CRF-Matching: Conditional Random Fields for Feature-Based Scan Matching

Fabio Ramos*   Dieter Fox†   Hugh Durrant-Whyte*

* ARC Centre of Excellence for Autonomous Systems   † Dept. of Computer Science & Engineering
Australian Centre for Field Robotics   University of Washington
The University of Sydney   Seattle, WA, USA
Sydney, NSW, Australia

Abstract—**Matching laser range scans observed at different points in time is a crucial component of many robotics tasks, including mobile robot localization and mapping. While existing techniques such as the Iterative Closest Point (ICP) algorithm perform well under many circumstances, they often fail when the initial estimate of the offset between scans is highly uncertain. This paper presents a novel approach to 2D laser scan matching. CRF-Matching generates a Condition Random Field (CRF) to reason about the joint association between the measurements of the two scans. The approach is able to consider arbitrary shape and appearance features in order to match laser scans. The model parameters are learned from labeled training data. Inference is performed efficiently using loopy belief propagation. Experiments using data collected by a car navigating through urban environments show that CRF-Matching is able to reliably and efficiently match laser scans even when no a priori knowledge about their offset is given. They additionally demonstrate that our approach can seamlessly integrate camera information, thereby further improving performance.**

I. INTRODUCTION

Many robotics tasks require the association of sensor data observed at different points time. For instance, in mobile robot mapping, a robot needs to be able to accurately determine the spatial relationship between different laser scans. While this task is rather straightforward if enough prior knowledge about the relative location of the scans is given, it becomes more challenging when no knowledge about the spatial relationship is available. Given two scans of sensor measurements, the matching problem can be defined as finding a transformation that best matches one scan to another. For example, when laser range finders are used in mobile robotics, such a transformation corresponds to the movement performed by the robot between scans. To find the transformation between two scans, it is necessary to associate the individual measurements in one scan with the corresponding measurements in the other scan.

The most widely used algorithm for matching range sensors in robotics is the Iterative Closest Point (ICP) algorithm [5]. ICP alternates between nearest neighbor association and least-squares optimization to compute the best transformation between the laser scans given the most recent association. Although ICP and its extensions are fast and in general produce good results, simple nearest neighbor association has a number of drawbacks. First, ICP frequently generates incorrect transformation estimates when the initial offset between the scans is large. This is in part caused by the association cost function that does not take into account higher-level information from the data, such as shape descriptions. Second, ICP does not provide adequate means of fusing data collected by multiple types of sensors to improve matching. Third, ICP provides only limited support for estimating the uncertainty of the resulting transformation. Such uncertainty estimates are important in the context of tasks such as robot localization or mapping using Bayes filters [21].

This paper presents CRF-Matching, an alternative procedure for laser scan matching based on Conditional Random Fields (CRFs) [8]. CRFs are undirected graphical models that are very powerful for modeling relational information (spatial data for example). By directly modeling the conditional probability of the hidden states given the observations rather than the joint probability, CRFs avoid the difficult task of specifying a generative model for observations, as necessary in techniques such as Hidden Markov Models (HMMs) or Markov Random Fields (MRFs). As a result, CRFs can handle arbitrary dependencies between observations, which gives them substantial flexibility in using high-dimensional feature vectors.

CRF-Matching focuses on the problem of data association between two laser scans. This is accomplished by converting the individual measurements of one laser scan into hidden nodes of a CRF. The states of each node range over all measurements in the other scan. The CRF models arbitrary information about local appearance and shape of the scans. Consistency of the association is achieved by connections between nodes in the CRF. CRF-Matching learns model parameters discriminatively from sets of aligned laser scans. When applied to a new pair of scans, maximum *a posteriori* estimation is used to determine the data association, which in turn specifies the spatial transformation between the scans.

Extensive experiments show that CRF-Matching significantly outperforms ICP when matching laser range-scans with large spatial offset. Furthermore, they show that our approach is able to reliably match scans without a priori information about their spatial transformation, and to incorporate visual information to further improve matching performance.

This paper is organized as follows. After discussing related

work in Section II, we provide an overview of Conditional Random Fields in Section III. CRF-Matching is introduced in Section IV, followed by an experimental evaluation in Section V. We conclude in Section VI.

II. RELATED WORK

ICP has been applied to robotics quite successfully, however, it does not explicitly account for sensor rotation since it uses the Euclidean distance to compute the nearest neighbor. To overcome this limitation, [12] combines the normal NN with angular constraints in the Iterative Dual Correspondence (IDC). The algorithm uses two types of correspondences (translation and rotation) and at each iteration performs two optimizations. [12] also proposes the interpolation of lines between laser points to improve robustness for large transformations. Although these extensions improve the basic ICP algorithm, they do not eliminate the chance the algorithm reaches a poor local minima.

Methods for computing the transformation uncertainty from ICP were also proposed [2], [17]. However, they do not take into account the association uncertainty between pairs of measurements. This can cause large errors in the uncertainty estimation for the transformation since weak associations can equally contribute to the overall estimate.

Shape matching has been a long-standing problem especially for the computer vision community (see [22] for a review). Various techniques have been applied to represent shape including Fourier descriptors [20], parametric curves [13], and geodesic distances [9]. We use some of these ideas in this work to encode shape and image properties.

A similar probabilistic model for 3D matching of non-rigid surfaces was proposed by [1]. The model is trained generatively and assumes a large number of range points in the objects for accurate results. The main feature employed is the geodesic distance which performs well when there is a well-defined object structure. However, in unstructured environments this assumption is not valid. For this reason we employ several features for shape description with the addition of image features when convenient.

Loop closure detection in outdoor environments was investigated in [15]. A combined laser and image matching was proposed with a vocabulary of features. In this paper, we also combine laser and camera information for loop closure detection as a possible application for our technique. The key benefit of our method is the deployment of a single probabilistic model able to fuse and provide uncertainty estimation in a natural manner.

III. PRELIMINARIES

A. Conditional Random Fields (CRF)

Conditional random fields (CRFs) are undirected graphical models developed for labeling sequence data [8]. CRFs directly model $p(\mathbf{x}|\mathbf{z})$, the *conditional* distribution over the hidden variables $\mathbf{x}$ given observations $\mathbf{z}$. This is in contrast to generative models such as Hidden Markov Models or Markov Random Fields, which apply Bayes rule to infer hidden states [18]. Due to this structure, CRFs can handle arbitrary dependencies between the observations $\mathbf{z}$, which gives them substantial flexibility in using high-dimensional feature vectors.

The nodes in a CRF represent hidden states, denoted $\mathbf{x} = \langle \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \rangle$, and data, denoted $\mathbf{z}$. The nodes $\mathbf{x}_i$, along with the connectivity structure represented by the undirected edges between them, define the conditional distribution $p(\mathbf{x}|\mathbf{z})$ over the hidden states $\mathbf{x}$. Let $\mathcal{C}$ be the set of cliques (fully connected subsets) in the graph of a CRF. Then, a CRF factorizes the conditional distribution into a product of *clique potentials* $\phi_c(\mathbf{z}, \mathbf{x}_c)$, where every $c \in \mathcal{C}$ is a clique in the graph and $\mathbf{z}$ and $\mathbf{x}_c$ are the observed data and the hidden nodes in the clique c, respectively. Clique potentials are functions that map variable configurations to non-negative numbers. Intuitively, a potential captures the "compatibility" among the variables in the clique: the larger the potential value, the more likely the configuration. Using clique potentials, the conditional distribution over hidden states is written as

$$p(\mathbf{x} \mid \mathbf{z}) \quad = \quad \frac{1}{Z(\mathbf{z})} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{z}, \mathbf{x}_c), \qquad (1)$$

where $Z(\mathbf{z}) = \sum_{\mathbf{x}} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{z}, \mathbf{x}_c)$ is the normalizing partition function. The computation of this partition function can be exponential in the size of $\mathbf{x}$. Hence, exact inference is possible for a limited class of CRF models only.

Potentials $\phi_c(\mathbf{z}, \mathbf{x}_c)$ are described by log-linear combinations of *feature functions* $\mathbf{f}_c$, *i.e.*,

$$\phi_c(\mathbf{z}, \mathbf{x}_c) \quad = \quad \exp\left(\mathbf{w}_c^T \cdot \mathbf{f}_c(\mathbf{z}, \mathbf{x}_c) \right), \qquad (2)$$

where $\mathbf{w}_c^T$ is a weight vector, and $\mathbf{f}_c(\mathbf{z}, \mathbf{x}_c)$ is a function that extracts a vector of features from the variable values. Using feature functions, we rewrite the conditional distribution (1) as

$$p(\mathbf{x} \mid \mathbf{z}) \quad = \quad \frac{1}{Z(\mathbf{z})} \exp\left\{ \sum_{c \in \mathcal{C}} \mathbf{w}_c^T \cdot \mathbf{f}_c(\mathbf{z}, \mathbf{x}_c) \right\} \qquad (3)$$

B. Inference

Inference in CRFs can estimate either the marginal distribution of each hidden variable $\mathbf{x}_i$ or the most likely configuration of all hidden variables $\mathbf{x}$ (*i.e.*, MAP estimation), as defined in (3). Both tasks can be solved using *belief propagation* (BP), which works by sending local messages through the graph structure of the model. Each node sends messages to its neighbors based on messages it receives and the clique potentials, which are defined via the observations and the neighborhood relation in the CRF.

BP generates exact results in graphs with no loops, such as trees or polytrees. However, since the models used in our approach contain various loops, we apply loopy belief propagation, an approximate inference algorithm that is not guaranteed to converge to the correct probability distribution [14]. In our experiments, we compute the MAP labeling of a scan match using max-product loopy BP. Fortunately, even when the algorithm failed to converge, our experiments showed reasonable results.
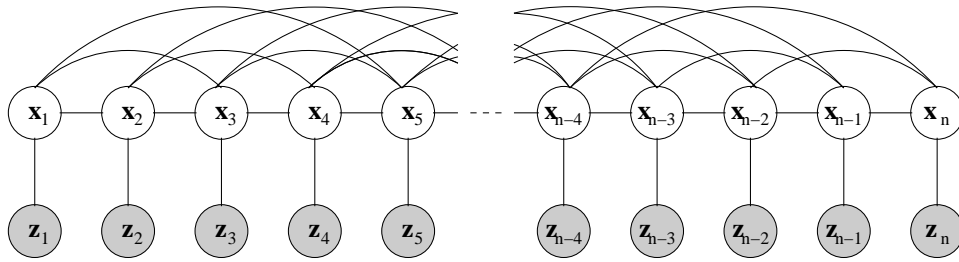
Fig. 1. Graphical representation of the CRF-Matching model. The hidden states $\mathbf{x}_i$ indicate the associations between the points in the different scans. The observations $\mathbf{z}_i$ corresponds to shape or visual appearance information extracted from the two laser scans.

C. Pseudo-Likelihood Parameter Learning

The goal of CRF parameter learning is to determine the weights of the feature functions used in the conditional likelihood (3). CRFs learn these weights discriminatively by maximizing the conditional likelihood of labeled training data. While there is no closed-form solution for optimizing (3), it can be shown that (3) is convex relative to the weights $\mathbf{w}_c$. Thus, the global optimum of (3) can be found using a numerical gradient algorithm. Unfortunately, this optimization runs an inference procedure at each iteration, which can be intractably inefficient in our case.

We therefore resort to maximizing the *pseudo-likelihood* of the training data, which is given by the sum of local likelihoods $p(\mathbf{x}_i \mid \mathrm{MB}(\mathbf{x}_i))$, where $\mathrm{MB}(\mathbf{x}_i)$ is the Markov blanket of variable $\mathbf{x}_i$: the set of the immediate neighbors of $\mathbf{x}_i$ in the CRF graph [4]. Optimization of this pseudo-likelihood is performed by minimizing the negative of its log, resulting in the following objective function:

$$L(\mathbf{w}) = -\sum_{i=1}^{n} \log p(\mathbf{x}_i \mid \mathrm{MB}(\mathbf{x}_i), \mathbf{w}) + \frac{(\mathbf{w}-\widetilde{\mathbf{w}})^T(\mathbf{w}-\widetilde{\mathbf{w}})}{2\sigma^2} \quad (4)$$

Here, the terms in the summation correspond to the negative pseudo log-likelihood and the right term represents a Gaussian shrinkage prior with mean $\widetilde{\mathbf{w}}$ and variance σ^2. Without additional information, the prior mean is typically set to zero. In our approach, we use unconstrained L-BFGS [11], an efficient gradient descent method, to optimize (4). The key advantage of maximizing pseudo-likelihood rather than the likelihood (3) is that the gradient of (4) can be computed extremely efficiently, without running an inference algorithm. Learning by maximizing pseudo-likelihood has been shown to perform very well in different domains; see [7], [19], [6].

IV. CRF-MATCHING

A. Model Definition

In order to find the association between a laser scan A and another scan B, CRF-Matching generates a CRF that contains a hidden node $\mathbf{x}_i$ for each laser point in scan A. Such a CRF is shown in Figure 1. For now, we assume that each point in scan A can be associated to a point in scan B; outliers will be discussed at the end of this section. To reason about associations between the two scans, each hidden state $\mathbf{x}_i$ of the CRF ranges over all points in laser scan B.

The nodes $\mathbf{z}_i$ in Figure 1 correspond to features associated with the individual laser points. These features describe local appearance properties of the laser scans. In order to achieve global consistency between the individual data associations, each hidden node is connected to other hidden nodes in the network. We now describe the individual features used in the clique potentials of the CRF model.

B. Local Features

CRF-Matching can employ arbitrary local features to describe shape, image properties, or any particular aspect of the data. Since our focus is on associating points in scan A to similar points in scan B, our features describe *differences* between data points. The learning algorithm provides means to weight each of the resulting features to best associate the data. The local features are described as follows:

Spatial distance: This feature measures the distance between points in one scan w.r.t. points in the other scan. This is the basic feature used in ICP, which rather than representing the shape of the scan, accounts for small position transformations. If we denote the locations of individual points in scan A and B by $z_{A,i}$ and $z_{B,j}$, respectively, then the feature $\mathbf{f}_d$ for point i in scan A can be defined as

$$\mathbf{f}_d(i,j,z_A,z_B) = \frac{\|z_{A,i} - z_{B,j}\|^2}{\sigma^2}, \quad (5)$$

where σ^2 is the variance of the distances in the training data. Note that the value of this feature depends on the point j to which i is associated. The reader may also notice that this feature is only useful if the initial offset and rotation between the two scans is small.

Shape difference: These features capture how much the local shape of the laser scans differs for each possible association. While local shape can be captured by various types of features, we chose to implement very simple shape features measuring distance, angle, and geodesic distance along the scans.

To generate distance features, we compute for each point in scan A its distance to other points in scan A. These other points are chosen based on their relative indices in the scan. The same distances can be computed for points in scan B, and the resulting feature is the difference between the distance values in A and B. To see, let k be an index offset for the distance feature. The feature value corresponding to points i

and j is computed as follows:

$$\mathbf{f}_{\text{dist}}\left(i, j, k, z_A, z_B\right) =$$
$$\frac{\left\|\left\|z_{A,i} - z_{A,i+k}\right\| - \left\|z_{B,j} - z_{B,j+k}\right\|\right\|^2}{\sigma^2}. \quad (6)$$

In our implementation this feature is computed for index offsets $k \in \{-1, 1, -3, 3, -5, 5\}$. The reader may notice that this feature is based on the assumption that the two scans have similar densities. However, even though this assumption is often violated, we found this feature to be very valuable to increase consistency of data associations.

Another way to consider local shape is by computing the difference between the angles of points in both scans w.r.t their neighbors. The angle of a point $z_{A,i}$ is defined as the angle between the segments connecting the point to its neighbors, where k indicates the index offset to the neighbor. This feature is defined as:

$$\mathbf{f}_{\text{angle}}\left(i, j, k, z_A, z_B\right) =$$
$$\frac{\left\|\angle\left(\overline{z_{A,i-k} z_{A,i}}, \overline{z_{A,i} z_{A,i+k}}\right) - \angle\left(\overline{z_{B,j-k} z_{B,j}}, \overline{z_{B,j} z_{B,j+k}}\right)\right\|^2}{\sigma^2} (7)$$

As with the distance feature, we compute the difference of angles for neighbors 1, 3 and 5 points apart.

The *geodesic distance* provides additional shape information as it is defined as the sum of Euclidean distances between points along the scan. As with the previous features, it can be calculated for different neighborhoods representing local or long-term shape information. Given points $z_{A,i}$ and $z_{B,j}$ and a neighborhood k, the geodesic distance feature is computed as:

$$\mathbf{f}_{\text{geo}}\left(i, j, k, z_A, z_B\right) =$$
$$\frac{\left\|\sum_{l=i}^{i+k-1}\left\|z_{A,l+1} - z_{A,l}\right\| - \sum_{l=j}^{j+k-1}\left\|z_{B,l+1} - z_{B,l}\right\|\right\|}{\sigma^2} (8)$$

[1] used such a feature for matching 3D laser scans.

Visual appearance: When camera data is available, it can be integrated with shape information from the scans to help with the association. The projection of laser points into the camera image is used to extract an image patch for each laser point. We use Principal Components Analysis to reduce the dimensionality of image features. The feature is then computed as the Euclidean distance between the principal components of a patch $z_{A,i}^I$ in scan A and $z_{B,j}^I$ in scan B:

$$\mathbf{f}_{\text{PCA}}\left(i, j, z_A^I, z_B^I\right) = \frac{\left\|z_{A,i}^I - z_{B,j}^I\right\|^2}{\sigma^2}. \quad (9)$$

All features described so far are *local features* in that they only depend on a single hidden state i in scan A (indices j and k in the features define nodes in scan B and neighborhood size). The main purpose of these features is to associate scan points that have similar appearance. However, in order to generate *consistent* associations it is necessary to define features that relate the hidden states in the CRF to each other.

C. Pairwise Features

The following features are used to define the clique potentials of nodes connected in the CRF.

Association: The main purpose of this feature is to ensure consistency (avoiding crossover labels for example) and enforce sequential labeling: If a measurement i in A is associated to measurement j in B, its neighbor $i + 1$ has a high chance of being associated to $j + 1$ in B. To measure consistency, we define a function $\Delta = \mathbf{x}_{i+k} - \mathbf{x}_i$ that determines the difference between the hidden states of two nodes $\mathbf{x}_i$ and $\mathbf{x}_{i+k}$. When $\Delta = k$, then the associations of $\mathbf{x}_i$ and $\mathbf{x}_{i+k}$ are fully consistent. We convert Δ into ten different binary features, corresponding to different Δ values.

Pairwise distance: This feature is very similar to $\mathbf{f}_{\text{dist}}$ described above. However, instead of being defined over a *single* hidden node $\mathbf{x}_i$ only, it measures the consistency between the associations of *two* hidden nodes $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$\mathbf{f}_{\text{pair}}\left(i, j, m, n, z_A, z_B\right) =$$
$$\frac{\left\|\left\|z_{A,i} - z_{A,j}\right\| - \left\|z_{B,m} - z_{B,n}\right\|\right\|^2}{\sigma^2}. \quad (10)$$

Here, i and j are the indices of points in scan A, and m and n are values of their hidden nodes. In essence, this feature assumes that the two scans have the same shape and scale. In contrast to $\mathbf{f}_{\text{dist}}$, however, it does not assume that the scans have the same spatial resolution. Such a feature has been shown to be very valuable for matching camera images [3].

D. Outlier Detection

So far we assumed that every point in scan A can be associated to a point in scan B. However, such an approach cannot handle changes in an environment or partially overlapping laser scans. In order to deal with such cases, the hidden state of each node contains one additional value that corresponds to "outlier". Unfortunately, most of the features discussed so far can and should not be computed if one of the involved hidden nodes is an outlier. For instance, the pairwise distance feature $\mathbf{f}_{\text{pair}}$ cannot be computed if the hidden value m is set to "outlier", since there is no associated laser point $z_{B,m}$ to which the distance can be computed. Fortunately, CRFs are extremely well suited for handling such cases. We simply define two additional binary feature functions that are true whenever the value of an involved hidden node is "outlier" (we get one function for local features and one function for pairwise features). Whenever such a feature is true, the values of all corresponding feature functions are set to zero. When learning the weights of the different feature values, the algorithm automatically learns weights for the binary outlier features that result in most consistent data associations.

E. Boosting Features

CRFs are able to directly incorporate the local, continuous features described above. However, in order to model more complex, nonlinear relationships between hidden states and feature values, it is advantageous to discretize the features. Recently, Friedman and colleagues [6] showed that it is

possible to learn AdaBoost classifiers from labeled training data and to use the resulting classifier outputs as features in a CRF. We apply their approach by learning boosted classifiers that combine the local features described above.

To train Adaboost, each pair of laser points is labeled as "associated" or "not associated". Adaboost then generates a discretization of the continuous features and an initial classification of pairs of laser points (whether or not they should be associated). The output value of Adaboost, which is a non-linear combination of the learned features, is used as an observation in the CRF. Note that this "observation" is different for each possible hidden state in the CRF (each hidden state gives a different association of laser points). The CRF then learns a weight for the Adaboost output in combination with the other weights (pairwise feature weights and outlier feature weight). For the experiments Adaboost was trained with 50 decision stumps.

F. Position estimation

To compute the spatial transformation between two laser scans, CRF-Matching uses the uncertainty over the association obtained. The uncertainty is incorporated in a least squared optimization problem for the non-outlying points:

$$Err = \sum_{i=1}^{n} w_i \left(z_{A,i} R + T - z_{B,a(i)} \right)^2,$$

where $a(i)$ is the point in scan B associated with point i, and R and T are the rotation and translation matrices respectively. The weight w_i corresponds to the probability of the association for point i obtained with belief propagation. Using this approach, points that are more likely to have the correct association have more importance on position estimation. The same does not occur in ICP where each point has the same contribution.

G. Algorithm Overview and Performance

CRF-Matching converts the matching problem into a probabilistic inference procedure on a CRF where the hidden states correspond to association hypotheses. Outliers are handled explicitly by an additional "outlier" state. We defined various local shape and appearance features. Instead of using these features directly in the CRF, we first train boosted classifiers in order to model more complex relationships between the features. The weights of all local and pairwise features are learned from labeled training scans using pseudo-likelihood maximization. During testing, we apply loopy belief propagation (LBP) to compute the maximum a posteriori association for the complete scan. This association can then be used to determine the spatial transformation between the two scans.

The computational complexity of loopy belief propagation is linear in the number of hidden nodes (points in scan A), and the complexity of computing each message is quadratic in the number of hidden states (points in scan B). In our experiments LBP converges in 0.1 second in scans with 90 beams and 1 second in scans with 180 beams.
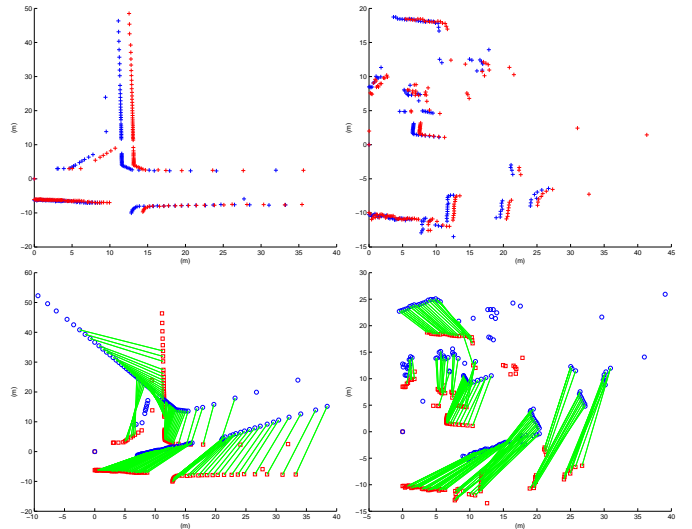

Fig. 2. Two examples of the scans used in the experiments. Original pairs of scans (top) are rotated and translated. CRF matching and ICP are applied to the new configuration. The bottom pictures show the MAP correspondence obtained with CRF-Matching.

Offset	1	2	3	4	5	6	7	8	9	10
X(m)	0.1	0.25	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Y(m)	0.1	0.25	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
θ(Deg.)	1	5	10	15	20	25	30	40	50	60

TABLE I

TRANSLATION AND ROTATION OFFSETS USED IN THE EXPERIMENTS.

V. EXPERIMENTS

We performed experiments with outdoor data collected with a modified car travelling at 0 to 40 km/h. The car performed several loops around the university campus which has structured areas with buildings, walls and cars, and areas less structured with bush, trees and lawn fields. 20 pairs of scans obtained at different points of the trajectory were used for training and 50 different pairs were used for testing. The scans of each pair are taken approximately 0.25 second apart while the car is in movement. To further evaluate the robustness of the algorithm, the scans are translated and rotated using 10 different offsets. The translation offsets ranged from 0.1m to 4m in x and y direction, and the rotations ranged from 1° to 60° degrees. Table I shows the ten different offsets used. Since we know the additional translation and rotation, the ground truth can be estimated from the association over the original scans by running ICP multiple times for different outlier thresholds. For each offset, a CRF model is trained using the training set with 20 pairs of scans. In all experiments, 90 beams per laser scan are associated (results achieved with 180 beams are virtually identical).

A. Laser-based Matching

In this experiment, we compare the robustness of ICP and CRF for laser scan matching. We have also tested the IDC algorithm with mutual closest point and trimming with similar results to ICP. The experiment is illustrated in Figure 2: a
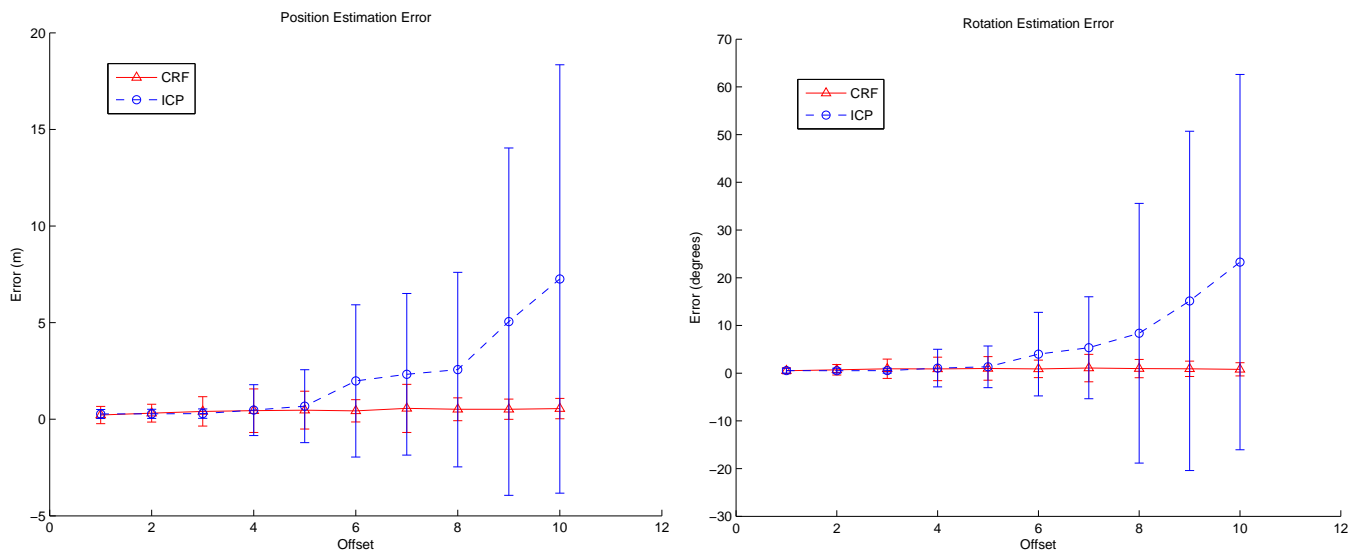
Fig. 4. Translation estimate error (top) and rotation estimate error (bottom) for 90-point scans. While ICP error increases significantly after the 5th offset, CRF matching error is constant for all configurations tested.
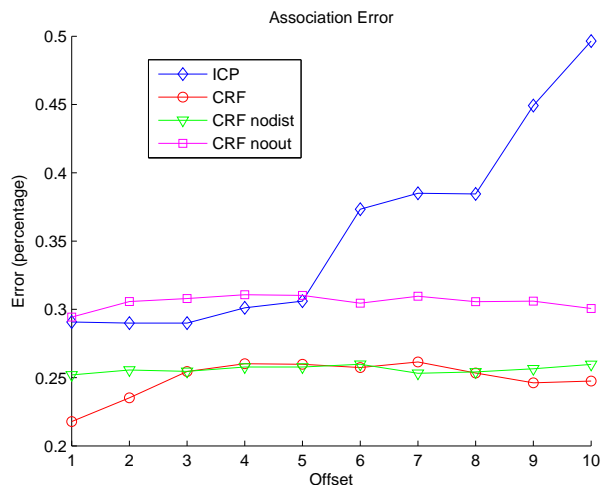

Fig. 3. Correspondence error for ICP and CRF matching for 90-point laser scans. The performance of ICP degrades after offset 5 while CRF matching remains constant.

pair of scans is obtained during the trajectory and artificially rotated and translated. ICP and CRF matching are then applied and compared. To guarantee ICP reaches its best output, the algorithm is computed twice, for a threshold of 20 metres and for a threshold of 5 metres (ICP treats nearest neighbor distances beyond the threshold as outliers). In each case 100 ICP iterations are performed.

Figure 3 summarizes the results of this experiment. For each offset, it shows the fraction of measurements that are not identical to the "ground truth" association computed with ICP on the non-shifted scans. It can be noticed that the performance of ICP degrades significantly for translations larger than 1.5 meter combined with rotation of 20 degrees (offset 5). In contrast, CRF matching keeps the same performance

independently of the offset. This indicates the main benefit of our approach; the algorithm is able to associate points by shape when it is necessary or by distance when this feature is more relevant.

Figure 4 shows the spatial transformation errors. As can be seen, for small offsets, the translation and rotation errors of ICP and CRF-Matching are very low (graphs labeled ICP and CRF). While CRF-Matching maintains this low error for larger offsets, the performance of ICP degrades significantly. In order to evaluate the ability of CRF-Matching to align scans without prior knowledge about the relative locations of laser scans, we removed the distance feature defined in (5) from the feature set of CRF-Matching. The resulting graph, labeled "CRF nodist" in Figure 3, indicates that CRF-Matching can globally match laser scans. The graph labeled "CRF noout" shows the performance when the CRF does not contain the outlier state. In this case, the association error increases, which shows that explicit outlier handling is important for real outdoor laser data.

We also performed experiments removing classes of features to evaluate the importance of each feature for the overall performance. When we removed the local shape features or the image features the performance was 10% worse (measured by the association error). Removing the pairwise features and making the matching entirely based on the output of Adaboost resulted in error rates three times as high as CRF-Matching. This indicates the importance of the joint reasoning of CRFs. Overall, the most important component of the network is the pairwise distance feature $\mathbf{f}_{pair}$, which models the rigidity of the transformation. Leaving this feature out results in 60% increase in the association error. Removing the pairwise association features resulted in a relative increase of 30%.

For trajectory estimation, the benefits of CRF matching are more evident. The key property of weighting each association
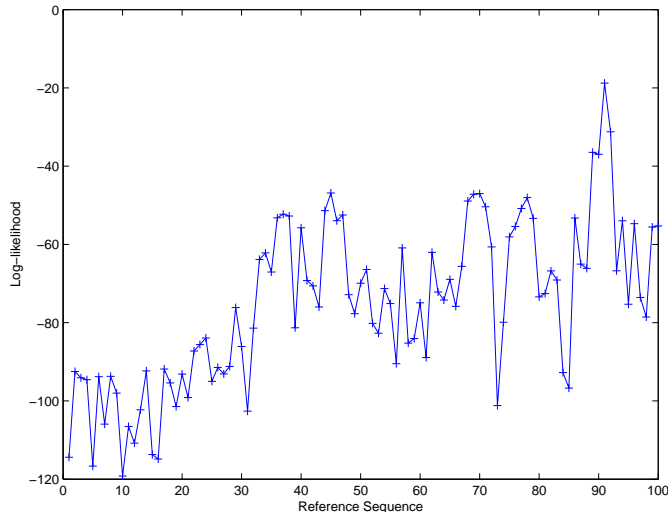
Fig. 5. Log-likelihood of the matching between a reference scene and scenes from a sequence used to detect loop closure. The log-likelihood was computed from the MAP association of CRF-Matching.

according to their matching probability significantly improves the performance. It also demonstrates that especially for rotation, ICP suffers a severe degradation for large offsets.

B. Loop Closure Detection with Laser and Camera Data

Loop closure detection is a key challenge for reliable simultaneous localization and map building (SLAM) [16], [15]. In this preliminary experiment, we show that CRF matching can be employed to detect loop closure in urban environments using both laser and camera. The data was acquired using an adapted car travelling at 30 km/h approximately. We use the approach described in [23] to compute the correct extrinsic calibration of these two sensors.

To detect the loop closure, a reference scene (camera and laser data) from the initial position of the vehicle is matched against a sequence of scenes. To compare different matches, the log-likelihood from the MAP association is computed. The maximum of this scoring function indicates the most probable location where the loop was closed. Figure 5 shows the log-likelihood for all scenes. The peak is clearly visible at scene 91 which indicates the best match for the reference scene. Figure 6 shows the reference image and image 91 where the loop closure was detected. It should be noted that even though the images look rather similar, the actual laser scans differ substantially between the two scenes (Figure 7).

VI. CONCLUSIONS AND FUTURE WORK

We introduced CRF-Matching, a novel approach to matching laser range scans. In contrast to existing scan matching techniques such as ICP, our approach takes shape information into account in order to improve the matching quality. Additionally, CRF-Matching explicitly reasons about outliers and performs a joint estimation over the data associations of all beams in a laser scan. By using a Conditional Random Field as the underlying graphical model, our approach is able to incorporate arbitrary features describing the shape and visual appearance of laser scans. The parameters of the model are learned efficiently from labeled training data. Maximum a posteriori data association computed via loopy belief propagation is used to determine the transformation between laser scans.

Our experiments show that CRF-Matching clearly outperforms ICP on scan matching tasks that involve large uncertainties in the relative locations of scans. Our approach is able to consistently match laser scans in real time without any information about their relative locations. Additional experiments using camera information indicate that the performance of our approach increases as richer sensor data becomes available.

We consider these results extremely encouraging. In future work we will investigate various extensions to the basic CRF-Matching described here. In addition to developing different feature functions, we plan to integrate CRF-Matching into an outdoor SLAM technique. Since CRF-Matching computes full distributions over data associations, it is possible to estimate the uncertainty of the resulting spatial transformations (for instance by sampling data associations). This uncertainty can be used directly by a SLAM approach.

Our current approach considers outliers as the only alternative to matching a measurement. However, one can extend CRF-Matching to additionally model different types of objects such as cars, buildings, trees, and bushes. Such a model could also perform improved data association for moving obstacles. The key advantage of our technique is the development of a framework for performing all these reasoning steps within a single model, in contrast to existing techniques.

We will also investigate the application of CRF-Matching to the more challenging problem of underwater mapping. Here, the observations are camera images annotated with 3D depth information provided by sonar. We hope that the ability of our approach to incorporate arbitrary features and to jointly reason about complete data associations will allow it to perform significantly better than current approaches. To make CRF-Matching applicable to 3D scans with many thousands of points, we intend to develop a hierarchical technique that automatically selects increasingly large subsets of points for matching. Such an approach could include classifiers that are trained not to associate but to select good points for matching. Finally, we will investigate the application of Virtual Evidence Boosting, which has recently been shown to provide superior feature selection capabilities for CRFs [10].

Fig. 6. Reference scene (left) and scene 91 (right) with projected laser point (gray/yellow crosses). CRF matching identifies this scene as the best match indicating the most probable position where the loop closure occurred. Note that the laser scans are substantially different.
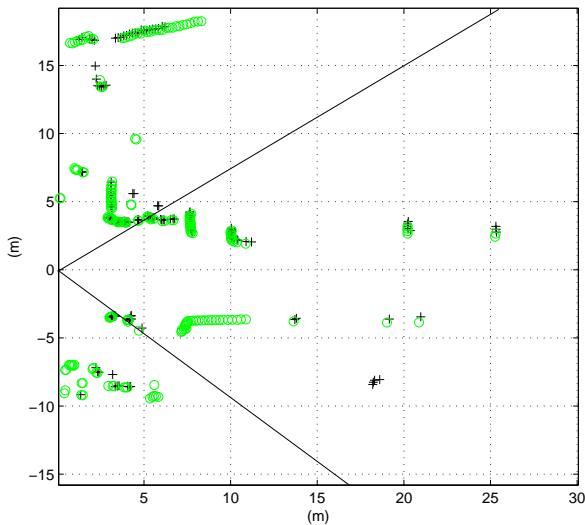

Fig. 7. Matched laser scans from Figure 6. The black crosses are points from the reference scene and the gray/green circles are points from scene 91. The dark lines indicate the camera field of view.

REFERENCES

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape completion and animation of people. *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, 24(3), 2005.

[2] O. Bengtsson and A. J. Baerveldt. Localization in changing environments - estimation of a covariance matrix for the IDC algorithm. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2001.

[3] A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[4] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24, 1975.

[5] P. J. Besl and McKay N. D. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(2):239–256, 1992.

[6] S. Friedman, D. Fox, and H. Pasula. Voronoi random fields: Extracting the topological structure of indoor environments via place labeling. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[7] S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proc. of the International Conference on Computer Vision (ICCV)*, 2003.

[8] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of the International Conference on Machine Learning (ICML)*, 2001.

[9] M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[10] L. Liao, T. Choudhury, D. Fox, and H. Kautz. Training conditional random fields using virtual evidence boosting. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[11] D. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Programming*, 45(3, (Ser. B)), 1989.

[12] F. Lu and E. Milios. Robot pose estimation in unknown environments by matching 2D range scans. *Journal of Intelligent and Robotic Systems*, 18, 1997.

[13] D. Metaxas. *Physics-Based Deformable Models*. Kluwer Academic, 1996.

[14] K. Murphy, Y. Weiss, and M. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 1999.

[15] P. Newman, D. Cole, and K. Ho. Outdoor SLAM using visual appearance and laser ranging. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, Orlando, USA, 2006.

[16] P. Newman and K. Ho. SLAM - loop closing with visually salient features. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, Barcelona, Spain, 2005.

[17] S. Pfister, K. L. Kriechbaum, S. I. Roumeliotis, and J. W. Burdick. Weighted range sensor matching algorithms for mobile robot displacement estimation. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.

[18] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*. IEEE, 1989. IEEE Log Number 8825949.

[19] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2), 2006.

[20] L. H. Staib and J. S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(11):1061–1075, 1992.

[21] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, Cambridge, MA, September 2005. ISBN 0-262-20162-3.

[22] R. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. Technical Report UU-CS-1999-27, Utrecht University, Netherlands, 1999.

[23] Q. Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sendai, Japan, 2004.