Markov Random Field Models in Computer Vision

S. Z. Li

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 2263
szli@ntu.ac.sg

Abstract. A variety of computer vision problems can be optimally posed as Bayesian labeling in which the solution of a problem is defined as the maximum *a posteriori* (MAP) probability estimate of the true labeling. The posterior probability is usually derived from a prior model and a likelihood model. The latter relates to how data is observed and is problem domain dependent. The former depends on how various prior constraints are expressed. Markov Random Field Models (MRF) theory is a tool to encode contextual constraints into the prior probability. This paper presents a unified approach for MRF modeling in low and high level computer vision. The unification is made possible due to a recent advance in MRF modeling for high level object recognition. Such unification provides a systematic approach for vision modeling based on sound mathematical principles.

1 Introduction

Since its beginning in early 1960's, computer vision research has been evolving from heuristic design of algorithms to systematic investigation of approaches for solving vision problems. In their search for solutions, researchers have realized the importance of contextual information in image understanding. In this process, a variety of vision models using context have been proposed. Among these are Markov Random Field (MRF) theory based models (of which analytic regularization theory based models are special cases).

MRF modeling is appealing for the following reasons (Preface of [4]): (1) One can systematically develop algorithms based on sound principles rather than on some *ad hoc* heuristics for a variety of problems; (2) It makes it easier to derive quantitative performance measures for characterizing how well the image analysis algorithms work; (3) MRF models can be used to incorporate various prior contextual information or constraints in a quantitative way; and (4) The MRF-based algorithms tend to be local, and tend themselves to parallel hardware implementation in a natural way.

Complete stochastic vision models based on MRF are formulated within the Bayesian framework. The optimal solution of a problem is defined as the maximum *a posteriori* (MAP) probability estimate of the truth, the best that one can get from random observations. Most of vision problems can be posed as one of labeling using constraints due to prior knowledge and observations. In this

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

case, the optimal solution is defined as the MAP labeling and is computed by minimizing a posterior energy. The posterior probability is derived, using the Bayesian rule, from a prior model and a likelihood model. The latter relates to how data is observed and is problem domain dependent. The former depends on how various prior constraints are expressed. Results from MRF theory provide us tools to encode contextual constraints into the prior probability. This is the main reason for MRF vision modeling.

MRF based approaches have been successful in modeling low level vision problems such as image restoration, segmentation, surface reconstruction, texture analysis, optical flow, shape from X, visual integration and edge detection (There are a long list of references. Readers may refer to collections of papers in [15, 4] and references therein). Relationships between low level MRF models are discussed in [16, 7] and those between MRF models and regularization models in [16]. The unifying theme of Bayesian modeling for low level problems appear for example, in [7, 2, 18]. A prototypical Bayesian formulation using MRF is that of Geman and Geman [8] for image restoration.

Investigation of MRF modeling in high level vision such as object matching and recognition, which is more challenging (Introduction of [15]), begins only recently. In an initial development of an MRF model for image interpretation [17], the optimal solution is defined as the MAP labeling. Unfortunately, the posterior probability therein is derived using heuristic rules instead of the laws of probability, which dissolves the original promises of MRF vision modeling. A coupled MRF network for simultaneous object recognition and segmentation is described in [5].

In a recent work [11], an MRF model for high level object matching and recognition is formulated based on sound mathematical principles. Mathematically, like the typical low level MRF model of Geman and Geman [8], the model utilizes MRF theory to characterize prior contextual constraints. This, plus an observation model for the joint likelihood, enables the derivation of the posterior probability. The model [11] is more general than the low level model [8] in that it makes use of contextual observations and allows non-homogeneous sites and non-isotropic neighborhood systems.

This makes it possible to formulate a larger number of low and high level problems in the single Bayesian framework in a systematic way. This is of significance in both theory and practice. It provides a rational approach on a sound basis. It implies some intrinsic properties or common mechanisms in seemingly different vision problems. It also suggests that these problems could be solved using a similar architecture.

This paper presents such a unified MRF modeling approach [10]. The systematic way to the MRF modeling is summarized as five steps:

1. Pose the vision problem as one of labeling in which a label configuration represents a solution (Sec.2).
2. Further pose it as a Bayesian labeling problem in which the optimal solution is defined as the MAP label configurations (Sec.3),
3. Use Gibbs distribution to characterize the prior distribution of label configurations (Sec.3.2),

4. Figure out the likelihood density of data based on an assumed observation model (domain dependent and exemplified in Sec.4) and
5. Use the Bayesian rule to derive the posterior distribution of label configurations, to measure the cost of a solution (Sec.3.3 and Sec.4).

(How to search for the MAP configuration is not discussed in this paper.) Two MRF models are described as cases in low and high level vision, respectively. The first is the prototypical Geman and Geman's low level model (Sec.4.1) and the second is the recent high level object recognition model [11] (Sec.4.2). The latter is described using the Geman-Geman's model as the reference point. The presentation is done in such as way that parallel concepts are seen clearly.

2 Vision Problems as Labeling

2.1 The Labeling Problem

A *labeling problem* is specified in terms of a set of sites and a set of labels. Let $\mathbf{d}$ be a set of m discrete *sites*.

$$\mathbf{d} = \{1, \ldots, m\} \tag{1}$$

The ordering of the sites is not important; their relationship is determined by a *neighborhood system* (the definition of neighborhood systems is central in MRF theory and will be introduced later). Let $\mathbf{D}$ be a set of *labels*. Labeling is to assign a label from $\mathbf{D}$ to each of the sites in $\mathbf{d}$.

A set of sites can be categorized in terms of their "homogeneity" and a set of labels in terms of their "continuity". Sites on a lattice such as those corresponding to an array of image pixels are considered as being spatially homogeneous whereas those corresponding to features extracted from images such as critical points, line segments or surface patches are considered as being inhomogeneous. Usually, homogeneous sites lead to an isotropic neighborhood system and inhomogeneous sites to an anisotropic neighborhood system.

A label can be continuous such as a continuous intensity or range value. The value can usually be confined to a real interval

$$f_i \in \mathbf{D} = [x_l, x_h] \tag{2}$$

In this case, there are an infinite number of labels. In the other case, a label may be discrete

$$f_i \in \mathbf{D} = \{1, \cdots, M\} \tag{3}$$

For example, a label may index to one of model object lines or regions.

Let $F = \{F_1, \ldots, F_m\}$ be a family of random variables defined on $\mathbf{d}$, in which each random variable F_i assumes a value in $\mathbf{D}$. A joint event $\{F_1 = f_1, \ldots, F_m = f_m\}$, abbreviated $F = f$, is a realization of F where $f = \{f_1, \ldots, f_m\}$ is called a *configuration* of F. A configuration may represent an image, an edge map, or a matching (mapping) from image features to object features. The set of all configurations is

$$\mathbf{S} = \mathbf{D}^m = \underbrace{\mathbf{D} \times \mathbf{D} \cdots \times \mathbf{D}}_{m \text{ times}} \tag{4}$$

The space of admissible solutions may be identical to $\mathbf{S}$ or if additional constraints are imposed, a subset of it. A configuration f can be interpreted in one of the two ways: It is a mapping $f: \mathbf{d} \longrightarrow \mathbf{D}$; or it is a labeling $\{f_1, \ldots, f_m\}$ of the sites.

2.2 Labeling Problems in Vision

In terms of the homogeneity and the continuity, we may classify a vision labeling problem into one of the following four categories:

LP1: Homogeneous sites with continuous labels.
LP2: Homogeneous sites with discrete labels.
LP3: Inhomogeneous sites with discrete labels.
LP4: Inhomogeneous sites with continuous labels.

The former two categories characterize low level processing performed on observed images and the latter high level processing on extracted token features. The following describes some vision problems in terms of the categories.

Restoration of grey scale images, or image smoothing, is an LP1. The set $\mathbf{d}$ of sites corresponds to image pixels and the set $\mathbf{D}$ of labels is a real interval. The restoration is to estimate the true image signal from a degraded or noise-corrupted image.

Restoration of binary or multi-level images is an LP2. Similar to the continuous restoration, the aim is also to estimate the true image signal. The difference is that each pixel in the resulting image here assumes a discrete value and thus $\mathbf{D}$ in this case is a set of discrete labels.

Image segmentation is an LP2. It partitions an observation image into mutually exclusive regions, each of which has some uniform and homogeneous properties whose values are significantly different from those of neighboring regions. The property can be for example grey tone, color or texture. Pixels within each region is assigned a unique label.

The prior assumption in these problems is that the signal is smooth or piecewise smooth. This is complimentary to the assumption about edges at which abrupt changes occur.

Edge detection is also an LP2. Each pixel (more precisely, between each pair of neighboring pixels) is assigned a label in {edge, non-edge} if along an arc passing through the pixel there are abrupt changes in some properties in the direction tangent to the arc. The property can be the pixel value or directional derivatives of pixel value function. Continuous restoration with discontinuities [8, 16, 3] is a combination of LP1 and LP2.

Perceptual grouping [14] is an LP3. The sites usually correspond to initially segmented features (points, lines and regions) which are inhomogeneously arranged. The fragmentary features are to be organized into perceptually meaningful groups. Between each pair of the features can be assigned a label in {connected,disconnected}, indicating whether the two features should be joined.

Feature-based object matching and recognition is an LP3. Each site indexes an image feature such as a point, a line segment or a region. Labels are discrete in nature and each of them indexes a model feature. The resulting configuration is a mapping from the image features to those of a model object. Stereo matching is a similar LP3.

Pose estimation from a set of point correspondences might be formulated as an LP4. Each label may assume the value of a real matrix, representing an admissible (orthogonal, affine or perspective) transformation. A prior (unary) constraint is that the label of transformation itself must be orthogonal, affine or perspective. A mutual constraint is that the labels $f_1, \cdots, f_m$ should be close to each other to form a consistent transformation. When outliers are present, a line process field [8, 16, 3] may be introduced to separate transformation labels which form a consistent cluster from those due outliers.

3 Bayesian Labeling based on MRF

3.1 Bayesian Labeling

Bayesian statistics is of fundamental importance in estimation and decision making. Let $\mathbf{D}$ be a set of truth candidates and $\mathbf{r}$ the observation. Suppose that we know both the *a priori* probabilities $P(f)$ of configurations f and the likelihood densities $p(\mathbf{r} \mid f)$ of the observation $\mathbf{r}$. The best estimate one can get from these is that maximizes the *a posteriori* probability (MAP). The posterior probability can be computed by using the Bayesian rule

$$P(f \mid \mathbf{r}) = p(\mathbf{r} \mid f)P(f)/p(\mathbf{r}) \tag{5}$$

where $p(\mathbf{r})$, the density function of $\mathbf{r}$, does not affect the MAP solution. The *Bayesian labeling problem* is that given the observation $\mathbf{r}$, find the MAP configuration of labeling $f^* = \arg\max_{f \in \mathbf{S}} P(F = f \mid \mathbf{r})$.

To find the MAP solution, we need to derive the prior probabilities and the likelihood functions. The likelihood function $p(\mathbf{r} \mid F = f)$ depends on the noise statistics and the underlying transformation from the truth to the observation. It will be discussed in conjunction with specific problems. Knowing the *a priori* joint probability $P(F = f)$ is difficult, in general. Fortunately, there exists a theorem which helps us specify the *a priori* probabilities of MRFs. This is the main reason for MRF modeling.

3.2 MRF Prior and Gibbs Distribution

MRF is a branch of probability theory which provides a tool for analyzing spatial or contextual dependencies of physical phenomena. Define a neighborhood system for $\mathbf{d}$

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathbf{d}\} \tag{6}$$

where $\mathcal{N}_i$ is the collection of sites neighboring to i for which (1) $i \notin \mathcal{N}_i$ and (2) $i \in \mathcal{N}_j \Longleftrightarrow j \in \mathcal{N}_i$. The pair $(\mathbf{d}, \mathcal{N})$ is a graph in the usual sense. A *clique* c for $(\mathbf{d}, \mathcal{N})$ is a subset of $\mathbf{d}$ such that c consists of a single site $c = \{i\}$, or a pair of neighboring sites $c = \{i, j\}$, or a triple of neighboring sites $c = \{i, j, k\}$, and so on. We denote the collection of single-site cliques, that of two-site cliques, $\cdots$, by $\mathcal{C}_1, \mathcal{C}_2, \cdots$, respectively. The collection of all cliques for $(\mathbf{d}, \mathcal{N})$ is $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots$.

A family F of random variables is said to be an MRF on $\mathbf{d}$ with respect to $\mathcal{N}$ if and only if the following two conditions are satisfied: (1) $P(F = f) > 0, \forall f \in \mathbf{S}$ (positivity), and (2) $P(F_i = f_i \mid F_j = f_j, j \in \mathbf{d}, j \neq i) = P(F_i = f_i \mid F_j = f_j, j \in \mathcal{N}_i)$ (Markovianity). Condition (1) above is for F to be a random field. Condition (2) is called the local characteristics. It says that the probability of a local event at i conditioned on all the remaining events is equivalent to that conditioned on the events at the neighbors of i. It can be shown that the joint probability $P(F = f)$ of any random field is uniquely determined by these local conditional probabilities [1]. However, it is usually difficult to specify the set of the conditional probabilities. Nonetheless, the Hammersley-Clifford theorem [1] of Markov-Gibbs equivalence provides a solution.

According to the Hammersley-Clifford theorem [1], F is an MRF on $\mathbf{d}$ with respect to $\mathcal{N}$ if and only if the probability distribution $P(F = f)$ of the configurations is a Gibbs distribution with respect to $\mathcal{N}$. A Gibbs distribution of the configurations f with respect to $\mathcal{N}$ is of the following form

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \tag{7}$$

In the above, Z is a normalizing constant, T is a global control parameter called the temperature and $U(f)$ is the *prior energy*. The prior energy has the form

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) = \sum_{\{i\} \in \mathcal{C}_1} V_1(f_i) + \sum_{\{i,j\} \in \mathcal{C}_2} V_2(f_i, f_j) + \cdots \tag{8}$$

where "$\cdots$" denotes possible higher order terms. The practical value of the theorem is that it provides a simple way of specifying the joint prior probability $P(F = f)$ of the configurations by specifying the prior potentials $V_c(f)$ for all $c \in \mathcal{C}$. One is allowed to choose appropriate potentials for desired system behavior. The potential functions contain the *a priori* knowledge of interactions between labels assigned to neighboring sites and reflect how individual matches affect one another — *a priori*.

3.3 Posterior MRF Energy

Let the likelihood function be expressed in the exponential form

$$p(\mathbf{r} \mid F = f) = Z_r^{-1} \times e^{-U(\mathbf{r} \mid f)} \tag{9}$$

where $U(\mathbf{r} \mid f)$ is called the *likelihood energy*. Then the posterior probability is a Gibbs distribution

$$P(F = f \mid \mathbf{r}) = Z_E^{-1} \times e^{-E(f)} \tag{10}$$

with *posterior energy*

$$E(f) = U(f \mid \mathbf{r}) = U(f)/T + U(\mathbf{r} \mid f) \tag{11}$$

Hence, given a fixed $\mathbf{r}$, F is also an MRF on $\mathbf{d}$ with respect to $\mathcal{N}$. The MAP solution is equivalently found by

$$f^* = \arg\min_{f \in \mathbf{S}} U(f \mid \mathbf{r}) \tag{12}$$

To summarize, the MRF modeling process consists of the following steps: Defining a neighborhood system $\mathcal{N}$, defining cliques $\mathcal{C}$, defining the prior clique potentials, deriving the likelihood energy, and deriving the posterior energy.

4 Two Cases of MRF Vision Modeling

In this section, the prototypical low level MRF model of Geman and Geman [8] for image restoration is described first and is taken as the reference point. It is prototypical because it can model problems falling in categories LP1 and LP2. It forms the basis for other low level problems such as edge detection, motion, stereo and texture [16, 7, 15, 4]. The high level MRF model for object matching [11] is described next as a prototype for LP3.

4.1 Image Restoration at Low Level

Low level processing is performed on images. The set of sites $\mathbf{d} = \{1, \ldots, m\}$ index image pixels in a 2D plane and the observation $\mathbf{r}$ represents the array of pixel values. The set $\mathbf{D}$ contains discrete label to be assigned to the pixels. The configuration $f = \{f_i \in \mathbf{D} \mid i \in \mathbf{d}\}$, or the state of labeling, is a realization of a Markov random intensity field.

Let the neighbors of pixel i consist of the four nearest pixels

$$\mathcal{N}_i = \{j \mid dist(\text{pixel}_i, \text{pixel}_j) \leq 1\} \tag{13}$$

where $dist(A, B)$ is the distance between A and B. For simplicity, here consider only two-site cliques

$$\mathcal{C} = \mathcal{C}_2 = \{\{i, j\} \mid j \in \mathcal{N}_i, \forall i \in \mathbf{d}\} \tag{14}$$

Examples of more complex cliques can be found in Fig.5 of [8].

Now define the prior clique potentials in Eq.(8). When only two-site cliques are considered, only second order prior potentials are nonzero. The second order potential is defined by

$$V_2(f_i, f_j) = v_{20}\; g(f_i - f_j) \tag{15}$$

where v_{20} is a real scalar and $g(\eta)$ is a function measuring the cost due to the smoothness violation caused by $f_i - f_j$. For continuous restoration with discontinuities [8, 3], $g(\eta) = \min(\eta^2, \alpha)$. For piecewise constant reconstruction with discontinuities [8, 9], $g(\eta) = [1 - \delta(f_i - f_j)]$ where $\delta(\eta)$ is the Dirichlet function. A general definition of g for discontinuity-adaptive restoration is given in [13].

Geman and Geman [8] describe a general degraded image model based on which the likelihood function is obtained. In an important special case, each observed pixel value is assumed to be $r_i = f_i + n$ where $n \sim N(0, \sigma)$ is independent Gaussian noise. In this case, the likelihood energy is

$$U(\mathbf{r} \mid f) = \sum_{i \in \mathbf{d}} (r_i - f_i)^2 / \sigma \tag{16}$$

The posterior energy $E(f) = U(f \mid \mathbf{r})$ can be computed from $U(f)$ and $U(\mathbf{r} \mid f)$ using (11)

$$U(f \mid \mathbf{r}) = \sum_{i \in \mathbf{d}} \sum_{j \in \mathcal{N}_i} v_{20}\, g(f_i - f_j)/T + \sum_{i \in \mathbf{d}} (r_i - f_i)^2 / \sigma \tag{17}$$

The above with $g(\eta) = \min(\eta^2, \alpha)$ is the notion of the weak string model [3] and that with $g(\eta) = [1 - \delta(f_i - f_j)]$ is the minimal length coding model [9].

4.2 Object Matching at High Level

High level processing is performed on token features extracted from images. A typical problem is (partial) matching from image features to those of a modeled object. Unlike the previous case, the observation $\mathbf{r}$ in this case include not only components describing each feature itself but also those describing contextual relations between them. Moreover, the neighborhood relationship between features is not isotropic as is in the image case.

Both an object and a scene are represented by a set of features, (unary) properties of the features and (bilateral or higher order) contextual relations between them. The features, properties and relations can be denoted compactly as a *relational structure* (RS). An RS describes a scene or a (part of) model object.

The scene RS is denoted by $\mathbf{g} = (\mathbf{d}, \mathbf{r})$ where $\mathbf{d} = \{1, \ldots, m\}$ indexes a set of m features and $\mathbf{r} = \{r_1, r_2, \ldots, r_H\}$ denotes the set of observation data of order 1 through order H (When $H = 2$, the RS is reduced to a *relational graph* (RG)). For order $n = 2$, $r_2(i,j) = [r_{2,1}(i,j), \ldots, r_{2,K_2}(i,j)]^T$ consists of K_2 *binary (bilateral) relations* between features i and j.

A model RS is similarly denoted as $\mathbf{G} = (\mathbf{D}, \mathbf{R})$ where $\mathbf{D} = \{1, \ldots, M\}$ and $\mathbf{R} = \{R_1, R_2, \ldots, R_H\}$. For particular n and k ($1 \leq k \leq K_n$; $1 \leq n \leq h$), $R_{n,k}$ represent the same type of constraint as $r_{n,k}$. Introduce a virtual NULL model $\mathbf{D}_0 = \{0\}$ to represents everything not modeled by $\mathbf{G}$. Then in matching the scene to the model object plus the NULL , the set of all labels is

$$\mathbf{D}^+ = \mathbf{D}_0 \cup \mathbf{D} = \{0, 1, \ldots, M\} \tag{18}$$

$\mathbf{S} = (\mathbf{D}^+)^m$ is the admissible space of label configurations.

In RS matching, the set $\mathcal{N}_i$ of neighbors of $i \in \mathbf{d}$ can comprises all related sites. But when the scene is very large, $\mathcal{N}_i$ needs to include only those which are within a spatial distance α from i.

$$\mathcal{N}_i = \{j \neq i \mid dist(\text{feature}_j, \text{feature}_i) < \alpha, j \in \mathbf{d}\} \tag{19}$$

The size α may reasonably be related to the size of the considered model object.

Now define the prior clique potentials in Eq.(8). The single-site potential is defined as

$$V_1(f_i) = \begin{cases} v_{10} & \text{if } f_i = 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where v_{10} is a constant. This definition says that if f_i is the NULL label, it incurs a penalty v_{10}; or otherwise no penalty. The two-sites potential is defined as

$$V_2(f_i, f_j) = \begin{cases} v_{20} & \text{if } f_i = 0 \text{ or } \; f_j = 0 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where v_{20} is a constant. This says that if either f_i or f_j is the NULL , it incurs a penalty v_{20}; or otherwise no penalty.

The joint likelihood function $p(\mathbf{r} \mid F = f)$ has the following properties: (1) It is conditioned on pure non-NULL matches $f_i \neq 0$; (2) It is regardless of the neighborhood system $\mathcal{N}$; and (3) It depends on how the model object is observed in the scene, which depends on the underlying transformations and noise. Assume that $\mathbf{R}$ and $\mathbf{r}$ consist of types of relations which are invariant under the group of underlying transformations and that the observation model is $\mathbf{r} = \mathbf{R} + \mathbf{n}$ where $\mathbf{n}$ is independent Gaussian noise. Then the likelihood energy is

$$U(\mathbf{r} \mid F = f) = \sum_{i \in \mathbf{d}, f_i \neq 0} V_1(\mathbf{r} \mid f_i) + \sum_{i \in \mathbf{d}, f_i \neq 0} \; \sum_{j \in \mathbf{d}, f_j \neq 0} V_2(\mathbf{r} \mid f_i, f_j) \tag{22}$$

Because the noise is independent, we have $U(\mathbf{r} \mid f_i) = U(r_1(i) \mid f_i)$ and $U(\mathbf{r} \mid f_i, f_j) = U(r_2(i,j) \mid f_i, f_j)$. The likelihood potentials are

$$V_1(r_1(i) \mid f_i) = \sum_{k=1}^{K_1} [r_{1,k}(i) - R_{1,k}(f_i)]^2 / 2\sigma_{1,k}^2 \tag{23}$$

and

$$V_2(r_2(i,j) \mid f_i, f_j) = \sum_{k=1}^{K_2} [r_{2,k}(i,j) - R_{2,k}(f_i, f_j)]^2 / 2\sigma_{2,k}^2 \tag{24}$$

where $\sigma_{n,k}^2$ ($k = 1, \ldots, K_n$ and $n = 1, 2$) are the standard deviations of the noise components. The vectors $R_1(f_i)$ and $R_2(f_i, f_j)$ is the "mean vector" for the random vectors $r_1(i)$ and $r_2(i,j)$, respectively. When the noise is correlated, there are correlating terms in the likelihood potentials. The assumption of independent Gaussian may not be accurate but offers a good approximation when the accurate likelihood is not available.

The posterior energy $E(f)$ can be computed from $U(f)$ and $U(r \mid f)$ using (11)

$$\begin{aligned} U(f \mid \mathbf{d}) = {} & \textstyle\sum_{i \in \mathbf{d}} V_{10}(f_i)/T + \sum_{i \in \mathbf{d}} \sum_{j \in \mathcal{N}_i} V_{20}(f_i, f_j)/T + \\ & \textstyle\sum_{i \in \mathbf{d}} V_1(r_1(i) \mid f_i) + \sum_{i \in \mathbf{d}} \; \sum_{j \in \mathbf{d}} V_2(r_2(i,j) \mid f_i, f_j) \end{aligned} \tag{25}$$

The MAP configuration f^* of (12) is the optimal labeling of the scene in terms of the model object. Matching to multiple model objects can be resolved after matching to each of the objects [11].

5 Conclusion

A variety of low and high level vision problems can formulated as Bayesian labeling using a unified MRF modeling approach. A labeling of an image, of an edge map or of a scene is considered as a configuration of an MRF. The solution to a problem is defined as the MAP label configuration which minimizes the posterior energy. The MRF modeling provides a systematic approach for vision modeling based on the rationale principles.

Related to the MRF modeling is estimation of involved parameters. In LP1 and LP2 at low level, the estimation can be done, for example, using the coding method [1] and least square error method [6]. A learning-from-example method for MRF parameter estimation in object recognition (LP3) is proposed in [12].

References

1. J. Besag. "Spatial interaction and the statistic analysis of lattice systems". *J. Royal. Statist. Soc. B*, 36:192–293, 1974.
2. J. Besag. "Towards Bayesian image analysis". *Journal of Applied Statistics*, 16(3):395–406, 1989.
3. A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.
4. R. Chellappa and A. Jain, editors. *Markov Random Fields: Theory and Applications*. Academic Press, 1993.
5. P. R. Cooper. "Parallel structure recognition with uncertainty: coupled segmentation and matching". In *Proceedings of IEEE International Conference on Computer Vision*, pages 287–290, 1990.
6. H. Derin and H. Elliott. "Modeling and segmentation of noisy and textured images using Gibbs random fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1):39–55, January 1987.
7. R. C. Dubes and A. K. Jain. "Random field models in image analysis". *Journal of Applied Statistics*, 16(2):131–164, 1989.
8. G. Geman and D. Geman. "Stochastic relaxation, gibbs distribution and bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.
9. Y. G. Leclerc. "Constructing simple stable descriptions for image partitioning". *International Journal of Computer Vision*, 3:73–102, 1989.
10. S. Z. Li. " Towards 3D vision from range images: An optimization framework and parallel networks". *CVGIP: Image Understanding*, 55(3):231–260, May 1992.
11. S. Z. Li. "A Markov random field model of object matching". *submitted*, 1993.
12. S. Z. Li. "Optimal selection of MRF parameters in object recognition". *in preparation*, 1993.
13. S. Z. Li. "On discontinuity adaptive regularization". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, accepted.
14. D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer, 1985.
15. K. V. Mardia. *Technical Editor*. Special Issue on Statistic Image Analysis. *Journal of Applied Statistics*, 16(2), 1989.
16. J. Marroquin, S. Mitter, and T. Poggio. "Probabilistic solution of ill-posed problems in computational vision". *Journal of the American Statistical Association*, 82(397):76–89, March 1987.
17. J. W. Modestino and J. Zhang. "A Markov random field model-based approach to image interpretation". In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 458–465, 1989. Also *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.6, pp.606-615, June 1992.
18. R. Szeliski. "Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, pages 271–301, 1990.