Novel Feature Based Outlier Rejection and Clustering Algorithm

Daniel Nix, Jeff McGough
Computational Sciences & Robotics
SD School of Mines & Technology
Rapid City, SD 57701
Daniel.Nix@mines.sdsmt.edu

Abstract

Feature based computer vision algorithms like those used in Simultaneous Localization and Mapping (SLAM) and stereo vision rely on accurately matching features between images to operate correctly. Even the best feature matching algorithms may still produce incorrect matches which must be discarded. These algorithms may also need to discern between many static and dynamic objects within a sequence of images.

Current solutions typically use Random Sample and Consensus (RANSAC) to determine which features have been incorrectly matched and must be thrown out. The same algorithm may group features in a sequence of images which share common dynamics.

We propose a new method using image mirroring, tiling, and clustering to reject incorrect feature matches and group features with common dynamics within an image.

1 Introduction

Computer vision frequently uses algorithms which operate on features, defined as points of interest which can be repeatably identified within subsequent images. Features are typically corners and intersections of lines within an image as opposed to curves which are not easily described by a single point. Good features are those which can be uniquely identified among other features within an image. Once features have been identified in an image they are used to extract spatial and temporal information from a scene.

Common issues for feature based algorithms are incorrect matches between features in different scenes and features moving with respect to one another. Presented is a method to reject incorrect feature matches and determine clusters of features moving between two images.

1.1 Outlier Rejection and RANSAC

For feature based algorithms to work features in subsequent images must be matched correctly. Algorithms like SIFT[1] and SURF[2] generally match features correctly but may still compute mismatches which limit the ability of other feature based algorithms to perform. Random Sampling and Consensus (RANSAC)[3] is commonly used to reject outliers. For a set of N feature matches K are selected in each iteration of RANSAC. The RANSAC algorithm computes a model based on these K data points from the sample. The model is then applied to the N feature matches in the set and matches not fitting the model are discarded for the iteration. If a sufficiently large number of matches fit, the model is considered accurate and the matches which did not fit are thrown out as outliers. While RANSAC may be linear with the number of data points it may require iterating over the set of feature matches many times before an adequate model is determined. The recommended upper bound for the number of iterations k is given by[3]:

$$k = \frac{\log(1-z)}{\log(1-w^n)} \tag{1}$$

where

- z = probability that at least one sample of the set will be error free

- w = probability that a selected data point fits the model within the desired tolerance

- n = the number of good data points required

As an example, if RANSAC is to determine the model with $n = 50$ data points with $w = 0.9$ probability that any point is an inlier with $z = 0.9$ confidence an upper bound of 446 iterations are required. However, if the number of desired points grows to 100 the upper bound should be set at $8.7 * 10^4$. Similarly, if the ratio of inliers decreases to $w = 0.75$ the upper bound on iterations grows to $4.1 * 10^6$.

For some applications a small number of points are acceptable but if many points are required for a mapping or the probability of selecting an inlier decreases RANSAC requires many more iterations.

1.2 Grouping Velocity Fields

Computer vision is commonly used for autonomous vehicle applications. These applications may be split into two problems. The vehicle must first determine its motion relative to a fixed world. This gives the vehicle odometry information which may be further used in Simultaneous Localization and Mapping (SLAM) algorithms. Next, the vehicle must react to a dynamic environment. This may consist of other moving vehicles, people, or any number of moving objects.

One application of the velocity field outlier rejection method is presented by Kitt et al in [4] for use with stereo based odometry. These frames may be determined by iteratively applying RANSAC to the matched feature set. With each iteration a new model is determined, each describing the motion of a different set of features. When a model fits one set of feature matches (say, features from objects in the fixed world), these points are removed from the set of all features and the algorithm is applied again. The process repeats until features contained in each different velocity frame are determined.

2 Tiled Intersection Grouping Analysis

We present a new approach to outlier rejection and grouping velocity fields. Tiled Intersection Grouping Analysis (TIGA) begins with two images Im_1 and Im_2 of equal dimension with the number of rows R and columns C. Images corresponding to each step of the process are provided. To present an optimal case where all features are correctly matched, $Im_1 = Im_2$ in this example. Figure 1 shows two images to be

(a) Image Im1 (b) Image Im2

Figure 1: Identical bicycle images used to demonstrate the clustering algorithm

(a) Im1 with features found (b) Im2 with features found

Figure 2: Bicycle images with detected features indicated

compared.

First, feature sets $F1$ and $F2$ of lengths M and N, respectively are computed in both images. These feature sets contain the row-column index of each feature along with the appropriate descriptor for each feature $f1_m$ and $f2_n$ in each set. For this example, a Harris feature detector with a SIFT descriptor was used. Features identified are presented in Figure 2.

Features from $F1$ are then matched to features in $F2$ to construct the matched feature set M. Each feature match is described by the row-column index of the matched feature in Im_1 and Im_2. Results of matching are shown in Figure 3.

For the first tiling, the matched features coordinates from F_2 in M are mapped to coordinates relative to Im_1, as if Im_2 were rotated 180° and placed to the right of Im_1. Line segments are then constructed from matched features in Im_1 to the newly mapped features in Im_2 as shown in Figure 4. The intersection of line segments are recorded.

Figure 3: Bicycle images with corresponding features matched

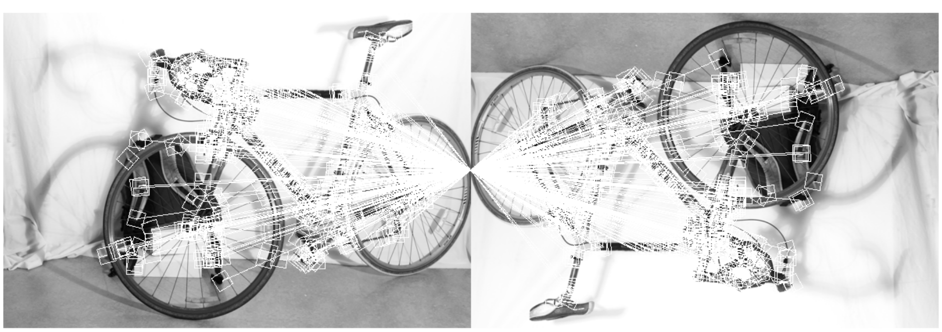

Figure 4: Bicycle images with corresponding features matched and second image rotated

The same process may be repeated placing Im_2 above, below, left, and right of the original, along with applying different transformations such as mirroring across the x or y axis to extract different intersection clusters.

2.1 TIGA for Outlier Rejection

Rejecting outliers in TIGA requires analysis of the clusters produced. Each cluster of intersection points represent a grouping of correctly matched image features. Feature correspondences included in these clusters represent inliers while those outside the clusters are considered outliers.

If Im_1 and Im_2 vary only slightly then the cluster of correct matches fall near the center of the image. That is, in the image clusters tend to fall in a row or column (depending on the tiling orientation) centered at the border of the $B \pm D$, where D is small with respect to the number of rows or columns the images. Incorrect feature matches do not fit this central cluster and may be discarded.

2.2 TIGA for Grouping Velocity Fields

A problem encountered in feature based algorithms, especially those used for odometry, is segregating groups of features which are moving inside an image. For example, a SLAM algorithm to map a city using a car with fixed camera will likely encounter other moving vehicles. Features detected on moving vehicles should not be included in the SLAM calculations and must be thrown out. Other algorithms may attempt to track moving objects or determine distance traveled based on discrepancies in feature movement.

Grouping features which are moving through the image together requires analyzing multiple clusters of feature correspondence intersections. One cluster is formed for each different velocity field. These varying velocity fields may arise from either object motion such as a moving person, or camera motion like a camera on an autonomous vehicle, or some combination of the two.

As objects move differently from Im_1 to Im_2 the centroid of their correspondence intersection clusters moves away from the center of the image. Larger positional changes of features between images equate to a greater deviation in the centroids of correspondence clusters. Analyzing the position of this centroid allows clustering of features moving together.

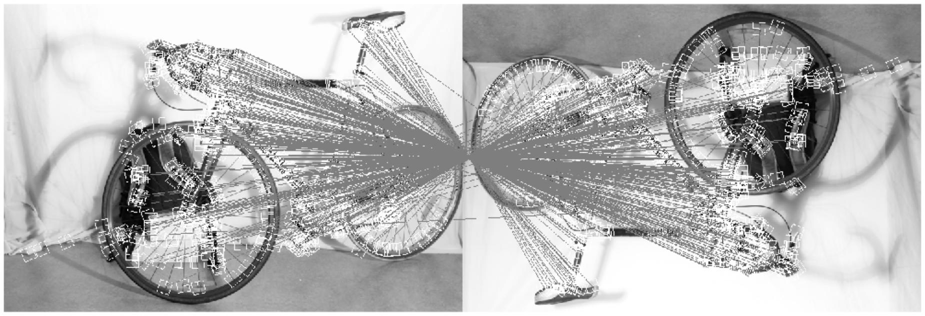

Figure 5: Bicycle image features matched. Inliers cluster near the center while outliers do not.

Figure 6: Indoor scene with features matched. Inliers cluster near the center while outliers do not.

3 Results

Preliminary results using TIGA for outlier rejection shows that iteration over all data points is not necessary as with RANSAC. It is also shown that TIGA is able to group features moving at different rates across the image. Detailed descriptions of both findings are provided.

3.1 Outlier Rejection

TIGA was able to discern between features incorrectly matched and those matched correctly by analyzing clusters formed by intersections between feature correspondences. Figures 5 and 6 demonstrate how line segments between correctly matched features form intersections in tight clusters while incorrectly matched features do not.

This is advantageous over RANSAC because outliers may be determined directly based on intersections. In scenes with large proportions of outliers the RANSAC

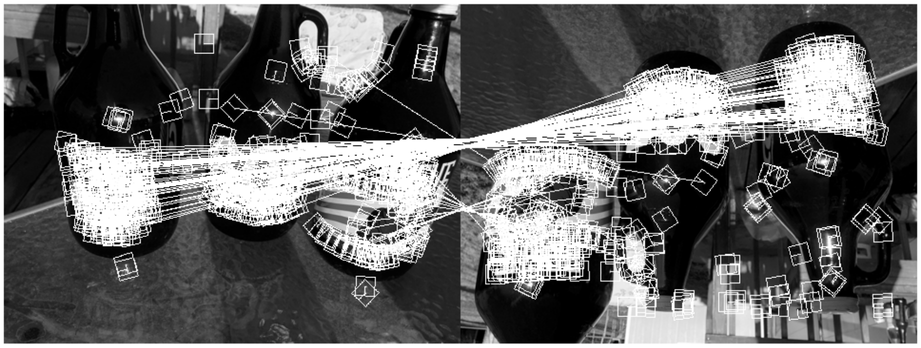

Figure 7: First tiling of subsequent images with one growler moved vertically with two remaining static

algorithm hopes to pick feature correspondences representing the true model but may not pick correctly. The larger the outlier proportion the greater the likelihood that RANSAC samples outliers to predict the model.

TIGA does not have this problem. Clustering is not dependent on the number of incorrect matches. Correct feature matches still form tight clusters of intersections while incorrect matches form only sparse ones.

Applying TIGA to outlier rejection also allows for fast rejection of incorrect feature matches when some objects are moving in the scene while others are not. Matched features present in individual solid objects produce line segments which cluster with other feature correspondences of that same object. Only when a feature correspondence does not cluster with any objects is it required to be considered an outlier, allowing the algorithm to selectively reject outliers based on which objects are of interest. This is closely related to using TIGA when grouping velocity fields.

The algorithm could be improved by predicting where clusters may form. For example, if the images are known to come from a stereo camera constraints may be placed to determine a point near the center of the concatenated images where segments are most likely to cross.

3.2 Grouping Velocity Fields

TIGA was able to discern between groups of features corresponding to different velocity fields without any prior knowledge of the scene and without computing a flow field for the entire image.

Figure 9 shows feature correspondences between two images with one of the three growlers moved. Before flipping it is difficult to discern movement. Figure 10 shows

Figure 8: Second tiling of subsequent images with one growler moved vertically with two remaining static

the correspondences after flipping. It is apparent that two of the growlers with line segments passing through the origin belong to one frame, and the remaining has moved a small amount based on its cluster left of center. If the camera is known to be static it is apparent that other static objects contain features correspondences which produce line segments through the center of the image.

Tiling the images horizontally as in Figure 7 and vertically as in Figure 8 show how different tilings make translations more apparent depending on the direction. Horizontal translations are more obvious in vertically stacked images while vertical translations are apparent when images are placed horizontally.

When an object moves with respect to the rest of the scene between two images, the intersection point of line segments connecting feature correspondences within that object moves as well. Each group of matched features corresponding to one or more objects forms intersections in the same cluster. In this way all features may be grouped with others moving in the same direction at the same velocity.

This improves on using RANSAC for this application because iteration to determine different velocity fields is not required. When RANSAC is applied points are selected at random to determine the model meaning the algorithm will likely select model points from the predominant velocity field. However, if there is no predominant field RANSAC is likely to take many iterations to randomly select points which describe

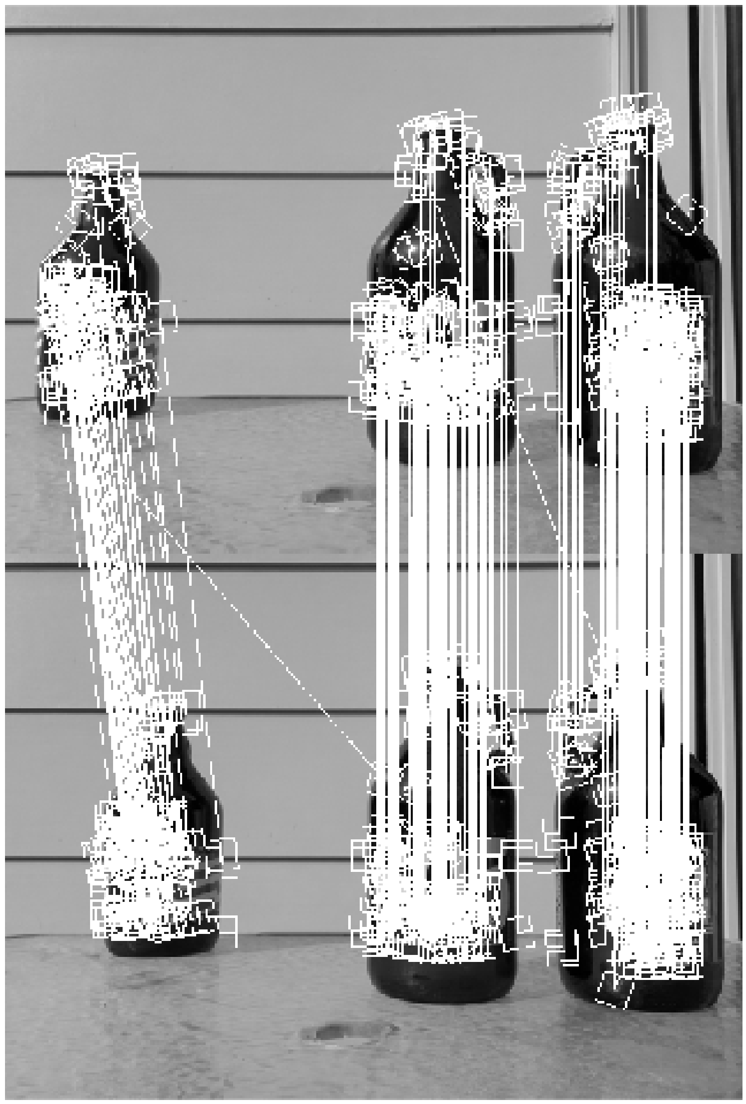

Figure 9: Feature matches before image flipping

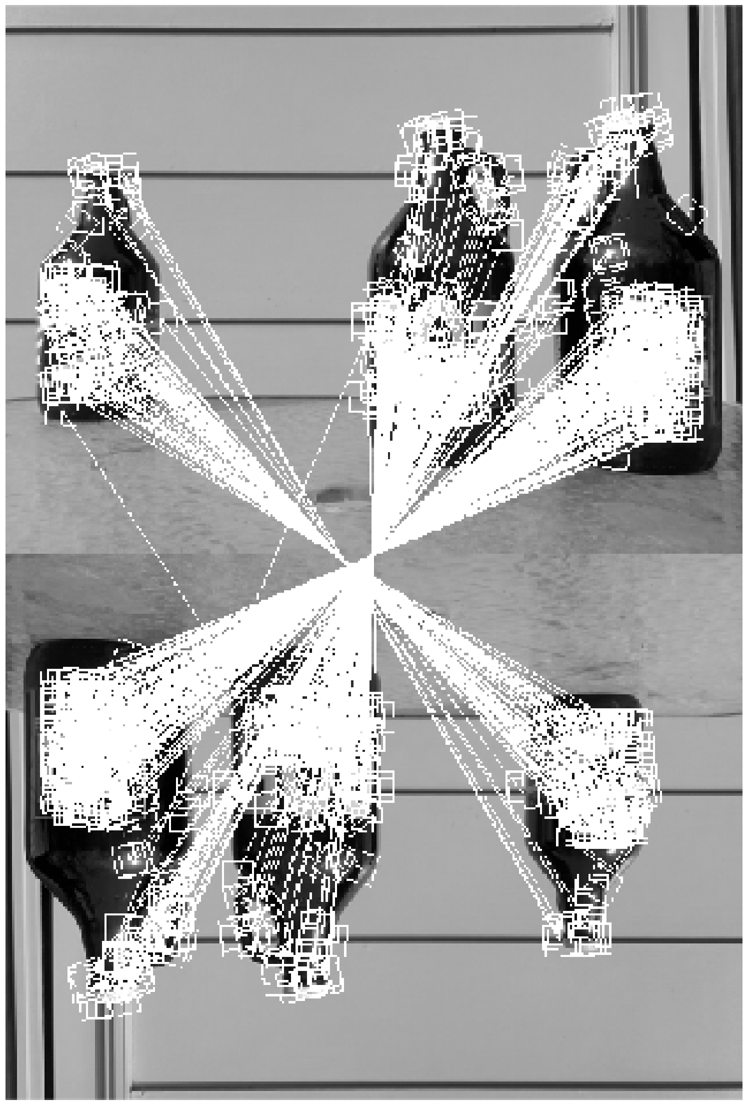

Figure 10: Feature matches after image flipping

one single field.

TIGA directly determines different velocity fields based on different intersection clusters. The static world clusters to one group while features belonging to other moving object cluster with those moving in the same direction.

It is possible that for a single tiling TIGA clusters intersections corresponding to different velocities around the same point. Computing clustering for multiple tilings (above, below, left, and right) helps break this degeneracy. It appears that assuming small differences between images prevents this degeneracy. This is typically the case for subsequent images in a stream or a stereo camera. Further analysis to confirm this quality should be done.

4 Conclusion

Tiled Intersection Cluster Analysis shows promise in outlier rejection and feature grouping. Qualitatively, images produced show that intersections of of line segments drawn between corresponding features cluster depending on change in feature position.

Finding the intersection points from line segments connecting matched features from one image to that of the other may allow for clustering and outlier rejection. Assuming small changes between images appears to imply the cluster of correct matches lies in a small region near the center of the concatenated image. This clustering may be mathematically described and used for fast outlier rejection.

Computing these same intersection points also appears to group features in the image with different velocities into different regions. Static objects have intersection points near the center of the concatenated images while moving objects cluster away from the center. Depending on the direction of object motion these clusters appear in different locations which may be used to quantify and determine velocity information.

Quantitative results showed promise for further development of the TIGA .

5 Future Work

Now that qualitative properties have been shown, the mathematical groundwork for a quantitative evaluation of these algorithms must be developed.

Modifying the algorithm for outlier rejection to account for stereo camera parameters may allow prediction of valid matched feature intersection points and ease outlier rejection. Temporal constraints on motion within an image may allow for quantitative evaluation of motion within an image which may prove useful in monocular SLAM algorithms or other algorithms requiring segregation of moving and static objects in an environment.

These features will be explored in depth for the author's thesis

References

[1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[3] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[4] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme.," in *Intelligent Vehicles Symposium*, pp. 486–492, 2010.