A NEW APPROACH TO THE CONSISTENT LABELING PROBLEM †

Haim Shvaytser and Shmuel Peleg

Dept. of Computer Science,
The Hebrew University of Jerusalem,
91904 Jerusalem, Israel.

Abstract

The problem of defining a goodness measure for object labeling problems is addressed. The optimal probabilistic measure is derived, but shown to be impractical for realistic problems. An alternative consistency measure is suggested, based on the accuracy with which each object can estimate its label. This measure is shown to be a generalization of the crisp idea of consistency and can be computed using statistical information about the problem. Results are shown on a triangle labeling example and on grey level picture segmentation and border detection.

Key words : consistency, consistent labeling problem, relaxation, segmentation, edge detection.

1. Introduction.

Consistency in labeling problems is usually discussed in connection with relaxation algorithms. It has been observed by many authors [1-6] that general information about a problem can help eliminate certain ambiguities in a suggested solution. In work on relaxation algorithms, consistency is usually defined as a fixed point of an operator [2] or a minimum point of a functional [5]. However, these operators or functionals are devised in such a way that the fixed points or minima points are "consistent", so that their definition of consistency is cyclic. In a more recent paper concerning relaxation [6], consistency is defined in terms of inequalities to be satisfied by the consistent labeling. These inequalities are assumed to be a part of the problem, but for many realistic problems there exists no simple way to determine them.

We present in this paper a definition of the "goodness" of solutions to the labeling problem. This involves a new definition of consistency that has a firm theoretical background and also agrees with intuition. The coefficients involved in this definition can be determined from a statistical model of the problem. Given a definition of the goodness of solutions, one can define a solution to the consistent labeling problem as one that optimizes the goodness measure. We show that a version of Peleg's discrete optimization technique [7] can be applied here and we prove some local convergence properties for this algorithm.

2. A formulation of the problem in terms of events.

We consider a world composed of objects $A = \{a_1, \cdots, a_n\}$. Each object can have one label out of $\Lambda = \{\lambda_1, \cdots, \lambda_m\}$. Our task is to assign to each $a_i \in A$ one label $\lambda_j \in \Lambda$. A solution to the labeling problem is an n-tuple $(k_1, \cdots, k_n)$ meaning that object a_i has label λ_{k_i}. There are m^n possible solutions.

We consider the labeling process as an experiment and its outcomes as events. We denote by $v_i^{k_i}$ the event that object a_i has label λ_{k_i}. The solution $(k_1, \cdots, k_n)$ means that the events $v_1^{k_1}, \cdots, v_n^{k_n}$ happened. We will sometimes refer to the set of events $(v_1^{k_1}, \cdots, v_n^{k_n})$ as a solution.

† This work has been supported by a grant from the Israel National Council for Research and Development.

A specific solution is to be chosen using two sources of information. The first is general information about possible solutions, and we denote it by Q. In its most general form, Q is the list of all solutions and their probabilities. We will consider cases where we have only partial information about these probabilities.

The second source of information consists of results of *local* measurements done on each object a_i in order to determine whether $v_i^{k_i}$ happened. We assume that the result is a set of real variables

$$\{ x_i^j \} \quad 0 \le x_i^j \le 1 \quad 1 \le i \le n \; . \; 1 \le j \le m$$

where the value of x_i^j is our "degree of certainty" that the event v_i^j happened, *according to the local measurements.* (These values are sometimes called *initial estimates*). An exact interpretation of the values of x_i^j will be given later. To make the notation simpler we introduce the notation

$$X_i = \{x_i^j\} \qquad 1 \le j \le m$$

$$\mathbf{X} = \{X_i\} \qquad 1 \le i \le n$$

Given $\mathbf{X}$ and Q we seek the best solution according to this information.

3. The optimal probability measure

In this section we derive the optimal goodness measure for the labeling problem. We denote it by *PDM* – Probabilistic Discrete Measure.

Consider a specific solution $(k_1, \cdots, k_n)$ to the labeling problem. We maintain that this is the optimal solution if the probability that $v_1^{k_1}, \cdots, v_n^{k_n}$ happened, given the information $\{\mathbf{X}, Q\}$, is maximal. The solution is optimal in the sense that it minimizes a very simple risk function, where there is no loss for being correct, and all errors are equally costly. For the proof and some other possible risk functions see [8].

Before computing $\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid \mathbf{X}, Q)$ we elaborate on the probabilities of several events: we assume that X_i, our initial degree of certainty about the labels of object a_i, is a result of measurements done *only on object* a_i, so that X_i depends only on v_i^j, $1 \le j \le m$. (See also Reference [8].) From this we have:

$$\mathrm{Prob}(\mathbf{X} \mid v_1^{k_1}, \cdots, v_n^{k_n}, Q) = \mathrm{Prob}(\mathbf{X} \mid v_1^{k_1}, \cdots, v_n^{k_n}) \quad (1)$$

$$\mathrm{Prob}(\mathbf{X} \mid v_1^{k_1}, \cdots, v_n^{k_n}) = \prod_{i=1}^{n} \mathrm{Prob}(X_i \mid v_i^{k_i}) \quad (2)$$

Consider the probability of the event $v_i^{k_i}$. This probability can be computed from our information about the world so that $\mathrm{Prob}(v_i^{k_i} \mid Q)$ and $\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q)$ are well defined. This is not so, however, when considering $\mathrm{Prob}(v_i^{k_i})$. By definition

$$\mathrm{Prob}(v_i^{k_i}) = \int_{\text{possible } Q} \mathrm{Prob}(v_i^{k_i} \mid Q) \; d\, \mathrm{Prob}(Q)$$

We assume that when averaging over all possible Q, all events have the same probability. Hence

$$\mathrm{Prob}(v_i^{k_i}) = c \tag{3}$$

Using these assumptions and the Bayesian rule we have:

$$\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid \mathbf{X}, Q) =$$

$$\frac{\mathrm{Prob}(\mathbf{X} \mid v_1^{k_1}, \cdots, v_n^{k_n}, Q) \cdot \mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q)}{\mathrm{Prob}(\mathbf{X} \mid Q)}$$

$$= \frac{\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q) \cdot \prod_{i=1}^{n} \mathrm{Prob}(X_i \mid v_i^{k_i})}{\mathrm{Prob}(\mathbf{X} \mid Q)}$$

Since $\mathrm{Prob}(\mathbf{X} \mid Q)$ is independent of a specific solution it is enough to consider

$$PDM^*(v_1^{k_1}, \cdots, v_n^{k_n}) =$$

$$\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q) \cdot \prod_{i=1}^{n} \mathrm{Prob}(X_i \mid v_i^{k_i}) \tag{4}$$

We must give now a precise interpretation of the variables X_i in terms of probabilities. We will show that the results are the same whether we interpret x_i^j as the probability that v_i^j has happened, or as a "relative" probability.

(i) If x_i^j is the probability that v_i^j happened,

$$x_i^{k_i} \equiv \mathrm{Prob}(v_i^{k_i} \mid X_i)$$

Then by the Bayesian rule

$$\mathrm{Prob}(X_i \mid v_i^{k_i}) = \frac{x_i^{k_i} \cdot \mathrm{Prob}(X_i)}{\mathrm{Prob}(v_i^{k_i})}$$

and by using (3)

$$\mathrm{Prob}(X_i \mid v_i^{k_i}) = \frac{1}{c} \cdot x_i^{k_i} \cdot \mathrm{Prob}(X_i) \tag{5}$$

Substituting (5) in (4) we have:

$$PDM^*(k_1, \cdots, k_n) =$$

$$\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q) \cdot \prod_{i=1}^{n} x_i^{k_i} \cdot \prod_{i=1}^{n} \frac{\mathrm{Prob}(X_i)}{c}$$

Neglecting expressions that are independent of k_i we obtain the following measure for the "goodness" of a given solution:

$$PDM(k_1, \cdots, k_n) = \mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q) \cdot \prod_{i=1}^{n} x_i^{k_i} \tag{6}$$

(ii) Let us assume that $x_i^j \neq 0$. Then, $\forall\, i, j$ we can have the following interpretation for the values of x_i^j:

$$\frac{x_i^{j1}}{x_i^{j2}} = \frac{\mathrm{Prob}(v_i^{j1} \mid X_i)}{\mathrm{Prob}(v_i^{j2} \mid X_i)} \qquad \forall\, j1, j2$$

In this case we can choose without loss of generality $j2 = 1$ and then

$$x_i^j = \frac{x_i^1}{\mathrm{Prob}(v_i^1 \mid X_i)} \cdot \mathrm{Prob}(v_i^j \mid X_i)$$

Letting

$$\frac{x_i^1}{\mathrm{Prob}(v_i^1 \mid X_i)} = c_i$$

we get

$$\mathrm{Prob}(v_i^{k_i} \mid X_i) = \frac{x_i^{k_i}}{c_i}$$

and then

$$\mathrm{Prob}(X_i \mid v_i^{k_i}) = \frac{x_i^{k_i} \cdot \mathrm{Prob}(X_i)}{c_i \cdot \mathrm{Prob}(v_i^{k_i})} = \frac{x_i^{k_i} \cdot \mathrm{Prob}(X_i)}{c_i \cdot c}$$

Substituting in (4) and neglecting expressions independent of $(k_1, \cdots, k_n)$ we obtain again the same measure *PDM* as in (6).

3.1. Difficulties in applying the Bayesian measure to real problems.

Given the *PDM* measure the following scheme can be applied to find an optimal solution to a given labeling problem: go over all possible solutions and choose as the optimal solution the solution whose *PDM* is maximal. There are two problems with this direct approach, which we refer to as the computational complexity problem and the partial information problem; both problems arise because of the exponential number of possible solutions.

The partial information problem is the problem of determining $\mathrm{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q)$ for all solutions. It is unlikely that such information is available for reasonable values of m and n. The information one usually has is partial information, the probabilities of pairs or triples of events. Given this partial information one can try to estimate the probability of a given solution by assuming a parameter-dependent distribution, or by making Markovian assumptions [9]. In many cases, however, these assumptions are too strong, and the actual probabilities cannot be computed. We will handle this problem by defining an alternative *consistency* measure, which will replace the probability measure of a given solution.

The computational complexity problem is the problem of determining the actual maximum among the m^n possible solutions. We will show that when given good initial estimates, one can use an iterative algorithm that converges to a local maximum.

4. Consistency and consistency measure.

In this section we suggest a formal definition for the intuitive concept of consistency.

Q, the general information, may contain constraints that every solution with positive probability must fulfill. In this case it is possible to define a consistent solution as a solution that meets all these constraints. Since we assume that Q contains probabilities of events, constraints will be partial information about events with probability zero. For example, we say that v_i^j is inconsistent with v_k^l if $\mathrm{Prob}(v_i^j, v_k^l) = 0$ (i.e. every solution containing both v_i^j and v_k^l has probability zero). Our aim is to generalize this concept of consistency to the case where there are not necessarily zero probability events.

4.1. A consistency measure for events.

Our consistency definition is given in terms of events. As before we denote by v_i^j the event that object a_i was given label λ_j. We refer to such events as elementary events.

Definition 1:

U is an event if one of the following is true:

1. U is the certain event.
2. $\exists\, i, j \quad U = v_i^j$ (U is an elementary event)
3. U_1 is an event and $U = not\ U_1$
4. U_1, U_2 are events and $U = U_1 \cap U_2$

Consider two events U_1 and U_2. We want to define the consistency of the event U_1 with the event U_2. If according to Q U_1 can not happen when U_2 happens, we say that U_1 is inconsistent with U_2. In other words, U_1 is inconsistent with U_2 if $\mathrm{Prob}(U_1 \mid U_2, Q) = 0$. Motivated by this we suggest the following definition for the consistency measure of two events:

Definition 2:

The consistency measure of the event U_1 with the event U_2 according to the general information Q is:

$$\text{Cons}(U_1 \mid U_2,Q) = \begin{cases} \text{Prob}(U_1 \mid U_2,Q) & \text{if Prob}(U_2 \mid Q) \neq 0 \\ 0 & \text{otherwize} \end{cases}$$

Notice that although the inconsistency relation is symmetric ($\text{Prob}(U_1 \mid U_2,Q) = 0$ implies that $\text{Prob}(U_2 \mid U_1,Q) = 0$) the consistency measure is not symmetric in general. We may consider the above measure as an approximation to $Prob\,(U_1\,,\,U_2 \mid Q)$. As was mentioned before, a very crude approximation to $\text{Prob}(U_1\,,\,U_2 \mid Q)$ will give 1 (consistent) if $Prob\,(U_1\,,\,U_2 \mid Q) > 0$, and 0 (inconsistent) otherwise. Since

$$Prob\,(U_1\,,\,U_2 \mid Q) = \begin{cases} Prob\,(U_1 \mid U_2,Q)\cdot Prob\,(U_2 \mid Q) & \text{if } Prob\,(U_2 \mid Q) \neq 0 \\ 0 & otherwize \end{cases}$$

4.2. Consistency of solutions to the labeling problem.

We proceed to define consistency of events in a given solution to the labeling problem.

Definition 3:

Given a solution $(v_1^{k_1}, \cdots, v_n^{k_n})$ to the labeling problem we define the local consistency of the elementary event $v_i^{k_i}$ as the consistency of this event with the other events of this solution:

$$\text{Lcons}(v_i^{k_i}) = \text{Cons}(v_i^{k_i} \mid \bigcap_{j \neq i} v_j^{k_j}, Q) = \text{Prob}(v_i^{k_i} \mid \bigcap_{j \neq i} v_j^{k_j}, Q)$$

Using the definition of local consistency we can define the global consistency of a given solution. Intuitively, a solution is consistent if all its elementary events are consistent. To define this property mathematically we use the terminology of fuzzy set theory [10]. A local consistency measure according to our definition is a real number in the range [0,1]. We can regard this number as a grade of membership in the set of consistent events. With this observation, the fact that all elementary events are consistent can be expressed using any of the intersection definitions of fuzzy set theory. We use here two definitions, the min operation and the algebraic product, that lead to two definitions of global consistency.

Definition 4:

$$\text{Cons}(v_1^{k_1}, \cdots, v_n^{k_n}) = \min_i \text{Lcons}(v_i^{k_i})$$

Definition 5:

$$\text{Cons}(v_1^{k_1}, \cdots, v_n^{k_n}) = \prod_i \text{Lcons}(v_i^{k_i})$$

4.3. Example

This example is meant to show the difference between probability and consistency. Assuming two objects and three labels α, β, γ, there are $3^2 = 9$ possible solutions. We assume that among them only five have positive probability and they all have the same probability. Q, the list of all these solutions, is:

$\alpha\alpha$
$\alpha\beta$
$\beta\alpha$
$\beta\beta$
$\gamma\gamma$

The solution $\alpha\alpha$ is the two events $(v_1^{\alpha}, v_2^{\alpha})$.

$$\text{Lcons}(v_1^{\alpha}) = \text{Cons}(v_1^{\alpha} \mid v_2^{\alpha}, Q) = \text{Prob}(v_1^{\alpha} \mid v_2^{\alpha}, Q) = 0.5$$

$$\text{Lcons}(v_2^{\alpha}) = \text{Cons}(v_2^{\alpha} \mid v_1^{\alpha}, Q) = \text{Prob}(v_2^{\alpha} \mid v_1^{\alpha}, Q) = 0.5$$

so

$$\text{Cons}(v_1^{\alpha}, v_2^{\alpha}) = \begin{cases} 0.5 & \text{first definition} \\ 0.25 & \text{second definition} \end{cases}$$

The solution $\gamma\gamma$ is the two events $(v_1^{\gamma}, v_2^{\gamma})$.

$$\text{Lcons}(v_1^{\gamma}) = \text{Cons}(v_1^{\gamma} \mid v_2^{\gamma}, Q) = \text{Prob}(v_1^{\gamma} \mid v_2^{\gamma}, Q) = 1$$

$$\text{Lcons}(v_2^{\gamma}) = \text{Cons}(v_2^{\gamma} \mid v_1^{\gamma}, Q) = \text{Prob}(v_2^{\gamma} \mid v_1^{\gamma}, Q) = 1$$

so

$$\text{Cons}(v_1^{\gamma}, v_2^{\gamma}) = 1$$

Notice that both solutions have probability 0.2.

4.4. The consistency goodness measure.

Given a measure for the consistency of a given solution we can have a goodness measure where consistency is used instead of probability. Replacing probability with consistency in (6) we get the following *CDM* (Consistency Discrete Measure) measure for the goodness of solutions:

$$CDM\,(k_1, \cdots, k_n) = \text{Cons}(v_1^{k_1}, \cdots, v_n^{k_n})\cdot \prod_{i=1}^{n} x_i^{k_i} \quad (7)$$

Using Definitions 4 and 5 we get the following measures:

$$CDM\,1(k_1, \cdots, k_n) = \min_i \text{Lcons}(v_i^{k_i})\cdot \prod_{i=1}^{n} x_i^{k_i} \quad (8)$$

$$CDM\,2(k_1, \cdots, k_n) = \prod_i \text{Lcons}(v_i^{k_i})\cdot \prod_{i=1}^{n} x_i^{k_i} \quad (9)$$

Notice that although it is not explicitly indicated, $\text{Lcons}(v_i^{k_i})$ depends on $k_1 \cdots k_n$ and Q.

5. Approximations to the consistency measure.

Our motivation in looking for a consistency measure was the partial information problem, the fact that we usually do not know $\text{Prob}(v_1^{k_1}, \cdots, v_n^{k_n} \mid Q)$ for every solution. It seems, however, that in order to determine local consistency as defined in Definition 3 we still need this information, since

$$\text{Lcons}(v_i^{k_i}) = \text{Prob}(v_i^{k_i} \mid \bigcap_{j \neq i} v_j^{k_j}, Q) = \frac{\text{Prob}(\bigcap_j v_j^{k_j} \mid Q)}{\text{Prob}(\bigcap_{j \neq i} v_j^{k_j} \mid Q)}$$

The advantage of the consistency measure over the probability measure lies in some useful approximations that can be derived and that will enable us to use only partial information for its computation. These approximations are discussed in this section.

5.1. Consistency as estimation of random variables.

In this section we develop a method of approximating consistency, based on the amount of partial information available. We do this by establishing a relation between events and certain random variables.

Definition 6:

Given an event U, its characteristic function ϕ is defined in the following way:

$$\phi(U) = \begin{cases} 1 & \text{if U happened} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $\phi(U)$ is a random variable.

Assume we have two events U_0, U_1. Let $u_0 = \phi(U_0)$, $u_1 = \phi(U_1)$, and let $\hat{u}_0 = g\,(u_1)$ be the estimate of the random variable u_0 in terms of the random variable u_1. We want our estimate to be "as good as possible". This is defined as the estimate that minimizes the error in the l_2 norm:

$$\text{Err} = E\,\{(\hat{u}_0 - u_0)^2 \mid Q\} \quad (10)$$

where E is the expectation operator, and the expectation is over all solutions.

Theorem 1:

If $\text{Prob}(U_1 \mid Q) \neq 0$) then the consistency measure of the event U_0 with the event U_1, according to the general information Q, is the estimate of u_0 by means of u_1 that minimizes (10).

Proof:

From Definition 5 we have:

$$\text{Cons}(U_0 \mid U_1, Q) = \text{Prob}(U_0 \mid U_1, Q)$$

From the theory of nonlinear estimation [11] it is known that the best estimate of u_0 that minimizes (10) is

$$\hat{u}_0 = E\{u_0 \mid u_1, Q\}$$

Thus,

$$E\{u_0 \mid u_1, Q\} = \sum_{u_0} u_0 \cdot \text{Prob}\{\phi(U_0)=u_0 \mid U_1, Q\}$$

$$= 0 \cdot \text{Prob}\{\phi(U_0)=0 \mid U_1, Q\} + 1 \cdot \text{Prob}\{\phi(U_0)=1 \mid U_1, Q\}$$

$$= \text{Prob}(U_0 \mid U_1, Q)$$

We proceed to define the consistency of an event U_0 with several events $U_1, \cdots U_r$, where we assume that not all these events happened. Without loss of generality we assume that the events $U_1, \cdots U_t$ happened, and the events $U_{t+1}, \cdots U_r$ did not happen, $(0 \leq t \leq r)$.

One straightforward way of defining this consistency is to define it as the consistency of U_0 with $\bigcap_{i=1}^{t} U_i$. This, however, will again require joint probabilities of several events. We suggest another definition for this consistency, motivated by the above theorem and results from linear estimation theory.

From linear estimation theory it is known that there is an extremely simple solution to the problem of linear estimation of a random variable by other random variables. Letting $u_i = \phi(U_i)$ and $\hat{u}_0$ the estimate for u_0, the linear estimation problem is to determine coefficients b_i such that

$$\hat{u}_0 = b_1 \cdot u_1 + \cdots + b_r \cdot u_r \tag{11}$$

and

$$\text{Err} = E\{(\hat{u}_0 - u_0)^2 \mid Q\} \tag{10}$$

is minimal.

The solution to the above problem is given by the Yule–Walker equations [11]:

$$\begin{aligned} R_{11} \cdot b_1 + R_{12} \cdot b_2 + \cdots + R_{1r} \cdot b_r &= R_{01} \\ R_{21} \cdot b_1 + R_{22} \cdot b_2 + \cdots + R_{2r} \cdot b_r &= R_{02} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots& \\ R_{r1} \cdot b_1 + R_{r2} \cdot b_2 + \cdots + R_{rr} \cdot b_r &= R_{0r} \end{aligned}$$

where $R_{ij} = E\{u_i \cdot u_j\}$

Since u_i, u_j are characteristic functions, R_{ij} can be expressed in terms of the probabilities of the corresponding events:

$$R_{ij} = \text{Prob}(U_i, U_j \mid Q)$$

and this is the information we assume we have in Q.

Because the linear estimate is an approximation to the nonlinear estimate, we can use it instead of the nonlinear estimate for approximating consistency. There is, however, a difficulty with such an approximation. Notice that when $\hat{u}_0$ is the best nonlinear estimate for u_0, we have $0 \leq \hat{u}_0 \leq 1$, since by the theorem it can be expressed as a probability. If $\hat{u}_0$ is the *linear* estimate for u_0, it can have values in $(-\infty, \infty)$. We therefore give the following definition for the estimated consistency:

Definition 7:

The estimated consistency of the event U_0 with the events $U_1, \cdots U_r$, where $U_1, \cdots U_t$ happened $(0 \leq t \leq r)$, is

$$\text{Econs}(U_0 \mid U_1 \cdots U_r, Q) = \begin{cases} 0 & \text{Prob}(\bigcap_{i=1}^{t} U_i \mid Q) = 0 \\ 1 - |\hat{u}_0 - 1| & |\hat{u}_0 - 1| \leq 1 \\ 0 & |\hat{u}_0 - 1| > 1 \end{cases}$$

where $\hat{u}_0$ is the linear estimate for $\phi(U_0)$ in terms of $\phi(U_1), \cdots, \phi(U_r)$.

Notice that when $0 \leq \hat{u}_0 \leq 1$, we have $1 - |\hat{u}_0 - 1| = \hat{u}_0$.

We can now apply the above definition to the labeling problem.

Definition 8:

The estimated consistency of the event $v_i^{k_i}$ in a solution to the labeling problem with the events $U_1, \cdots U_r$, where these events are composed of $v_j^{k_j}$ $j \neq i$ in the sense of Definition 1 is:

$$\text{Elcons}(v_i^{k_i} \mid U_1 \cdots U_r, Q) = \text{Econs}(v_i^{k_i} \mid U_1 \cdots U_r, Q)$$

For the above consistency definition to be practical, we still have to show how to check the condition $\text{Prob}(\bigcap_{i=1}^{t} U_i) = 0$ of Definition 7. We maintain that this information can also be estimated from the Yule–Walker equations. Consider the error expression of the estimate:

$$\text{Err} = E\{(\hat{u}_0 - u_0)^2 \mid Q\} = \sum (\hat{u}_0 - u_0)^2 \cdot \text{Prob}(U_1 \cdots U_r \mid Q)$$

where the sum is over all $U_1, \cdots, U_r$. If the above probability is 0, it does not contribute to the total error. Therefore, the corresponding estimate $\hat{u}_0$ that minimizes Err is not unique. Since the Yule–Walker equations give *all* the optimal estimates, $\hat{u}_0$ will be parameter dependent.

5.2. Summary

In order to estimate the local consistency of the event $v_i^{k_i}$, one chooses a set of events $U_1, \cdots, U_r$. The joint probability of all pairs of events must be known.

In an extreme case, this set will include a single event $U_1 = \bigcap_{j \neq i} v_j^{k_j}$.

Another possibility is to choose this set as $v_j^{k_j}$, and then one needs to know only joint probabilities of elementary events. In many applications of the labeling problem, it is evident that an object is mostly affected by its close neighbors. In these cases, these events can be taken as the set of events from which the local consistency is to be estimated.

When the joint probabilities of the events are given, one can use the Yule–Walker equations and determine the coefficients b_i. Then, using the actual values of the characteristic functions of the events, and the coefficients, the local consistency can be determined using (11) and Definition 8.

Notice that the computation of the coefficients b_i can be done in advance, and then computing the local consistency by (11) requires only additions (no multiplications at all).

6. Experimental Results.

In this section we describe the results obtained using the consistency goodness measures for two problems. The first problem is to label a triangle line drawing using Huffman labels, and has only $4^3 = 64$ possible solutions. Therefore, the computational complexity is very small, and a comparison can be made with the optimal probability measure. The second problem is to segment a gray level picture into background, objects and borders. If we consider pictures of 64×64 pixels, the total number of possible solutions is $3^{64 \times 64}$. Therefore, the actual maximum cannot be obtained, but we describe a simple algorithm that converges to a local maximum in a sense to be defined later.

6.1. Labeling a triangle.

Consider a triangle as shown in Figure 1. The objects to be labeled are the triangle edges a_1, a_2, a_3, and the labels are the Huffman labels [1]. This problem was first introduced by Rosenfeld, et al. in [2].

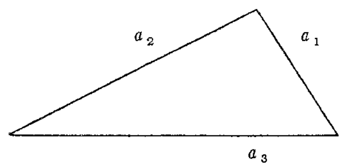

Figure 1: A triangle

The original Huffman labels are:

↑ = occluding edge, object above background.

↓ = occluding edge, object below background.

+ = convex fold.

− = concave fold.

We denote these four labels by α, β, γ, and δ, respectively

Since there are three objects and four possible labels, the total number of possible solutions is $4^3 = 64$. However, it is easily seen that most of them cannot describe a "real" object. There are only eight solutions that describe a "real" object and they are listed in Table 1. We assume that every possible labeling of Table 1 has the same probability $\frac{1}{8}$. (We assume solutions not listed in the table to have zero probability). The set of events from which local consistency is estimated is chosen here to be the other elementary events. The joint probabilities can be determined from Table 1, and the resulting Yule–Walker equations can be solved. The results for the estimates for v_1^α are listed below.

$$\hat{\phi}(v_1^\alpha) = (1-q\,1)\cdot\phi(v_2^\alpha) - q\,2\cdot\phi(v_2^\beta) + (1.5-q\,1)\cdot\phi(v_2^\gamma) - q\,2\cdot\phi(v_2^\delta)$$

$$+ (-0.5+q\,1)\cdot\phi(v_3^\alpha) + q\,2\cdot\phi(v_3^\beta) + q\,1\cdot\phi(v_3^\gamma)$$

where $q\,1, q\,2$ are arbitrary constants.
Consider several examples:

v_2^α , v_3^α happened, the estimate is 0.5 , $\mathrm{Elcons}(v_1^\alpha) = 0.5$

v_2^α , v_3^γ happened, the estimate is 1 , $\mathrm{Elcons}(v_1^\alpha) = 1$

v_2^β , v_3^β happened, the estimate is 0 , $\mathrm{Elcons}(v_1^\alpha) = 0$

v_2^α , v_3^β happened, the estimate is $1-q\,1+q\,2$, $\mathrm{Elcons}(v_1^\alpha) = 0$

Experiments with the goodness measures were done for the same initial estimates that were tried in [2]. These initial estimates are listed in Table 2 and for each case, the best solutions according to the various goodness measures are listed in Table 3. The *PDM* is the probability measure, which we consider the "true" solution. *CDM* 1 and *CDM* 2 are the consistency measures of equations (8) and (9). Notice that the results of *CDM* 1 are *optimal*, although it uses less information than *PDM*, and the results of *CDM* 2 are very close.

Table 1: Real object solutions for the triangle example.

	a_1	a_2	a_3
1	α	α	α
2	α	α	γ
3	α	γ	α
4	γ	α	α
5	β	β	β
6	β	β	δ
7	β	δ	β
8	δ	β	β

Table 2: Initial estimates.

Case	A	B	C	D	E	F	G	H
x_1^α	.25	.5	.5	.5	.3	.2	.3	.3
x_1^β	.25	.0	.0	.0	.0	.0	.2	.2
x_1^γ	.25	.5	.5	.5	.7	.8	.3	.3
x_1^δ	.25	.0	.0	.0	.0	.0	.2	.2
x_2^α	.25	.5	.4	.3	.3	.3	.3	.25
x_2^β	.25	.0	.0	.0	.0	.0	.2	.25
x_2^γ	.25	.5	.6	.7	.7	.7	.3	.25
x_2^δ	.25	.0	.0	.0	.0	.0	.2	.2
x_3^α	.25	.5	.5	.5	.5	.5	.3	.2
x_3^β	.25	.0	.0	.0	.0	.0	.2	.2
x_3^γ	.25	.5	.5	.5	.5	.5	.3	.4
x_3^δ	.25	.0	.0	.0	.0	.0	.2	.2

Table 3: Best solutions for the initial cases of Table 2.

X	best PDM	best CDM1	best CDM2
A	$\alpha\alpha\alpha$ $\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\beta\beta\beta$ $\beta\beta\delta$ $\beta\delta\beta$ $\gamma\alpha\alpha$ $\delta\beta\beta$	$\alpha\alpha\alpha$ $\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\beta\beta\beta$ $\beta\beta\delta$ $\beta\delta\beta$ $\gamma\alpha\alpha$ $\delta\beta\beta$	$\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\beta\beta\delta$ $\beta\delta\beta$ $\gamma\alpha\alpha$ $\delta\beta\beta$
B	$\alpha\alpha\alpha$ $\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\gamma\alpha\alpha$	$\alpha\alpha\alpha$ $\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\gamma\alpha\alpha$	$\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\gamma\alpha\alpha$
C	$\alpha\gamma\alpha$	$\alpha\gamma\alpha$	$\alpha\gamma\alpha$
D	$\alpha\gamma\alpha$	$\alpha\gamma\alpha$	$\alpha\gamma\alpha$
E	$\alpha\gamma\alpha$ $\gamma\alpha\alpha$	$\alpha\gamma\alpha$ $\gamma\alpha\alpha$	$\alpha\gamma\alpha$ $\gamma\alpha\alpha$
F	$\gamma\alpha\alpha$	$\gamma\alpha\alpha$	$\gamma\alpha\alpha$
G	$\alpha\alpha\alpha$ $\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\gamma\alpha\alpha$	$\alpha\alpha\alpha$ $\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\gamma\alpha\alpha$	$\alpha\alpha\gamma$ $\alpha\gamma\alpha$ $\gamma\alpha\alpha$
H	$\alpha\alpha\gamma$	$\alpha\alpha\gamma$	$\alpha\alpha\gamma$

6.2. Segmentation and border detection in gray level pictures.

In this section we show how the ideas of this paper can be applied to the problem of segmenting objects in gray level pictures.

Two major problems in low level vision are a segmentation of a given gray level picture into background and object(s), and edge detection [12]. In their extreme case, these two problems are identical, since given perfect edges of a picture, it can be immediately segmented, and given a segmentation of a picture into background and object, the borders can be immediately found. This idea has been used to develop some powerful segmentation techniques [13].

6.2.1. Consistency of a segmented picture.

We consider the problem of segmenting a gray level picture (into background, object, and border) as a consistent labeling problem. The set of solutions to this problem is all the possible ways in which a scene can be labeled using these three labels. We suggest using the relations between these labels to define a consistency measure over the set of segmented pictures. Given a binary picture, its border pixels are defined as object pixels that have a 4–neighbor in the background [12]. Therefore, most of the combinations of these three labels are inconsistent with the border definition. For example, all segmentations containing object pixels that are 4–connected to background pixels are inconsistent.

The problem with such a definition of consistency is that it does not reflect our intuitive concept of a "good" segmentation. For example, noisy pictures may still be considered consistent. Since the concept of a "good segmentation" is intuitive, we suggest defining consistency as a learning process. Given a set of good segmentations, we use it to learn what a good segmentation is. In our case, where the labels are background, object, and border, we use the good segmentations to learn the possible relations between these three labels. In this case, it is clear that the close neighbors of a pixel will have the greatest effect on its local consistency. Using Definition 3 we see that the local consistency of a given label with the labels of its neighbors is the probability of having such a label with such neighbors *among the good segmentations.*

6.2.2. Events for determining local consistency.

We restrict ourselves to considering only the eight closest neighbors of a given pixel for computing its local consistency. Even so, since there are three possible labels, the total number of different events is $3^8 = 6561$. It is clear therefore that Definition 3 cannot be applied, and we have to consider the estimated local consistency of Definition 8. If we choose the events $U_1 \cdots U_r$ of Definition 8 as the elementary events of the neighbors, the estimate may not be good enough. We suggest events that make use of the symmetry of the problem.

In this problem we assume that if a given labeling of a picture is consistent, so will be the mirror image of this labeling and its four rotations. This reduces the number of events, and for computational efficiency instead of considering the actual labels of the neighbors, one can consider invariant functions of these labels under rotation and mirror imaging. Such invariants were developed by Hu in [14] and are given in terms of algebraic moments. To obtain the moments of the eight neighbors, they must be encoded into numbers. A simple way of doing this is to assign 0 to background, 1 to border, and 2 to object pixels. The moments can thus be computed, and we take the events $U_1 \cdots U_r$ as the values of the six invariants of [14] with the additional invariant m_{00} (the sum of all neighbors). Each event therefore consists of seven invariant values.

Since the values of these invariants are distinct, the corresponding events are orthogonal, i.e. $R_{ij} = \text{Prob}(U_i, U_j) = 0,\ i \neq j$. Substituting this into the Yule-Walker equations we obtain the trivial solution

$$b_i = \frac{R_{0i}}{R_{ii}} = \frac{\text{Prob}(v_j^{k_j}, U_i)}{\text{Prob}(U_i)} = \text{Prob}(v_j^{k_j} \mid U_i)$$

Substituting this in (11) and noticing that only $u_i = 1$ we get

$$\hat{u}_0 = \text{Prob}(v_j^{k_j} \mid U_i)$$

and since $0 \leq \hat{u}_0 \leq 1$ we get from Definition 7

$$\text{Elcons}(v_j^{k_j}) = \text{Prob}(v_j^{k_j} \mid U_i)$$

where U_i is the sole event among $U_1 \cdots U_r$ that happened. This means that the estimated local consistency of a label is the probability of this label given the invariants of its neighbors.

6.2.3. Determining the initial estimates.

To determine the goodness measure of a given segmentation we must first decide how to compute the initial estimates x_i^j from the gray level picture. One way of obtaining these estimates is to apply three local operators, one to detect background, one to detect objects, and one to detect edges. Trivially, the first two operators can be identical. Considering the initial estimates for borders, we may try one of the edge detection operators [12]. The problem with such an approach is that it is well known that these operators are very sensitive to noise. Hence, in a noisy picture, the border estimates might be very bad compared with the estimates of object and background.

We suggest the following initial estimates. The estimate for borders will be a small number regardless of the pixel's gray level (or its neighbors). This number can be chosen as the relative number of border pixels in the good segmentations. Denoting object labels by α, background labels by β, and border labels by γ, we have $x_i^\gamma = p\ \ \forall i$. x_i^α, the local estimate that a pixel is in an object, can be taken as the pixel's normalized gray level, and then x_i^β can be taken as $1 - x_i^\alpha$.

6.2.4. A heuristic algorithm for maximizing the goodness measure.

Because of the huge number of possible solutions, it is unlikely that the global maximum can be found. We suggest therefore a less robust definition for maximum, that corresponds to a local maximum for continuous functions.

Definition 9:

> A local maximum of order l is an assignment of labels to objects such that no change of l or fewer labels will increase the goodness measure.

Clearly, a local maximum of order $l1$ is also a local maximum of order $l2$ if $l1 > l2$, and the global maximum is a local maximum of order n.

We proceed to develop an algorithm that achieves a local maximum of order l. We denote by $t(v_{i1}^{k_{i1}}, \cdots, v_{il}^{k_{il}})$ the contribution of the labeling $k_{i1}, \cdots, k_{il}$ of the objects $a_{i1}, \cdots, a_{il}$ to the global consistency measure. There are $\binom{n}{l}$ different subsets of l objects. For each such subset there exist m^l possible labelings. We can thus try to maximize the consistency measure by maximizing $t(v_{i1}^{k_{i1}}, \cdots, v_{il}^{k_{il}})$ for all the $\binom{n}{l}$ subsets. For each subset, the optimization involves choosing a maximum among the m^l possible labelings. The total number of labelings of size l to be considered when each subset maximizes itself is therefore $\binom{n}{l} \cdot m^l$, and for small l it is approximately $(m \cdot n)^l$. Peleg in [7] showed the effectiveness of parallel local optimizations. This method however, is not immediately applicable in our case, since the subsets to be optimized are not mutually exclusive. Furthermore, even for mutually exclusive subsets, changing the labeling of one subset may affect the consistency of another subset so that convergence is not guaranteed. A sequential optimization of the subsets, however, does converge to a local maximum, since the sequential optimization can only increase the global consistency which has an upper bound of 1. When the process converges, its convergence point trivially fulfills the local consistency condition. We must consider, however, two disadvantages of the sequential scanning of the subsets. It cannot be implemented on a parallel computer, and the result depends on the order in which the subsets are scanned. When the consistency of a given object is determined by some *local* events, i.e. events involving its close neighbors, it is possible to partially eliminate these disadvantages. We suggest the following parallel/sequential algorithm: divide all subsets into distinct independent classes. By this we mean that changing the labeling of one subset will not affect the consistency of other subsets in the same class. Thus, it is possible to optimize in parallel all the subsets in one class, and go over all classes sequentially.

To implement these ideas to an algorithm that achieves local consistency of order 1, consider the consistency goodness measure $CDM\,2$ of (9):

$$CDM\,2(k_1, \cdots, k_n) = \prod_i \text{Lcons}(v_i^{k_i}) \cdot \prod_{i=1}^{n} x_i^{k_i}$$

The label k_i for a_i contributes the multiplicative term $\text{Lcons}(v_i^{k_i}) \cdot x_i^{k_i}$ to the goodness measure. It also affect the local consistency of the neighbors of a_i, so that the total contribution of the label k_i to the goodness measure is

$$t(k_i) = \text{Lcons}(v_i^{k_i}) \cdot x_i^{k_i} \cdot \prod_{j\,=\,\text{neighbor of }i} \text{Lcons}(v_j^{k_j})$$

Given the labels of the neighbors of a_i, $t(k_i)$ can have three values, corresponding to the three possible values of k_i. In this case all the pixels that do not have mutual neighbors can belong to the same class. So, there are four independent classes: pixels with both coordinates even, pixels with both coordinates odd, pixels with x coordinate odd and y coordinate even, and pixels with x coordinate even and y coordinate odd.

6.2.5. Results.

Results of experiments done with the above interlacing algorithm for consistency of order 1 are shown in Figures 2-5. Figure 2 is a "good" segmentation from which statistics were gathered. It was found that there are only 22 different invariant values, so that the consistency measure is found by determining which of these 22 events happened. This involves five comparisons in the worst case, and less than three comparisons on the average. Figure 3 is a noisy gray level picture of a tank. Figure 4 is the initial segmentation obtained by thresholding, and Figure 5 is the result of our algorithm ,which converged in this case after five iterations. In all these figures the background labels are represented by dots, the object labels are represented by the symbol " • ", and the border labels by the symbol " ◇ ".

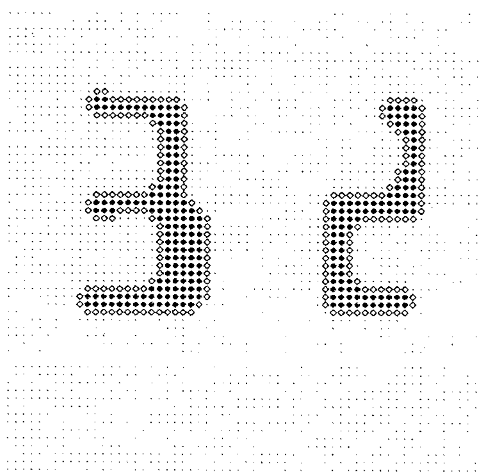

Figure 2: The learning example from which statistics for the background, boundary, and object labels were gathered.

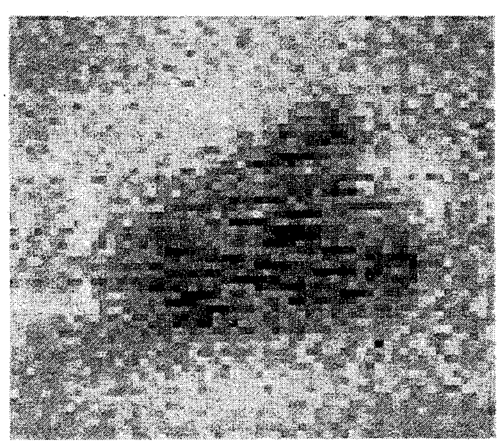

Figure 3: A noisy FLIR image of a tank.

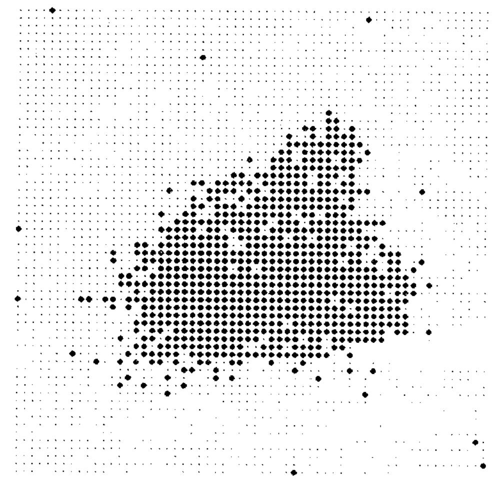

Figure 4: Initial segmentation of the tank image obtained by thresholding.

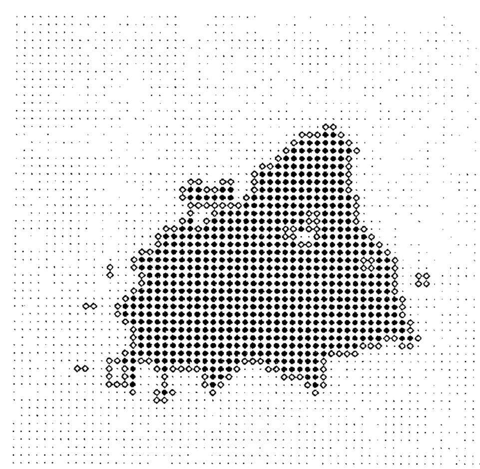

Figure 5: Applying the segmentation algorithm of Section 6.2.4, with initial labeling as Figure 4.

7. Concluding remarks.

A new approach to the consistent labeling problem has been described. Labeling is being regarded as events in a probability space. This representation enables a derivation of an optimal Bayesian measure. This measure, however, has been shown impractical from both computational and information aspects. The major novelty of this paper is in the suggestion of an alternative measure that corresponds to the intuitive concept of consistency. Although the new measure is also impractical, since it needs the same amount of information as the probability measure, we were able to derive approximations that use limited probabilistic information which is usually available.

The consistent labeling problem considered here is sometimes called the "continuous consistent labeling problem" to distinguish it from the "discrete labeling problem". The discrete problem differs from its continuous version by considering only possible and impossible solutions. This means that all possible solutions are equally "good" and the "goodness" measure has only two values: possible and impossible. It seems that the approach of this paper can also be applied to the discrete case. A probability space can be artificially created by considering all legal solutions as having the same probability, and the goodness measure is then defined as an event that happens with probability 1 for all the possible solutions. Preliminary results show that in order to get such a goodness measure one can use information which is probabilistic, i.e. not necessarily crisp.

References

1. D. A. Huffman, Impossible objects as nonsense sentences, in *Machine Intelligence* , vol. 6 , B. Meltzer and D. Mitchi (ed.), Edinburgh University Press, 1971 , 295 - 323.
2. A. Rosenfeld, R. Hummel and S. W. Zucker, Scene labeling by relaxation operations, *IEEE Trans. Systems Man Cyber. 6* , (1976), 420-433.
3. S. Peleg, A new probabilistic relaxation scheme , *IEEE Trans. Pattern Anal. Machine Intell. 2*, (1980), 362-369.
4. D. Waltz, Understanding line drawings of scanes with shadows, in *The Psychology of Computer Vision*, P. Winston (ed.), McGraw-Hill, 1975.
5. O. Faugeras and M. Berthod, Improving consistency and reducing ambiguity in stochastic labeling: an optimization approach, *IEEE Trans. Pattern Anal. Machine Intell. 3* , (1981), 412-424.
6. R. Hummel and S. Zucker, On the foundation of relaxation labeling processes, *IEEE Trans. Pattern Anal. Machine Intell. 5* , (1983), 267-287.
7. S. Peleg, Classification by discrete optimization, *Computer Vision, Graphics, Image Processing 25*, (1984), 122-130.
8. R. M. Haralick, Decision making in context, *IEEE Trans. Pattern Anal. Machine Intell. 5*, (1983), 417-428.
9. J. Raviv, decision making in Markov chains applied to the problem of pattern recognition, *IEEE Trans. Inform. Theory 13*, (1967), 536-551.
10. L. A. Zadeh, Fuzzy sets, *Information Control 8*, (1965), 338-353.
11. A. Papoulis, *Probability, Random Variables, and Stochastic processes*, (second edition), McGraw-Hill, New York, 1984.
12. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, (second edition), Academic Press, New York, 1982.
13. D. L. Milgram, Region extraction using convergent evidence, *Computer Vision, Graphics, Image Processing 11*, (1979), 1-12.
14. M. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Inform. Theory 8*, (1962), 179-187.