Joint Spectral Correspondence for Disparate Image Matching

Mayank Bansal and Kostas Daniilidis*
GRASP Lab, University of Pennsylvania
Philadelphia, PA, USA
{mayankb,kostas}@cis.upenn.edu

Abstract

We address the problem of matching images with disparate appearance arising from factors like dramatic illumination (day vs. night), age (historic vs. new) and rendering style differences. The lack of local intensity or gradient patterns in these images makes the application of pixel-level descriptors like SIFT infeasible. We propose a novel formulation for detecting and matching persistent features between such images by analyzing the eigen-spectrum of the joint image graph constructed from all the pixels in the two images. We show experimental results of our approach on a public dataset of challenging image pairs and demonstrate significant performance improvements over state-of-the-art.

1. Introduction

In this paper, we focus on matching images with disparate appearance. Such images might be taken during day and night or in different times in history, and they differ at the local pixel level in the sense that neither intensity nor gradient distributions are locally comparable. Thus, we cannot rely on pixel-level feature descriptors like SIFT. Instead, we propose to use the joint image graph spectrum to detect and match persistent features which robustly encode the appearance similarity we perceive when we look at such images.

Consider the images in Fig. 1 where we have the same scene captured at different times of day[1]. Visual comparison reveals the large amount of appearance change that occurs in the scene due to the illumination variation. Numerous SIFT features are detected in these images and they show good repeatability (blue bars in the plot) as well. However, the average precision (AP) of the SIFT descriptors computed directly from these images significantly degrades as the illumination difference between the matched image pairs is increased as is visible from the blue polyline in the plot. In contrast, the spectral features we propose in this paper are comparable in their repeatability (red bars) and they behave significantly better in the Average Precision (red polyline) even for the most challenging pair: night vs. day.

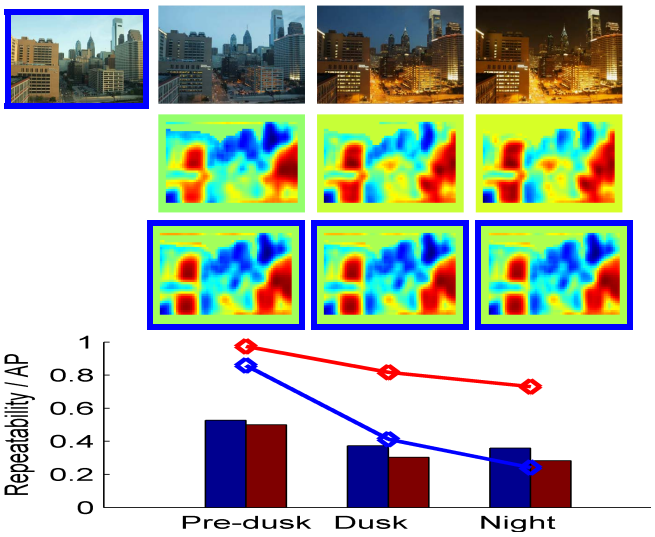

Figure 1. Day to night persistence matching. The spectrum of the joint image graphs is computed. The first row shows a day-time query-image (blue box) which is matched pair-wise against the pre-dusk, dusk and night images respectively from left to right. The second and third rows show the second eigen-vectors $J_1^{(2)}$ and $J_2^{(2)}$ for each pair of images (pre-dusk : day), (dusk : day), (night : day). The eigen-vectors corresponding to the query have a blue box to ease visualization. The plot compares the repeatability (bars) and average-precision (AP)(polyline) of the SIFT detector (blue) with the spectral method (red).

Spectral methods on the image graph laplacian have been used extensively in the literature for applications like clustering, segmentation [1, 11] etc. The extracted eigen-functions are either discretized to obtain the desired number of clusters or segments in the image or they are used directly as the spectral space coordinates of the pixels in an

*The authors are grateful for support through the following grants: ARL MAST-CTA W911NF-08-2-0004, and ARL RCTA W911NF-10-2-0016, NSF-DGE-0966142.

[1]Frames extracted from the time-lapse sequence at: http://www.youtube.com/watch?v=0OK4CdQ-haU

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.361

embedded space representation. These coordinates are then further clustered using K-means to obtain discrete clustering solutions. In contrast, in this paper we propose to use the individual eigen-functions themselves as a feature representation of the image pair from which interesting and useful feature correspondence can be derived. We show how such a representation captures **persistent regions** in the image pair even when the appearance difference between them is substantial (day-night, historic-new etc.). Moreover, we propose a new definition of the joint image graph: all pixels of both images are nodes and the corresponding edge weights depend only on the difference of the local image structures and not on the proximity between the pixels. Although a partitioning of such a graph might cluster together distant regions, these regions even though disconnected in the image space are persistent across images.

Our most significant contributions in this paper are: (1) a new representation between two images: the joint image graph defined only based on the affinity between image structures in the dense set of pixels from both images without considering the proximity between two image positions; (2) a new definition of persistent regions as the stable regions of the eigen-functions of that graph considered in their "soft" form without any discretization. We show that such persistent features are both repeatable across images and similar in terms of SIFT descriptors computed in the eigen space itself, in a variety of cross domain setups.

We show experimental results of our approach on the challenging dataset from [3] which contains image pairs exhibiting dramatic illumination, age and rendering style differences. Our results clearly indicate the substantial matching improvement possible by looking at features derived from a joint image spectrum rather than relying on features detected individually in the two images to match in their descriptors. Unlike standard local-features approaches which detect features on each image independently, our method relies on computing features using both images simultaneously. However, we believe that the global information encoded in the joint image graph and its eigen-functions is the new cue that enables a better performance than approaches relying only on local neighborhoods.

2. Related Work

Not many approaches exist that can handle the discrepancy between two images at the level that we address in this paper. Shechtman and Irani [7] proposed an approach for matching disparate images using patterns of local self-similarity encoded into a shape-context like descriptor. However, for the kind of disparate images we consider, the local self-similarity pattern itself can be significantly different between corresponding points in the image pair. Shrivastava *et al*. [8] recently proposed an approach for cross-domain image matching using data-driven learning techniques. Using a linear classifier, they learn the relative importance of different features (specifically, components of the global image HoG descriptor in the paper) for a given query image and then use the weight vector to define a matching score. In contrast, our focus is on extracting local features that are persistent between a pair of images instead of deriving a global image descriptor that can be used for retrieval. Recently, Hauagge and Snavely [3] have focused on the task of matching such images by defining "local-symmetry" features which capture various symmetries like bilateral, rotational etc. at the local level. This approach addresses matching rather than retrieval and the discrepancy level is similar to the level our approach handles, hence, we decided to use the dataset [3] as our main test set.

Our methodology is most related to the works of [11] and [9]. The spectral analysis of the joint matrix between two images appeared first in [11] where the affinity matrices of object model patches and the input image are combined with a non-diagonal matrix associating object patches and image pixels. Toshev *et al*. [9] proposed an approach to determine co-salient regions between two images using a spectral technique on the joint image graph constructed from the images. Their joint image graph was constructed with all the pixels in the two images by defining separate affinity functions for intra and inter image terms. The intra image affinity was defined using the intervening contour cue while the inter image term was based entirely on the initial set of feature correspondences between the images. In this paper, we also use a joint image graph but we differ in (i) using no proximity information in the affinity matrix, (ii) using SIFT descriptor information instead of intensity differences, and (iii) defining both the intra and inter image terms densely using all the image pixels with a uniform affinity function that successfully captures the persistent regions shared between the images. The density of the inter-image term allows us to apply spectral decomposition directly instead of requiring us to use the subspace technique in [9]. Thereafter, we show how we can use the extracted eigen-functions to construct features that are invariant to the large appearance changes between the input images.

Glasner *et al*. [2] proposed a shape-based approach to jointly segment multiple closely-related images by combining contour and region information. They show examples of image pairs with illumination differences where their joint segmentation approach achieves better co-clustering than what is possible by using intra-image constraints alone. They start from a super-pixel segmentation of the images and then use contour-based constraints to drive their intra-image affinities. The inter-image constraints are derived from a comparison of HoG-like features only on contour segments. Our approach also uses gradient-based descriptors to enforce inter-image constraints – however, we do so

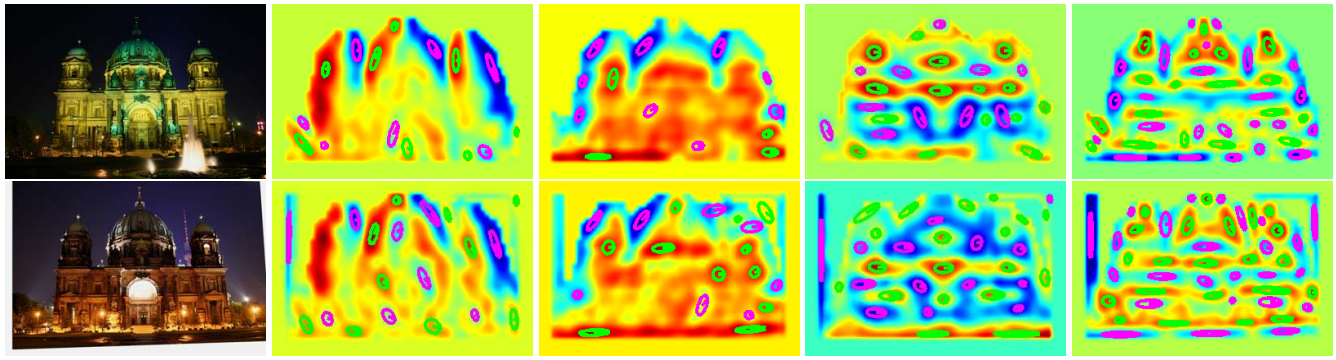

Figure 2. First column shows an image-pair from the dataset in [3]. Second through fifth columns show eigen-function pairs $(J_1^{(2)}, J_2^{(2)}), \ldots, (J_1^{(5)}, J_2^{(5)})$ along with the detected MSER feature-ellipses. The green and magenta colors denote whether the features correspond to maxima or minima.

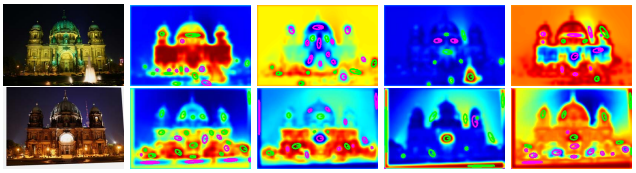

Figure 3. For the image pair in the first column, the successive columns show the second-through-fifth eigen-function pairs obtained using a pixel-color based joint image graph. In this case, the eigen-functions do not suggest any significant correlation with the region correspondence in the original images.

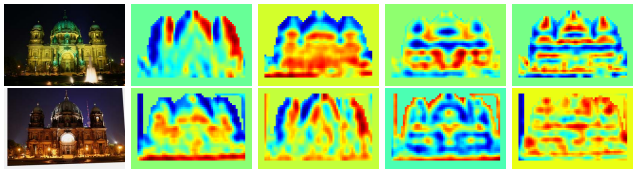

Figure 4. For the image pair in the first column, the successive columns show the second-through-fifth eigen-functions obtained using dense-SIFT-based image graphs. The eigen analysis is performed on each image graph independently. In this case, the eigen-functions show correlations but the correlated regions are distributed across several different eigen-functions.

densely between every pair of pixels in the input images. Thus, we do not need any prior segmentation and we are not prone to errors due to misdetection of contours, particularly since contour detection would be very challenging for the kind of disparate images we focus on. Contours of the soft version of the eigenvectors of a single image affinity matrix computed following the Normalized Cuts criterion have also been used in [1] to include global relationships into the probabilistic boundary feature vector.

3. Technical Approach

Consider the image pair in the first column of Fig. 2 depicting the same monument under significantly different illumination conditions. Each local facet of the monument is illuminated differently leading to dissimilar contrast and color characteristics. It is evident that finding features that are repetitive between the two pictures is a daunting task; in fact, the problem of finding descriptors that can account for the appearance differences at geometrically matching locations is itself quite challenging. On the other hand, we humans find it quite easy to ascertain by visual inspection that these two images correspond to the same monument. The kind of features we use to make such a judgment are the more inherent "persistent" features in the scene like the contours, salient regions, the local shapes, patterns of contrast etc. One can argue, then, that shape-based image matching techniques should be applicable for matching these images. However, the contrast variations make it very difficult to detect the image contours robustly. Most of the dominant contours in the scene are very low energy and the intensity at which corresponding contours would get detected varies between different regions in the two images. Therefore, we propose a spectral approach that detects these persistent image features using the eigen-spectrum of the joint image graph computed from appropriate local gradients in the two images.

Before going into the details of the way the graph is constructed, let us focus on the images in the second to fifth columns of Fig. 2. In each column, the top and bottom images correspond to one of the eigen-functions of the joint graph reshaped back to the size of the images. The red and blue shades represent the maxima and minima respectively, of the eigen-function. It is clear that, for each eigen-function pair (i.e. images in a single column) the distribution and shapes of these eigen extrema correspond well between the two images and the image regions where this correspondence is strong is in agreement with the actually corresponding image regions. Thus, by computing features that encode these extrema (both in their shape and the eigen-energy profile), we can more robustly match these images without relying on descriptors computed directly from the images.

The technical approach is organized as follows. First,

we will review basic fundamentals of the image graph construction and its spectrum, followed by a look at the actual features we use to build the joint image graph. Then, we will characterize the eigen-function extrema as persistent regions and discuss algorithms to detect and match these extrema.

3.1. Image Graph

The spectral analysis of the content of an image is carried out on a weighted image graph $G(V, E, W)$ which contains all the image-pixels as vertices in the vertex-set V of cardinality n. The edge-set E contains all pair-wise relationships between every pair of vertices (pixels) in the set V thus 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 pixels represented by vertices v_i and v_j. We can collect these weights into an $n \times n$ affinity matrix $W = (w_{ij})_{i,j=1,\ldots,n}$. The degree matrix D of this graph is defined as a diagonal matrix D with $D(i,i) = \sum_{j=1}^{n} w_{ij}$. Using W and D, we can now compute the normalized graph laplacian $\bar{L}$ as $\bar{L} = I - D^{-1/2} W D^{-1/2}$. We are interested in the eigen-spectra U of this laplacian matrix which can be computed by eigen-value decomposition $\bar{L}\bar{U} = \lambda\bar{U}$ and setting $U = D^{-1/2}\bar{U}$. The eigen-vectors $u_1, \ldots, u_K$ corresponding to the K smallest eigen-values are related to the structure of the graph [10] and are extensively used in the literature to obtain a $K-$partition of an image based on appropriately defined weight values. However, in this paper, we will study the individual eigen-vectors directly to ascertain useful persistent regions in the image.

The formulation above can be easily extended from a single image to a pair of images 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 $\mathcal{I}_1$ and $\mathcal{I}_2$. Then the joint image graph $G(V, E, W)$ is defined such that $V = V_1 \cup V_2$, $E = E_1 \cup E_2 \cup V_1 \times V_2$ where $V_1 \times V_2$ is the set of edges connecting every pair of vertices in (V_1, V_2). The affinity matrix W is given by:

$$W = \begin{pmatrix} W_1 & C \\ C^t & W_2 \end{pmatrix}_{(n_1+n_2)\times(n_1+n_2)} \tag{1}$$

where $|V_1| = n_1$, $|V_2| = n_2$ and C is the $n_1 \times n_2$ matrix encoding the affinities of edges in $V_1 \times V_2$. The eigen-spectra for the joint graph can be computed exactly as before by defining the normalized laplacian $\bar{L}$ and carrying out its eigen-value decomposition.

3.2. Image Features and the Joint Spectrum

Consider first an experiment where we perform spectral analysis of the joint image graph $G(V, E, W)$ with the matrix W defined directly in terms of the pixel color values in the two images, i.e. both the intra-image weights W_1, W_2 as well as the inter-image weights C are defined by a function of the perceptually uniform Lab-space color difference between the pixel pair. Now, we compute the eigen-spectra of this graph's laplacian to see if the corresponding eigen-functions show any patterns of correspondence. Fig. 3 shows the second through fifth eigen-function pairs (reshaped back into a matrix) for the same image pair as in Fig. 2 obtained using the above Lab-based graph. It is clear that we do not see much correspondence between the eigen-functions in this case – this motivates the need for features stronger than just the individual pixel colors. Features that encode local image gradients are good candidates as they provide a soft metric on the salient regions without needing a hard-thresholding step needed by edge and contour detectors. In this paper, we propose to use SIFT [4] descriptors computed densely on the image at a fixed spatial sampling σ (we use $\sigma = 5$ pixels for experiments in this paper).

To capture local image gradients at multiple scales, at each location we compute the SIFT descriptors at two different scales (size of the SIFT spatial bin) s_1 and s_2. The resulting feature vectors are concatenated to result in a $256-$D feature vector $f_i(x)$ at each location x in image $\mathcal{I}_i$. Taking into account the spatial sampling of the features σ, let n_1 and n_2 be the number of feature vectors obtained from images $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively. Then the affinity matrices W_1, W_2 and C are defined as follows:

$$(W_i)_{x,y} = \exp\left(-\frac{\|f_i(x) - f_i(y)\|^2}{\sigma_f^2}\right) \tag{2}$$

$$(C)_{x,y} = \exp\left(-\frac{\|f_i(x) - f_j(y)\|^2}{\sigma_f^2}\right) \tag{3}$$

where $f_i(x)$ and $f_i(y)$ are features at locations x and y in image $\mathcal{I}_i$. We use the cosine-distance as the feature distance function $\|.\|$ with tuning parameter σ_f set to 1.0 in all our experiments. The scales s_1 and s_2 were set to 10 and 6 respectively for all the experiments. Note that unlike most image-domain spectral approaches in the literature, we do not use a spatial affinity term to reduce the influence of spatially separated pixels. In fact, supporting long range interactions is a key component of our approach as this allows us to obtain more distinctive profiles in the computed eigen-functions. With a spatial proximity term in the affinity matrix, we run the risk of artificially limiting the spatial extent of an eigen-function extrema and thus rendering the derived features less distinctive.

Given the joint graph affinity matrix W from eqns-(1), (2) and (3), it is straightforward to compute the eigen-spectra. But before we do that, let us see if we can determine any correspondence information between image regions by extracting the spectra from each image graph separately. Fig. 4 shows the eigen-functions obtained by spectral analysis of the image graphs of the top and bottom row im-

ages independently. Even though the eigen-functions correctly represent the grouping of gradient information as is expected from our gradient features, one cannot infer useful correspondence information between image regions from the corresponding pair of eigen-functions directly.

Now, we can go back and look at the eigen-function pairs in Fig. 2. These were obtained as the eigen-vectors $u_1, \ldots, u_5$ corresponding to the smallest $K = 5$ eigen-values of the eigen-value decomposition $\bar{L}U = \lambda U$. From each $n_1 + n_2$ dimensional eigen-vector u_k, we compute an eigen-function pair $(J_1^{(k)}, J_2^{(k)})$ as follows. The first n_1 entries of u_k are reshaped to the dimensions of $\mathcal{I}_1$ by assigning its component values to the sampling locations where the features were extracted from and then interpolating the values in between. Similarly, the next n_2 entries of u_k are reshaped to the dimensions of $\mathcal{I}_2$ leading to the eigen-function $J_2^{(k)}$.

3.3. Characterization of persistent regions

As discussed before, the extrema of the eigen-function pairs $(J_1^{(2)}, J_2^{(2)}), \ldots, (J_1^{(5)}, J_2^{(5)})$ represent persistent features that can serve well as means of finding correspondences across these difficult pairs of images. We want to characterize these extrema in terms of their location, their region of support as well as the variation of the eigen-energy in the vicinity of each extrema. Since the extrema can commonly exhibit elongated ridge-like shapes, an isotropic blob-detector would not work well. The continuous nature of the eigen-functions suggests that a water-shed like algorithm would serve as a good detector that might find both the location as well as the support region for these extrema. Therefore, we found the well known feature detection algorithm – the Maximally Stable Extremal Region (MSER) detector [5] – to be suitable for this purpose.

The intensity-based MSER detector is typically used to find affine-covariant regions in an image by looking for water-shed areas that remain stable as an image intensity threshold is varied. Each detected region is a set of connected pixels to which an ellipse is typically fit to represent the support region. To apply the MSER detector, we first normalize each eigen-function $J_1^{(k)}$ (and $J_2^{(k)}$) to a range of $[0, 255]$ by scale and offset correction. Then, we run intensity-based MSER along with ellipse fitting to detect stable affine regions. All the eigen-function figures in this paper depict these regions as green or magenta ellipses corresponding to maxima and minima respectively.

To represent the eigen-energy variation around each detected MSER region, a number of different descriptors can be used. Through empirical evaluation, we have found the SIFT [4] descriptor to work well. Each detected MSER ellipse is affine corrected to a circular region and a SIFT descriptor is computed for a region five times the ellipse size by computing gradients on the eigen-function. The large spatial extent of the SIFT window allows us to capture the eigen-energy profile more distinctly while still finding corresponding features between the eigen-function pairs. We will use the term **JSPEC** to refer to this feature which combines MSER ellipse keypoint with the eigen-space SIFT descriptor.

Algorithm 1 JSPEC Algorithm

1. Compute features $f_1(x)$ and $f_2(x)$ at a spatial sampling σ for $\mathcal{I}_1$ and $\mathcal{I}_2$.
2. Compute affinity matrix W using eqns-(1), (2) and (3).
3. Compute the K smallest eigen-vectors $u_1, \ldots, u_K$ for W.
4. Extract eigen-function pairs $(J_1^{(k)}, J_2^{(k)})$ from each u_k.
5. Detect MSER features and compute SIFT descriptors for each $(J_1^{(k)}, J_2^{(k)})$.
6. Match features from each $(J_1^{(k)}, J_2^{(k)})$ by bi-directional SIFT matching.
7. Collect matches from all K eigen-functions to get the final match set.

3.4. Eigen-function feature matching

The centroids of the MSER ellipses along with their associated SIFT descriptors can be treated as image features in a traditional sense. Therefore, we adopt a simple approach to matching these features by using the nearest-neighbor criterion coupled with the ratio-test [4]. However, we match the descriptors from each pair of eigen-functions $(J_1^{(k)}, J_2^{(k)})$ independently i.e. for each descriptor in $J_1^{(k)}$, the nearest and second-nearest descriptors are searched **only** in $J_2^{(k)}$ and the association to the nearest descriptor is accepted only if its euclidean descriptor distance is less than τ times the distance to the second-nearest descriptor. To enforce a stronger match criterion, we perform matching from $J_1^{(k)}$ to $J_2^{(k)}$ and from $J_2^{(k)}$ to $J_1^{(k)}$ and keep the matches which are mutually consistent. This gives us a set of correspondences $\mathcal{C}_k$ from the eigen-function pair $(J_1^{(k)}, J_2^{(k)})$. It should be noted that unlike traditional SIFT feature matching, our constraint on being able to match between individual eigen-function pairs results in a much stronger match criterion.

Algorithm. We present our method in a reproducible algorithmic form in Alg. 1.

4. Experiments

We evaluate our approach on the dataset of challenging image pairs from [3]. This dataset contains 46 pairs of images exhibiting dramatic illumination, age and rendering style differences. Some image pairs are pre-registered with a homography to focus on appearance differences, while others exhibit both geometric and photometric variation. For each image pair, a manually extracted ground-truth homography H_{12} is included with the dataset.

Hauagge and Snavely [3] evaluated their local symmetry features first, in terms of the detector repeatability and second, in terms of descriptor mean-average-precision performance. In our evaluation, we follow their methodology and evaluation metrics closely and provide a thorough comparison of our JSPEC features with their SYM-I and SYM-G

	Scale		Score	
	100	200	100	200
MSER	0.087	0.103	-	-
SIFT (DoG)	0.144	0.153	0.050	0.078
SYM-I	0.135	0.184	0.173	0.206
SYM-G	0.173	0.228	0.227	0.281
JSPEC	**0.287**	**0.292**	-	-

Table 1. Detector repeatability compared with [3].

	GRID	SIFT	SYM-I	SYM-G	JSPEC
Self-Sim.	0.29	0.14	0.12	0.16	
SIFT	0.49	0.21	0.28	0.25	**0.61**
SYMD	0.41	0.22	0.20	0.25	
SIFT-SYMD	0.58	0.28	0.35	0.36	

Table 2. Descriptor mean average precision (mAP) evaluation and comparison with [3].

features.

4.1. Detector repeatability

To evaluate the repeatability of the eigen-space MSER features for a given image pair, we consider all the detections before the SIFT matching step. We collect all the features from across all eigen-functions into two sets of keypoints K_1 and K_2 for images $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively. Each keypoint has a centroid and an ellipse associated with it. Therefore, we can directly apply the repeatability metric from [6] which we briefly review next. Each keypoint $k_1 \in K_1$ is warped into $\mathcal{I}_2$'s coordinate frame using the ground-truth homography H_{12} and its (warped) support region is compared with the support region of each keypoint $k_2 \in K_2$ to obtain an overlap score. If the overlap score is more than 0.6 (i.e. less than 40% overlap error), then we count the associated keypoint-pair as a correct detection. The ratio of the correct detections to the smaller of the number of keypoints in either image is used as the measure of detector "repeatability". To be invariant to absolute keypoint scales, the keypoints in $\mathcal{I}_1$ are rescaled by a factor s to a fixed area A before applying the homography. The same scale s is applied to the keypoints in $\mathcal{I}_2$ before determining the overlap score.

Hauagge and Snavely [3] computed the repeatability scores of their features by considering subsets of top-100 and top-200 detections ordered by either feature scale or score. This was done to avoid a bias when comparing to detectors that produce a large number of keypoints. Our MSER detector does not output a detection score and so we only present repeatability numbers based on ordering by scale. These are shown in Table-1 where we have also reproduced numbers from [3] for ease of comparison. We observe that our JSPEC features achieve slightly better repeatability than what SYM-G achieved using the top scoring 200 detections.

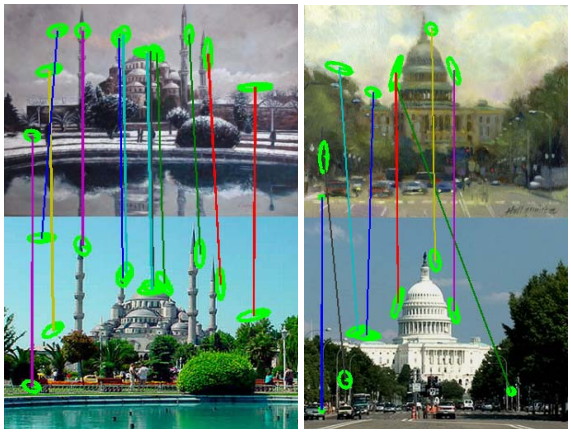

Figure 5. Painting to image matching. The painting images (top-row) have been taken from the dataset in [8].

4.2. Descriptor Evaluation

We evaluate the discriminative power of our eigen-space SIFT descriptors in a manner similar to [3]. Here again, we consider the descriptors associated with all the features (i.e. before the matching step) collected together from all the eigen-functions. Then, we match these descriptors using the standard ratio test [4] on the top two nearest neighbor distances. For a given choice of the ratio threshold, we get a set of candidate correspondences which are evaluated with the ellipse overlap criterion of [6] using the ground-truth homography H_{12} to compute a point on the precision-recall curve. By varying the ratio-test threshold, we can trace the full precision-recall curve.

Fig.6 compares the performance of our JSPEC features against each of the features studied by [3][2]. Note that the JSPEC plots across each column are exactly the same since we do not vary the detector. We have plotted them four times to allow comparison with the individual plots from [3]. The first four image pairs show a substantial improvement in performance over other competing methods. The first row of plots ("Grid") represents a synthetic detector experiment in [3] where locations on a uniform grid in $\mathcal{I}_1$ are chosen as K_1 and these locations warped by H_{12} into $\mathcal{I}_2$ are chosen as K_2. This is meant to test how well the descriptor matches appearance of perfect geometrically matching locations. Even though we do not use the grid-detector, a comparison of the JSPEC PR-curves with other curves in the "Grid" row clearly indicate that SIFT features computed on the eigen-functions match better across the extreme day-night appearance changes. The graffiti image pair (fifth column) shows that we perform similar to the SYMD descriptor on SIFT features but, as expected, worse than the SIFT detector-descriptor pair. Finally, the Taj example (last column) shows a failure case where our method fails because large parts of the scene have changed com-

[2]The precision-recall data plotted here was obtained from the authors directly.

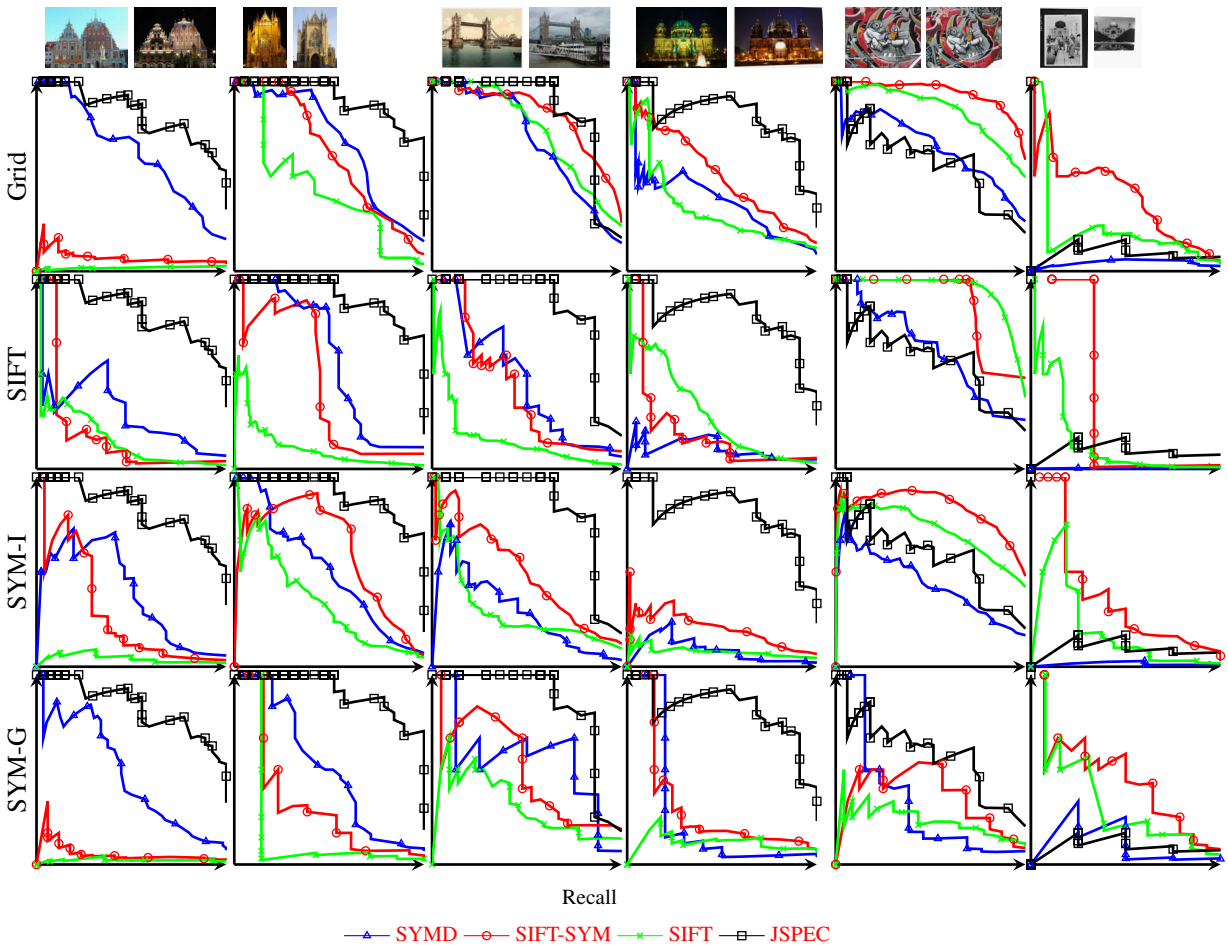

Figure 6. Precision-Recall curves comparing performance of the spectral approach (JSPEC) with the features evaluated in [3]. Each column shows plots for the image pair in the top row. For each image pair, the JSPEC curve is repeated in the four rows to show comparison with the four different detectors in [3].

pletely. In this case, a combination with SIFT features is likely to give better performance. Also note that we have not applied either the bi-directional matching criterion or the "match only within each eigen-function pair" criterion to obtain these precision-recall curves for a fair comparison with other methods (which also do not apply the bi-directional constraint). The performance is expected to be higher with these additional criteria.

Table-2 compares our mean average precision with [3] on the entire dataset. We achieve an overall mAP of 0.61 which is higher than the synthetic grid-detector (combined with SIFT-SYMD descriptor) based mAP of 0.58 achieved by [3].

4.3. Qualitative Results

In Fig. 5 and Fig. 7, we show some qualitative outputs of our algorithm. The matches overlaid on the images are the final matches obtained after bi-directional SIFT matching on the JSPEC features at a ratio-threshold of 0.8. Fig. 5 shows two examples of paintings from the dataset shared by Shrivastava *et al*. [8] in the top row. We downloaded similar looking images from the web and tested our algorithm on this difficult painting to image matching task. The examples show the quality of our matches. Fig. 7 shows three more different kinds of examples with the correspondences detected in each of the four eigen-function pairs collected together and overlaid in the first column. In the first row, we have a difficult day-night pair where we find valid matches despite the poor contrast. The second row shows matching between a historic picture and a drawing. The third row shows the standard graffiti image pair with uncorrected perspective distortion that our method can easily cope with. Note the accuracy of the feature detections on the original images.

5. Conclusion

Image matching across different illumination conditions and capture times has been addressed in the past by comparing descriptors of local neighborhoods or employing discriminative learning of local patches. In this paper we intro-

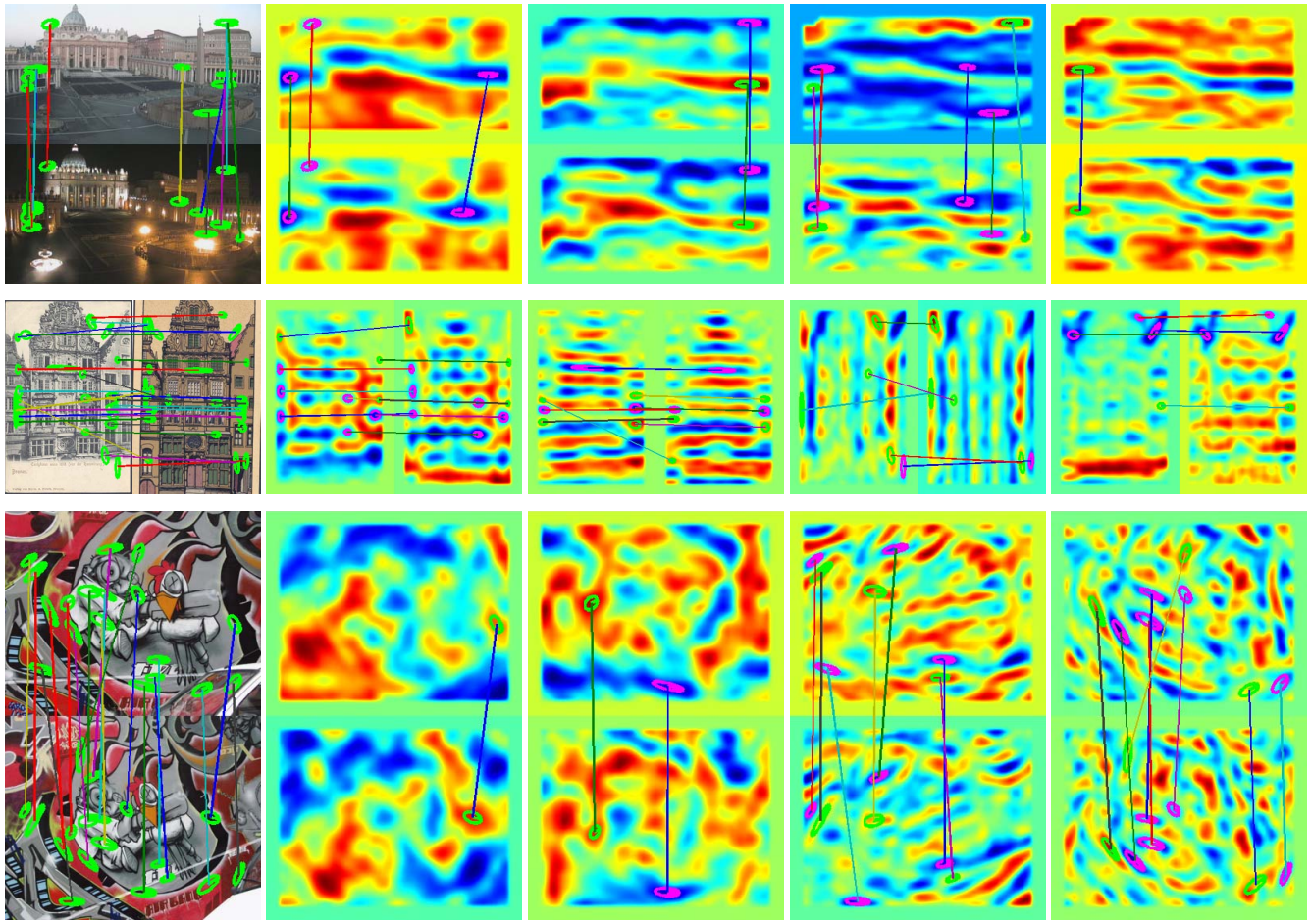

Figure 7. The first column shows an image-pair from the dataset in [3] along with the correspondences assembled from the individual eigen-functions. Second through fifth columns show eigen-function pairs $(J_1^{(2)}, J_2^{(2)}), \ldots, (J_1^{(5)}, J_2^{(5)})$ along with the MSER feature-ellipses that have been matched using the SIFT-bidirectional matching criterion ($\tau = 0.8$).

duced global image information into the matching process by computing the spectrum of the graph of all pixels in both images associated only by the similarity of their neighborhoods. Significantly, the eigen-functions of this joint graph exhibit persistent regions across disparate images which can be captured with the MSER characteristic point detector and represented with the SIFT descriptor in the resulting stable regions. Such characteristic points exhibit surprisingly high repeatability and local similarity.

In our ongoing work, we study how such persistent features can be used for testing geometric consistency, a task impossible when no correspondences can be established in the raw image domain. Establishing the two view geometry using these persistent features would also allow re-photography: establishing the same view for a photo today given a reference photo from the past.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *PAMI*, 33(5):898–916, 2011. 1, 3

[2] D. Glasner, S. Vitaladevuni, and R. Basri. Contour-based joint clustering of multiple segmentations. In *CVPR*, 2011. 2

[3] D. Hauagge and N. Snavely. Image matching using local symmetry features. In *CVPR*, 2012. 2, 3, 5, 6, 7, 8

[4] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 4, 5, 6

[5] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004. 5

[6] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *IJCV*, 65(1):43–72, 2005. 6

[7] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *CVPR*, 2007. 2

[8] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros. Data-driven visual similarity for cross-domain image matching. *ACM Transaction of Graphics (TOG) (Proceedings of ACM SIGGRAPH ASIA)*, 30(6), 2011. 2, 6, 7

[9] A. Toshev, J. Shi, and K. Daniilidis. Image matching via saliency region correspondences. In *CVPR*, 2007. 2

[10] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007. 4

[11] S. Yu, R. Gross, and J. Shi. Concurrent object recognition and segmentation by graph partitioning. In *Proc. NIPS*, pages 1383–1390, 2002. 1, 2