An Ensemble Approach to Image Matching Using Contextual Features

Brittany Morago* Giang Bui* Ye Duan
University of Missouri-Columbia
209 Engineering Building West
Columbia, MO 65211
duanye@missouri.edu

Abstract

We propose a contextual framework for 2D image matching and registration using an ensemble feature. Our system is beneficial for registering image pairs that have captured the same scene but have large visual discrepancies between them. It is common to encounter challenging visual variations in image sets with artistic rendering differences or in those collected over a period of time during which the lighting conditions and scene content may have changed. Differences between images may also be caused by using a variety of cameras with different sensors, focal lengths, and exposure values. Local feature matching techniques cannot always handle these difficulties, so we have developed an approach that builds on traditional methods to consider linear and histogram of gradient information over a larger, more stable region. We also present a technique for using linear features to estimate corner keypoints, or Pseudo Corners, that can be used for matching. Our pipeline follows this unique matching stage with homography refinement methods using edge and gradient information. Our goal is to increase the size of accurate keypoint match sets and align photographs containing a combination of man-made and natural imagery. We show that incorporating contextual information can provide complimentary information for SIFT and boost local keypoint matching performance, as well as be used to describe corner feature points.

Index Terms

Keypoint matching, 2D registration, contextual features, ensemble features, linear features, histogram of gradients.

I. Introduction

Matching photographs and finding image correspondences is necessary for a variety of applications in computer vision from creating structure from motion point clouds to image classification. While several methods already exist for detecting and matching sets of pixels using image intensity patterns, many will encounter difficulties when images

*The first two authors contributed equally to this work.

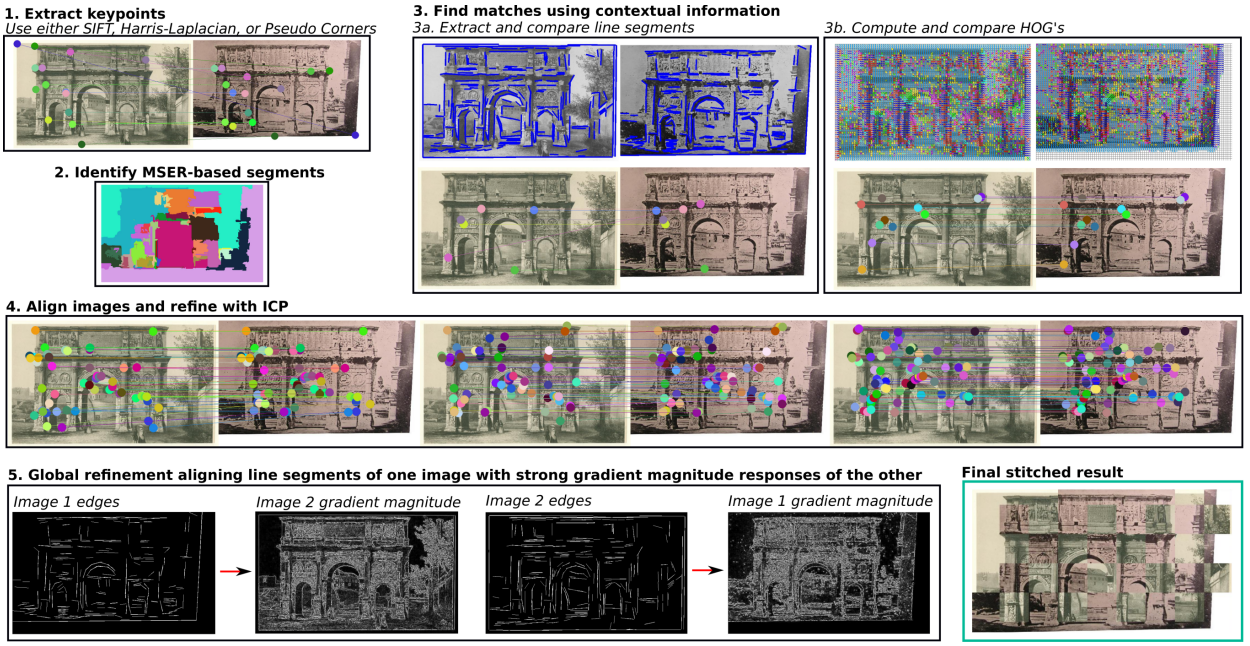

Fig. 1. Outline of our registration pipeline. Keypoint features are first extracted. SIFT keypoints, Harris-Laplacian corners, or our proposed Pseudo-Corners can be used. If SIFT is used, keypoints are matched using the SIFT descriptor. We then search for matches using contextual features. I_1 is segmented into MSER-based neighborhoods which are used to determine neighborhood sizes for studying line segments and HOG's. (HOG information in top right image is colored according to the dominant gradient direction for each cell.) Linear and HOG information is used to verify new matches. An aligning homography is calculated and refined using ICP and gradient-based methods.

are captured under highly varying conditions. Changes in lighting, camera exposure values, scale, scene content, and artistic styles can greatly affect local gradient information surrounding keypoints. Popular keypoint detection and matching methods such as Harris Laplacian [1] and SIFT [2] that rely on such information may not perform well under these circumstances. In difficult situations, they may provide too few matches or so many incorrect matches that a correct image alignment may not be identified after applying robust methods like Random Sample Consensus (RANSAC) [3]. Standard detector techniques may not even find a sufficient amount of repeatable features to match keypoints successfully. We propose a pipeline that builds upon traditional keypoint matching techniques and uses contextual information to handle these challenging scenarios. We hypothesize that using this extra regional information can help clear up a great deal of the ambiguity presented by small scale gradient information to find a larger set of matches for registration purposes. We also present a method for estimating corner keypoint locations for situations where other repeatable features may be hard to detect and use this same contextual information to match them.

All features have their own strengths and weaknesses and each may be more well-suited for specific tasks than others [1]. On one hand, defining features on a very local scale can create a relatively large set of keypoints that are highly informative for matching and invariant to certain amounts of perspective and content changes. However, limiting the amount of visual information accounted for in a descriptor may obscure the differences in these local

keypoint descriptors [4]. This is especially true when highly different viewing conditions, clutter, and occlusions are present, causing local visual properties to change, making it difficult to match locally-defined keypoints correctly. Using a larger scale can help define more discriminative features. By combining local and regional information into one ensemble feature, we feel we can achieve the best of both worlds by taking advantage of both precise local information and distinct regional data.

A. Background

Images can be aligned using information from locally defined feature descriptors, regional descriptors, direct methods that rely on global pixel-pixel correspondences, or structural features such as line segments. Locally defined keypoints use data, such as gradient direction and magnitude, from neighboring pixels within a relatively small window size. Regional descriptors use information over a larger portion of the image to describe an area and potentially match it to corresponding regions in other images. Global matching techniques, such as optical flow and Fourier-based alignment, can provide a very accurate image alignment, but generally require a good initialization to be successful [5].

We try to take advantage of the strengths of several of these methods by combining them together into one ensemble feature matching and alignment refinement approach that uses local, regional, global, and structural information.

1) Local Matching Methods: The Harris detector and use of the Hessian matrix are two classical methods for finding local interest points in images [5] [6] [1]. Both detectors use the second moment matrix to study the gradients of pixels surrounding an interest point. The scales for points' feature descriptors are often chosen using the Laplacian [7]. These methods work to identify local areas with significant visual changes, such as corners, and are fairly invariant to lighting changes. However, these early methods were shown to be sensitive to large scale changes and had limited affine invariance, which led to the development of more sophisticated detectors and descriptors. For example, the Scale Invariant Feature Transforms method (SIFT) [2] is a commonly used and robust local feature matching technique that assigns a scale and orientation to each interest point. A descriptor is constructed based on local image gradients that is generally unique to a single area in the image. The neighborhood size used to construct the descriptor is usually around 16x16 pixels at the appropriate image scale. Several groups have expanded on the SIFT descriptor to develop feature matching methods that are faster and more robust to various transformations [8] [9] [10].

The relatively small window sizes used by these methods to create keypoint descriptors can be beneficial in many cases. They will be able to identify locally distinct areas accurately. However, cases where image pairs have large scale, lighting, and/or content changes may require a larger window size to correctly pinpoint very distinct keypoint matches since the images will have so many visual differences.

2) Regional Matching Methods: Several methods for matching and aligning photographs have expanded on these locally defined feature descriptors to use visual information covering a larger region within images. Yang et al. [13] designed a pipeline using bootstrapping and region growing to search for a dense set of corner and edge features.

Fig. 2. Registration results after applying our pipeline to several challenging image pairs from our dataset as well as from the datasets provided in [11] and [12]. Original images are shown above registered images.

They use these features to compute a global homography relating two images. Since this method uses a single homography to overlay an image pair, it is best-suited for planar-approximate scenes. Hacohen et at. [14] also identifies dense matches across image pairs. They align image patches by searching across a set of transformations such as rotations and translations that overlay pairs of patches as well as possible and use groupings of matches to calculate the coherence of a region.

Other groups have developed region detectors and descriptors that rely on less-structured features. MSER's are used to identify image neighborhoods that are visually similar and contrast greatly with adjacent regions. The SIFT descriptor can be used to describe and match corresponding MSER's in image pairs [15]. MSER is affine invariant but has been shown to be biased to "round" shapes [16]. If an image has repetitive patterns, such as a row of windows, it is possible that the MSER feature will be ambiguous if its size is not large enough. Shape contexts [17] describe feature points that are uniformly sampled along contours based on nearby keypoint's relative locations. These groupings of points are used to find optimal image correspondences and are useful for object recognition. Contours throughout the image must be correctly extracted, which is not always an easy task.

Measuring the similarity of all overlapping pixels is another approach for region matching. Normalized cross correlation, the sum of squared differences, and the sum of absolute differences all perform a pixel-by-pixel comparison and can be used to find the best alignment for an image pair [18]. Histograms of gradients (HOG) have also been used as unstructured region descriptors for image alignment verification [19] and object recognition [20]. While these methods can be very informative, scale and orientation information should be known in order to use them efficiently. To maximize their robustness, and to avoid exhaustively searching for overlapping regions, they

can be combined with other information or descriptors to provide this information.

Another class of mid-level, regional matching approaches is to identify symmetrical elements and self-similarities in images. This is especially useful when working with architectural data that often times contains repetitive structures [21]. Hauagge and Snavely search for horizontal, vertical, and rotational symmetries about various axises and scales across images [12]. Self-similarities can even be identified in patterns of colors, edges, and repeated visual elements in non-architectural imagery by measuring how similar a small region is to parts of the larger surrounding region [22].

Machine learning techniques can also be used to identify unique regions in photographs, paintings, and 3D models to match imagery with highly varying properties and to perform object recognition [11], [23], [24]. These mid-level discriminative features ideally occur often enough in the data to be learned, but are different enough from the rest of the imagery to be clearly located in photographs and models with visual dissimilarities. Alignment estimates can even be made between multi-modal data such as paintings and photo-realistic 3D models by using the gist scene descriptor and edge information of different regions [25].

3) Linear Features: Extracting long linear features is another approach used for matching images. This can be a very difficult task as line features are relatively fragile and their lengths may vary with changing viewpoints and lighting conditions. Wang et al. [26] match clusters of line segments, or line signatures, in image pairs using the assumption that local neighborhoods of strong structural features will maintain the same relative arrangements throughout baseline and lighting changes. Fan et al. [27] also match lines by using spatially proximate features. They consider two lines to be a matching pair if they have similar distance ratios to neighboring keypoints that are known matches. Since each potential linear match requires nearby correctly matched keypoints, this method is dependent on having a relatively large number of correct feature point correspondences. Color histograms of the color profiles surrounding a line segment have also been used for line matching under the observation that the changes in color between two intersecting planes is relatively invariant to camera motion [28].

B. Overview and Contributions

In this paper we present a framework for using contextual information for image matching with the goal of increasing the number of keypoint matches found between challenging image pairs in order to find an aligning transformation relating the images. We mainly work with piece-wise planar man-made scenes that need not be time adjacent, examples of which are shown in Figure 2. For planar-approximate image patches, we believe that using relatively large regions for keypoint comparison helps features remain more distinct throughout such condition changes. Our contextual framework for registering images can take three forms. The first is to build upon and complement existing sparse local feature matching techniques such as SIFT. We describe the surrounding regions of keypoints in terms of salient structural features via line segment extraction and their rich texture information using histograms of gradients (HOG). These neighborhood sizes are determined adaptively using maximally stable extremal regions (MSER) [29]. To eliminate the requirement that we incorporate an existing matching algorithm, we can also use our regional contextual information to describe and match Harris-Laplacian corners. The third

option is to estimate corner locations using extracted line segments, which we call Pseudo Corners, freeing this algorithm from any dependencies on existing detectors or descriptors. After combining several local keypoint and regional template matching techniques to make an educated estimate about the image alignment, we follow up with iterative and global refinement stages using corner, edge, and gradient information across the entire image planes. Our pipeline is outlined in Figure 1.

The contributions of our registration method lie in the combination of methods used including the structure of our contextual matching framework and the way we incorporate this framework with local keypoint matching to create an ensemble feature. We propose using robust SIFT feature points or plentiful corner keypoints as anchor points for exploring potentially matching regions by comparing both HOG's and line segments. The MSER's that encompass each SIFT or corner keypoint, and that are often used to identify homogeneous regions, give us a general idea of the dimensions of local planar patches and what the neighborhood size should be for stable matching. We match both hard and soft edge features (lines and HOG's) in these regions to increase the matching feature set size between images with visual discrepancies. Line segments represent the salient features and structures found in an image and are relatively invariant to perspective and lighting changes. However, using line segments creates a very minimalist representation of an image. It does not provide as much distinct information as HOG, which picks up curved structures and textures. HOG, on the other hand, is sensitive to clutter in the image. For line segment extraction, we use the Line Segment Detector method [30]. This algorithm has been shown to accurately identify existing lines in an image without also including many false detections due to its incorporation of line-support regions that require pixels on or near a line segment share very similar gradient orientations. Including this method in our work helps us to match stable regions across lighting and time changes in the images.

One of the main issues encountered in line-based matching is the lack of a robust distinctive line-based feature descriptor. Lines in general have no area and the sizes of their neighborhoods are not known or are hard to determine. The length of a line segment is fragile and therefore is not a stable attribute for comparison. However, the relative orientations and locations of line segments can be computed if a proper anchor point is available. In this paper, we combine SIFT and corner points with line segments to conduct line matching. By exploring multiple scales of regions and using dominant gradient information in a local neighborhood, we are able to develop a coordinate system which can be used to encode the relative locations and orientations of the line segments in local neighborhoods. We avoid including any information dependent on the unstable end points of the line segments and focus on matching groupings of segments as opposed to individual lines. Our experimental results show that making use of line segments in this fashion can provide accurate information for matching on a large variety of imagery.

By using a variety of methods to find image pair correspondences and making no assumptions about a scene's structure, we have developed a robust pipeline that is capable of registering images with varying subjects including highly regularized architecture with repeating patterns and more natural, unorganized subjects combining buildings with curved features and non-standard designs and added foliage.

C. Notation

For simplicity during our discussion, we will denote an image pair as I_1 and I_2. A feature point in I_1 is represented as A and its two closest matches in I_2 are B and C. When I_2 is transformed to I_1's coordinate system using a homography, H, we will represent it as I_2'.

Throughout our discussion, we use the term "contextual" to refer to the set of matches verified using both linear and HOG contextual information. "Ensemble" will refer to the set of features combining both SIFT matches and contextually-verified matches.

II. Contextual Feature Matching

The goal of our method is to extract highly distinct interest points and obtain a large number of correspondences that can be used to correctly register image pairs [4]. Our pipeline begins by using regional contextual information to match keypoints. We explore using three different types of keypoints as anchors for matching contextual descriptors. The first approach we study is matching existing locally defined keypoint descriptors, such as SIFT, and expanding upon them with our contextual information to create an ensemble feature. The second keypoint we test is Harris-Laplacian using only our contextual work for matching. The third, Pseudo Corners, is a new keypoint type we propose which is identified using line segments. We then identify inliers in the match set from any of these methods using geometric verification, globally estimate and refine the image alignment, and increase our set of matches.

A. Adaptive MSER Neighborhoods

We use MSER's to determine the neighborhood that is considered during contextual matching. MSER's are often used to find homogeneous areas in images that contain distinctive information that is useful for matching [31], [32]. An ideal neighborhood needs to contain enough regional information to clear up any ambiguity at the local keypoint level but not include so much information that no regions actually match. We also want every pixel in an image to belong to an MSER so that we have a "stable" neighborhood ready to select for every potential keypoint match. To find potential MSER's that satisfy these criteria, we create a Gaussian pyramid for an image and extract MSER's at each level, as shown in Figure 4. At the original image resolution, we get many MSER's that tend to be relatively small. Also, initially, not all pixels are assigned to an MSER. As we move down the pyramid and extract MSER's on increasingly blurred images, we end up with fewer, larger regions and more pixels assigned to a region. Just to ensure that every pixel is indeed assigned to a region, we have one post-processing step after MSER's are extracted at each level. Connected components of pixels that have no region label are identified. Each component is merged with the region that touches the highest number of pixels along the component's perimeter. In general, segmentation is a very difficult task, but this adaptive MSER labeling gives us a reasonable approximation of the regions we need.

To help achieve our goal of assigning every pixel to a mid-sized MSER, we allow different pixels in the image to be assigned to MSER's from different Gaussian pyramid levels. Initially, every pixel is assigned its corresponding

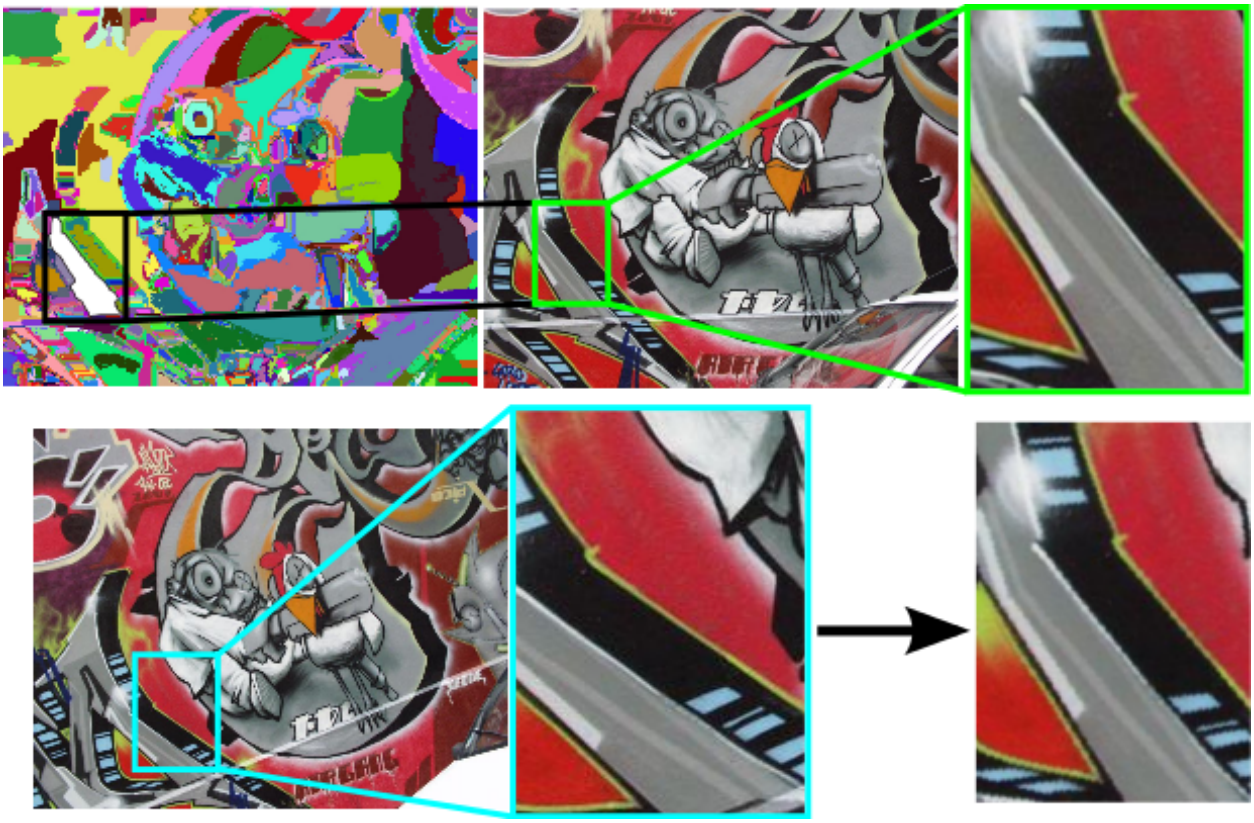

Fig. 3. Mapping bounding box of MSER neighborhood in I_1 to I_2. *Top:* MSER segments and MSER defined neighborhood in I_1. MSER segment being used is shown in white. *Bottom:* Original MSER neighborhood in I_2 and I_2's patch after it has been transformed to the same coordinate system as I_1's patch via a scaling and rotation. The patch in I_2 is extracted at multiple scales and each patch is rotated such that its x-axis points along the dominant gradient direction in the patch.

label at the bottom level (original image resolution) of the pyramid. Then we begin moving up the pyramid. If a pixel's corresponding region in the next level of the pyramid has an area that is no more than 10% of the image area (a threshold determined experimentally for our work), the pixel takes on the higher level region area.

The keypoint's neighborhood in I_1 used during contextual matching is an orthogonal bounding box that encompasses an entire chosen MSER. Using this bounding box essentially grows the region slightly and allows us to take into account boundary information surrounding the distinctive area. During the matching stage, the MSER bounding box in I_1 is mapped to I_2 at multiple relative scales to take into account scale changes between the images. The bounding boxes are oriented such that each one's x-axis is aligned along the dominant gradient direction within the image patch. An example of using this information to find a corresponding region between I_1 and I_2 is shown in Figure 3.

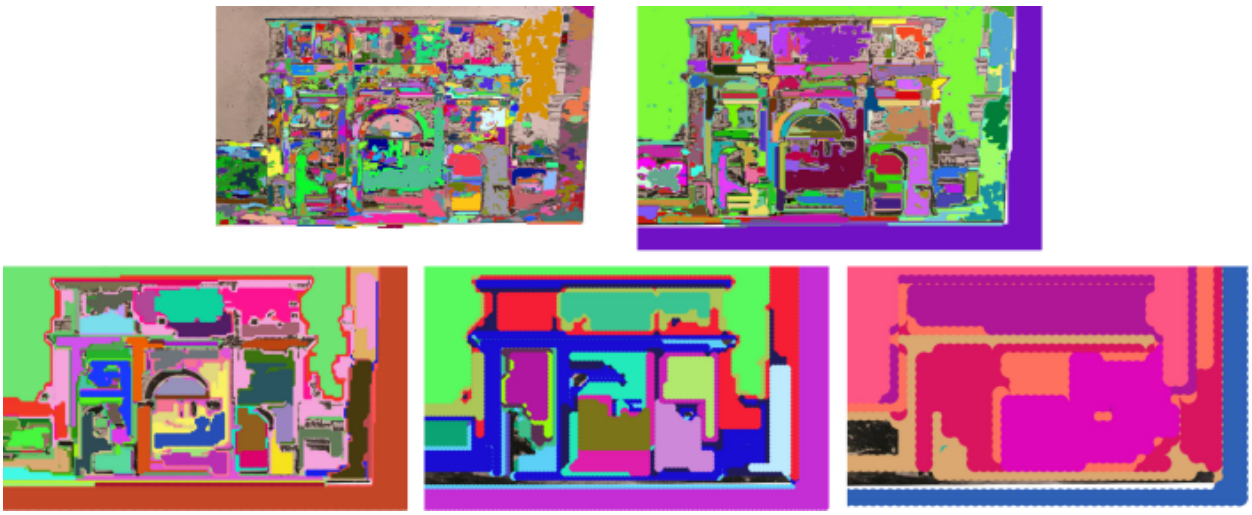

Fig. 4. MSER's extracted at different levels of a Gaussian pyramid. Segments are combined from different levels to create adaptive MSER segments shown in Figure 1.

B. Contextual Line Segments

We use linear contextual matching to describe potential matches by searching for and matching line segments near keypoints. Stable line segments in the MSER-based neighborhood of potentially matching keypoints are identified using the Line Segment Detector method [30]. Line segments are extracted from I_1 and I_2 at three different levels of a Gaussian pyramid. The lines found at each Gaussian level are all combined into one set of lines. Extracting lines from multiple image scales helps to address the fragile nature of line segments and gives us more information for matching. Each line segment must have a length of at least $0.03 * max(image_{width}, image_{height})$ and its midpoint must be located within the selected MSER bounding box to be taken into account. All the stable line segments in this neighborhood are transformed to a polar coordinate system determined by the neighborhood's dominant gradient orientation and normalized by the neighborhood's dimensions and are represented as (ρ, θ). We then match groupings of line segments in corresponding neighborhoods using Hungarian graph matching [33]. Figure 5 outlines this process. Two line segments (l_A, l_B) are considered to be similar if they have a low dissimilarity score according to Equation 1.

$$dissim(l_A, l_B) = (\rho_A - \rho_B)^2 + \omega(\theta_A - \theta_B)^2 \quad (1)$$

ω is a weight parameter that normalizes the range of θ ($\theta \in [0, 2pi]$) to that of ρ which is based on the contextual neighborhood dimensions. This dissimilarity measure is only calculated if at least three lines are identified in the contextual neighborhood of the keypoint. If a required percentage of the line segments have low dissimilarity scores, the keypoint match is saved. Table I shows how the precision of linear contextually verified matches changes as the required percentage of matching lines in a region and the dissimilarity score thresholds change. These values were obtained by running our method on the symmetry dataset in [12]. Figure 6 shows a set of matching lines.

TABLE I

AVERAGE PRECISION ON SYMMETRY DATASET FOR VARYING GRAPH MATCHING PARAMETERS - DISSIMILARITY SCORE AND THRESHOLD ON REQUIRED % OF MATCHING LINES

-	$dissim = 1.0$	$dissim = 2.0$	$dissim = 3.0$
10%	0.552	0.617	0.412
30%	0.371	0.426	0.365
60%	0.000	0.081	0.081

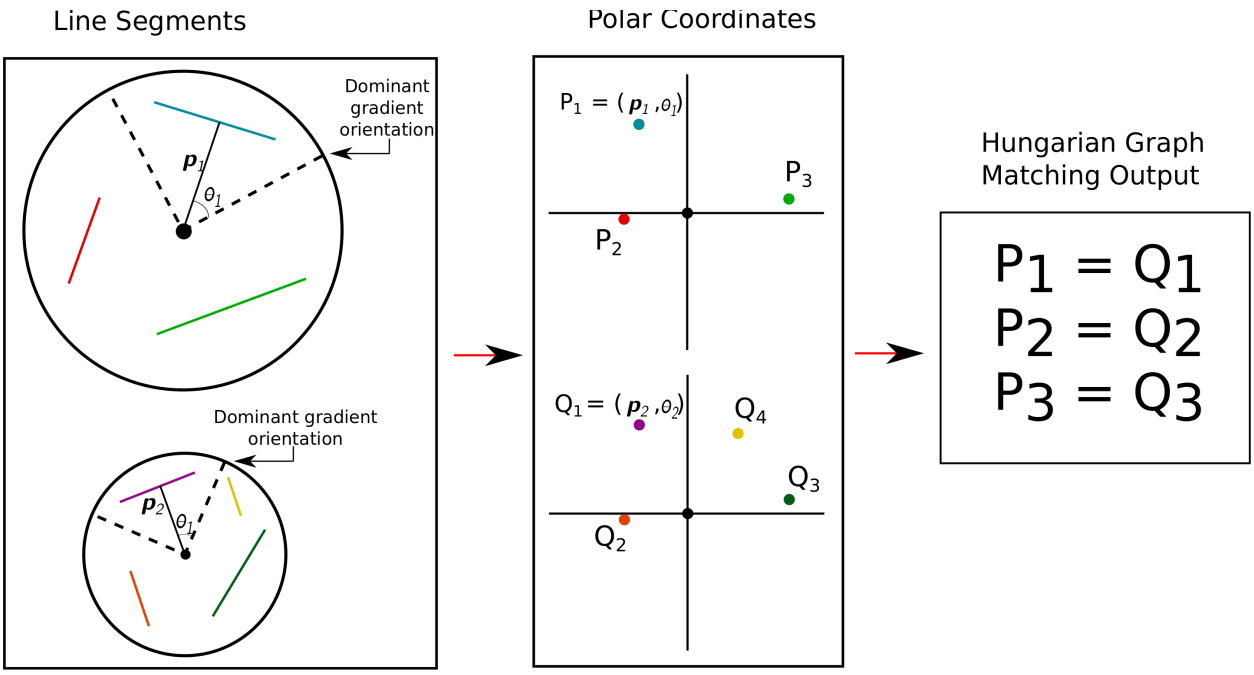

Fig. 5. Matching lines in the neighborhood of a feature point. *Left:* Line segments are identified within neighborhoods centering a potential match. The keypoint is shown in black in both neighborhoods and the relative scales are represented by black circles. *Middle:* Line segments are converted to polar coordinates based on their distance from the keypoint and their orientation in relation to neighborhood's dominant gradient direction. *Right:* Hungarian graph matching is applied to the set of polar coordinates to figure out the optimal set of matching lines.

C. Contextual HOG

Our second approach for contextual matching takes advantage of every pixel in MSER-based neighborhoods surrounding potential keypoint matches. The potentially matching neighborhood in the second image is calculated at multiple scales and its orientation is determined using the region's dominant gradient direction, giving us a region that contains the same content and is based in the same coordinate system.

To measure how well the two neighborhoods match, we begin by computing histograms of gradients for each image patch. Just as was discussed in Section II-B, a three-level Gaussian pyramid is built for the corresponding neighborhoods. A vector is constructed for each HOG cell by concatenating its nine bin values and normalizing using the vector's magnitude. We calculate the L2 distance between the vectors for corresponding cells in the

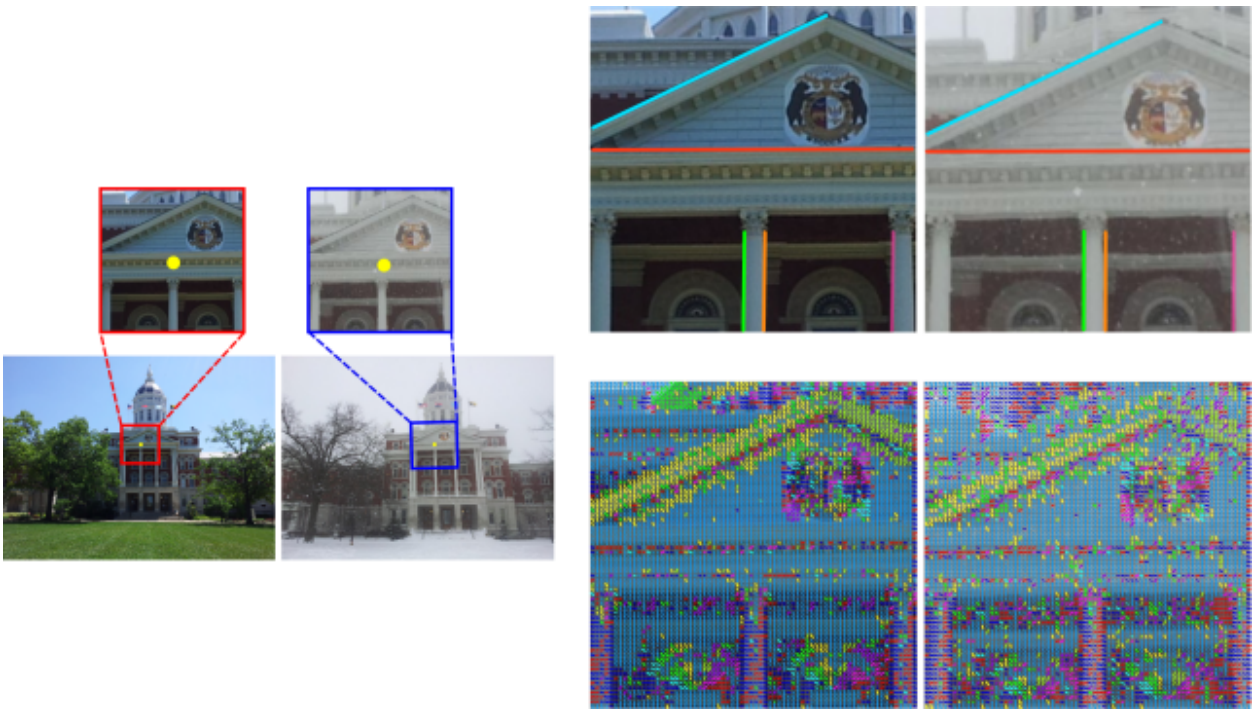

Fig. 6. Identifying matching neighborhoods for contextual matching. This example uses a SIFT match as an anchor for exploring contextual neighborhoods. *Left:* Two images are matched and share a corresponding point, shown in yellow. This keypoint pair has a distinctiveness ratio between 0.7 and 0.9 and is not verified by SIFT. *Top Right:* Matching line segments within neighborhood. *Bottom Right:* HOG descriptors for cells in the neighborhood. Cells are colored according to dominant gradient direction. The line orientation in each cell shows the cell's dominant gradient direction.

matching neighborhoods across all levels of the pyramid. If the average of all these distances is under a defined threshold ($threshold = 0.5$ for our experiments) we say that the neighborhoods surrounding the match have a strong correspondence and save the keypoint match. By requiring a region to have matching HOG's along multiple scales, we are placing a stricter requirement on the matching criteria thereby removing noise that may be present when only one scale is used. Figure 6 shows matching HOG cells for two corresponding neighborhoods.

This method makes no assumption about the structure of a region. It can be used to find features in both natural, highly varied regions such as areas of images containing trees as well as in areas containing uniform architecture.

D. Ensemble SIFT

The first way we incorporate contextual information in our pipeline begins with obtaining an initial set of matches for an image pair using SIFT, a very robust and scale invariant blob detector that describes the rich and varied gradient information surrounding an interest point. In the traditional SIFT pipeline, A and B are identified as a true match if the distance between their descriptors passes a distinctiveness ratio test (i.e. $\frac{||A_{des}-B_{des}||}{||A_{des}-C_{des}||} < 0.7$). When matching photographs with highly varying image properties, we may not obtain enough keypoint matches, or enough correct matches, that are very distinct to unearth a transformation mapping the images to each other.

In order to identify correct point matches that may not pass the ratio test (and increase our pool of matches), we also take into account contextual information. This entails measuring the similarities of the larger neighborhoods surrounding a potential match. Sample matching neighborhoods are shown in Figure 6. We can use our two different types of contextual descriptors to describe these larger neighborhoods, taking advantage of hard linear features and soft histograms of gradients (HOG). Our SIFT + line segment method is robust to lighting and content changes and helps by identifying salient linear features. Our SIFT + HOG method is useful for matching regions that lack dominant lines but are rich in texture. These methods are both used to clear up some of the ambiguity between possible keypoint matches. If the neighborhoods of the matches with relatively higher distinctiveness ratios ($0.7 < ratio < 0.9$) match well using one of our measures, we include them in our list of putative correspondences that are used for future registration. Figure 7 outlines how we use the ratio test for collecting matches. An example of how using our contextual features in addition to SIFT can help improve alignment is shown in Figure 8.

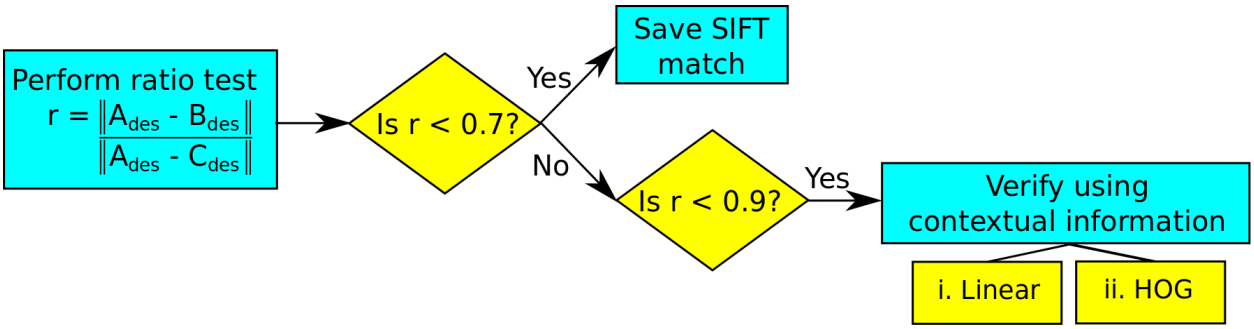

Fig. 7. Outline of how we use the distinctiveness ratio test to collect matches and to determine when to use contextual matching in conjunction with the SIFT matching scheme.

E. Contextual Harris-Laplacian

To eliminate our dependency on SIFT so that we can handle cases where an insufficient number of potential SIFT matches exists, we can also apply our contextual matching scheme to corner matching. In this scenario, we extract Harris-Laplacian corners from a pair of images and the Line Segment and HOG neighborhoods surrounding each detected keypoint. For every Harris-Laplacian corner in I_1, we compare its Line Segment neighborhood and its HOG neighborhood to the respective contextual descriptors of every corner in I_2. Neighborhoods are extracted at multiple scales as discussed in Section II-B. Pairs of corners with a sufficiently low Euclidean distance between either their Line Segment or HOG-based descriptors are saved as matches.

F. Identifying Pseudo Corners

We may also encounter cases where the scale, lighting, content, or rendering style changes between an image pair are so great that it is difficult to extract repeatable SIFT keypoints or Harris-Laplacian corners in both images, leaving us with insufficient keypoint information to work with for matching and registration. For these scenarios,

Fig. 8. Image alignment improvement using our ensemble SIFT method. *Left:* Image pair to be aligned. *Center:* Image alignment using only SIFT features. *Right:* Image alignment using our ensemble features.

we propose using line segments extracted from the images to estimate corner locations, which we refer to as Pseudo Corner keypoints. Line segments are extracted at multiple scales and merged together. Two segments are connected into one larger line segment if they are near collinear and are in close proximity. Similarly, two line segments will be merged into a single line segment if they are near parallel and are in close proximity. We can also incorporate a saliency requirement on the line segments used by thresholding out lines with average gradient magnitude values below a certain value as done in [27]. Our new corner keypoints are the intersection points between nearby segments. We avoid using the inaccurate intersection points of near parallel segments by restricting that any Pseudo Corner must be within a certain distance of each of the line segments used to find it. This process is especially tuned to planar, architectural scenes where we can have a high degree of confidence that the intersection point of segments laying on the same plane will correspond to a real corner. We match our Pseudo Corners using the same approach described in Section II-E by comparing the contextual neighborhoods of keypoints in the image pair. An example of our results using this approach to align a difficult image pair entitled Vatican is shown in Figure 9 Bottom. We can see from Figure 15 that this image pair is not registered well using the SIFT-based ensemble method or the contextual Harris-Laplacian method that are discussed earlier in this paper. Figure 9 Top shows the matches identified using our ensemble SIFT method. The middle rows show the line segments extracted from the images

and the Pseudo Corners identified. The bottom image shows the image pair aligned using matched Pseudo Corners to estimate the pair's homography.

III. Homography Estimation, Refinement, and Verification

All 2D matches identified using any form of our ensemble approach are used together to calculate a homography relating the image pair. The initial homography matrix is computed using RANSAC and Direct Linear Transform (DLT) [34]. All 2D matches found so far are geometrically verified by calculating the homography relating the image pair and identifying inlying matches. Multiple homographies are calculated for each image pair to take into account various 3D planes that may be represented in the images [35]. Each homography matrix H_i is computed using RANSAC and the chosen solution is the transformation matrix supported by the largest consensus set among the 2D matches. Matches that support a particular homography are removed and are not considered in future homography calculations.

Our iterative approach tries to find a balance between single, global homography calculations that are limited to planar-approximate scenes and multiple-homography methods that explicitly require a planar segmentation of photographs [36], [37] or more complicated image deformation models [38], [39], [40], [41], [42]. Generally, our iterative approach chooses matches belonging to different planes on different iterations as shown in Figure 10, but we do not complicate our method by enforcing such constraints. We conduct iterative homography fitting for the remaining features (i.e. features that are considered outliers based on the homographies estimated in the previous loops).

We take a two step approach to refine our image alignment that deviates from the popular approach of reestimating each H_i based on the reprojection error of the match set. We begin by searching for a dense set of matches on a relatively local scale. This first stage gives sets of keypoints the flexibility to search for correct matches along a variety of directions and distances. This is followed up by computing an optimal global alignment that is more rigid, but represents a general consensus for H_i throughout the image planes.

For refinement with local information, we apply the iterative closest point (ICP) algorithm [43] on the edge and corner pixels of I_1 and find their best matches in I_2. Edge points are identified by extracting line segments and uniformly sampling them. Ideally, by finding edge points using the Line Segment Detector algorithm, we are working with "strong" edges. Edges and corners in I_1 are transformed to the matching image plane using H_i and are compared to all of I_2's keypoints within a specified search radius. Pairs of keypoints with SIFT descriptor distances passing the traditional distinctiveness ratio test are saved. H_i is re-estimated using RANSAC and DLT. The search radius for matching keypoints plays an important role in our optimization algorithm. If the current homography estimation is far away from the optimum, we would like to have a large initial radius. On the other hand, if the current estimation is close to the optimum, a smaller radius is all that is necessary. However, this information is not known a priori, so we start a relatively large radius (10% of $max(image_{width}, image_{height})$) and divide it in half for each iteration of our refinement.

After reestimating H_i, we check the new alignment by comparing HOG's across the entire overlapping region

of the two images within the bounds set by the matches used to compute H_i. If the relative HOG distance after performing ICP is higher than that of the alignment before running ICP, the new alignment is rejected. This is essentially a quality control step to ensure that the refinement stage does not lead to an alignment that diverges from the best solution we can find.

We next use gradient information from the image pair to perform one last global alignment. The gradient magnitude field for I_1 and the binary edge image for I_2 are computed. We wish to find the homography matrix that aligns the edge points of each image with the strongest gradient responses in the matching image as much as possible. The varying strengths of I_1's gradient field are used to guide I_2's edges to their ideal alignment and vice versa. To accomplish this, H_i is perturbed using Levenberg-Marquardt Optimization to maximize Equation 2. Only edge points that are within the bounded region of the image that H_i represents are considered in Equation 2.

$$E = \arg\max_{H_i} \left(\sum_{j=0}^{\#\ I_2\ \text{edges}} \| \bigtriangledown I_1(H_i' pt_j) \| + \sum_{k=0}^{\#\ I_1\ \text{edges}} \| \bigtriangledown I_2(H_i pt_k) \| \right) \tag{2}$$

Figure 11 displays gradient information that can guide our alignment and the result of using this approach. This second stage of refinement is applied at the end of our pipeline after we have already established a relatively strong estimate of the image pair alignment. Line segments in one image need to be transformed closely to a linearly shaped area of relatively strong gradient responses in the matching image for these pixels to contribute to the refining of the image alignment. Equation 2 tries to find a general agreement amongst the transformation of all the line segments in the overlapping regions of the images. The combination of using the Line Segment Detector method, which is robust to noise, and searching for a general consensus amongst all the line segments in the images helps our optimization avoid being overly influenced by noise and clutter in the images. Figures 1, 9, and 11 show examples of lines extracted using the Line Segment Detector method. We can see in these cases that the lines picked up mainly correspond to structures that exist in both images despite very different qualities in the photograph pairs. We pick up very few lines along the ground, trees, people, or other clutter or noise in the images.

The benefit of using these types of refinement techniques instead of trying to simply minimize a match set's reprojection error is apparent when the initial point correspondences are very sparse or confined to one section of the image pair. In this situation, we are very limited in how much we can actually refine our transformation. On the other hand, our refinement approach that considers edges, corners, lines, and gradient information takes into account visual properties spanning the majority of the image planes regardless of the distribution of our point correspondences.

We have several different criteria for verifying each H_i. We first make sure that the points used to calculate the homography matrix are not collinear. Once the matrix is computed, we also take into account the layout of the transformed images and the numerical properties of the homography matrix. To check against our layout criteria, we project the four corners of I_2 to I_1's image plane using H_i. The top left and bottom left corners must remain

to the left of the top and bottom right corners, and the top left and right corners must remain above the bottom left and right corners after the transformation. If this condition is not met, H_i is discarded. We also calculate the determinant of our normalized H_i. Experiments have shown that this value should not be close to zero [35].

IV. Results and Discussion

We ran our pipeline on many of the image pairs provided in the architectural symmetry dataset [12] which contains images with a variety of visual discrepancies. This dataset includes paintings, drawings, and historical photographs matched to modern day images and several image pairs taken at very different times. Some image pairs are obtained under different lighting conditions, including changes in season and time of day, and have scale and perspective changes. The dataset includes ground truth homographies for aligning image pairs. We ran several tests on this dataset to determine the effectiveness of the first version of our pipeline which builds upon a local feature matching technique such as SIFT. The first of these tests is shown at the top of Figure 13. The goal of these tests and Figure 13 is to show how we can increase the accuracy and robustness of a widely used feature matching technique with our contextual framework. This test consisted of stitching the images using the homography calculated from four different matching techniques. We show results for using traditional SIFT matches with a 0.7 distinctiveness ratio, SIFT matches with a 0.9 distinctiveness ratio, our ensemble features, and our ensemble features with our refinement pipeline. The contextual features identified in conjunction with our ensemble features are based on SIFT keypoints that have distinctiveness ratios between 0.7 and 0.9. We use the first homography calculated with our pipeline since only one ground truth transformation is provided for each image pair. We have included both standard SIFT results to show that just increasing our search range to potential matches with ratios up to 0.9 does not account for the increased match pool and accuracy produced by our pipeline. The overall accuracy of using all SIFT matches with a distinctiveness ratio below 0.9 is very low.

For each of these four matching technique tests, we translate every pixel in I_1 to I_2 using both the ground truth homography and the first homography calculated using the method being tested. We compute the percentage of pixels that are transformed to the same coordinate. To allow for rounding errors and small inaccuracies in the ground truth data, we set a threshold on what the 2D distance between projected pixels can be to label a pixel as having been transformed correctly. This threshold is 0.6% of $max(image_{width}, image_{height})$ which is the same threshold used on 2D symmetrical transfer errors for identifying inliers in [44].

Our proposed ensemble feature compares favorably to SIFT across this dataset. From our tests, we also have determined that blindly raising the distinctiveness ratio allows too many inaccurate matches for RANSAC to find a correct homography between the image pairs and yields very unstable results. However, our approach for looking through these more ambiguous matches and considering contextual information appears to be a very valuable source of information. We increase the size of our match pool, making it easier to find a correct alignment, without letting in so much noise that RANSAC is overwhelmed by the outliers.

We also performed a test to observe what percentage of the keypoint matches identified in the different techniques are indeed correct. These results are shown in Figure 13 Center. From this chart we can see that, in general, adding in

contextually verified matches increases the percentage of accurate correspondences in the pool of matches. In theory this should increase the odds of a correct aligning transformation being calculated using RANSAC. In conjunction with this, we also charted the number of each type of match that was used in Figure 13 Bottom.

In Figure 12, we show the precision-recall curves for four different image pairs in the symmetry dataset. These curves were obtained by increasing the distinctiveness ratio (r) for accepting matches. r increases from left to right. We show the curves for using just SIFT features, using our ensemble features, and using only our contextually-verified features. The ensemble feature contains SIFT features with ratios under r and contextually-verified features found within a distinctiveness ratio range of $[r,r+0.2)$. The contextual curve shows only SIFT features that passed either the linear or HOG requirements for all matches with ratios under r. The contextual and ensemble features perform quite well under this metric. Combining the precision-recall measurements with those shown in Figure 13 demonstrates that the contextual information we propose using can complement local keypoint matching techniques very well.

In Figure 15, we also provide the image alignment accuracy results for each of the three versions of our pipeline. These consist of building on SIFT and creating an ensemble feature of SIFT and contextual descriptors, matching Harris-Laplacian corners with contextual descriptors, and matching our proposed Pseudo Corners with contextual descriptors. This chart shows the alignment accuracy using the initial set of matches from each of these methods and the accuracy after refinement is applied. We can see here that each of the images tested can be aligned very well using one of our techniques. Our contextual descriptor framework can be used successfully in combination with a variety of keypoint detectors.

Further tests of our contextual corner methods are shown in the supplementary material in which we used both our Harris-Laplacian-based method and Pseudo Corner method to align images in the png-ZuBuD dataset [45], the Stanford Mobile Visual Search dataset, and the dataset provided in [11]. Visual results for several of these image pairs are shown in Figure 2 and in the supplementary material. For these tests, we compute a homography using the matches calculated from both tested techniques and report the number of matches found and the average symmetrical transfer error [34] for all of the matches using the estimated homography. These tests show that we can obtain sufficient numbers of matches between image pairs with a wide variety of visual properties to find an aligning transformation. They also show, via their low average symmetrical transfer error, that the majority of the matches found agree with each other about the value of the solved homography matrix. This indicates that the matches found are structurally consistent.

To test our proposed feature detector, Pseudo Corners, we computed its repeatability scores on the Oxford Affine Covariant Region Detectors Dataset [5] which provides image sets demonstrating changes in perspective, lighting, scale, and JPEG compression. We compared the repeatability of our Pseudo Corners to that of SIFT and Harris-Laplacian. These values are shown in Figure 14. Our Pseudo Corners method tends to out perform SIFT and Harris-Laplacian on all of the image sets except “bark” and “wall”. Given the nature of Pseudo Corners, which looks for the intersections of line segments, it is reasonable that it does not perform as well on these natural images.

We have also included several visual examples of image registration using our pipeline (Figure 2) for photographs

from our own dataset, the symmetry dataset, and several other public datasets (see supplementary material). These image pairs present challenges including changes in season and content, lighting, scale, camera orientation, blur, and rendering styles. The runtimes to register many of these image pairs using our pipeline are also provided in the supplementary material. Both quantitative and qualitative comparisons of our method to other feature extraction and matching techniques including MSER, SIFT Flow [46], and local symmetry features [12] are presented in the supplementary material as well.

A. Limitations and Future Work

Depending on the type of data used, our pipeline may face a few roadblocks. If only a small percentage of the identified ensemble features are actually inliers, it is very unlikely that we will find a correct image alignment using RANSAC. To address this in the future, we will explore using techniques that have expanded on RANSAC to work with extremely noisy data such as [47].

One of the reasons we may encounter this problem is that fact that, currently, our pipeline does not incorporate affine invariant descriptors and is limited to relatively small baselines. One option for addressing this is to include an affine invariant descriptor such as ASIFT [8] to expand our work. Any keypoint matching technique or combination of techniques can be inserted into this pipeline.

V. Conclusion

We have presented a contextual framework for describing and matching keypoints' neighborhoods for the purpose of aligning image pairs. Our framework can be used to build upon and complement local keypoint matching techniques, creating an ensemble feature, or can be used independently to describe and match keypoints. In addition to potentially incorporating raw local keypoint matches provided by existing techniques, we can use various keypoint detectors as anchors to study larger regions surrounding keypoints in terms of salient line segments and texture-rich HOG information. We have studied, tested, and presented results for using our framework in combination with several different keypoint descriptors, including our newly proposed Pseudo Corners, to demonstrate that it is very flexible and not dependent on any one method or type of information. By using both local and regional descriptors, we are able to identify small unique regions surrounded by clutter and to clear up potential ambiguity during feature matching by looking at a larger subset of the image. This pipeline has been designed to combine these various feature extraction and matching methods in such a way that it is robust under lighting, content, and rendering changes. We have shown quantitatively that using this information can increase the accuracy of identified matches and image alignment over using SIFT alone as well as can be used to accurately describe Harris-Laplacian corners and our proposed Pseudo Corners. Our technique can work on images containing a variety of architectural styles, man-made features, and natural content without making assumptions about the general structure of a scene and can handle different artistic representations of the same location.

ACKNOWLEDGMENT

This work is supported in part by the NSF CC-NIE award #1245795, NSF CMMI award #1039433, and the NSF Graduate Research Fellowship award #0943941. The authors would like the thank the reviewers of this paper for all of their helpful comments and suggestions. The authors are grateful to the researchers who provided the public datasets in [45], [5], [13], [48], [11], [49], [12] and code [46], [12] that we used for testing and evaluating the work in this paper. The authors also would like to thank the engineers and researchers who designed and implemented the OpenCV [50] and VLFeat [51] libraries which were incorporated into this work.

REFERENCES

[1] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. Journ. of Comp. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journ. of Comp. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[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] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends® in Comp. Graphics and Vis.*, vol. 3, no. 3, pp. 177–280, 2008.

[5] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. Journ. of Comp. Vis.*, vol. 65, no. 1-2, pp. 43–72, 2005.

[6] C. Harris and M. Stephens, "A combined corner and edge detector." in *Proc. Alvey Vis. Conf.*, vol. 15. Manchester, UK, 1988, p. 50.

[7] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *IEEE Proc. Int. Conf. on Comp. Vis.*, vol. 1, 2001, pp. 525–531.

[8] J. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journ. on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Euro. Conf. on Comp. Vis.* Springer, 2006, pp. 404–417.

[10] C. Wu, F. F. J.-M., Frahm, and M. Pollefeys, "3d model search and pose estimation from single images using vip features," in *IEEE Comp. Vis. and Pattern Recongnit. Workshops*, 2008, pp. 1–8.

[11] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros, "Data-driven visual similarity for cross-domain image matching," in *ACM Trans. on Graphics*, vol. 30, no. 6, 2011, p. 154.

[12] D. Hauagge and N. Snavely, "Image matching using local symmetry features," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2012, pp. 206–213.

[13] G. Yang, C. V. Stewart, M. Sofka, and C. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 29, no. 11, pp. 1973–1989, 2007.

[14] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," in *ACM ACM Trans. on Graphics*, vol. 30, no. 4, 2011, p. 70.

[15] P. Forssen and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *IEEE Int. Conf. on Comp. Vis.*, 2007, pp. 1–8.

[16] R. Kimmel, C. Zhang, A. M. Bronstein, and M. M. Bronstein, "Are mser features really interesting?" *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 33, no. 11, pp. 2316–2320, 2011.

[17] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[18] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends® in Comp. Graphics and Vis.*, vol. 2, no. 1, pp. 1–104, 2006.

[19] Q. Li, G. Qu, and Z. Li, "Matching between sar images and optical images based on hog descriptor," in *Radar Conf.* IET Int., April 2013, pp. 1–4.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Comp. Vis. and Pattern Recongnit.*, vol. 1, 2005, pp. 886–893.

[21] G. Schindler, P. Krishnamurthy, R. Lublinerman, Y. Liu, and F. Dellaert, "Detecting and matching repeated patterns for automatic geo-tagging in urban environments," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2008, pp. 1–7.

[22] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2007, pp. 1–8.

[23] S. Singh and A. G. A. A. Efros, "Unsupervised discovery of mid-level discriminative patches," in *Euro. Conf. on Comp. Vis.* Springer, 2012, pp. 73–86.

[24] M. Aubry, B. C. Russell, and J. Sivic, "Painting-to-3d model alignment via discriminative visual elements," *ACM Trans. on Graphics*, vol. 33, no. 2, p. 14, 2014.

[25] B. C. Russell, J. Sivic, J. Ponce, and H. Dessales, "Automatic alignment of paintings and photographs depicting a 3d scene," in *IEEE Int. Conf. on Comp. Vis. Workshops*, 2011, pp. 545–552.

[26] L. Wang, U. Neumann, and S. You, "Wide-baseline image matching using line signatures," in *IEEE Int. Conf. on Comp. Vis.*, 2009, pp. 1311–1318.

[27] B. Fan, F. Wu, and Z. Hu, "Line matching leveraged by point correspondences," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2010, pp. 390–397.

[28] H. Bay, V. Ferrari, and L. V. Gool, "Wide-baseline stereo matching with line segments," in *IEEE Comp. Vis. and Pattern Recongnit.*, vol. 1, 2005, pp. 329–336.

[29] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Distinguished regions for wide-baseline stereo," *Center for Machine Perception-K333 CVUT, Praha, Tech. Rep.*, 2001.

[30] V. Gioi, R. Grompone, J. Jakubowicz, J. Morel, and G. Randall, "Lsd: a line segment detector," *Image Processing On Line*, 2012.

[31] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering objects and their location in images," in *IEEE Int. Conf. on Comp. Vis.*, vol. 1, 2005, pp. 370–377.

[32] Z. Wu, Q. Ke, M. Isard, and J. Sun, "Bundling features for large scale partial-duplicate web image search," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2009, pp. 25–32.

[33] C. Stachniss, "C implementation of the hungarian method," 2004. [Online]. Available: http://www.informatik.uni-freiburg.de/ stachnis/resources.html

[34] R. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge, United Kingdom: Cambridge University Press, 2010.

[35] T. Vincent and R. Laganiére, "Detecting planar homographies in an image pair," in *IEEE Proc. Image and Signal Processing and Anal.*, 2001, pp. 182–187.

[36] F. Fraundorfer, K. Schindler, and H. Bischof, "Piecewise planar scene reconstruction from sparse correspondences," *Image and Vis. Computing*, vol. 24, no. 4, pp. 395 – 406, 2006.

[37] D. S. Kumar and C. V. Jawahar, "Robust homography-based control for camera positioning in piecewise planar environments," in *Comp. Vis., Graphics and Image Processing*. Springer, 2006, pp. 906–918.

[38] W. Lin, S. Liu, Y. Matsushita, T. Ng, and L. Cheong, "Smoothly varying affine stitching," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2011, pp. 345–352.

[39] J. Zaragoza, T. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving dlt," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2013, pp. 2339–2346.

[40] Y. Lipman, S. Yagev, R. Poranne, D. W. Jacobs, and R. Basri, "Feature matching with bounded distortion," *ACM Trans. on Graphics*, vol. 33, no. 3, p. 26, 2014.

[41] W. D. Lin, M. Cheng, J. Lu, H. Yang, M. N. Do, and P. Torr, "Bilateral functions for global motion modeling," in *Euro. Conf. on Comp. Vis.* Springer, 2014, pp. 341–356.

[42] H. Yang, W. Lin, and J. Lu, "Daisy filter flow: a generalized discrete approach to dense correspondences," in *IEEE Comp. Vis. and Pattern Recongnit.*, 2014, pp. 3406–3413.

[43] P. Besl and N. McKay, "Method for registration of 3-d shapes," in *Proc. SPIE Sensor Fusion IV: Control Paradigms and Data Structures*. Int. Society for Optics and Photonics, 1992, pp. 586–606.

[44] N. Snavely, S. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *Int. Journ. of Comp. Vis.*, vol. 80, no. 2, pp. 189–210, 2007.

[45] H. Shao, T. Svoboda, and L. V. Gool, "Zubud-zurich buildings database for image based recognition," *Comp. Vis. Lab, Swiss Federal Institute of Technology, Switzerland, Tech. Rep*, vol. 260, 2003.

[46] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Pattern Anal. and Mach. Intell.*, vol. 33, no. 5, pp. 978–994, 2011.

[47] L. Moisan, P. Moulon, and P. Monasse, "Automatic homographic registration of a pair of images, with a contrario elimination of outliers," *Image Processing On Line*, vol. 10, 2012.

[48] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Comp. Vis. and Pattern Recongnit.*, 2007.

[49] V. R. Chandrasekhar, D. M. Chen, S. S. Tsai, N. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeszczuk, and J. Bach, "The stanford mobile visual search data set," in *ACM Multimedia systems*, 2011, pp. 117–122.

[50] G. Bradski, "Opencv library," *Dr. Dobb's Journ. of Software Tools*, 2000.

[51] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of comp. vis. algorithms," http://www.vlfeat.org/, 2008.

Brittany Morago Brittany Morago received a BS degree in Digital Arts and Sciences from the University of Florida in 2010. She is currently working towards her PhD in Computer Science at the University of Missouri-Columbia and is a recipient of NSFGRF and GAANN Fellowships. Her research interests include computer vision and graphics.

Giang Bui Giang Bui received the BS and MS degrees from Vietnam National University of Hanoi city in 2004 and in 2007 respectively. He is currently a graduate student of the University of Missouri - Columbia. He works as a research assistant in the Computer Graphics and Image Understanding laboratory mentored by Dr. Ye Duan. His research interests include image and video processing, 3D computer vision, and machine learning.

Ye Duan Ye Duan is an Associate Professor of Computer Science at University of Missouri-Columbia. He received his BA degree in Mathematics from Peking University in 1991. He received his MS degree in Mathematics from Utah State University in 1996. He received his MS and PhD degree in Computer Science from the State University of New York at Stony Brook in 1998 and 2003. From September 2003 to August 2009, he was an Assistant Professor of Computer Science at University of Missouri-Columbia. His research interests include Computer Graphics and Visualization, Biomedical Imaging and Computer Vision.

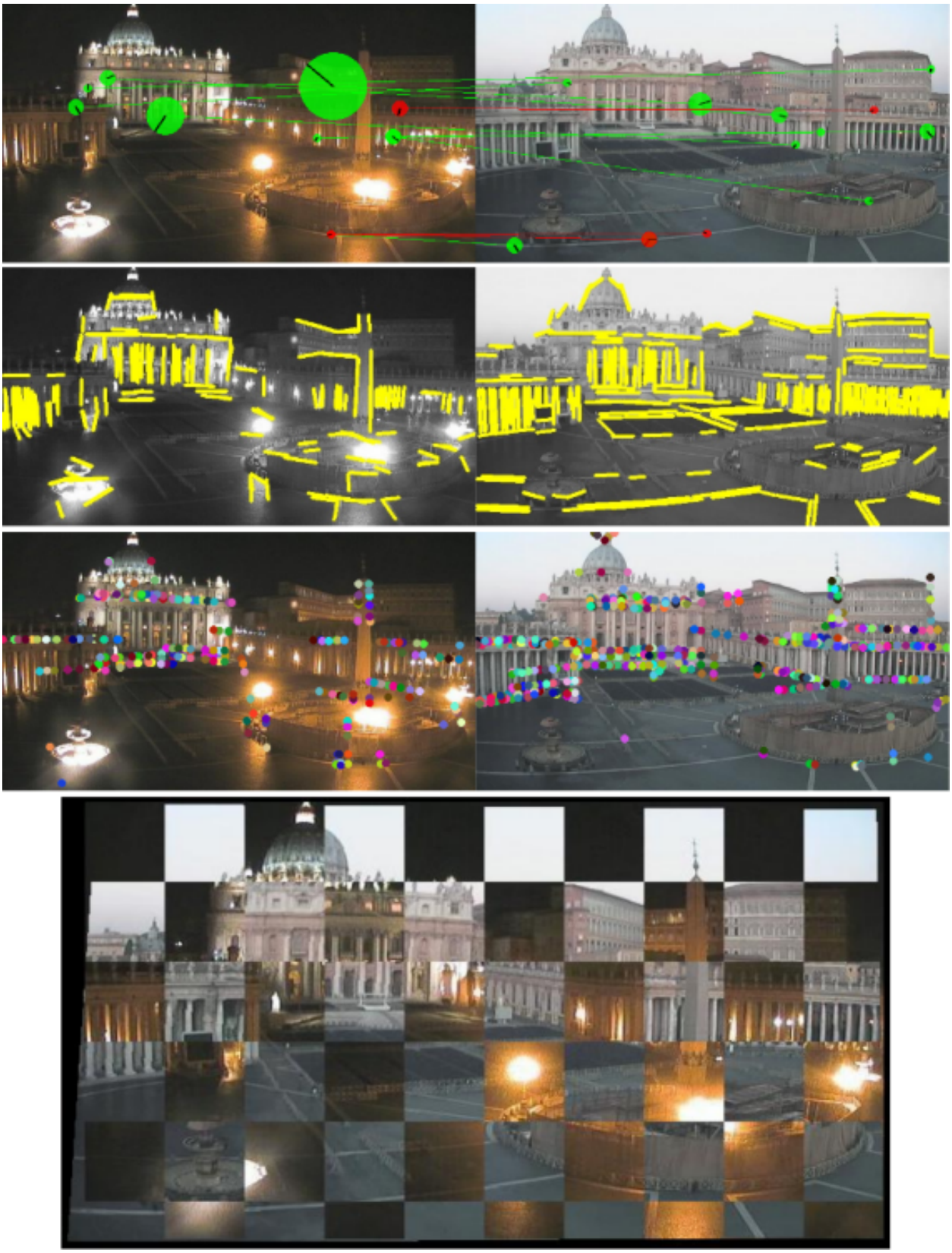

Fig. 9. Using Pseudo Corners to align a difficult image pair, called Vatican. *Top:* The majority of the SIFT (green) and linearly-verified (red) keypoint matches for this pair are incorrect and cannot be used to align the images. No SIFT+HOG matches were identified in this example. The relative SIFT scales are represented by the size of the circles and the features' orientations are denoted by the black lines. *2nd Row:* Line segments extracted from images (shown in yellow) that are used to find Pseudo Corners. *3rd Row:* Pseudo Corners identified after intersecting line segments. *Bottom:* Image pair is aligned using the homography calculated from the matched Pseudo Corners.

Fig. 10. Matches identified using multiple homographies. Each color corresponds to a different homography. We can see how each homography tends to find matching points in different regions and planes of the image.

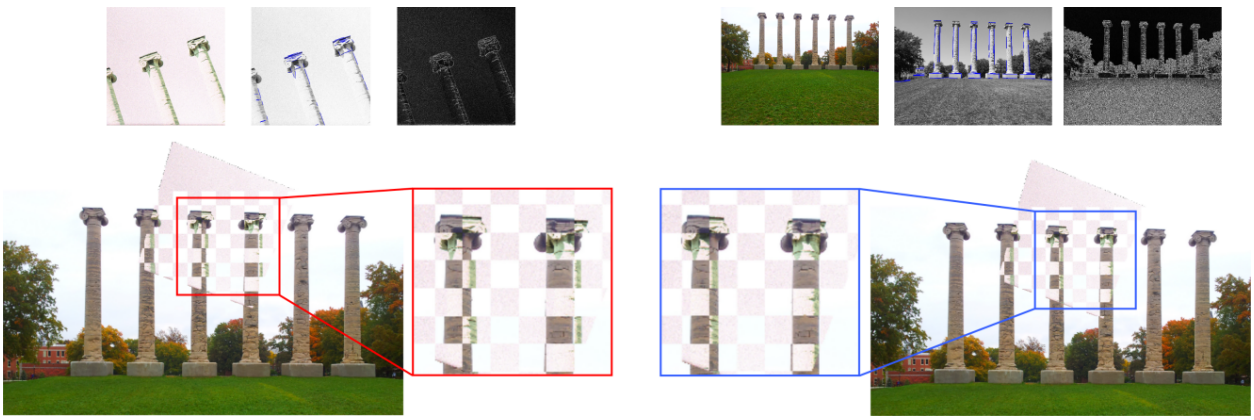

Fig. 11. Using gradient information to refine homography. *Top:* Two original images and their edge images and gradient field images. We aim to align the edge and gradient field images as well as possible. *Bottom Left:* Stitching result before our refinement method is applied. *Bottom right:* Result after our refinement is applied. *Bottom middle:* Zoomed in views of the area where the alignment is corrected through refinement.

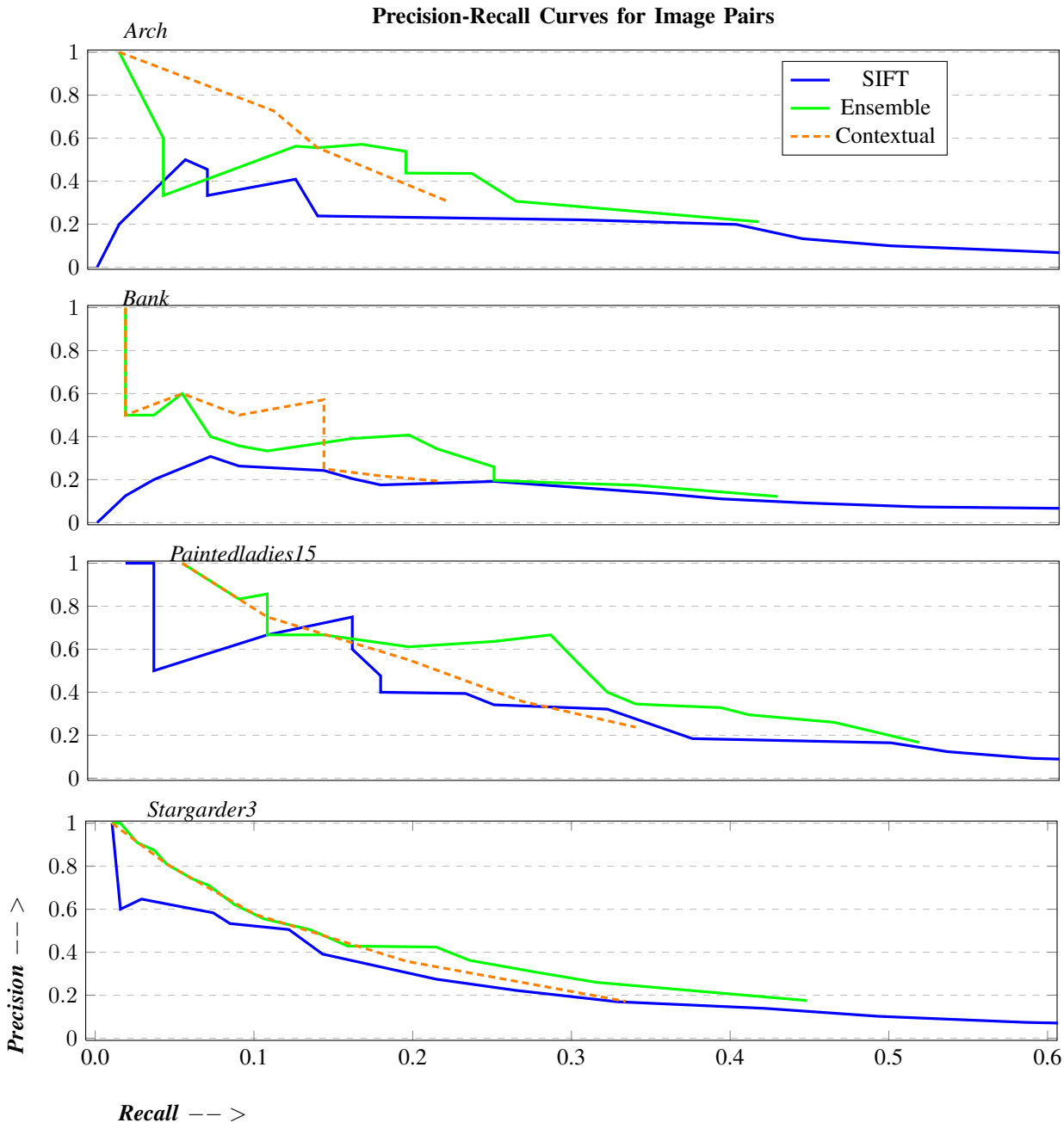

Fig. 12. Precision-recall curves for four different image pairs from [12]. Curves are provided for using SIFT matches, our ensemble feature, and just the contextually verified SIFT features identified using our method. Data points correspond to increasing distinctiveness ratios moving from left to right.

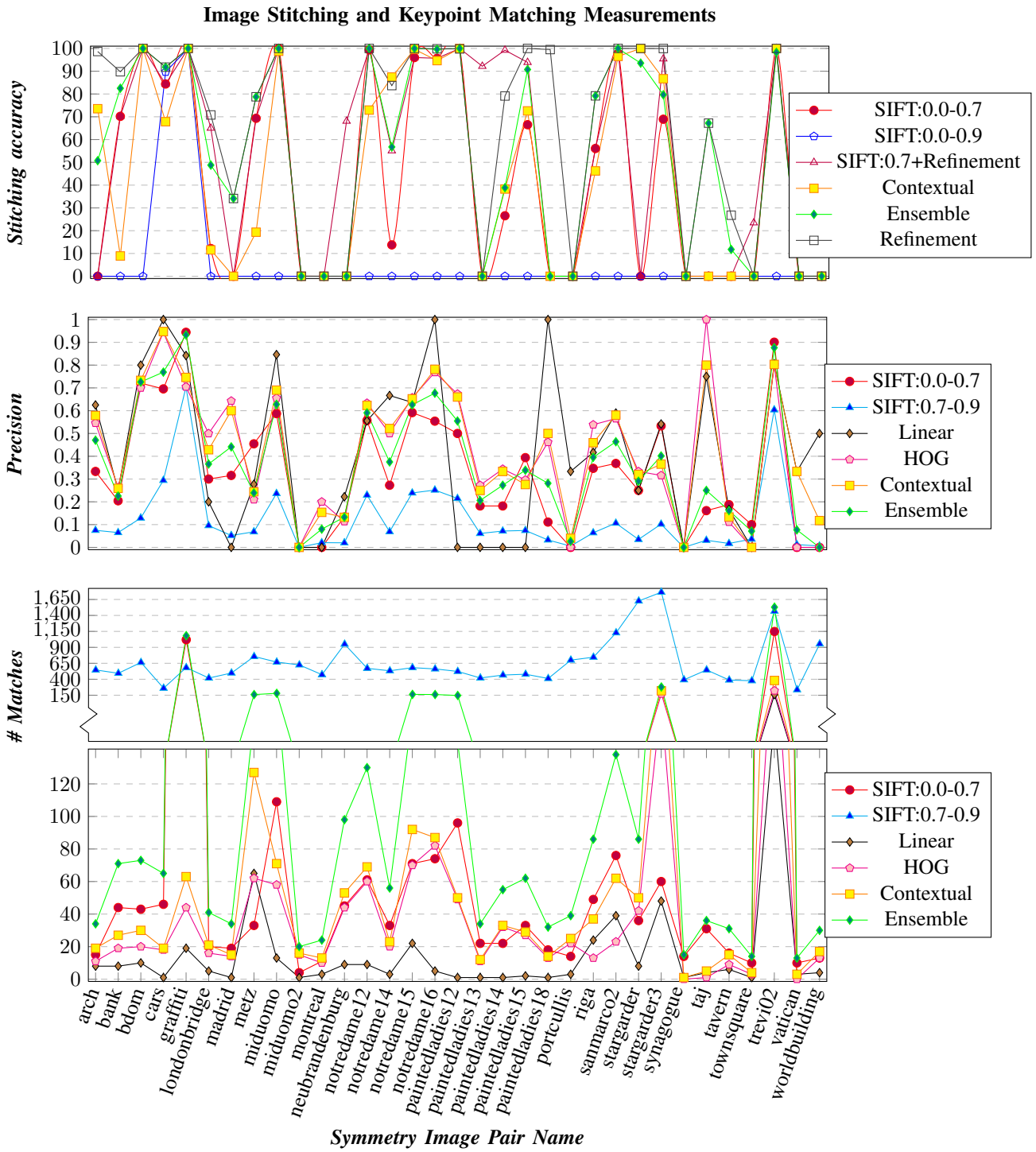

Fig. 13. Quantitative results for stitching image pairs. *Top:* Percentage of pixels in image plane that are correctly aligned after using solved homography. *Center:* Percentage of matches that are identified using different methods that are correct. *Bottom:* Number of matches identified using several different methods.

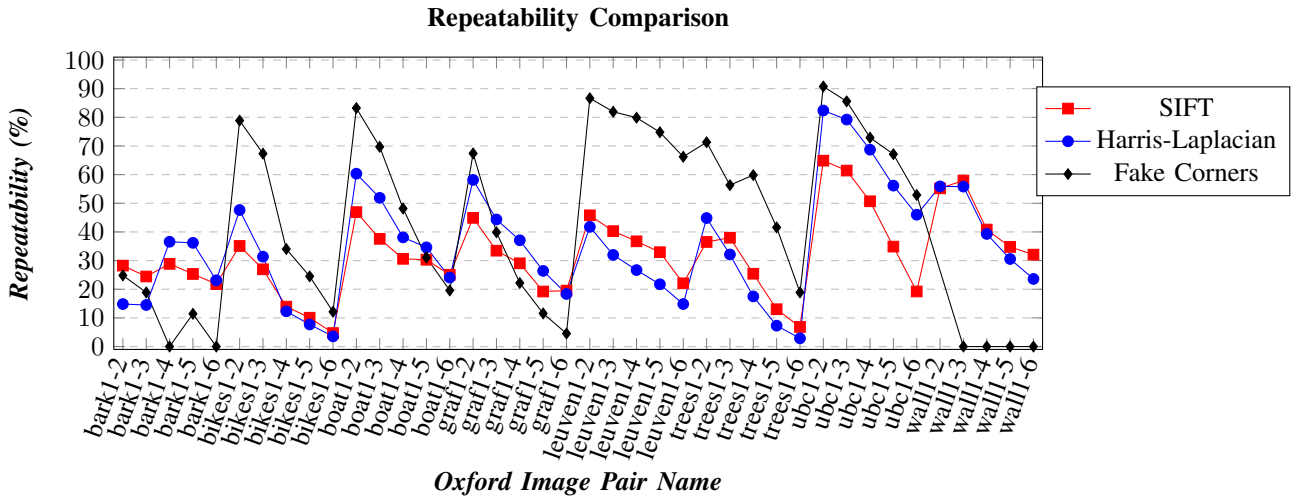

Fig. 14. Repeatability measurements for SIFT, Harris-Laplacian, and Pseudo Corners on Oxford Affine Covariant Region Detectors Dataset [5]. The sets of images test the following changes in image properties: bark - scale and image rotation changes on textured scene, bikes - increased blur on a structured scene, boat - scale changes, graf - viewpoint changes, leuven - illumination changes, trees - increased blur on a structured scene, ubc - changes in JPEG compression, wall - viewpoint angle changes on textured scene.

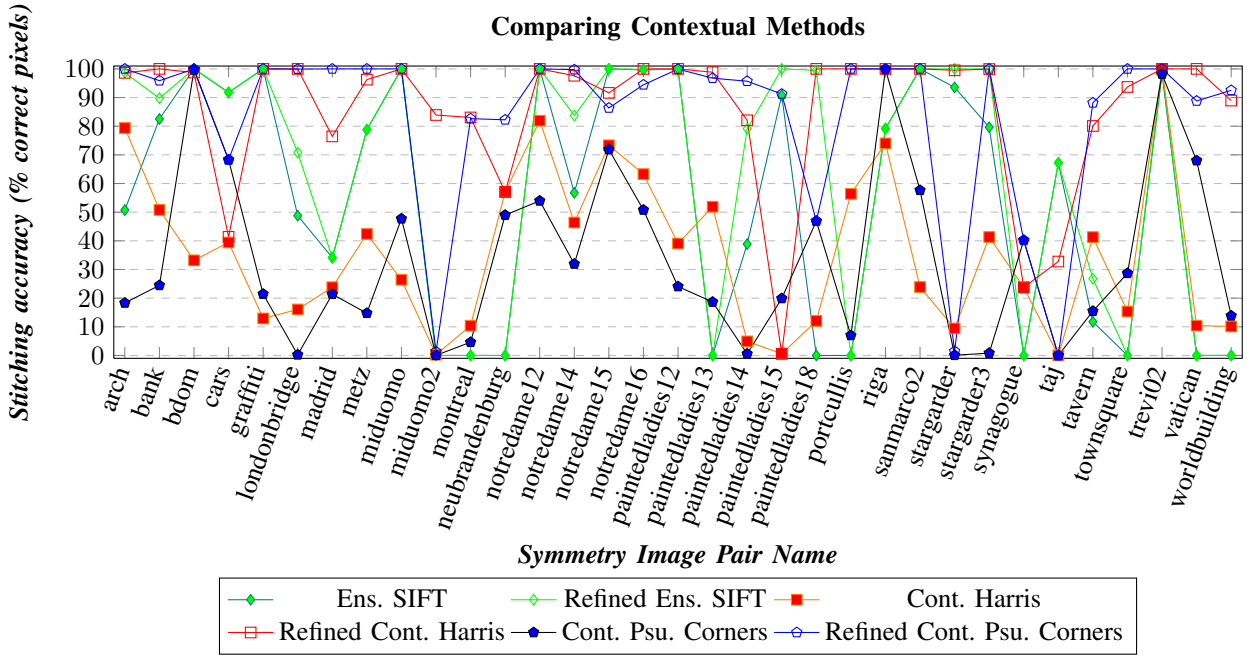

Fig. 15. Comparing our contextual methods. Using the symmetry dataset [12], we show the alignment accuracy using our ensemble SIFT features before and after refinement, contextual Harris-Laplacian before and after refinement, and contextual pseudo corners before and after refinement. As mentioned in Section I-C, ensemble features refer to the version of our method that uses an existing keypoint descriptor, such as SIFT, in addition to our contextual linear and contextual HOG information for matching. Contextual features use only our contextual descriptor for matching. The accuracy is measured in terms of the percentage of pixels in I_1 that are transformed to the correct location in I_2 using the homography estimated from the match set.