# Learning Temporal Co-Observability Relationships for Lifelong Robotic Mapping

Nicholas Carlevaris-Bianco and Ryan M. Eustice

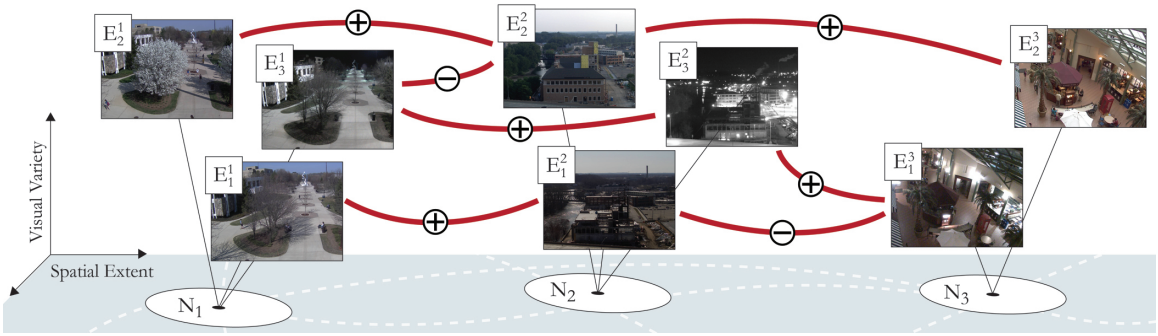



Fig. 1: Illustration of an "exemplar" based map representation. Locations are modeled as "metric" neighborhoods, $N_i$ for $i = [1 \dots n]$. The visual appearance of each neighborhood is represented by a set of exemplar views, $E_j^i$ for $j = [1 \dots m_i]$, which seek to represent the possible variation in appearance of a location due to temporal changes in the environment. **Our proposed method learns the temporal observability relationships between exemplars allowing one to predict which exemplars will be observed based on recent measurements.** The temporal observability relationships are illustrated as red links with the sign indicating the sign of the correlation.

***Abstract*—This paper reports on a method that learns the temporal co-observability relationships for exemplar views of a dynamic environment collected during long-term robotic mapping. These relationships are efficiently captured using a Chow-Liu tree approximation and allow one to predict which exemplars will be observed by the robot given the robot's recent observations. For example, these learned relationships can encode scene dependent changes in lighting due to time of day and weather, without explicitly modeling them. Preliminary experimental results are shown using images from 17 fixed locations collected hourly over the course of 116 days.**

## I. Introduction

Recent work has sought to address the challenge of robotic simultaneous localization and mapping (SLAM) in dynamic environments by using a map representation that encodes both spatial and temporal information. Promising examples of these methods elect to represent locations in the map as a collection of views [1], environmental states [2], or view sequences [3], corresponding to how the environment appeared at various observation instances. These "exemplar" based methods are capable of representing many different types of temporal variation. It is our opinion that these methods provide a promising way forward toward developing SLAM systems for long-term autonomy in dynamic environments.

Current methods do not attempt to understand the temporal relationship between map exemplars. Data association is performed by comparing the current sensor measurement against all possible exemplars to determine which best explains the current measurement. Essentially, the robot considers all possible temporal variations of a location every time it attempts to measure the environment. This results in a map that grows both with spatial extent and with the amount of dynamic variation in the environment.

In this paper we present a method that learns the temporal observability relationships between exemplars in the map. These relationships can be used to predict which exemplars will be observed given what the robot has recently observed. Conceptually this can be thought of as producing a maximum likelihood (ML) temporal map.

The learned relationships encode many different types of temporal changes, without explicitly modeling them. For example, the proposed method learns groupings of exemplars caused by changes in lighting between morning, noon, and night, even though the time of day is not a modeled input to the system. The structure of the correlation learned by the proposed method is determined by the properties of the sensor system. In the case of vision the correlation structure is primarily defined by changes in lighting. This would not be true in the case of light detection and ranging (LIDAR), which is invariant to changes in illumination and would be more affected by changes in geometric structure and occlusion.

Our motivation for this work is a centralized multi-agent system in which a common map is shared between many

This work was supported in part by the National Science Foundation under award IIS-0746455, and in part by the Naval Sea Systems Command (NAVSEA) through the Naval Engineering Education Center under award N65540-10-C-0003.

N. Carlevaris-Bianco is with the Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI 48109, USA carlevar@umich.edu

R. Eustice is with the Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA eustice@umich.edu

agents. We advocate in Section V that the proposed algorithm is particularly well suited for this scenario. The remainder of the paper is outlined as follows. In Section II we review related work. The proposed method is described in Section III. Finally, our experimental results are presented and discussed in Sections IV and V, respectively. Section VI concludes.

## II. Related Work

The majority of SLAM solutions, to date, make the assumption that the environment is static. For many robotic applications this assumption is not extremely detrimental. This is especially true for systems that exploit robust data association methods, notably image and laser range registration, as these methods are capable of successful measurements in the presence of short-term dynamic effects, such as partial scene occlusion and moving objects.

Several methods have been proposed that try to filter out dynamic elements of the environment while maintaining the assumption that the underlying environment is static, for example [4]–[7], among others. It has also been shown that it is possible to explicitly identify and track specific classes of dynamic objects in the environment such as people [8].

Recently, several works have proposed methods that explicitly model dynamic changes in the map. In [9], [10], Biber and Duckett represent the environment by a collection of sample-based maps, each of which incorporates new samples and forgets old samples at a different rate. This rate determines the timescale of each map. During localization the robot tests all timescales to determine which best agrees with the current observation and then performs measurement updates against that map. In some respects this representation is similar to the exemplar-based methods that this paper expands upon, as each map could be thought of as a type of exemplar. They are not exactly equivalent though, due to the fact that in this case each timescale-sampled map is a collection of data measured at multiple instances in time, and not a collection of a single-time-instance example views of the environment. The idea of sample-based maps is continued by Dayoub and Duckett in [11], which uses a short-term versus long-term memory model to update the collection of visual features that represent the appearance of a location. In [12] a method is presented which uses a "dynamic occupancy grid" based on a hidden Markov model to capture dynamic changes in the environment.

The works most relevant to our proposed method can be considered as "exemplar-based" methods [1]–[3]. In [2] Stachniss and Burgard propose a method to learn exemplar configurations of an indoor environment from 2D laser scan data using fuzzy k-means clustering. They then use these exemplar configurations in a particle-filter based localization framework. Konolige and Bowman present a vision-based method in [1]. This method works within the context of vision based pose-graph SLAM [13], where the pose-graph is divided into metric neighborhoods of views bounded by physical location and view attitude. Within these neighborhoods each view is an example of how the environment looked at the time it was collected. They then present a least-recently-used view deletion algorithm, which limits the number of exemplars per neighborhood to a fixed number and encourages a long-term equilibrium with the minimum set of exemplars that explains the visual variation of that neighborhood. In [3], Churchill and Newman use sequences of views, termed "experiences", as the basic unit of the temporal map instead of individual views. New experiences are added to the map when the existing experiences are unable to explain the current observations.

The map representation used in this paper, including the idea of exemplars and their use in metric neighborhoods, was first introduced in [1]. Our contribution beyond existing exemplar-based methods, [1]–[3], is an algorithm to learn the temporal observability relationships between exemplars throughout the map. This allows one to *predict which exemplar within a yet unseen neighborhood is most likely to be observed* given the most recent exemplar observations from other neighborhoods, which avoids having to compare the robot's current view against all exemplars at each neighborhood.

## III. Learning Temporal Observability Relationships Between Exemplars

The proposed algorithm uses a map representation that models each location as a "spatial" or "metric" neighborhood, $N_i$ for $i = [1 \dots n]$. Each neighborhood represents a physical space in the environment. The nature of this space is defined by the choice of sensing modality so that measurements within the neighborhood should overlap sufficiently for registration. For example, in the case of omnidirectional vision or 2D LIDAR a neighborhood may be defined as a max translation between any two views. For monocular vision, one would also need to include the vantage point of the camera. The appearance, geometric shape, or state of each neighborhood is represented by a set of "exemplar" views, $E_j^i$ for $j = [1 \dots m_i]$, which seek to represent the possible variation in appearance or structure of a location due to temporal changes in the environment, as illustrated in Fig. 1.

In order to learn the temporal observability relationships between map exemplars, and to predict the observability of each exemplar in the map, we wish to estimate the joint distribution of the probability that each exemplar will be observed during a given time window of length $\lambda$. Let $Z_k$ be a binary random variable representing if the $k^{th}$ exemplar was observed in a given time window. Therefore, we wish to estimate,

$$P(\mathbf{Z}) = P(Z_1, Z_2, \dots, Z_M), \tag{1}$$

where $M = \sum_{i=1}^{n} m_i$ is the total number of exemplars in the map.

As the robot builds the map over time, it records each of its attempts to localize against the exemplars in the map. This produces a vector of observations $\mathbf{z}$. We subdivide this set of all previous observations into two groups

$$\mathbf{z} = \mathbf{z}_p \cup \mathbf{z}_c \tag{2}$$

$$\mathbf{z}_p = [z_i : t_i \neq t] \tag{3}$$

$$\mathbf{z}_c = [z_i : t_i = t] \tag{4}$$

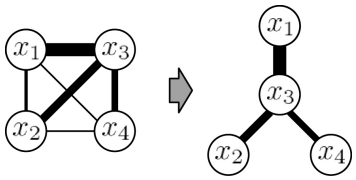



Fig. 3: Illustration of the Chow-Liu tree approximation. The magnitude of mutual information between variables is indicated by line thickness. The original distribution $P(x_1, x_2, x_3, x_4)$ is approximated as $P(x_1)P(x_3|x_1)P(x_2|x_3)P(x_4|x_3)$.

where $\mathbf{z}_p$ includes all previous measurements observed in a time window, $t_i$, other than the current, $t$. Conversely, $\mathbf{z}_c$ represents all current observations collected in the current time window. Our aim then is to learn the joint distribution, (1), from $\mathbf{z}_p$, and then, as the robot moves through the environment, infer the probability of observing each exemplar in the current time window given the current observations, $\mathbf{z}_c$.

The time window parameter, $\lambda$, can be thought of as a lower bound on the timescale of the map. Two exemplars, $E_a$ and $E_b$, are considered co-observed if they were observed within the same time window $t_i$ of length $\lambda$. Therefore, our method learns the temporal correlation only for events with a timescale greater than $\lambda$. In our experiments $\lambda$ is 1 hour, low enough to capture changes in lighting throughout the day, but long enough to exclude short-term dynamics such as people and moving vehicles.

Unfortunately, as the size of the map increases so will the number of exemplar views, and estimating the full distribution with all possible conditional dependencies quickly becomes intractable. Simply treating the probability of observing each exemplar as independent would solve the tractability problem. However, this would prevent any understanding of the temporal correlation in observability between exemplars. Without this correlation the most recent observations, $\mathbf{z}_c$, will not provide information about future observations in the same time window.

As a compromise between tractability and the ability to capture the temporal correlation between exemplar views, we have elected to approximate the joint distribution with a Chow-Liu tree [14]. The Chow-Liu tree approximates the full distribution as the product of pairwise conditional distributions. The pairwise conditional distributions are selected such that the Kullback-Leibler divergence between the original distribution and the Chow-Liu tree approximation is minimized. Therefore, we can approximate (1) as

$$P(\mathbf{Z}) \approx P(Z_r) \prod_{i=2}^{M} P(Z_i|Z_{p(i)}) \tag{5}$$

where $Z_r$ is the root variable of the Chow-Liu tree and $Z_{p(i)}$ is the parent of $Z_i$.

The Chow-Lui tree approximation is constructed by finding the mutual information between each pair of variables. This produces a fully connected graph where the weight of each edge is determined by the mutual information between the connected variables. From this graph the Chow-Liu tree is computed by finding the maximum spanning tree (Fig. 3). The mutual information between two variables is defined as

$$I(Z_i, Z_j) = \sum_{z_i \in 0,1} \sum_{z_j \in 0,1} P(z_i, z_j) \log \frac{P(z_i, z_j)}{P(z_i)P(z_j)}. \tag{6}$$

Therefore, the Chow-Liu approximation can be learned directly from the previous observation data, $\mathbf{z}_p$, by estimating each marginal distribution and each pairwise joint distribution.

Finding the maximum *a posteriori* (MAP) estimate of these quantities amounts to frequency counting based on the number of times in which each exemplar was observed, and the number of times in which each pair was successfully co-observed in the same time window. For the marginal distributions

$$\hat{P}(Z_i = z) = \frac{N_{[Z_i=z]} + (\alpha - 1)}{N_{[Z_i]} + (\alpha + \beta - 2)}, \tag{7}$$

where $N_{[Z_i=z]}$ is the number of observations where $Z_i = z$, $N_{[Z_i]}$ is the number of times $Z_i$ was observed either positively or negatively, and $\alpha$ and $\beta$ are parameters of a beta prior that encourages the estimate away from the extremes, $0.0$ and $1.0$, and toward an empirically determined target mode. In our experiments the beta prior's mode was set to $0.2$, with the constraint that $\alpha + \beta = 4$. The joint distributions were estimated as

$$\hat{P}(Z_i = z_i, Z_j = z_j) = \frac{N_{[Z_i=z_i \,\&\, Z_j=z_j]} + (\alpha - 1)}{N_{[Z_i \,\&\, Z_j]} + (\alpha + \beta - 2)}, \tag{8}$$

where $N_{[Z_i=z_i \,\&\, Z_j=z_j]}$ is the number of times $Z_i = z_i$ and $Z_j = z_j$ in the same time window, and $N_{[Z_i \,\&\, Z_j]}$ is the number of times both exemplars were co-observed in the same time window, regardless of value. In this case the beta prior parameters are automatically chosen so that the mode is equal to $\hat{P}(Z_i = z_i)\hat{P}(Z_j = z_j)$, with the constraint that $\alpha+\beta = 4$. This essentially assumes independence until there is sufficient data to indicate otherwise. Similarly, the conditional distributions that form the Chow-Liu tree were estimated as

$$\hat{P}(Z_i = z_i|Z_j = z_j) = \frac{N_{[Z_i=z_i \,\&\, Z_j=z_j]} + (\alpha - 1)}{N_{[Z_i \,\&\, Z_j=z_j]} + (\alpha + \beta - 2)}. \tag{9}$$

Again the beta prior parameters are automatically chosen to assume independence, with the mode of the prior equal to $\hat{P}(Z_i = z_i)$ with the constraint that $\alpha + \beta = 4$.

It is important to note that the robot may not encounter all neighborhoods during each time window. Therefore, for the exemplars within an unvisited neighborhood we have no evidence of their observability. It is important not to count this absence of observation as a negative observation. Finally, the robot may never have attempted to co-observe some pairs in the same time window. In this case we again assume their independence with $\hat{P}(z_i, z_j) = \hat{P}(z_i)\hat{P}(z_j)$.

When building the Chow-Liu tree we enforce an additional constraint that no edges should be included between exemplars within the same neighborhood. This forces the tree to only include dependencies between neighborhoods. Our focus is to make predictions between neighborhoods, not within them.


Fig. 4: Depiction of webcam locations at GVSU.

Having learned an approximation of the underlining observability distribution using observations from all previous time windows, $\mathbf{z}_p$, we now wish to predict the probability of observing each exemplar given the previous observations made during the current time window, $\mathbf{z}_c$. We can use the sum-product algorithm [15], [16] to efficiently perform inference because the joint distribution (1) has been approximated using a tree structure (5).

## IV. Experimental Results

### A. Experimental Setup

In order to evaluate the proposed algorithm we performed experiments using data collected at 17 fixed locations over the course of 116 days (Fig. 4). Data was collected approximately hourly from March 9th, 2012 through July 3rd, 2012 by automatically downloading images from web-accessible security cameras at Grand Valley State University's campus [17]. The data includes both indoor and outdoor scenes under dynamic changes including lighting, weather, people, cars, moving furniture, seasonal changes and construction. Additionally, for each set of images the time of day and weather conditions [18] were automatically recorded for later comparison. This time and weather meta data was not in any way used by the algorithm, and was only used after-the-fact as ground-truth to examine what types of relationships were being learned by the algorithm.

Considering each of the 17 locations as a metric neighborhood, approximately 1000 simulated robot trajectories were generated by visiting each neighborhood in a random order. The time between each synthetic trajectory was also randomized uniformly between 1 and 5 hours. In this experiment we assume that the neighborhood for each view is known. For the experiment the map was built as follows. The first trajectory was used to initialize a set of metric neighborhoods, each with a single visual exemplar view. Each following trajectory was then processed according to Algorithm 1.

Prior to processing each trajectory, we calculate the Chow-Liu approximation [14] of the joint distribution of exemplar observations, $P(\mathbf{Z})$. We then consider each view in the trajectory. At each neighborhood, prior to performing matching, the probability of observing each of the neighborhood's exemplars is estimated given the observations made so far in this trajectory, $P(\mathbf{Z}|\mathbf{z}_c)$. We then perform visual matching between the current view and each of the neighborhood's exemplars in order to compare the results with the prediction and to add more observations to $\mathbf{z}_c$.

**Algorithm 1** Experiment Map Update Algorithm

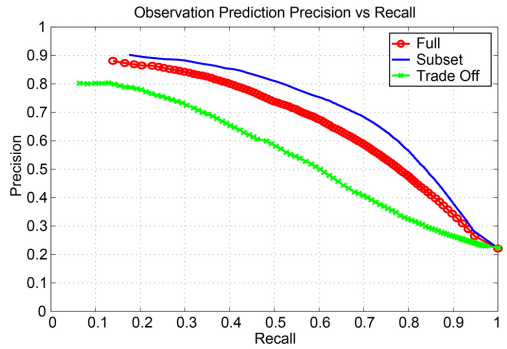
Given current map, observation history $\mathbf{z}_p$, and new trajectory $T$ consisting of a set of views $V_i$ for $1 = [1 \dots v]$
*// Localization Loop*
Calculate Chow-Liu approximation of $P(\mathbf{Z})$ using $\mathbf{z}_p$
**for all** Views, $V_i$, in $T$ **do**
  Evaluate $P(\mathbf{Z}|\mathbf{z}_c)$ using Sum-Product
  Predict matching results using $P(\mathbf{Z}|\mathbf{z}_c)$
  Visual matching between $V_i$ and exemplars in $N_i$
  Update $\mathbf{z}_c$ with matching results
**end for**
*// Map Update Loop*
**for all** Views, $V_i$, in $T$ **do**
  **if** no match between $V_i$ and exemplars in $N_i$ **then**
    **if** $m_i == m_{\max}$ **then**
      Forget least-recently-used exemplar, $m_i = m_i - 1$
    **end if**
    Add $V_i$ as new exemplar in $N_i$, $m_i = m_i + 1$
  **end if**
**end for**

After processing each trajectory, the neighborhoods in the map are updated using a least-recently-used update rule, similar in concept but simplified with respect to that proposed in [1]. In each neighborhood, if the new view does not match against any of the existing exemplars, then the new view is added as a new exemplar. If the number of exemplars in a neighborhood exceeds a threshold (in our experiments $m_{\max} = 9$), the exemplar that has been least recently matched against is removed.

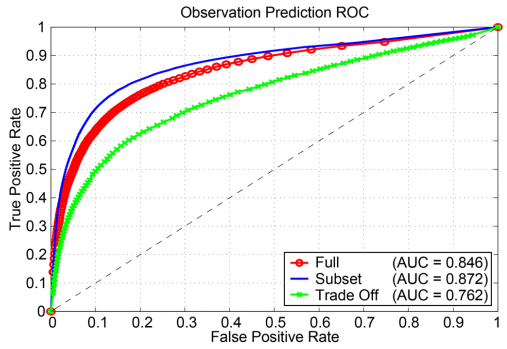
For this experiment a match occurs if the number of inlier features is greater than 20, and the mean squared reprojection error is less than 10 pixels. Note that it is possible for the current view to match more than one exemplar in a neighborhood. The exact definition of a match depends strongly on the sensing modality and SLAM front-end implementation used. Matching should be defined by the same criteria used in determining which constraints are added to the SLAM solution and in a full implementation may include additional criteria such as consistency with SLAM-derived motion priors.

### B. Prediction Accuracy

Precision-recall and receiver operating characteristic (ROC) curves for the predicted observability in this experiment are shown in Fig. 5. These curves are generated by varying the probability threshold used to predict if an exemplar will be observable. Ground-truth is known as matching was performed

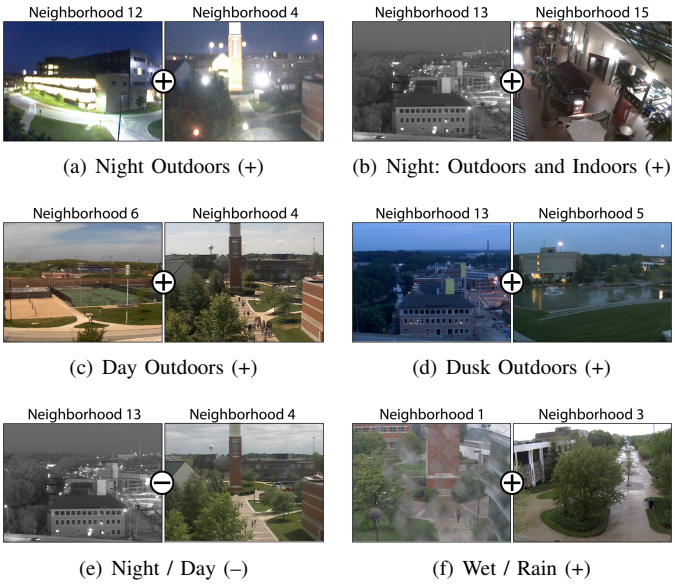



(a) Precision-Recall Curve




(b) ROC Curve

Fig. 5: Precision-recall curve and ROC curve for observability prediction. Results for all predictions made throughout the experiments are shown in red and labeled "Full." Results that exclude predictions made on newly added exemplars (less than 2 days old), and predictions made with a small number of current observations (less than 4 neighborhoods observed), are shown in blue and labeled "Subset." Results that consider all predictions, but only use training data from the top two most likely and two randomly chosen "test" exemplars, are shown in green and labeled "Trade Off." AUC indicates the area under the curve.




(a) Night Outdoors (+)

(b) Night: Outdoors and Indoors (+)

(c) Day Outdoors (+)

(d) Dusk Outdoors (+)

(e) Night / Day (–)

(f) Wet / Rain (+)

Fig. 6: Sample exemplar pairs with high positive or negative correlation that were included in the Chow-Liu tree approximation. Sign between pairs indicates sign of correlation. These pairs are representative of the types of relationships automatically found in the Chow-Liu approximation.

after the prediction to verify the results. It is interesting to note that there are two situations that occur during the experiment where one would not expect the algorithm to perform well. The first occurs when a new exemplar is added to the map. The estimated probabilities for new exemplars with few observations are based largely on beta priors that assume independence from other exemplars. Therefore, we do not expect the algorithm to accurately predict the observability of exemplars until sufficient observations have been made to estimate the inter-exemplar correlation. The second occurs at the start of each trajectory where, before any current observations have been made, the predictions are based on the learned probabilities of observation for each exemplar.

For comparison in Fig. 5, curves are shown for all predictions made throughout the experiment, shown in red and labeled "Full," and for a subset that excludes predictions made under the aforementioned conditions where less observation data is available, shown in blue and labeled "Subset." As one would expect, excluding the predictions made with few observations improves performance. More interestingly, this illustrates the interaction between the proposed observation prediction algorithm and the exemplar updating scheme. Exemplar updating schemes that quickly settle to a stable set of exemplars per neighborhood are ideal, while update rules that continuously add new and forget old exemplars will be detrimental.

It is important to note that there is an inherent trade-off between reducing the data association effort and producing more data to learn the Chow-Liu approximation. In the "Full" and "Subset" experiments, visual matching was performed between the current view and each exemplar in the neighborhood regardless of the predicted probability of observation. All of these observations were then provided to subsequent predictions. Therefore, there was no gain in terms of the effort required for data association. To investigate this trade-off, an experiment was performed in which the match data from only the top two most likely and two randomly chosen "test" exemplars were used to train the Chow-Liu tree and to perform inference. These results are shown in green and labeled "Trade Off" in Fig. 5. As only a maximum of four exemplars are tested per neighborhood, this represents a data association effort reduction of greater than $50\%$ at steady-state (when each neighborhood has reached its maximum of nine exemplars).

### C. Learned Temporal Relationships

In order to visualize the type of co-observability relationships being learned by the Chow-Liu approximation several representative examples of highly correlated exemplar pairs (from different neighborhoods) that were learned by the Chow-Liu tree approximation are depicted in Fig. 6. It appears

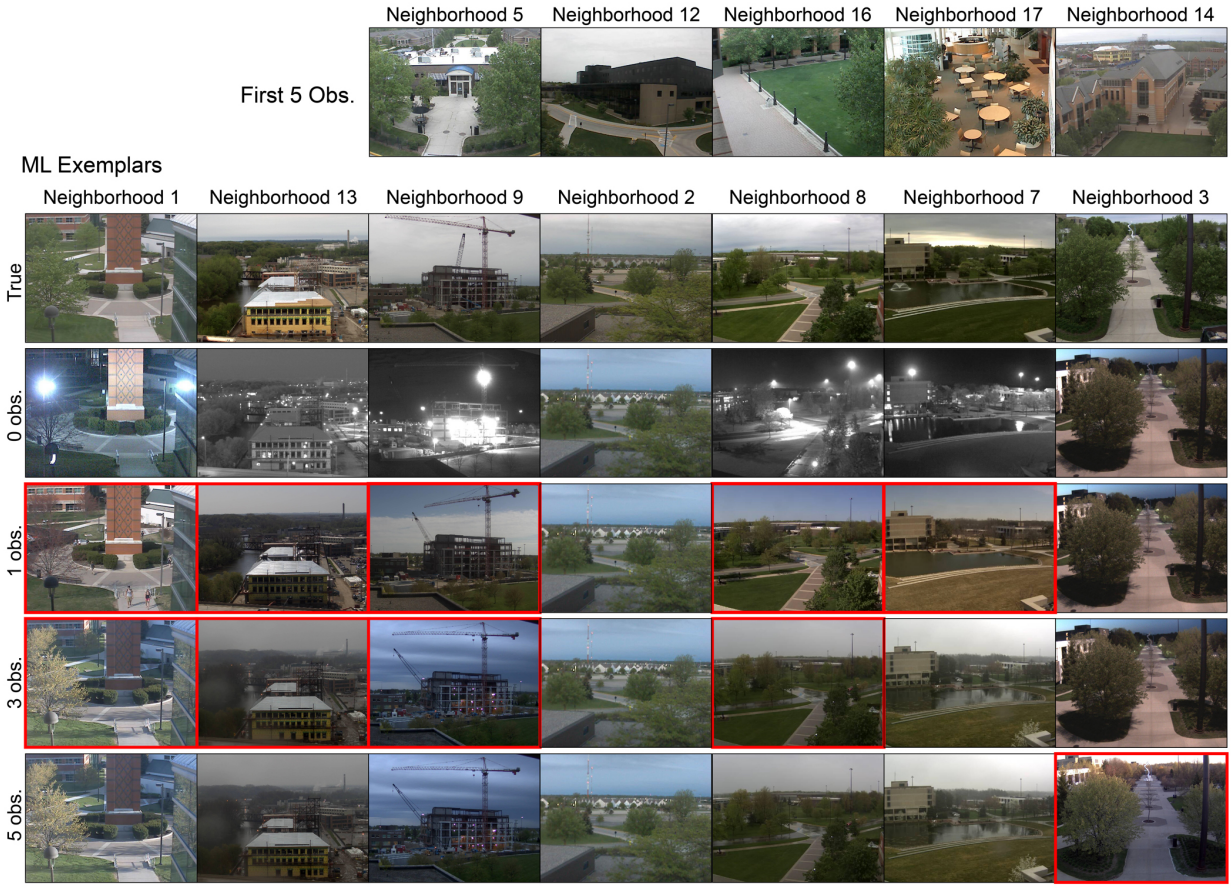
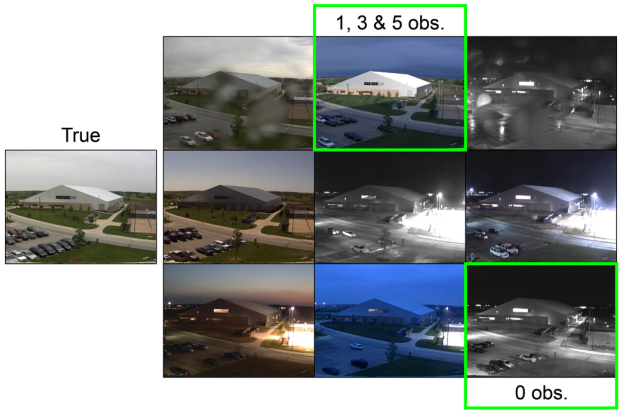
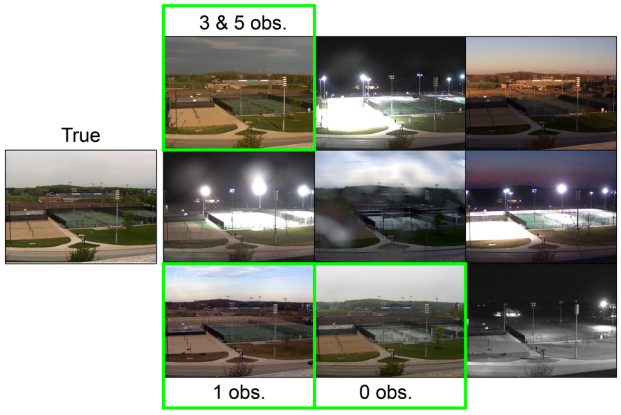



Fig. 7: Sample imagery for ML exemplars for 7 not-yet-observed neighborhoods (1,2,3,7,8,9,13) as other neighborhoods in the map are observed (5,12,14,16,17). The upper right shows the first 5 observations (in order from left to right) made during this trajectory. In the grid below, the first row shows the true realization of each of the neighborhoods, which have yet to be observed. The remaining rows show the maximum likelihood exemplars after the first 0, 1, 3, and 5 neighborhoods are observed (0 being the prior). Red boxes highlight exemplars that have changed given the new observations for that row.




(a) Neighborhood 11




(b) Neighborhood 6

Fig. 8: Sample imagery showing the nine exemplars from two not-yet-observed neighborhoods. To the left of the exemplars the true realization of each of the neighborhoods is shown. The ML exemplars after 0, 1, 2, and 5 other neighborhoods are observed (0 being the prior) are highlighted with green boxes.

that the structure of Chow-Liu tree is driven primarily by lighting variation, with weather and sky conditions playing a much smaller role. Many of the strongest positive correlations were found between night exemplars, most likely due to the consistent artificial lighting in these scenes and their complete visual dissimilarity with daytime exemplars. Almost all cases of negative correlation were between day and night exemplars.

The imagery in Fig. 7 and in Fig. 8 shows examples of how the ML exemplars for several sample neighborhoods change as new observations are made. With no observations the ML exemplars are based on the learned prior probability of observation and in this example contain a mixture of night and day exemplars. After observing the first neighborhood, the night exemplars are replaced with daytime exemplars. As more neighborhoods are observed, the ML exemplars change more subtly toward less sunny exemplars, which better match the true realization.

For illustration purposes, Fig. 7 shows the ML exemplars for multiple neighborhoods at the same time. However, during the actual experiment the ML exemplar for each neighborhood is only calculated immediately before attempting to match exemplars in that neighborhood, and the liklihood is maximized only locally over the exemplars in that neighborhood. To truly estimate a maximum likelihood map over multiple neighborhoods, it would be more appropriate to use the max-product algorithm [16] instead of the sum-product to find the set of exemplars that *jointly* maximize the likelihood.

The temporal changes learned in the Chow-Liu approximation are further illustrated in Fig. 9. By considering all observations that meet a specific time-of-day condition, we can calculate the ratios of the time-of-day conditions of the successfully matched exemplars for these observations. As one would expect, we see in Fig. 9(a) that visual matching is dependent on lighting conditions, which change throughout the day. Observations are more likely to match against exemplars collected under similar time-of-day conditions. At no time does the proposed algorithm use any external information about the time of day, yet it effects visual matching; we would like the algorithm to propose appropriate exemplars, following the relationships in Fig. 9(a). In Fig. 9(b) we see that this is the case. The ratios of conditions of the predicted ML exemplars closely match those in Fig. 9(a) indicating that the algorithm has learned the underlying co-observability relationships caused by changes in lighting. For observations made during the morning, midday, afternoon and night, our algorithm learns to propose exemplars from the appropriate time of day.

Similarly, in Fig. 9(c) we see that the current sky condition also effects visual matching, though to a lesser extent than the time-of-day. Fig. 9(d) indicates our algorithm learns to predict exemplars according to the underlying co-observability relationships caused by changes between clear and overcast skies.

Again it is important to note, that the algorithm does not use any external knowledge of lighting or weather conditions, and that correlation between the observability of exemplars could be developed by many different types of dynamic changes that affect visual matching.

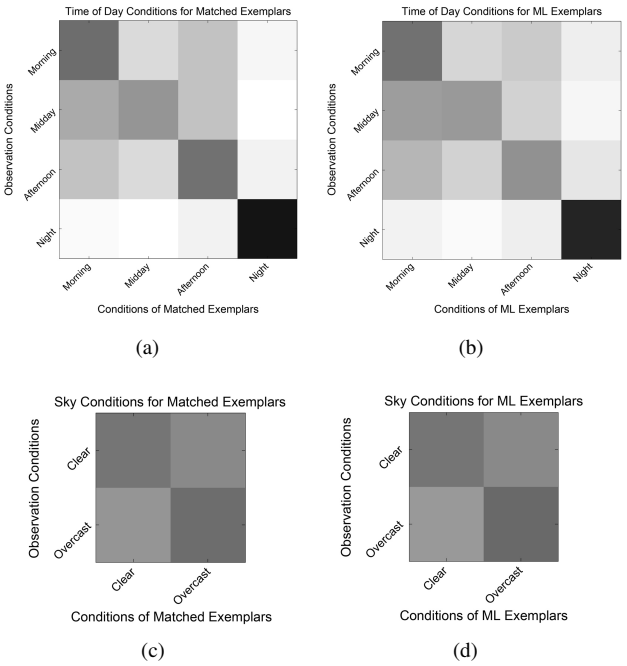



Fig. 9: Comparison between the conditions of the matched exemplars and the ML exemplars for a given observation condition. In (a) and (b) each row corresponds to the set of observations that occurred at a specific time of day. The shade of each cell represents the ratio of exemplars from that time of day, where white is 0.0 and black is 1.0. Similarly, in (c) and (d) exemplars collected under clear and overcast sky conditions are compared.

## V. Discussion and Future Work

There are several important aspects to consider with respect to the applicability and utility of the proposed algorithm:

- There is an inherent trade-off between the data association effort and producing more data to learn the Chow-Liu approximation. The more attempts made to match against exemplars, the more accurate the predictions become. However, the fewer the matching attempts, the larger the reduction in data association effort.
- Once built, performing inference using the Chow-Liu approximation is very efficient due to its tree structure. Unfortunately, building the Chow-Liu approximation itself is computationally expensive (i.e., $\mathcal{O}(M^2 \log M)$ where $M$ is the number of exemplars).
- One might argue that using data association methods based on fast bag-of-words (BoW) place recognition techniques [19]–[21], as demonstrated in [1], reduces the cost of producing data association hypotheses sufficiently such that it is not computationally necessary to predict which exemplars will be observed. Instead, one can simply consider all exemplars at every data association attempt.

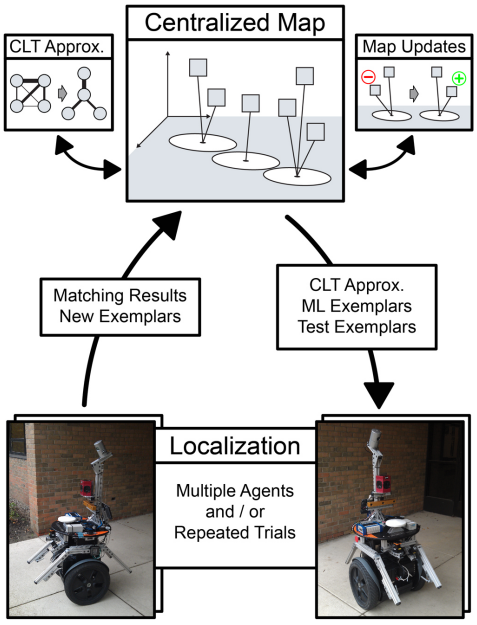



Fig. 10: Illustration of a centralized multi-agent system. In this scenario a common map is shared between many agents. The centralized map provides the current Chow-Liu approximation, ML exemplars, and possibly additional test exemplars to the localizing agents. The localizing agents return match results and possible new exemplars. Updating the Chow-Liu approximation and performing map updates are carried out as offline batch processes.

Based on these concerns we understand that the present algorithm may not seem practical for online, single-agent, map building. It is our belief, though, that if we consider a centralized multi-agent scenario, where a common map is shared between many agents (Fig. 10), then the proposed algorithm could provide substantial benefits:

- By sharing a map between multiple agents the burden of testing exemplar observability can be spread out among all agents while improving the reduction in data association effort.
- Building the Chow-Liu approximation can be done as a batch offline process and the result transferred to all agents where it can be used for efficient inference.
- Being able to predict the ML exemplars allows only a small subset of exemplars to be transferred between the centralized map and the individual agents.

Much work remains to fully demonstrate the effectiveness of the proposed algorithm. Most importantly, we intend to apply it to a large-scale real-world robotic data set. Additionally, exploring methods to balance the trade-off between data association effort and training data production, such as actively selecting test exemplars, and exploring the interaction between exemplar update rules and the ability to predict exemplar observability, could also improve the algorithm further.

## VI. Conclusions

We presented a method that learns the temporal observability relationships between views of a dynamic environment. We have demonstrated that these relationships can be used to predict which views of the environment will be observed by the robot given what the robot has recently observed. We have demonstrated that the learned relationships can encode many different types of temporal changes, without explicitly modeling them, including changes in lighting due to time of day and weather.

## References


[1] K. Konolige and J. Bowman, "Towards lifelong visual maps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2009, pp. 1156–1163.

[2] C. Stachniss and W. Burgard, "Mobile robot mapping and localization in non-static environments," in *Proc. AAAI Nat. Conf. Artif. Intell.*, 2005, pp. 1324–1329.

[3] W. Churchill and P. Newman, "Practice makes perfect? managing and leveraging visual experiences for lifelong navigation," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2012, pp. 4525–4532.

[4] D. Fox, W. Burgard, and S. Thrun, "Markov localization for mobile robots in dynamic environments," *J. Artif. Intell. Res.*, vol. 11, pp. 391–427, Nov. 1999.

[5] D. Hähnel, R. Triebel, W. Burgard, and S. Thrun, "Map building with mobile robots in dynamic environments," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2003, pp. 1557–1563.

[6] C.-C. Wang, C. Thorpe, and A. Suppe, "LADAR-based detection and tracking of moving objects from a ground vehicle at high speeds," in *IEEE Intelligent Vehicles Symp.*, June 2003, pp. 416–421.

[7] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "Experiences with an interactive museum tour-guide robot," *Artificial Intelligence*, vol. 114, pp. 3–55, 2003.

[8] M. Montemerlo, S. Thrun, and W. Whittake, "Conditional particle filters for simultaneous mobile robot localization and people-tracking," in *Proc. IEEE Int. Conf. Robot. and Automation*, 2002, pp. 695–701.

[9] P. Biber and T. Duckett, "Dynamic maps for long-term operation of mobile service robots," in *Proc. Robot.: Sci. & Syst. Conf.*, 2005, pp. 17–24.

[10] P. Biber and T. Ducket, "Experimental analysis of sample-based maps for long-term SLAM," *Int. J. Robot. Res.*, vol. 28, no. 1, pp. 20–33, 2009.

[11] F. Dayoub and T. Duckett, "An adaptive appearance-based map for long-term topological localization of mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2008, pp. 3364–3369.

[12] G. D. Tipaldi, D. Meyer-Delius, M. Beinhofer, and W. Burgard, "Lifelong localization and dynamic map estimation in changing environments," in *RSS Workshop on Robots in Clutter*, 2012.

[13] K. Konolige, J. Bowman, J. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-based maps," *Int. J. Robot. Res.*, vol. 29, no. 8, pp. 941–957, 2010.

[14] C. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Trans. on Info. Theory*, vol. 14, pp. 462–467, 1968.

[15] J. Pearl, "Reverend Bayes on inference engines: A distributed hierarchical approach," in *Proc. AAAI Nat. Conf. Artif. Intell.*, 1982, pp. 133–136.

[16] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Info. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[17] Grand Valley State University Webcams. Grand Valley State University. Allendale, MI, USA. Accessed March 9, 2012 through July 3, 2012. [Online]. Available: http://www.gvsu.edu/webcams.htm

[18] Weather Undergound API. Weather Undergound. Ann Arbor, MI, USA. Accessed March 9, 2012 through July 3, 2012. [Online]. Available: http://www.wunderground.com/weather/api/

[19] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Oct. 2003, pp. 1470–1477.

[20] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, 2006, pp. 2161–2168.

[21] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Int. J. Robot. Res.*, vol. 27, no. 6, pp. 647–665, Jun. 2008.