Joint Geometric Graph Embedding for Partial Shape Matching in Images

Anirban Mukhopadhyay*
Zuse Institute Berlin
Berlin, Germany
anirban.akash@gmail.com

Arun CS Kumar*
Suchendra M. Bhandarkar
Department of Computer Science, The University of Georgia
Athens, Georgia, 30602-7404, USA
{aruncs@uga.edu, suchi@cs.uga.edu}

(* indicates equal contribution)

Abstract

A novel multi-criteria optimization framework for matching of partially visible shapes in multiple images using joint geometric graph embedding is proposed. The proposed framework achieves matching of partial shapes in images that exhibit extreme variations in scale, orientation, viewpoint and illumination and also instances of occlusion; conditions which render impractical the use of global contour-based descriptors or local pixel-level features for shape matching. The proposed technique is based on optimization of the embedding distances of geometric features obtained from the eigenspectrum of the joint image graph, coupled with regularization over values of the mean pixel intensity or histogram of oriented gradients. It is shown to obtain successfully the correspondences denoting partial shape similarities as well as correspondences between feature points in the images. A new benchmark dataset is proposed which contains disparate image pairs with extremely challenging variations in viewing conditions when compared to an existing dataset [18]. The proposed technique is shown to significantly outperform several state-of-the-art partial shape matching techniques on both datasets.

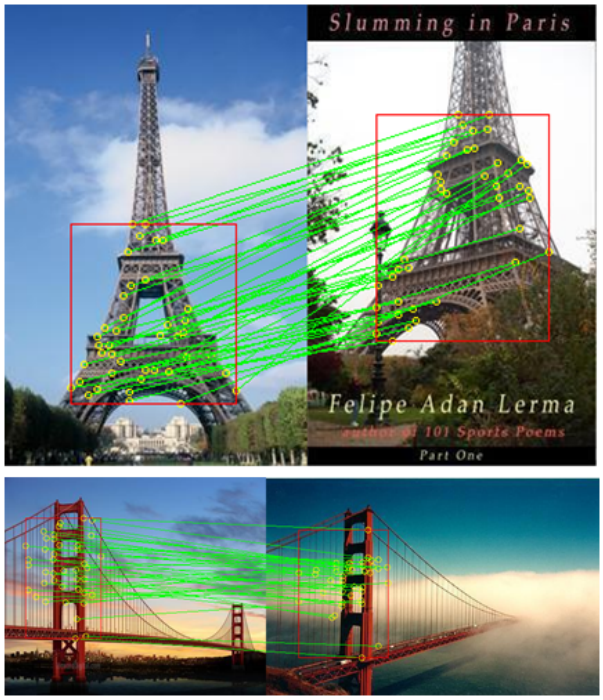

Figure 1. Partial shape matching in real world images. The result of the proposed technique is denoted by a red rectangle. Feature points are denoted using yellow circles whereas the feature point correspondences in the partial matches are shown using green lines.

1. Introduction

In spite of extensive research on shape matching over the years, obtaining a reliable and accurate correspondence between shapes in images that exhibit considerable variation in viewing conditions continues to pose an extremely difficult problem. In this paper, we propose a multicriteria optimization-based framework for matching partially visible shapes in images exhibiting significant variations in viewing conditions. Given a set of features extracted from the images under consideration, our goal is to obtain accurate and optimal region-based correspondences for partial shape matching. Formal characterization of the partial shape matching problem is of particular importance because of its wide applicability in several computer vision problems such as structure-from-motion, object localization, fine-grained object categorization, shape-based image retrieval and image registration, to name a few.

Classical approaches to the partial shape matching problem can be broadly categorized as *global* or *local* based on the granularity of the image features employed. Most global shape matching techniques are seen to lack the ability to handle strong articulation, deformation or occlusion of objects (or their parts). Local feature-based shape matching techniques, on the other hand, are more robust to articula-

tion, deformation or occlusion of objects but fare poorly in the face of significant variations in illumination, scale, orientation and viewpoint and often fail to provide a strong global description for accurate shape alignment. Also, in most existing shape matching techniques, the underlying shape under consideration is often that of a very distinct object in an image captured in a highly controlled environment; an assumption that may not hold in many real-world images.

In this paper, we focus on the problem of partial shape matching in natural real-world images where the object shapes exhibit significant variations on account of variations in viewing conditions and do not exhibit the same degree of saliency as they would if imaged in a highly controlled environment. In the proposed framework, we do not rely completely on either local shape matching methods based entirely on feature correspondences [12], [21] or on global shape matching techniques based on optimization of a global cost function [7]. Instead, we propose to use a multicriteria optimization technique for matching of persistent image features, obtained via geometric embedding of the joint image graph, that encode the appearance similarity of objects or regions in a robust manner. Accurate and robust partial shape matches are obtained across images that exhibit considerable variation in viewing parameters via optimization of a suitably defined distance measure between jointly embedded local feature vectors coupled with a global regularization constraint. Since the proposed optimization criterion incorporates both, global information encoded in the joint image graph and the regularization constraint, and local information encoded in the mean pixel intensity and intensity gradient information computed in a local neighborhood of the feature point, it is shown to not only retrieve the region-wise partial similarity matching between the shapes, but also provide robust and accurate pixel-level correspondences between the shapes under consideration.

The primary contributions of this paper are:

(1) The formulation of a novel multicriteria optimization framework based on joint geometric graph embedding (JGGE) coupled with a global regularization constraint to address the hitherto unsolved problem of partial shape matching across images where the variations in imaging and viewing parameters are truly challenging.

(2) The introduction of a new benchmark dataset wherein the variability in the images ranges over several imaging and viewing parameters such as illumination (day versus night), viewpoint, age of structures (historic versus new), presence of occlusion, presence of partially constructed structures, and inclusion of sketches and/or paintings of objects along with their captured images.

The proposed technique is experimentally verified on the aforementioned benchmark dataset and an existing benchmark dataset [18]. The experimental results clearly demonstrate the substantial improvement in the matching of partial and complete shapes obtained using the proposed multicriteria optimization technique.

2. Related work

Conventional global shape matching techniques, which include contour-based and region-based methods, compare two shapes by computing a suitably defined similarity or distance measure and a global matching cost function which is then optimized to determine the best match. Contour-based methods exploit primarily the boundary information for matching the underlying shapes. *Shape context*-based contour matching [7], chamfer distance measure-based contour matching [23], [32], dynamic programming-based contour matching [24], set matching-based contour similarity measurement [33], and *shape skeleton*-based contour matching [3] are examples of global contour-based shape matching techniques. Other works in contour-based global shape matching include the triangle area representation [1], and segment-based shape matching methods such as the shape tree method [16] and the hierarchical procrustes matching procedure [22]. Global contour-based shape matching methods though capable of capturing the global shape of the object, are unable to handle strong articulations in the object shapes.

Region-based approaches, in contrast, derive the global shape descriptors using pixel-level information within a shape region. Some region-based global shape matching methods are based on the computation of invariant moments such as the Zernike moments [19]. Skeleton-based shape descriptors [28], [30] have proven better than conventional contour-based methods at capturing shape articulation but their performance deteriorates when dealing with complex shapes due to the absence of region-based descriptors. The general inability of global shape matching techniques to handle shape deformation arising from strong articulation or occlusion of objects (or their parts) has motivated the design of local shape matching techniques.

Local shape matching techniques attempt the address the problem of shape deformation in their underlying formulation [12], [21]. Though robust to modest shape deformation and capable of handling limited extent of shape articulation while providing an accurate measure of local similarity, local shape matching techniques are typically unable to provide a strong global description for accurate shape alignment. Also, most local matching techniques call for prior knowledge of the underlying shape when dealing with the matching of highly articulated object shapes, thus severely limiting their scalability to real-world problems [12], [21].

Although the proposed method is designed primarily to address the problem of partial shape matching in real world images with extreme variations in viewing and imaging

conditions, it has the added advantage of improving the quality of feature point correspondences. Despite the several invariant feature-based matching techniques that have been proposed for improving the correspondence between feature points, obtaining accurate and reliable point correspondences is challenging using only local versions of invariant feature descriptors such as the rotation-invariant descriptor [27] or scale- and rotation-invariant descriptor [20]. Feature point correspondences could be improved by identifying and localizing interest points such that the solutions for the positions, orientations and scale of the matching feature points are stable [31]. Patch matching techniques attempt to improve feature point correspondences by incorporating mid-level cues from region patches and their nearest neighboring patches using random sampling [5] or from multiple views of region patches [11]. However, extreme variations in viewing conditions often encountered in real-world images render the patch matching techniques infeasible. Since the proposed technique incorporates multicriteria optimization of the spectral embedding of local geometric features of the joint image graph coupled with region-based regularization, it inherently restricts the features to a shape or region resulting in improved correspondences between feature points.

Spectral methods on the Laplacian of suitably defined image graphs have been presented in recent research literature in the context of feature clustering and image segmentation [2]. The proposed technique is motivated by the recent work on the use of the eigenspectra of scale-invariant feature transform (SIFT) features [20] of the joint image graph as descriptors of image structure [4]. It has been shown that eigenspectral analysis of the joint image graph constructed using dense pixel-level SIFT features extracted from a pair of images, can yield matching features that are robust and persistent across illumination changes [4]. In particular, features that encode the extrema of the eigenfunctions of the joint image graph are shown to be stable, persistent and robust across wide range of illumination variations [4]. Although the formalism in [4] bears some resemblance to the formalism underlying the proposed technique, it is important to note that in the proposed technique, the partial shape matching is based on features that are not only robust to illumination variations but also to variations in perceived shape geometry resulting from changes in viewpoint and changes in local orientations and occlusions of object subparts. The proposed technique is also loosely motivated by the partial shape matching technique described in [10]; however, an important difference between the work in [10] and the proposed techniques is that the former deals primarily with objects whose surfaces are described by well-defined triangular meshes generated in an artificial environment, whereas the proposed technique tackles the more general and challenging domain of matching objects in real-world images that exhibit significant variability in several imaging and viewing parameters, including cases where the underlying 2D object geometry is not consistent across the images under consideration.

3. Theoretical Derivation

The proposed technique focuses on image pairs containing a *single dominant object* where the dominant object is either completely or partially visible. We denote the images constituting the image pair under consideration as X and Y. We assume that the images X and Y can each be decomposed into subparts, where the subparts are modeled as subsets $X' \subset X$ and $Y' \subset Y$ respectively. The *degree of dissimilarity* between subparts X' and Y' can be expressed as a non-negative geometric feature distance function $d : \Sigma'_X \times \Sigma'_Y \rightarrow \mathbb{R}+$, where Σ defines the geometry of the subpart. Let $\varphi : X' \rightarrow Y'$ be defined as the underlying function that is responsible for transforming the subpart X' to subpart Y'. For every point $x \in X'$, $\varphi(x) \in Y'$ represents a unique point with similar local geometry. In the continuous case, the corresponding optimization problem can be represented as:

$$(X'^*, Y'^*, \varphi^*) = \underset{X',Y';\varphi:X'\rightarrow Y'}{\arg\min} \left[\int_{X'\times Y'} d(x, \varphi(x)) dx \right] \quad (1)$$

where, the goal of the optimization in eq. (1) is to determine the optimal subparts X'^* and Y'^* as well as the optimal underlying transformation φ^*. However, in the discrete setting, X' and Y' are represented as rectangular image patches and the corresponding optimization problem is formulated as:

$$(X'^*, Y'^*, \varphi^*) = \underset{X',Y';\varphi:X'\rightarrow Y'}{\arg\min} \left[\sum_{X'\times Y'} d(x, \varphi(x)) \right] \quad (2)$$

In eq. (2) instead of minimizing the geometric feature distance $d(\cdot, \cdot)$ for all pixels of a rectangular patch, we have opted to minimize the overall $d(\cdot, \cdot)$ value between feature points. This approximation can be justified by an important result in machine learning which states that correct partial point correspondences between two manifolds can be used to infer the complete alignment between all points on the two manifolds [17]. The geometric feature distance $d(\cdot, \cdot)$ between two feature points is obtained using the JGGE framework described in Section 3.1. We introduce a regularization term $r(\cdot, \cdot)$ in the objective/energy function in eq. (2) that comprises of an intensity- or texture-based similarity measure which encapsulates the global structure of the subparts. The final multicriteria optimization framework is given by:

$$(X'^*, Y'^*, \varphi^*) = \underset{X',Y';\varphi:X'\to Y'}{\arg\min} \left[\lambda \left\{ \sum_{X'\times Y'} d(x, \varphi(x)) \right\} + (1-\lambda) r(X', Y') \right] \tag{3}$$

where $0 \leq \lambda \leq 1$.

3.1. Joint Geometric Graph Embedding (JGGE)

The spectral analysis of the contents of an image is typically performed on a weighted image graph $G(V, E, W)$ [2]. The vertices in the vertex set V denote the pixel-level features of the image. The edge set E denotes the pair-wise relationships between each pair of vertices in the set V, making G a complete graph. The weight $w_{ij} \geq 0$ associated with an edge $(v_i, v_j) \in E$ encodes the affinity between the corresponding pixel-level features represented by vertices v_i and v_j. The edge weights are represented by an $n \times n$ affinity matrix $W = [w_{ij}]_{i,j=1,2,\ldots,n}$.

The above formulation is extended for a joint graph as follows: Let $G_1(V_1, E_1, W_1)$ and $G_2(V_2, E_2, W_2)$ be the image graphs for images I_1 and I_2, respectively. The joint image graph $G(V, E, W)$ is defined by the vertex set $V = V_1 \cup V_2$ and edge set $E = E_1 \cup E_2 \cup (V_1 \times V_2)$ where $V_1 \times V_2$ is the set of edges connecting every pair of vertices in (V_1, V_2). The resulting affinity matrix W is given by:

$$W = \begin{bmatrix} W_1 & C \\ C^T & W_2 \end{bmatrix}_{(n_1+n_2)\times(n_1+n_2)} \tag{4}$$

where the affinity submatrices W_1, W_2 and C are defined as follows:

$$(W_i)_{x,y} = \exp(-(\|f_i(x) - f_i(y)\|)^2) \tag{5}$$

$$C_{x,y} = \exp(-(\|f_1(x) - f_2(y)\|)^2) \tag{6}$$

where, $f_i(x)$ and $f_i(y)$ are pixel-level features at locations x and y respectively in image I_i and $\|\cdot\|$ is the Euclidean norm or distance measure.

We propose the use of geometric blur (GB) features [8] as an alternative to SIFT features, for generating the joint image graph. The inherent ability of GB to focus on feature points on the dominant objects within the image has prompted their use, instead of SIFT features, for determining the partial geometric similarities between objects. GB features yield image descriptors that are robust to small transformations while focusing on the geometric structure of the dominant object within the image. GB averages the underlying signal over small transformations and then samples the signal at fixed locations to construct the desired robust descriptor [8]. For oriented images, the edge response has been shown to be a good underlying signal. In practice, the averaging of transformations can be modeled by convolving the signal with a kernel that weighs the contribution of neighboring signals at a given point or pixel location. Gaussian kernels where the support (σ) increases linearly with the distance from the origin are commonly used for this purpose [8].

We compute the *joint geometric graph embedding distance* (JGGED) between two feature points x and y in two different images (comprising the image pair) using the first m non-trivial eigenvectors (ϕ_k) corresponding to the m smallest non-trivial eigenvalues of the joint graph as follows:

$$d^2_{JGGED}(x, y) = \sum_{k=1}^{m} (\phi_k(x) - \phi_k(y))^2 \tag{7}$$

Bansal and Daniilidis [4] have shown that the joint geometric graph embedding (JGGE) procedure ensures very good correspondence between the shapes and distributions of the eigen extrema obtained from the two images being matched. The JGGE procedure is shown to reduce the divergence between features derived from the corresponding regions of the two images, so that regions in the two images that exhibit strong correspondence in the JGGE space are noted to be in visual agreement with the correspondence results in image space [4]. Consequently, the JGGED measure d_{JGGED} computed in the JGGE space is more robust to the variations in imaging and viewing parameters commonly encountered in real-world images than a simple feature distance measure computed in the GB space. The JGGED measure $d_{JGGED}(\cdot, \cdot)$ is used as the geometric feature distance $d(\cdot, \cdot)$ in the first term of the energy function in eq. (3).

3.2. Regularization

The quality of the match between subparts X' and Y' in the image pair under consideration can be measured using an appropriately defined region-based *irregularity function* $r(X', Y')$, where the subparts X' and Y' are rectangular image patches selected around the feature points extracted at varying resolutions from images I_1 and I_2 respectively. Minimization of the region-based irregularity term ensures a sufficiently high-quality match between the two regions instead of just the feature points, i.e., $r(X', Y')$ serves as a regularization term in the multicriteria optimization framework (eq. (3)). In particular, we compute the difference in the values of the mean pixel intensity (MPI) and histogram of oriented gradient (HOG) features between the pixels of subparts X' and Y'. The popular use of HOG features in recent computer vision literature on object detection and recognition motivated us to consider it as a potential regularizer for the regions X' and Y' in eq. (3) to improve the quality of match between the corresponding subparts in the image pair.

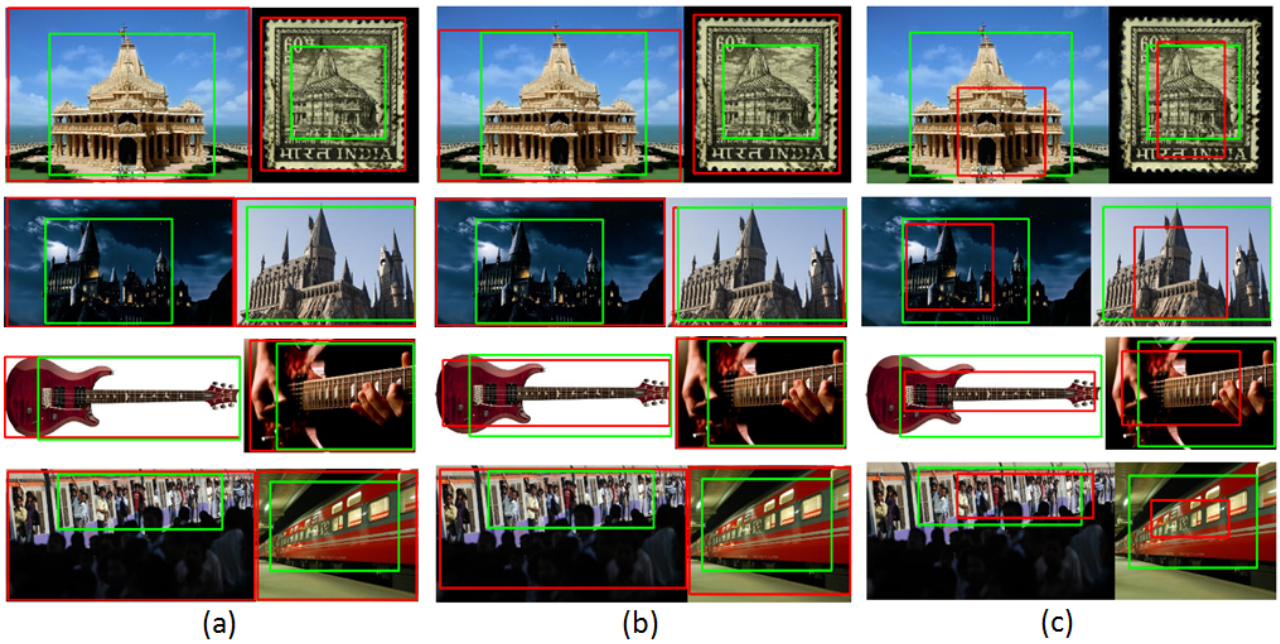

Figure 2. Partial shape matching on real world images under varying conditions. The green rectangle shows the ground truth partial match whereas the results of image matching using (a) SIFT features in isolation and (b) GB features in isolation, and (c) the proposed JGGE technique are indicated by the red rectangle. Note that the techniques in (a) and (b) determine matching regions without taking into consideration the underlying region shape. Since they fail to capture the region shape, they treat the entire image as a match resulting in a bounding box that covers the entire image and several false partial shape correspondences. In contrast, the proposed approach (c) yields partial shape correspondences that are very close to the ground truth bounding box.

3.3. Energy Minimization

The minimization of the energy function in eq. (3) is performed using a multiresolution gradient descent technique similar to the one described in [9]. Due to the non-convex nature of the energy function in eq. (3), conventional gradient descent-based optimization techniques typically converge only to a local minimum. In contrast, multiresolution gradient descent schemes have been observed to exhibit reduced sensitivity to the presence of local minima in the solution space in the context of image registration [13], [15]. The gradient descent procedure is performed at each level of resolution on equal-sized rectangular patches. A set of overlapping patches are sampled across the entire image and a random patch is used for initialization. The negative gradient direction is computed by differentiating the objective function in eq. (3). The optimization is initially performed at a coarse level, using large equal-sized rectangular patches to localize the object of interest within the image. The coarse-level solution is then interpolated to the next, finer level of resolution and used to initialize the optimization at the finer level. The optimization procedure is repeated at successively finer levels of resolution for refinement, with the solution at the previous (coarser) level of resolution used to initialize the optimization at the next (finer) level of resolution, until the convergence criterion at the finest level of resolution is reached. In particular, in our implementation, we have used 3 different image patch resolutions, i.e., 15×15, 11×11 and 7×7. Algorithm 1 describes the workflow of the overall multiresolution multicriteria optimization scheme.

Algorithm 1 JGGE algorithm

1: Extract GB features from images I_1 and I_2 respectively.
2: Compute weighted image graph $G_i(V_i, E_i, W_i)$ from image i using pixel-level features V_i as vertices, pairwise relationships between V's as edges E_i and affinity submatrix W_i which represents the edge weights $(W_i)_{x,y} = \exp(-(\|f_i(x) - f_i(y)\|)^2)$, where $f_i(x)$ and $f_i(y)$ are pixel-level GB features at locations x and y respectively.
3: Compute inter-image weights $C_{x,y} = \exp(-(\|f_1(x) - f_2(y)\|)^2)$.
4: Compute joint image graph $G(V, E, W)$ where $V = V_1 \cup V_2$, $E = E_1 \cup E_2 \cup (V_1 \times V_2)$ and the resulting affinity matrix is given by W in eq. (4).
5: Compute eigenvectors (ϕ_k) for the affinity matrix W. Use the top m non-trivial eigenvectors corresponding to the m smallest non-trivial eigenvalues of the joint image graph G to compute the joint geometric graph embedding distance (JGGED) using eq. (7).
6: Use the proposed multiresolution gradient descent technique on the multicriteria energy function in eq. (3) to achieve region-based matching.

4. Experimental Results

We designed a set of experiments to evaluate the proposed partial shape matching technique. We evaluated the performance of the proposed technique using two different metrics. The first metric evaluates the repeatability of the proposed technique for different image pairs whereas the second metric compares the reliability of the proposed technique against a set of state-of-the-art features on a partial shape matching benchmark dataset proposed and designed by us and an existing benchmark dataset proposed by Hauagge and Snavely [18]. Since certain correspondence determination techniques [21] require the contour model to be known a priori, it was not feasible to evaluate and compare the results of these techniques using our benchmark dataset since it includes wide variations in viewing conditions.

For the purpose of evaluation, we collected a set of image pairs with a *single dominant object* that is either completely or partially visible in the image. These image pairs focus on architectural scenes and, in some cases, musical instruments. We believe these image pairs will prove extremely challenging for current feature matching methods. The image pairs in the proposed dataset exhibit dramatic variations in object appearances, arising due to occlusion, presence of structures still under construction and variations in illumination, age of the objects and rendering styles (such as paintings, drawings and sketches). None of the image pairs are pre-aligned, thus ensuring that factors such as geometry and appearance variations play an important role in the matching process. To construct the proposed dataset, we gathered 40 image pairs using Google Image Search. For each image pair in both the benchmark datasets, we manually annotate the dominant object using a bounding box (denoted by X'_g and Y'_g) in each of the images in order to compute the ground truth matches. Examples of ground truth matches between the image pairs from the proposed benchmark dataset are shown in Figure 1 and Figure 2.

4.1. Evaluation of the Matching Procedure

For objective evaluation of the performance of the proposed technique, we computed a *relevance score* R for each image pair as follows:

$$R = \frac{TP}{TP + TN + FP} \tag{8}$$

where,

$$TP = (X' \cap X'_g) \cup (Y' \cap Y'_g) \tag{9}$$

$$TN = (X'_g - (X' \cap X'_g)) \cup (Y'_g - (Y' \cap Y'_g)) \tag{10}$$

$$FP = (X' - (X' \cap X'_g)) \cup (Y' - (Y' \cap Y'_g)) \tag{11}$$

The mean relevance score (MRS) and the false positive rate (FPR) values were considered as the performance evaluation metrics for all image pairs in both benchmark datasets.

Table 1. Parameter tuning for maximizing the MRS and minimizing the FPR across 40 image pairs.

		HOG		MPI	
n_{GB}	m	MRS	FPR	MRS	FPR
100	5	0.70	0.28	0.71	0.27
100	10	0.72	0.26	0.72	0.26
100	20	0.74	0.24	0.76	0.22
100	30	0.76	0.22	0.76	0.22
200	5	0.72	0.26	0.75	0.23
200	10	0.73	0.24	0.77	0.21
200	20	0.78	0.20	0.79	0.19
200	30	0.78	0.20	0.80	0.18
300	5	0.71	0.27	0.74	0.24
300	10	0.74	0.24	0.78	0.20
300	20	0.78	0.20	0.82	0.17
300	30	0.79	0.19	0.81	0.18

4.2. Parameter Selection

The two main tunable parameters in the proposed method are n_{GB}: the number of GB feature points and m: the number of top non-trivial eigenvectors resulting from the JGGE procedure (eq. (3)). We computed the MRS and FPR values over all the image pairs for different combinations of (n_{GB}, m) values. The results are reported in Table 1. An interesting observation from Table 1 is that MPI-based regularization outperformed HOG-based regularization in our proposed dataset. This could be attributed to the fact that the HOG features did not provide adequate discriminative texture orientation information in many cases of extreme variation in viewing conditions arising from change in scale, orientation and viewpoint. The MPI values, on the other hand, were observed to be more consistent across variations in viewing conditions.

Since the combination of parameter values $n_{GB} = 300$, $m = 20$ in eq. (3) were observed to yield the best classification accuracy, we used this set of parameter values for all experiments reported in the remainder of the paper. Likewise, we observed $\lambda = 0.5$ in eq. (3) to yield the best results and hence used this value in all subsequent experiments.

4.3. Repeatability of the Matching Procedure

The random initialization step in the procedure for extraction of the GB feature points necessitates an evaluation of repeatability of the proposed partial shape matching technique on both benchmark datasets. For each image pair, a *repeatability score* R_1 was computed as follows:

$$R_1 = \frac{(X' \cap X'_g) \cup (Y' \cap Y'_g)}{X' \cup Y'} \tag{12}$$

We ran the proposed partial shape matching procedure 50 times for each image pair across both datasets using mean intensity-based regularization. The value of the repeatability score R_1 was computed for each run for each image pair

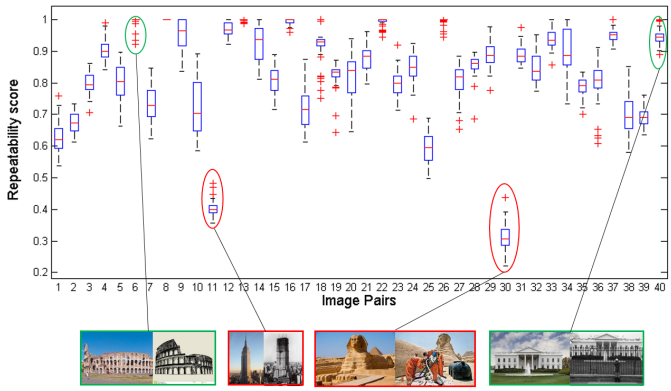

Figure 3. Repeatability score R_1 for all the image pairs after 50 runs per image pair in the proposed benchmark dataset. Two instances of image pairs with high repeatability scores (green) and two instances of outliers (red) are also shown.

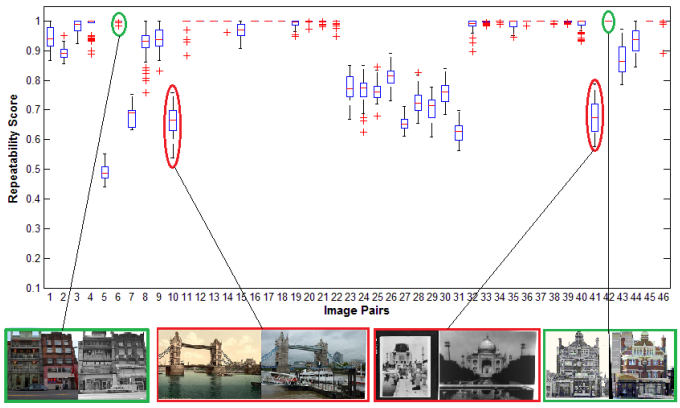

Figure 4. Repeatability score R_1 for all the image pairs after 50 runs per image pair in the benchmark dataset proposed by Hauagge and Snavely [18]. Two instances of image pairs with high repeatability scores (green) and two instances of low repeatability scores (red) are also shown.

and plotted in Figure 3 for the proposed dataset and in Figure 4 for the existing dataset [18]. It is important to note that in the case of the first outlier in the proposed dataset, i.e., the *Empire State Building*, the scarcity of prominent features in the left image resulted in a very low repeatability score; whereas in the case of the second outlier, i.e., the *Sphinx*, the GB features considered the people occluding the Sphinx in the right image to constitute the dominant object, resulting in low repeatability score. The outliers in the existing dataset [18] resulted from the partial occlusion of the common object as shown in Figure 4. The average repeatability score over the entire proposed dataset (after excluding the two outliers mentioned above) was observed to be 0.8255, whereas for the existing dataset [18], it was observed to be 0.9011 (without excluding any outliers). Both these scores are quite high and show the effectiveness of the proposed partial shape matching procedure.

4.4. Experimental Evaluation and Comparison

We evaluated the results of the proposed JGGE technique and compared them with those obtained using shape matching techniques that use standard features such as SIFT [20], speeded up robust features (SURF) [6], features from accelerated segment test (FAST) [25], oriented FAST and oriented BRIEF (ORB) [26] and GB [8] in isolation. We also compared the performance of the proposed JGGE technique with a variant of the maximally stable extremal region-based feature-ellipse matching (MSER-FEM) technique described in [4]. In particular, we reimplemented the MSER-FEM technique [4], replacing the portion pertaining to MSER-based selection with the proposed multiresolution gradient descent technique in the interest of fair comparison. To emphasize the advantages of JGGE we considered both variants of the proposed technique, i.e., JGGE-GB-MPI and JGGE-GB-HOG which use JGGE with GB features coupled with MPI- and HOG-based regularization respectively. To evaluate the effectiveness of the proposed regularization scheme, we also evaluated the performance of the proposed JGGE technique with GB features but without the regularization term (i.e., JGGE-GB-NR) on both datasets. Table 2 summarizes the results of the quantitative evaluation in terms of MRS and FPR for both datasets. In addition we also considered incorporating alternative features such as SIFT and ORB in the JGGE procedure with MPI- and HOG-based regularization (the resulting variants are termed as JGGE-SIFT-MPI, JGGE-SIFT-HOG, JGGE-ORB-MPI and JGGE-ORB-HOG respectively) and compared their performance with that of JGGE-GB-MPI and JGGE-GB-HOG in terms of MRS and FPR for both data sets as summarized in Table 3.

Table 2 shows that the proposed JGGE technique with MPI- and HOG-based regularization, performed significantly better than the standard feature-based matching techniques and the MSER-FEM technique [4]. In almost all cases, the standard feature-based matching techniques were observed to regard the entire image as a partial match, thus rendering them ineffective in solving the partial shape matching problem. In the case of most image pairs, using standard features in isolation was observed to result in failure to localize the partial shape of the object. Figure 2 depicts the matches between the image pairs from the proposed benchmark dataset obtained using (a) SIFT features in isolation, (b) GB features in isolation, and (c) the proposed JGGE-GB-MPI technique. Note that the standard SIFT- and GB-based matching techniques obtained the matching regions without taking into consideration the underlying region shape. Since they failed to capture the region shape, they treated the entire image as a match resulting in a bounding box that covered the entire image and several false partial shape correspondences. In contrast, the proposed JGGE-GB-MPI approach was observed

Table 2. Comparison of standard feature-based shape matching techniques, the MSER-FEM technique and three variants of the proposed JGGE technique on both datasets using the Mean Relevance Score (MRS - higher is better) and False Positive Rate (FPR - lower is better) as performance measures.

Results on Proposed Benchmark Dataset

	SIFT	ORB	SURF	FAST	GB	MSER-FEM	JGGE-GB-MPI	JGGE-GB-HOG	JGGE-GB-NR
MRS	0.49	0.63	0.59	0.40	0.36	0.68	**0.82**	0.79	0.60
FPR	0.48	0.34	0.37	0.54	0.61	0.31	**0.17**	0.19	0.40

Results on Existing Benchmark Dataset by Hauagge and Snavely [18]

	SIFT	ORB	SURF	FAST	GB	MSER-FEM	JGGE-GB-MPI	JGGE-GB-HOG	JGGE-GB-NR
MRS	0.66	0.68	0.76	0.50	0.73	0.72	0.90	**0.91**	0.87
FPR	0.31	0.28	0.21	0.47	0.24	0.27	0.10	**0.08**	0.11

Table 3. Comparison of the proposed JGGE technique using SIFT, ORB and GB features with MPI- and HOG-based regularization using the MRS and FPR performance measures.

Results on Proposed Benchmark Dataset

	JGGE-SIFT-MPI	JGGE-SIFT-HOG	JGGE-ORB-MPI	JGGE-ORB-HOG	JGGE-GB-MPI	JGGE-GB-HOG
MRS	0.74	0.73	0.78	0.78	**0.82**	**0.79**
FPR	0.26	0.27	0.21	0.22	**0.17**	**0.19**

Results on Existing Benchmark Dataset by Hauagge and Snavely [18]

	JGGE-SIFT-MPI	JGGE-SIFT-HOG	JGGE-ORB-MPI	JGGE-ORB-HOG	JGGE-GB-MPI	JGGE-GB-HOG
MRS	0.75	0.74	0.82	0.83	**0.90**	**0.91**
FPR	0.24	0.25	0.17	0.16	**0.10**	**0.08**

to yield partial shape correspondences that are very close to the ground truth bounding box (Figure 2). Thus, the incorporation of JGGE was seen to result in a significant improvement in the outcome of the partial shape matching procedure.

As seen in Table 2, in the case of JGGE, MPI-based regularization was observed perform slightly better than HOG feature-based regularization on the proposed benchmark dataset whereas on the existing benchmark dataset [18], the situation was reversed. This suggests that for simpler shape matching instances, as in the existing dataset [18], texture-based regularization is the better choice, whereas more complex instances, as encountered in the proposed dataset, are better served by intensity-based regularization. Also, the omission of the regularization term (i.e., JGGE-GB-NR) is seen to result in a significant deterioration in performance (in terms of MRS and FPR) in the case of both datasets (Table 2). This shows that the proposed regularization term is essential to ensure a high-quality match between corresponding regions of the image pair. The impact of the regularization term is seen to be more significant in the case of the proposed dataset than the existing benchmark dataset [18], suggesting that the proposed dataset is indeed more challenging and that region-based regularization is more critical when dealing with complex instances of partial shape matching.

Table 3 shows that the proposed JGGE technique with GB features, performs better than the JGGE technique with SIFT or ORB features for both datasets with MPI- and HOG-based regularization. This shows that the GB features are more effective in capturing the characteristics of the dominant object in the image than the SIFT or ORB features and justifies our choice of GB features in the proposed JGGE technique.

5. Conclusions and Future Work

In this paper, we presented a novel multicriteria optimization framework for partial shape matching in general real-world images. The proposed framework is shown to significantly outperform state-of-the-art feature-based partial shape matching techniques on two challenging datasets, one specifically designed by us to challenge the proposed technique and an already existing one [18]. It is shown that the proposed technique is able to handle significant variability along multiple imaging and object parameters ranging from illumination, viewpoint, age of structures, degree of occlusion, structures under partially construction and inclusion of paintings and sketches of objects. The success of our method can be attributed to the region-based optimization in the proposed JGGE framework rather than determination of point correspondences alone using state-of-the-art features in isolation. In future, we plan to extend the proposed framework to more complex computer vision problems such as structure-from-motion, object recognition and shape-based image retrieval.

Acknowledgment: This material is based upon work supported by the National Science Foundation under Grant No. 1442672. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] N. Alajlan et al., Shape retrieval using triangle-area representation and dynamic space warping, *Pattern Recog.*, 40, pp. 1911 - 1920, 2007. 2

[2] P. Arbelaez et al., Contour detection and hierarchical image segmentation, *IEEE TPAMI*, 33(5), pp. 898 - 916, 2011. 3, 4

[3] X. Bai et al., Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE TPAMI*, 29(3), pp. 449 - 462, 2007. 2

[4] M. Bansal and K. Daniilidis, Joint spectral correspondence for disparate image matching, *Proc. IEEE CVPR*, pp. 2802 - 2809, 2013. 3, 4, 7

[5] C. Barnes et al., PatchMatch: a randomized correspondence algorithm for structural image editing, *ACM TOG*, 28(3), pp. 24, 2009. 3

[6] H. Bay et al., SURF: Speeded up robust features. *Proc. ECCV*, pp. 404 - 417, 2006. 7

[7] S. Belongie et al., Shape matching and object recognition using shape contexts, *IEEE TPAMI*, 24(4), pp. 509 - 522, 2002. 2

[8] A. Berg and J. Malik, Geometric blur for template matching, *Proc. IEEE CVPR*, Vol. 1, pp. 607 - 614, 2001. 4, 7

[9] M.M. Bronstein et al., Multigrid multidimensional scaling. *Numerical Linear Algebra with Applications (NLAA)*,13, pp. 149-171, 2006. 5

[10] A.M. Bronstein et al., Partial similarity of objects, or how to compare a centaur to a horse, *IJCV*, 84(2), pp. 163 - 183, 2008. 3

[11] M. Brown et al., Multi-image matching using multi-scale oriented patches, *Proc. IEEE CVPR*, Vol. 1, pp. 510 - 517, 2005. 3

[12] L. Chen et al., Efficient partial shape matching using Smith-Waterman algorithm, *Proc. NORDIA Wkshp. at CVPR*, 2008. 2

[13] A. Cole-Rhodes et al., Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient, *IEEE Trans. Image Processing*, Vol. 12(12), pp 1495-1511, December 2003. 5

[14] N. Dalal, and B. Triggs, Histograms of oriented gradients for human detection, *Proc. IEEE CVPR*, Vol. 1, pp. 886 - 893, 2005.

[15] R. Eastman and J. Le Moigne, Gradient-descent techniques for multi-temporal and multi-sensor image registration of remotely sensed imagery, *Proc. 4th Intl. Conf. Information Fusion* (FUSION 2001), pp. 7-10, 2001. 5

[16] P.F. Felzenszwalb and J. Schwartz, Hierarchical matching of deformable shapes. *Proc. IEEE CVPR*, pp. 1 - 8, 2007. 2

[17] J. Ham et al., Semisupervised alignment of manifolds. *AISTATS*, pp. 120-127, 2004. 3

[18] D. Hauagge, and N. Snavely, Image matching using local symmetry features. *Proc. IEEE CVPR*, pp. 206-213, 2012. 1, 2, 6, 7, 8

[19] W.Y. Kim and Y.S. Kim, A region-based shape descriptor using Zernike moments, *Signal Process.: Image Commun.*, 16(1-2), pp. 95-102, 2000. 2

[20] D.G. Lowe, Distinctive image features from scale-invariant keypoints *IJCV*, 60(2), pp. 91 - 110, 2004. 3, 7

[21] T. Ma and L.J. Latecki, From partial shape matching through local deformation to robust global shape similarity for object detection, *Proc. IEEE CVPR*, pp. 1441 - 1448, 2011. 2, 6

[22] G. McNeill and S. Vijayakumar, Hierarchical procrustes matching for shape retrieval, *Proc. IEEE CVPR*, pp. 885 - 894, 2006. 2

[23] A. Opelt et al., A boundary-fragment-model for object detection, *Proc. ECCV*, pp. 575 - 588, 2006. 2

[24] S. Ravishankar et al., Multi-stage contour based detection of deformable objects, *Proc. ECCV*, pp. 483 - 496, 2008. 2

[25] E. Rosten and T. Drummond, Machine learning for high-speed corner detection, *Proc. ECCV*, pp. 430 - 443, 2006. 7

[26] E. Rublee et al., ORB: an efficient alternative to SIFT or SURF, *Proc. ICCV*, pp. 2564 - 2571, 2011. 7

[27] C. Schmid and R. Mohr, Local gray value invariants for image retrieval, *IEEE TPAMI*, 19(5), pp. 530 - 535, 1997. 3

[28] T.B. Sebastian and B.B. Kimia, Curves vs skeletons in object recognition, *Proc. ICIP*, pp. 22 - 25, 2001. 2

[29] T. Sebastian and P.N. Klein, On aligning curves, *IEEE TPAMI*, 25(1), pp. 116 - 125, 2003.

[30] T. Sebastian et al., Recognition of shapes by editing their shock graphs, *IEEE TPAMI*, 26(5), pp. 550 - 571, 2004. 2

[31] J. Shi and C. Tomasi, Good features to track, *Proc. IEEE CVPR*, pp. 593 - 600, 1994. 3

[32] J. Shotton et. al., Multi-scale categorical object recognition using contour fragments, *IEEE TPAMI*, 30(7), pp. 1270 - 1281, 2008. 2

[33] Q. Zhu et al., Contour context selection for object detection: A set-to-set contour matching approach. *Proc. ECCV*, pp. 774 - 787, 2008. 2