Joint Spectral Correspondence for Disparate Image Matching

Shivin Yadav
201401072

Anshuman Majumdar
201401005

Joyneel Misra
201401074

Abstract

In this work, we address the problem of matching images with disparate appearance arising from factors such as illumination, age and rendering style differences. Due to lack of local intensity or gradient patterns in the images, descriptors such as SIFT are not that useful for matching. This work is done as a part of Computer Vision project for Spring-2017.

1. Introduction

In this project, we focus on matching images with disparate appearance which differ at a local pixel level in the sense that neither their intensity nor gradient distributions are locally comparable. For example, consider the images in Fig. 1 where the images are of the same scene but, differ in terms of lighting conditions. A large amount of appearance change occurs in the scene as a result of illumination variation. For the day scenes in the first row, multiple highlights and shadows can be observed clearly whereas, for the night scenes in the second row, the highlights are produced as a result of street lights and the top part of the images are more or less dark.

Spectral methods on the image graph Laplacian are generally used for clustering, segmentation 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 embedded space representation. These coordinates are then further clustered using K-means to obtain discrete clustering solutions. In this project we use the individual eigen-functions themselves as a feature representation of the image pair from which the feature correspondence can be derived. Such a representation captures persistent regions in the image pairs even when there are substantial differences in lighting.

2. Algorithm & Procedure

Given any pair of images in Fig. 1, we are interested in finding some inherent persistent features in the scene like contours, salient regions, local shapes, contrast patterns etc. A shape based matching approach would be useful in matching such images. However, the contrast variations make it very difficult to detect the image contours robustly. Therefore, we use 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.

2.1. 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 having a 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. These weights can be represented as a $n \times n$ combined affinity matrix given by,

$$W = [w_{ij}] \quad \forall\ i, j \in 1, \cdots, n$$

The degree matrix D of this graph is a $n \times n$ diagonal matrix given by,

$$D(i, i) = \sum_{j=1}^{n} w_{ij} \quad \forall\ i \in 1, \cdots, n$$

Using W and D, the normalized graph Laplacian is given by,

$$\bar{L} = I - D^{-1/2} W D^{-1/2}$$

We are interested in the eigen-spectra U of this Laplacian matrix which can be computed by the eigen-value decomposition $\bar{L}\bar{U} = \lambda\bar{U}$ and setting $U = D^{-1/2}\bar{U}$. The eigenvectors $u_1, u_2, \cdots, u_K$ corresponding to the K smallest eigen-values are related to the structure of the graph and are extensively used in the literature to obtain a K-partition of an image based on appropriately defined weight values.

This formulation can be used for a pair of images also. Let, $G_1(V_1, E_1, W_1)$ and $G_2(V_2, E_2, W_2)$ be the image

Figure 1. Multiple images of the same scene with different natural and artificial lightings.

graphs for images I_1 and 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,

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

where, $|V_1| = n_1$, $|V_2| = n_2$ and C is the $n_1 \times n_2$ matrix containing 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.

2.2. Image Features and the Joint Spectrum

Here, we compute the eigen-spectra of this graphs Laplacian to see if the corresponding eigen-functions show any patterns of correspondence. But, we do not see much correspondence between the eigen-functions in this case which motivates the need for features stronger than just the individual pixel colors. Therefore, we use SIFT descriptors computed densely on the image at a fixed spatial sampling.

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-dimension feature vector $f_i(x)$ at each location x in image 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 I_1 and I_2 respectively. Then the affinity matrices W_1, W_2 and C are defined as,

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

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

where, $f_i(x)$ and $f_i(y)$ are features at locations x and y in image I_i. Here, cosine distance is used as the feature distance function. 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.

2.3. Characterization of persistent regions

the extrema of the eigen-function pairs $(J_1^{(2)}, J_2^{(2)}), \cdots, (J_1^{(5)}, J_2^{(5)})$ 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, the Maximally Stable Extremal Region (MSER) detector is 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

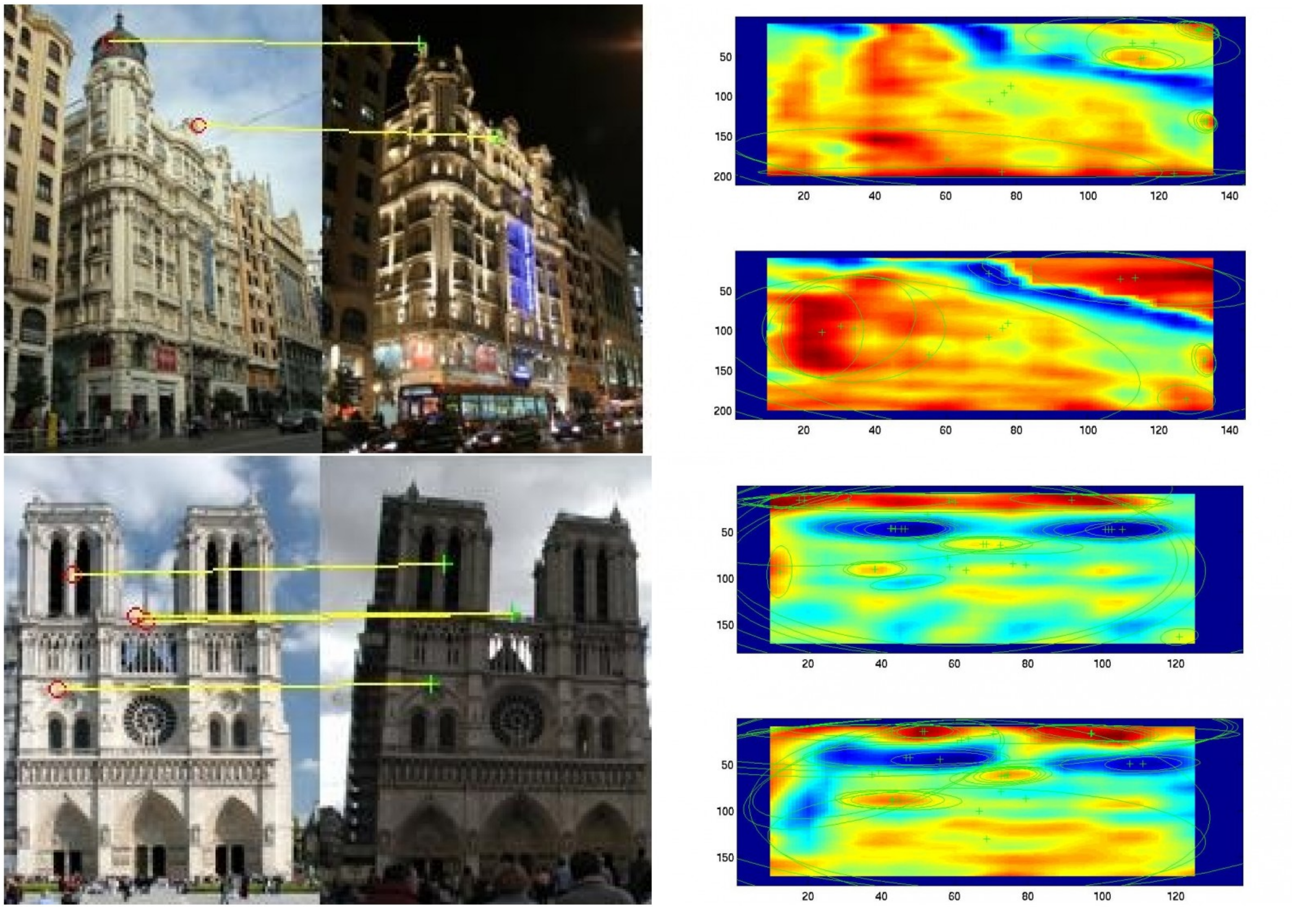

Figure 2. Results of matchings on the left with their MSER features on the right.

an ellipse is typically fit to represent the support region. To apply the MSER detector, each eigen-function $J_1^{(k)}$ and $J_2^{(k)}$ is normalized to a range of $[0, 255]$ by scale and offset correction. Then, intensity-based MSER is run along with ellipse fitting to detect stable affine regions.

2.4. 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. However, the descriptors from each pair of eigen-functions $(J_1^{(k)}, J_2^{(k)})$ are matched 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, a matching from $J_1^{(k)}$ to $J_2^{(k)}$ and from $J_2^{(k)}$ to $J_1^{(k)}$ is performed and only the matches which are mutually consistent are kept. This gives a set of correspondences C_k from the eigen-function pair $(J_1^{(k)}, J_2^{(k)})$. It should be noted that unlike traditional SIFT feature matching, the constraint on being able to match between individual eigen-function pairs results in a much stronger match criterion.

3. Conclusion

In this project, we explored the 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. 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 high repeatability and local similarity.

4. References

[1] Bansal, Mayank and Daniilidis, Kostas. "Joint spectral correspondence for disparate image matching." CVPR 2013.