Chapter 4

High-Level MRF Models

High-level vision tasks, such as object matching and recognition and pose estimation, are performed on features extracted from images. The arrangements of such features are usually irregular, and hence the problems fall into categories LP3 and LP4. In this chapter, we present MAP-MRF formulations for solving these problems.

We begin with a study on the problem of object matching and recognition under contextual constraints. An MAP-MRF model is then formulated following the systematic approach summarized in Section 1.3.4. The labeling of a scene in terms of a model\(^1\) object is considered as an MRF. The optimal labeling of the MRF is obtained by using the MAP principle. The matching of different types of features and multiple objects is discussed. A related issue, MRF parameter estimation for object matching and recognition, will be studied in Chapter 7.

We then derive two MRF models for pose computation, pose meaning the geometric transformation from one coordinate system to another. In visual matching, the transformation is from the scene (image) to the model object considered (or vice versa). In derived models, the transformation is from a set of object features to a set of image features. They minimize posterior energies derived for the MAP pose estimation, possibly together with an MRF for matching.

4.1 Matching under Relational Constraints

In high-level image analysis, we are dealing with image features, such as critical points, lines, and surface patches, that are more abstract than image pixels. Such features in a scene are not only attributed by (unary) properties about the features themselves but also related to each other by relations

\(^1\)In this chapter, the word “model” is used to refer to both mathematical vision models and object models.
between them. In other words, an object or a scene is represented by features constrained by the properties and relations. It is the bilateral or higher-order relations that convey the contextual constraints. They play a crucial role in visual pattern matching.

4.1.1 Relational Structure Representation

The features, properties and relations can be denoted compactly as a relational structure (RS) (Fischler and Elschlager 1973; Ambler et al. 1973; Cheng and Huang 1984; Radig 1984; Li 1992c; Li 1992a). An RS describes a scene or (part of) a model object. The problem of object recognition is reduced to that of RS matching.

Let us start with a scene RS. Assume there are \( m \) features in the scene. These features are indexed by a set \( S = \{ 1, \ldots, m \} \) of sites. The sites constitute the nodes of the RS. Each node \( i \in S \) has associated with it a vector \( d_1(i) \) composed of a number of \( K_1 \) unary properties or unary relations, \( d_1(i) = [d_1^{(1)}(i), \ldots, d_1^{(K_1)}(i)]^T \). A unary property could be, for example, the color of a region, the size of an area, or the length of a line. Each pair of nodes \( (i, i' \in S, i' \neq i) \) are related to each other by a vector \( d_2(i, i') \) composed of a number of \( K_2 \) binary (bilateral) relations, \( d_2(i, i') = [d_2^{(1)}(i, i'), \ldots, d_2^{(K_2)}(i, i')]^T \). A binary relation could be, for example, the distance between two points or the angle between two lines. More generally, among \( n \) features \( i_1, \ldots, i_n \in S \), there may be a vector \( d_n(i_1, \ldots, i_n) \) of \( K_n \) \( n \)-ary relations. This is illustrated in Fig. 4.1. An \( n \)-ary relation is also called a relation, or constraint, of order \( n \). The scope of relational dependencies can be determined by a neighborhood system \( N \) on \( S \). Now, the RS for the scene is defined by a triple

\[
G = (S, N, d) \tag{4.1}
\]

where \( d = \{ d_1, d_2, \ldots, d_H \} \) and \( H \) is the highest-order. For \( H = 2 \), the RS is also called a relational graph (RG). The highest-order \( H \) cannot be lower than 2 when contextual constraints must be considered.

The RS for a model object is similarly defined as

\[
G' = (L, N', D) \tag{4.2}
\]

where \( D = \{ D_1, D_2, \ldots, D_H \} \), \( D_1(I) = [D_1^{(1)}(I), \ldots, D_1^{(K_1)}(I)]^T \), \( D_2(I, I') = [D_2^{(1)}(I, I'), \ldots, D_2^{(K_2)}(I, I')]^T \), and so on. In this case, the set of labels \( L \) replaces the set of sites. Each element in \( L \) indexes one of the \( M \) model features. In addition, the “neighborhood system” for \( L \) is defined to consist of all the other elements, that is,

\[
N'_I = \{ I' \mid \forall I' \in L, I' \neq I \} \tag{4.3}
\]

This means each model feature is related to all the other model features. The highest-order considered, \( H \), in \( G' \) is equal to that in \( G \). For particular \( n \) and