Implicit Data Association from Spectrally Clustered Local Matches

Edwin Olson

Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

Email: eolson@mit.edu

Abstract— Place recognition is a fundamental but challenging perceptual problem. One common application is robot mapping: recognizing places allows the trajectory of a robot to be over-constrained, allowing better maps to be produced. Similar looking (but physically distinct) environments can result in false positives, a problem exacerbated by the large positional uncertainties that can accumulate from dead-reckoning. A classic approach is based on associating observations with specific landmarks, i.e., explicit global data association. Good results, however, typically require jointly assigning several observations, leading to computationally expensive algorithms.

In this paper, we describe an alternative approach that is based on *local* matching. These local matches are derived from a set of easily computed pose-to-pose matches. Outliers and ambiguous cases are rejected using SCGP, a spectral clustering algorithm. We argue that when the spatial extent of a locally unambiguous match is large in comparison to the robot's positional uncertainty, the match becomes a reliable place recognition. We illustrate our method on a real dataset with significant perceptual ambiguity, evaluating the robustness of the method using ground-truth data.

I. Introduction

Place recognition is a key problem for many mobile robot applications. It plays a central role in map building: each place recognition gives information about the robot's trajectory, allowing optimization algorithms to compute better maps.

Many map building systems recognize places by recognizing landmarks. As the robot makes new observations of landmarks, the robot must associate the observation data with a particular landmark. This is known as *data association*. When landmarks are highly-distinctive (such as wireless networking base stations that broadcast a globally unique id [**?**]), data association is straight-forward. More commonly, however, observations of different landmarks can be similar. To a laser scanner, for example, all the corners in a room look about the same.

Unfortunately, data association errors can be catastrophic: even a single error can create a "wormhole", leading to a poor map. The reliability of data association is commonly increased by jointly considering an ensemble of observations: a "constellation" of observations is significantly less ambiguous than a single observation in isolation. While reliability increases with the size of the constellation, so does the computational cost. Given N observations, each of which could be any of M different landmarks, there are M^N possible joint data associations to consider. Even with efficient search methods
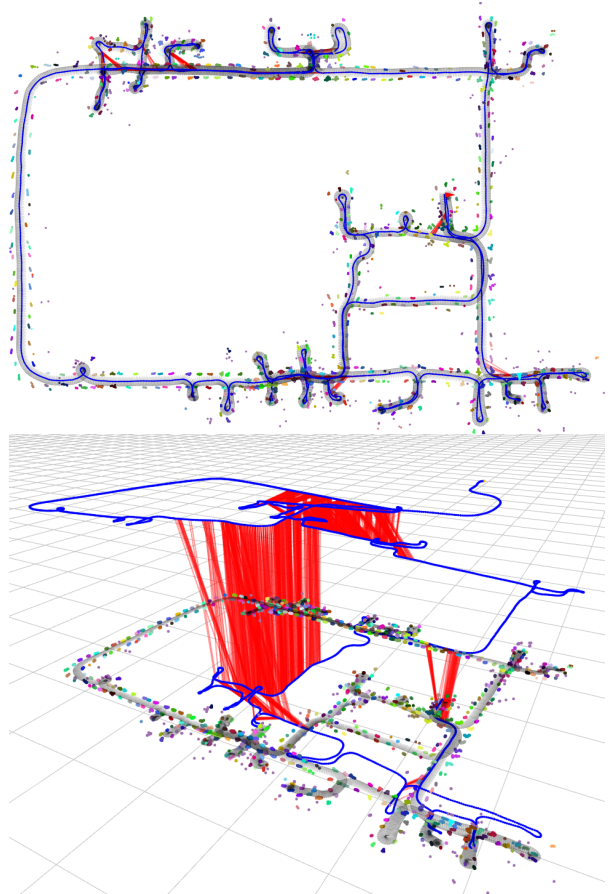

Fig. 1. Posterior Map. Once loop closures have been obtained using our method, a posterior map can be easily computed. Top: an overhead view of the graph. Bottom: the trajectory is plotted with height corresponding to time. Each red line represents a loop closure; the high density of loop closures significantly over-constrains the robot trajectory, allowing a good posterior to be computed.

like JCBB [**?**], the cost quickly becomes prohibitive. Further, these approaches generally require a full covariance matrix to be maintained so that the probability of joint assignments can be computed. In large maps, this can be expensive both in terms of memory and computational requirements.

In this paper, we describe an alternative way of recognizing places that avoids performing explicit global data association and thus avoids the challenges associated with it. Our method

finds parts of the environment that *locally* resemble each other. These local matches are inexpensive to compute in comparison to global data association. Pose-to-pose matches are a special case of local matches; in this paper, we will build spatially-extended local matches by combining sets of pose-to-pose matches.

Local matching methods are susceptible to false positives due to the fact that different places can have similar appearances. The main contribution of this paper is a principled means of determining whether a local match is globally correct. Our approach incorporates outlier and ambiguity testing based on Single Cluster Graph Partitioning (SCGP) [**?**]. Like CCDA [**?**], SCGP uses a pair-wise consistency graph, but is algorithmically more similar to other spectral methods [**?**], [**?**].

Our approach does not explicitly associate observations to globally-known features. Instead, the local matches encode correspondences from which the global identity of features can be determined. Consequently, in the context of data association, our method is best described as an implicit method. In general, however, the data associations are not of direct interest: it is the posterior map that is desired.

We are not the first to attempt to circumvent the difficult data-association problem. Cummins and Newman describe a scheme that is based solely on local sensor data [**?**] as do Ho and Newman [**?**]. In the absence of *some* knowledge of position, however, it is difficult to determine what constitutes absolute evidence of being in a particular place. In this paper, we argue that the amount of evidence required to reliably determine that two places are identical should scale with the robot's positional uncertainty.

Our approach is similar in several ways to that of Gutmann and Konolige [**?**], in that we attempt to find local matches within a search are provided by a prior and combine multiple sensor scans to yield larger (and less ambiguous) local matches. Rather than correlating matches over relatively large submaps, we show how ambiguity can be detected from an ensemble of smaller pose-to-pose matches. We also address the issue of how large a local match must be in order for a loop closure to be reliable.

We demonstrate our place recognition method on the DLR "circles" dataset [**?**]. In this dataset, sensor observations consist solely of sparse (and indistinguishable) point features (see Fig. 2). The limited sensing scheme makes false matches very common: it is easy to find two or three circles that seem to "line up". Even in the presence of this extreme environmental ambiguity, our method is able to reject incorrect local matches. The dataset also provides ground-truth data associations, making it possible to quantitatively evaluate the reliability of our local matching system.

This paper begins with a high-level overview of our approach (Section II). Next, we describe our pose-to-pose matching algorithm, which uses a RANSAC variant incorporating negative information (Section III). These pose-to-pose matches are grouped into local matches, allowing the rejection of outliers and the detection of multiple ambiguous solutions (Section IV). We subject these local matches to a geometrical
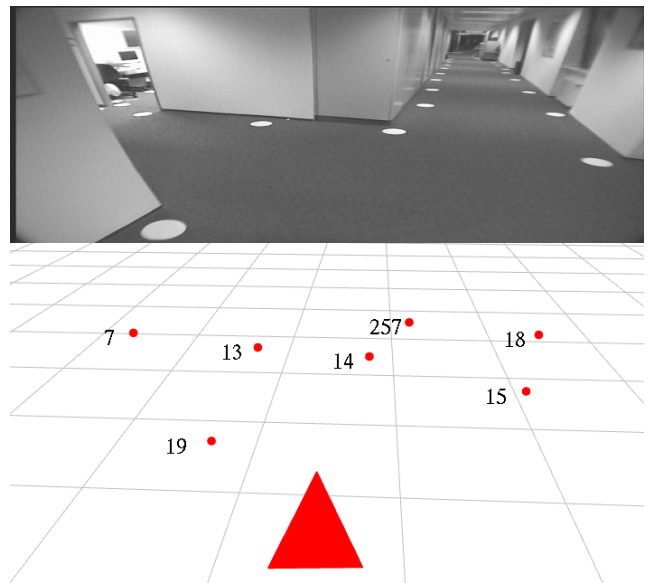

Fig. 2. DLR "circles" example data. White circles were manually distributed in an office environment (top). These circles were automatically extracted using computer vision methods (bottom). The resulting dataset was annotated with the ground-truth identity of each detected circle [**?**]. Note that ground-truth *positions* were not available.

test that ensures that the matches are large enough to resolve the positional uncertainty of the robot (Section V). Finally, Section VI presents results of our method applied to the DLR circles dataset.

II. METHOD OVERVIEW

Suppose that a robot has constructed two local maps of its environment (A and B) at different points along its trajectory, and suppose that these two maps are locally consistent (i.e., they match). We wish to determine whether A and B represent the same location, or whether they are two physically distinct (but similar-looking) places.

We begin by assuming that we have some prior knowledge of the relative position of the two environments, as would be provided by a simultaneous mapping and localization (SLAM) algorithm. This allows us to compute a bounding ellipse relative to area B in which area A *must* be found. Naturally, if B is not contained within this ellipse, B cannot be the same location as A.

The more interesting case is when local map B *is* within the ellipse that contains local map A. The main idea of this paper is that a local match is globally consistent if two conditions are satisfied. The first condition requires that the local match cover a spatial area that is comparable in size to the uncertainty ellipse. If the uncertainty ellipse is large enough to contain two or more identical-looking regions that might match area A, then we cannot be sure that area A is the same as area B. Conversely, if the uncertainty ellipse is only large enough to hold one "copy" of the locally matched region, then A and B must be the same location, since two distinct regions of that size could not both exist within the uncertainty ellipse.

In order for the above argument to hold, a second (more subtle) condition must be satisfied. Suppose, for example, that the area being matched is self-similar: it is then possible for multiple *overlapping* matches to exist within the uncertainty ellipse. A simple example would be an environment containing a picket fence: many overlapping local matches are possible due to the repeated fence posts, and each of these local matches can be large. In other words, the local match must be *locally unambiguous*: within the area of the local match, there must not be other conflicting matches of similar size.

III. POSE-TO-POSE MATCHING

When a new set of observations arrive, we begin by identifying all the previous poses that might have overlapping sensor data. Rather than maintain a covariance matrix over all poses, we use a Dijkstra projection [**?**], [**?**] which computes the minimum uncertainty path from the robot's current position to all earlier nodes. These projections, along with conservative upper bounds on their uncertainty, can be rapidly computed. Since a covariance matrix is not needed by our method, applications are free to use highly-efficient non-linear SLAM methods [**?**], [**?**] that do not estimate it.

For each earlier node that could plausibly have overlapping sensor data (using a Mahalanobis distance threshold of 3 and a nominal sensor range of 4m), we attempt to compute a local match. Let us denote the two robot poses being matched as a and b.

The pose-to-pose is generated via RANSAC: two points are randomly selected from both a and b, and using Horn's algorithm [**?**], a rigid-body transformation is computed that optimally aligns those points. For each point in a, we compute the distance to the closest point in b, and vice versa. Let the minimum distance for point a_i be $\mathrm{dist}(a_i)$. We also incorporated a negative information penalty P, described below. The probabilistically motivated consensus score S is then:

$$S = \sum_{i \in a} e^{-\beta \mathrm{dist}(a_i)^2} + \sum_{j \in b} e^{-\beta \mathrm{dist}(b_j)^2} - P \qquad (1)$$

The parameter β was set to 10.0 in our experiments. We also reject any rigid-body transformations that are more than three Mahalanobis distances away from the Dijkstra-projection prior. The rigid-body transformation achieving the highest consensus score S becomes a pose-to-pose match *hypothesis*, and is subject to further filtering as subsequent sections will describe. We call them hypotheses in order to emphasize the fact that they may be incorrect.

When there is environmental ambiguity (like a picket fence), the pose-to-pose matching will generate many different and incompatible solutions. This is due to the fact that each pose-to-pose match is independently computed by a greedy local optimization. The ambiguity of the environment is implicitly encoded in the dissonances of the pose-to-pose matches, and is detected in Section IV.

A. Negative Information

The data in the circles dataset is highly ambiguous: individual features are indistinguishable and it is often possible to find erroneous rigid-body transformation that align two or more points reasonably well. However, it is rare to *fail* to detect a feature if one is present. Consequently, we estimate which landmarks should be visible and penalize alignments that result in missing features.

To do this, we must first estimate the size and shape of the overlapping views from the two poses. Of course, we would need to know the relative positions of the two poses in order to compute this exactly, and this is information is unknown.

The RANSAC method computes scores assuming that two pairs of features match. In other words, if a given RANSAC iteration is correct, the sensor range overlap must include at least the points upon which the alignment is conditioned. Thus, we model the overlapping sensor range as the smallest circle that includes both points. Each unmatched observations within these circles incurs a penalty of 1.0 (equivalent to the score resulting from a single *good* match.) In our experiments, the use of negative information significantly improved hypothesis generation.

IV. LOCAL MATCHING

Pose-to-pose match hypotheses are grouped together to form hypothesis sets such that each set contains topologically similar hypotheses, as described in [**?**]. In short, the hypotheses in a single set relate two small pieces of the trajectory— i.e., a single loop closure. From this hypothesis set, we will construct a single local match. By combining multiple pose-to-pose matches into a single local match, we increase the spatial extent of the match: this is necessary in order to resolve large positional uncertainties.
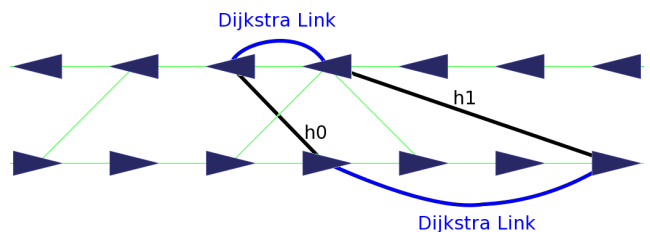

Fig. 4. Pair-wise hypothesis test. A loop of rigid-body constraints can be formed given two hypotheses and two short segments from dead-reckoning. If the hypotheses are consistent, the composition of the four rigid-body transformations should be close to the identity transform.

For each pair of hypotheses, we can construct a loop of rigid-body constraints (see Fig. 3). This loop incorporates two additional rigid-body constraints derived from dead-reckoning. Since the constraints form a loop, the composition of their rigid-body transforms should be the identity matrix. Each rigid-body transformation is associated with a covariance matrix, allowing us to compute the probability that the loop is the identity matrix: high probabilities indicate pair-wise consistency of the two hypotheses.

We construct a *consistency matrix* A by computing the pair-wise consistency of each pair of hypotheses. Using SCGP [**?**], we can compute subsets of the hypotheses that are mutually self-consistent. Each of these subsets represents a local match.
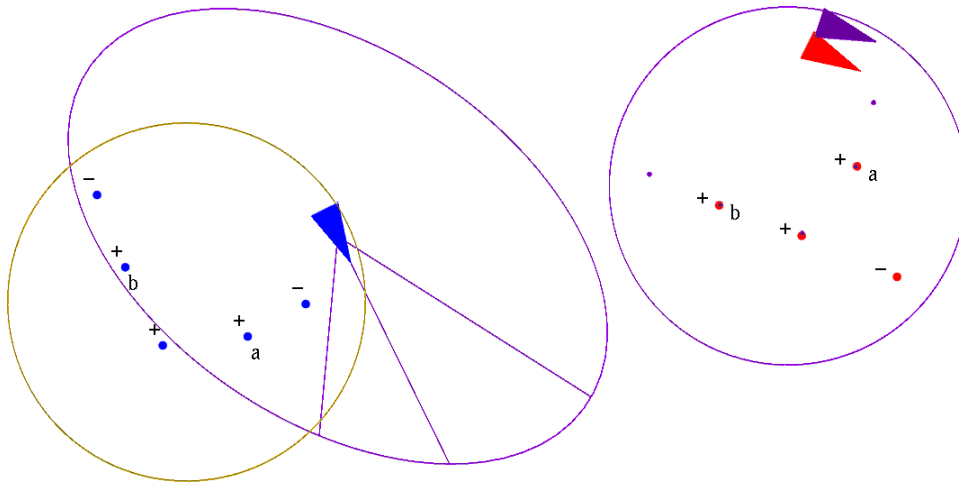
Fig. 3. Pose-to-Pose matching. We use RANSAC to find a pair of points from each pose (labeled a and b) that best align the ensemble of points. The two large circles indicate the estimated sensor range. The ellipse indicates the positional uncertainty of the two poses. Points which are successfully matched are labeled with a plus sign; those that have no counterpart incur a penalty, and are labeled with a negative sign.

SCGP produces a confidence metric for each subset. Environmental ambiguity can be detected when the best local match has a similar confidence measure as the next best match. In the case of a picket fence, for example, multiple local matches with similar confidence will be detected. In our experiments, we require that the confidence of the best match be twice that of the next best match: this ensures that the local match is locally unambiguous.

In our approach, we do not attempt to fuse the various pose-to-pose matches comprising a local match into a single rigid-body constraint. Instead, the local match is represented by the entire set of pose-to-pose matches. If a local match is ultimately accepted by our algorithm, all of its pose-to-pose matches are individually incorporated into the posterior map.

V. GLOBAL CONSISTENCY

In the previous step, SCGP identified local matches that are locally unambiguous. This satisfies one of the two conditions described in Section II. Before accepting a local match, however, we must satisfy the other condition, namely that only one local match can fit within the positional uncertainty ellipse.

In principle, testing this condition would require computing the worst-case packing of the local maps. The required size of the local match would thus be a complicated function of the *shape* of that local match.

In practice, we have used a simple (but easier to compute) approximation instead. We compare the maximum dimension of this local match to the length of the major axis of the covariance ellipse. If the matched region is over half as large as the uncertainty, then it is unlikely that the uncertainty ellipse could contain two matching environments. This approximation could erroneously accept local matches in some cases. Long skinny hallways, for example, can be packed very closely together: more than one "copy" of a hallway might fit within the uncertainty ellipse even though the match has a large

bounding dimension. In practice, this theoretical problem does not appear to be an issue, but a better means of estimating the worst-case packing remains an area of future work.

If a local match is too small in comparison to the positional uncertainty, we continue to add pose-to-pose matches until it is satisfies the global consistency condition. In practice, it is necessary to limit the number of hypotheses in a single set in order to bound the computational complexity of the SCGP filtering. If the limit (40 in our experiments) is exceeded before the match is sufficiently large, we discard every other hypothesis in order to make room for additional hypotheses. This "sparsifies" the hypothesis set without significantly shrinking its physical size.

When the prior uncertainty is very large, the local matches must also be large. Finding local matches becomes more difficult, in part because our current approach assumes that the dead-reckoning is of reasonable quality over the scale of local matches. Worse, a local match may be missed if the robot takes substantially different paths through an environment during different visits. However, we believe our method is applicable to many practical scenarios. Even with relatively poor odometry, sufficiently large local matches can be achieved for most indoor environments. Outdoors, even intermittent GPS reception will bound the prior uncertainty to reasonable levels.

VI. RESULTS

We tested our algorithm on the DLR "circles" dataset, using it to identify loop closures. These loop closures classically represent difficult data association problems.

A total of 4644 pose-to-pose hypotheses belonging to 278 potential local matches were generated, 2043 of which were accepted. We instrumented our algorithm to remember which data associations were used to generate each hypothesis. Using ground truth data, we can determine whether each hypothesis (generated by RANSAC) was the result of "good" or "bad" data associations. The performance of our algorithm on a

Hypothesis Outcome	Good Assoc.	Bad Assoc.
Accepted Hypotheses	2043	0
Rejected (small set)	714	32
Rejected (ambiguous)	1168	105
Rejected (inconsistent)	544	38

Fig. 5.   Hypothesis Error Rates. The "good" and "bad" columns denote hypotheses arising from correct (and incorrect) data associations, based on ground truth data. Most importantly, the algorithm exhibits no false positives. However, a large number of false positives are present due to the noise in the observations: this noise can result in poor hypotheses even when the data associations are correct. The numbers above reflect individual pose-to-pose matches; these were grouped into a total of 278 local matches.

typical run is tabulated in Fig. 5. Notably, there were *no* false positives.

Since our algorithm includes randomized components, the performance varies from run to run. False positive rates are consistently low (usually zero), though failures do occasionally occur. These appear to be the result of repeated iterations of RANSAC yielding similar (but incorrect) pose-to-pose matches. These incorrect matches can form a self-consistent local match. We believe these failures can be avoided by preventing multiple pose-to-pose matches from using the same pair of data associations. Even when a failures occurs, it generally has a minuscule effect on the quality of the posterior map. This is because the errors are rare, small in magnitude, and generally surrounded by dozens of correct matches that mitigate their effect.

Our false negative rate, based on Fig. 5 appears to be quite high. Indeed, a significant number of good hypotheses *are* rejected. However, the numbers are not as bad as they appear: observation noise can cause hypotheses to be poor even if the data association is correct. Since our pair-wise consistency test is based on the agreement of the rigid-body transformations (and *not* the data associations), these poor hypotheses are frequently rejected.

A total of 199 seconds of CPU time (on a 2.4GHz Intel processor) were required for feature matching (including RANSAC) and hypothesis filtering. Similarly good loop-closing results can be obtained even when the original dataset is decimated, but with much lower computational costs.

VII. CONCLUSION

This paper presented a place recognition algorithm based on local matches rather than explicit data association. We have described specific conditions under which these local matches can be trusted to be globally consistent. We evaluated our algorithm on a perceptually challenging dataset, showing that our algorithm can reliably close loops.

A disadvantage of the local matching approach is that it requires that matches can be computed using only locally-available information. This in turn requires that observation noise be sufficiently low to generate decent matches and that it be possible to generate a rigid-body transformation given a pair of observations. When these conditions are satisfied, however, our local matching approach is fast, reliable, and fairly easy to implement.

REFERENCES

[1] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking.* New York, NY, USA: ACM, 2004, pp. 70–84.

[2] J. Neira and J. D. Tardos, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 890–897, December 2001.

[3] E. Olson, M. Walter, J. Leonard, and S. Teller, "Single cluster graph partitioning for robotics applications," in *Proceedings of Robotics Science and Systems*, 2005, pp. 265–272.

[4] T. Bailey, "Mobile robot localisation and mapping in extensive outdoor environments," Ph.D. dissertation, Australian Centre for Field Robotics, University of Sydney, August 2002.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 888–905, August 2000.

[6] C. Ding, X. He, H. Zha, M. Gu, and H. D. Simon, "A MinMaxCut spectral method for data clustering and graph partitioning," Lawrence Berkeley National Laboratory, Tech. Rep. 54111, 2003.

[7] M. Cummins and P. Newman, "Probabilistic appearance based navigation and loop closing," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Rome, April 2007.

[8] K. L. Ho and P. Newman, "Detecting loop closure with scene sequences," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 261–286, 2007.

[9] J. Gutmann and K. Konolige, "Incremental mapping of large cyclic environments," in *Proceedings of the IEEE International Symposium on Computational Intelligence (CIRA)*, 2000.

[10] U. Frese, "Deutsches zentrum für luft- und raumfahrt (DLR) dataset," 2003.

[11] M. Bosse, P. Newman, J. Leonard, and S. Teller, "Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework," *International Journal of Robotics Research*, vol. 23, no. 12, pp. 1113–1139, December 2004.

[12] E. Olson, "Robust and efficient robotic mapping," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, June 2008.

[13] E. Olson, J. Leonard, and S. Teller, "Fast iterative optimization of pose graphs with poor initial estimates," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2006, pp. 2262–2269.

[14] M. Kaess, A. Ranganathan, and F. Dellaert, "Fast incremental square root information smoothing," in *International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad; India, 2007, pp. 2129–2134. [Online]. Available: http://www.cc.gatech.edu/ dellaert/pub/Kaess07ijcai.pdf

[15] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America. A*, vol. 4, no. 4, pp. 629–642, Apr 1987.