated with this cluster. For the most prominent compare and extract similar images, here we introduce a new image descriptor, the label based descriptor which summarizes the semantic information of a labeled image. Here, labeling and segmentation results are shown for specific images.*

***Keywords***– CRF, Label Descriptor, position, appearance, structural

## I. INTRODUCTION

### 1.1 Background

Segmentation is the decomposition of an image into these objects and regions by associating every point with the object that it corresponds to. Most humans can easily segment an image. Computer automated segmentation is a difficult problem. Labeling identifies an object record of an information page based on the labeling of object elements within an object record and labels object elements based on the identification of an object record that contains the object elements.

### 1.2 Motivation

Among the various high level and low level computer vision tasks, semantic scene interpretation or image understanding is considered an important and challenging task. For example, suppose we need to design a system capable of discriminating an office scene from a kitchen scene (Figure 1.1). This involves identifying the objects of the scene as some meaningful entities. The initial step would be to parse the image, segment it into meaningful regions and label those segments as known entities. For classifying the scene as an office area, we need to first segment objects like a computer, keyboard, books etc. Similarly for classifying it into a kitchen scene we need to identify the kitchen objects like utensils, microwave, and gas stove etc. The system should also be able to capture the context relationship between the various objects like keyboard and computer. This gives us a motivation to build an image modeling system, which would provide us with a different representation of the underlying image; representation which will be useful in carrying out high level tasks like scene interpretation or image understanding. In this thesis we propose a probabilistic hierarchical image modeling system based on Conditional Random Fields The model efficiently learns the complex class dependencies and produces labels over the input image.

Fig. 1.1: Images showing office and kitchen scene. The discrimination can be done only by first identifying the objects contained in the scene like microwave or computer.

The performance of model has been evaluated and verified by experimenting with two key problems of computer vision, namely image labeling (a multiclass problem where each pixel belongs to one of the predefined classes) and object recognition( a two class problem where object of interest are identified). The images considered for the two tasks are natural images encountered commonly in our surroundings. These images may contain both man-made as well as other natural objects such as sky, vegetation etc. It is assumed that the images are static in nature and no motion information is included.

## II. REVIEW OF LITERATURE

A variety of signal processing, image processing and machine learning methods exist for meaningful and efficient image modeling tasks. We can broadly classify those methods as non-probabilistic and probabilistic. The framework is categorized as non-probabilistic if the overall labeling objective is not given by a consistent probabilistic formulation, even if the framework utilizes probabilistic methods to address parts of it. Rule-based context (Ohta, 1980) and relaxation labeling (Rosenfeld et al., 1976) are two main techniques among the non-probabilistic approaches, other than using weak measures to capture spatial smoothness of natural images using filters with local neighborhood supports. Ohta used a rule-based approach to assign labels to regions obtained from a single-pass segmentation. The stumbling block in case of rule based approaches was their inability to handle the statistical variations in the data. Singhal (Singhal et al., 2003) proposed the use of conditional histograms to make a local decision regarding assigning a label to a new region given the previous regions labels, to avoid the absolute constraints imposed by the rule-based approaches.

However, such a sequential implementation of context will suffer if an intermediate region is assigned a wrong label. In late-1970s, the VISIONS schema system was proposed by Hanson and Riseman (1978), which provides a framework for building a general interpretation system based on the output of several small special purpose experts. Each scheme is an 'expert' at recognizing one type of object. The schema instances run concurrently and communicate asynchronously. Each expert gives its prediction about the presence and location of objects in the image. Based on hand a coded if-then rule, the system analyzes the consistency among the different hypotheses in order to arrive at reliable decisions. Later, a similar idea was presented by Strat (1992) in a system called CONDOR to recognize natural objects for the visual navigation of an autonomous robot. The system used hand-written rules to encode the knowledge database of the system. A collection of rules (context sets) defines the conditions under which it is appropriate to use an operator to identify a candidate region or object. While analyzing generic 2D images, manually defining the context information is not a very convenient task. Instead, one needs to derive the context directly from the input image itself. A comprehensive review of the use of context for recognizing natural objects in color images of outdoor scenes is given in (Batlle et al., 2000). Torralba and Sinha (2001) proposed a framework for modeling the relationship between context and object properties based on the correlation between the statistics of low-level features across the entire scene and the objects that it contains. In Summary, the non probabilistic models suffered because they tried to manually visualize the context and extract them using rule based expert.

Image analysis application have embedded uncertainty (Rao and Jain, 1988; Winston, 1970) in them, which needs a more robust and principled approach. Even though efforts were made to represent global uncertainty using graph structures, the tools available for learning and inference over these structures were limited. These adhoc procedures for resolving ambiguities using rules systems were unreliable or constrained to a narrow domain. Image classification methods consider both spectral statistics and uncertainties in the dependencies for the classification task, which intuitively support probabilistic models. Probabilistic models, model the uncertainties in the form of an underlying probability distribution. The problem then reduces to a problem of learning the relevant dependency parameters, computing the probability distribution and inference using the distribution. The probabilistic techniques fall largely under the paradigm of probabilistic graphical models. Graphical models are very efficient and intuitive frameworks for building probabilistic models for a set of random variables which represent a particular domain. Quoted from (Jordan, 2004)- Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering – uncertainty and complexity. Graphical models combine ideas from graph theory and probability theory in an elegant manner to represent complex systems. Efficient training and inference procedures make them a popular choice for many problem domains. They provide a powerful yet flexible framework for representing and manipulating global probability distributions defined by relatively local constraints. The nodes of the graph represent the random variables and the edges represent the constraints among the random variables. The undirected graphical models widely used in the vision literature are generally termed as Random Fields.

Parametric methods usually involve optimizing a Conditional Random Field (CRF) model which evaluates the probability of assigning a particular label to each pixel, and the probability of assigning each pair of labels to neighboring pixels. A parametric method usually has a learning phase where the parameters of the CRF models are optimized from training examples and an inference phase where the CRF model is applied to label a test image. In contrast to parametric methods, nonparametric methods do not involve any training at all. The basic idea of these methods is to transfer labels from a retrieval set which contains semantically similar images.

Nonparametric methods tend to be more scalable than parametric methods because it is easy for nonparametric methods to incorporate new training examples and class labels. To introduce a hierarchical two-stage CRF model this combines the ideas used in both parametric and nonparametric image labeling methods. In addition to learning a global CRF model from all the training images, group training data into clusters of images with similar spatial object class layout and object appearance, and train a separate CRF model for each cluster. Given a test image, first run the global CRF model to obtain initial pixel labels. We then find the cluster with most similar images, Finally, relabel the input image by the CRF model associated with this cluster. To effectively compare and extract similar images, we introduce a new image descriptor: The label-based descriptor which summarizes the semantic information of a labeled image.

## III. RESEARCH AND COLLECT IDEA

**Positional information:**

To calculate position score is important factor to label and segment image in this paper. Let us consider eye portion of Human face. Eye portion gets divided into 2 clusters say cluster1and cluster2.

Considering position for Cluster1 say P1 = 1
Position for cluster2 say P2 = 0.9

So the position score (PS1) for Eye portion can be calculated by summation of P1 and P2 and Let us consider there are two more portion in Human Face as Hair and Nose. Following the same procedure as that of for Eye, Calculate the position score for Hair and Nose portion. Then the mean position score for Human face calculated as

$$(PS1 + PS2 + PS3) / 3$$

**Appearance Score**

The appearance score calculated on the basis of LUV [15] points. Say There is 2 clusters in the Eye portion of Human Face. Each cluster having LUV points. The formula for LUV point is as follows

The LScore for C1(cluster 1) calculated as

$$1- (L – L_i ) / L$$

where L is Lpoints of trained image and Li is L1Points of test image

The UScore for cluster1 calculated as

$$1-(U – U_i ) / U$$

where U is Upoints of trained image and Ui is U1Points of test image

The VScore for cluster1 calculated as

$$1-(V – V_i ) / V$$

where V is Lpoints of trained image and Vi is V1Points of test image then the

$$\text{Total LUV Score} = (\text{LScore} + \text{UScore} + \text{VScore}) / 3.0$$

This calculated LUVScore represents appearance score for cluster1 say APS1. Similarly For cluster2 say appearance score is APS2. The selection of maximum appearance score say APS1 gives appearance score for Eye portion.

**Structural Score**

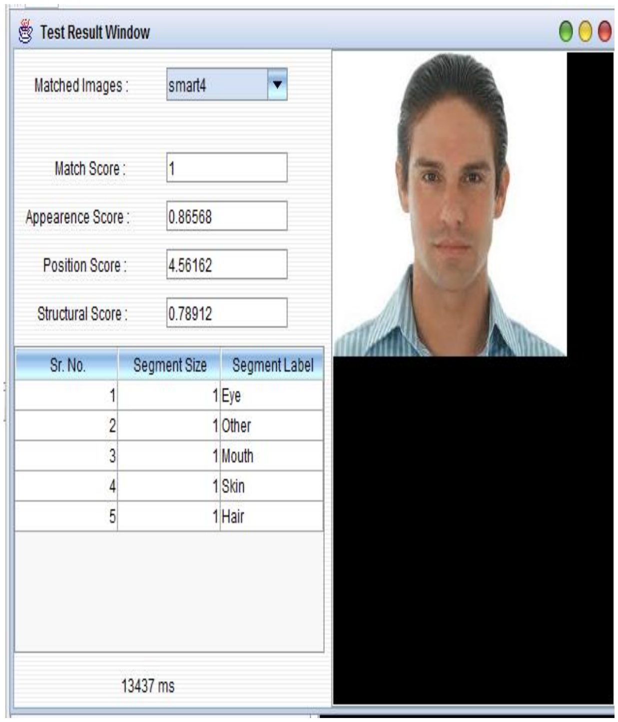
Structural score plays a very important role in image labeling it means means calculating the distance between two parts of the same object. Consider the distance (di) between Eye and Nose is 20 units of trained image.

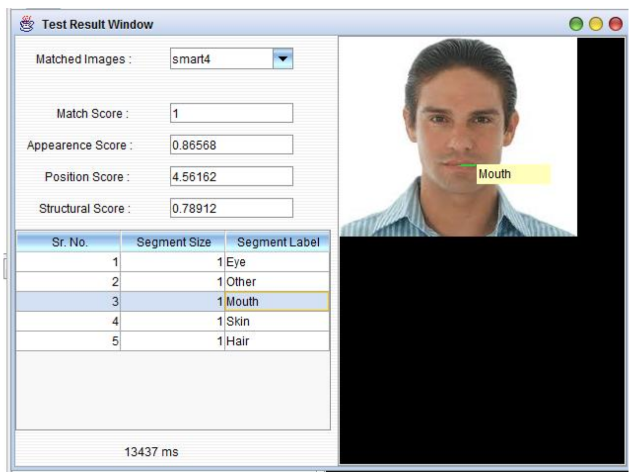
The distance (d1) between Eye and Nose is 18 units of test image.

Then the structural score = 1-(d-d1)/d
=1-(20-18)/20
=0.9

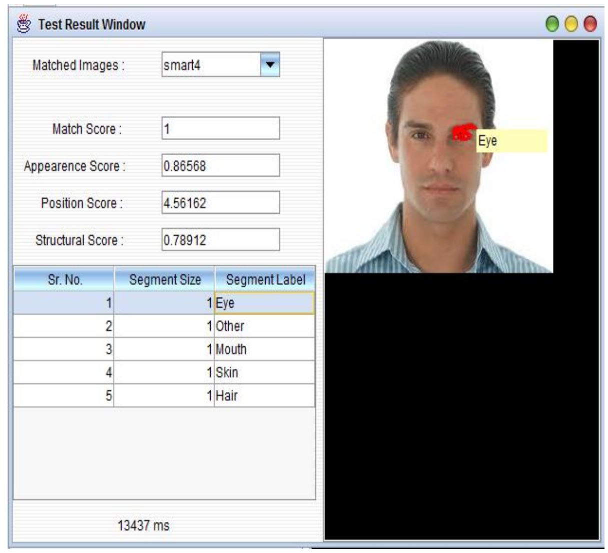
Say 0.9 is the structural score1 for Eye and Nose.

Similarly Say 0.8 is the structural score2 for Eye and hair Then Structural score for
Human Face = (structural score1 + structural score2) / 2
The experimental results are shown below:








## IV. CONCLUSION

This paper, presents an approach to segment the image according to label generated during training process. The trained image gets compared with the test image to shows position score, appearance score and structural score according to input image. Here we implement a approach to select desired label using supervise learning process that obtained at the time of testing, then that part of the image gets segmented. The generated result is helpful to distinguish different parts of the image on the basis of label.

## REFERENCES

[1] C. Galleguillos, A. Rabinovich, and S. Belongie “Object categorization using co-occurrence, location and appearance in Computer Vision and Pattern Recognition”, IEEE 2008

[2] Thomas Deselaers and vittorio,”Visual and semantic similarity in ImageNet”,IEEE 2010

[3] Gemma Roig,Xavier Boix,Joan serrat,” Hierarchical CRF with Product Label Spaces for Parts-based Models”,IEEE 2010

[4] Jun Tang,” A color Image Segmentation Based on Region Growing”,2010

[5] Hichem Sahbi,”Superpixel-Based Object Class Segmentation Using Conditional Random Fields”,IEEE 2011

[6] Yushi Jing,Michele Covell,”Comparision Of Clustering Approaches For Summarizing Large Populations Of Images”,IEEE 2010

[7] Jordan Reynolds and Kevin Murphy,” Figure-ground segmentation using a hierarchical conditional random field”,IEEE 2007.

[8] Lubor Ladickly and Chris Russell,"Associative Hierarchical CRF for Object Class Image Segmentation",IEEE 2009
[9] Qixing Huang, Mei Han, Bo Wu and Sergey Ioffe," A Hierarchical Conditional Random Field Model for Labeling and Segmenting Images of Street Scenes",2011
[10] Ce Liu and Jenny Yuen,"Nonparametric scene Parsing:Label Transfer via Dense Scene Alignment",IEEE 2009
[11] Peng Lu and He Ren,"Hierarchical Conditional Random Fields For Chinese Named Entity Tagging",IEEE 2009
[12] William Freeman and Mark Rubin," Context-based vision system for place and object recognition",IEEE 2008
[13] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman.LabelMe: A Database and Web-Based Tool for Image Annotation. International Journal of Computer Vision.
[14] Gustavo Carneiro, Antoni B. Chan, Pedro J. Moreno, Nuno Vasconcelos,," Supervised Learning of Semantic Classes for Image Annotation and Retrieval", IEEE 2007
[15] Luo Ronghua, Min Huaqing," Hybrid Conditional Random Fields for Multi-object Tracking with a Mobile Robot", IEEE 2010
[16] Luis von Ahn and Laura Dabbish," Labeling Images with a Computer Game", IEEE 2004
[17] Paul Schnitzspan, Mario Fritz, Stefan Roth, Bernt Schiele," Discriminative Structure Learning of Hierarchical Representations for Object Detection",IEEE 2008
[18] Gorsevski, P. V., P. E. Gessler, and P. Jankowski (2003). Integrating a fuzzy k-means classification and a bayesian approach for spatial prediction of landslide hazard. Journal of Geographical Systems, 5, 223–251.