LAPLACIAN OBJECT: ONE-SHOT OBJECT DETECTION BY LOCALITY PRESERVING PROJECTION

Sujoy Kumar Biswas and Peyman Milanfar

Electrical Engineering Department
University of California, Santa Cruz
1156 High Street, Santa Cruz, CA, 95064

ABSTRACT

One shot, generic object detection involves detecting a single query image in a target image. Relevant approaches have benefitted from features that typically model the local similarity patterns. Also important is the global matching of local features along the object detection process. In this paper, we consider such global information early in the feature extraction stage by combining local geodesic structure (encoded by LARK descriptors) with a global context (i.e., graph structure) of pairwise affinities among the local descriptors. The result is an embedding of the LARK descriptors (extracted from query image) into a discriminatory subspace (obtained using locality preserving projection [1]) that preserves the local intrinsic geometry of the query image patterns. Experiments on standard data sets demonstrate efficacy of our proposed approach.

Index Terms— object detection, locality preserving projection, manifold learning, principal component analysis

1. INTRODUCTION

In this paper, we focus on a particular variety of fine-grained detection [2] where the objective is to take a single query image as input and locate the pattern of the query image in a (usually bigger) target image as illustrated in Fig. 1. Past research work [3, 4] has shown that exemplar based detection strategies can work with laudable success, sometimes as good as training based approaches in coping with specific pose variations. Relevant approaches benefitted from descriptors that typically model the local similarity patterns. Shechtman *et al.* [4] proposed a pattern matching scheme based on *self-similarity* [5, 6, 7, 8, 9, 10], the premise of which is that local internal layout of self-similar pixels are shared by visually similar images. In [3], Seo *et al.*, modeled local geometric layout with *Locally Adaptive Regression Kernel* (LARK) [11] descriptors, and computed similarity between query and target features at each pixel of the target image by sweeping the query window over the target. One common aspect that emerges from such competing approaches is the emphasis on

Thanks to XYZ agency for funding.

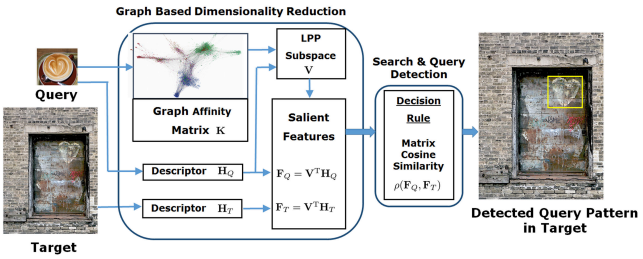


Fig. 1. An overview of the query based object detection methodology where a query pattern is detected in a visually similar part of the target image (on right).

the role of i) local geometry in developing features, and ii) global context for robust pattern matching. For example, in [3], Seo *et al.* have shown that projecting LARK descriptors on principal components results in salient features contributing to improved performance. But we argue that while projecting descriptors on a discriminatory subspace as in [3], it is imperative to take into account which descriptor comes from what spatial location so as to preserve the intrinsic local geometry (something that PCA can't do). To put local information in such global context, we employ a graph based dimensionality reduction technique where a global graph structure comprising pairwise affinities (among neighboring pixels) embeds the local descriptors into a low dimensional manifold in order to preserve locality for superior pattern detection.

The overall strategy to detect a query image in a target image follows a three step process — extracting local descriptors, computing salient features from them, and detecting the features in the target image (see Fig. 1). To avoid ambiguous matching between sets of features, we propose a global graph structure that incorporates spatial information corresponding to each descriptor, and we build this graph with pairwise affinities between neighboring descriptors. Section 2 discusses graph based dimensionality reduction, Section 3 introduces the notion of geodesic affinity, Section 4 presents detection strategy, Section 5 includes experimental results, and Section 6 draws conclusions.

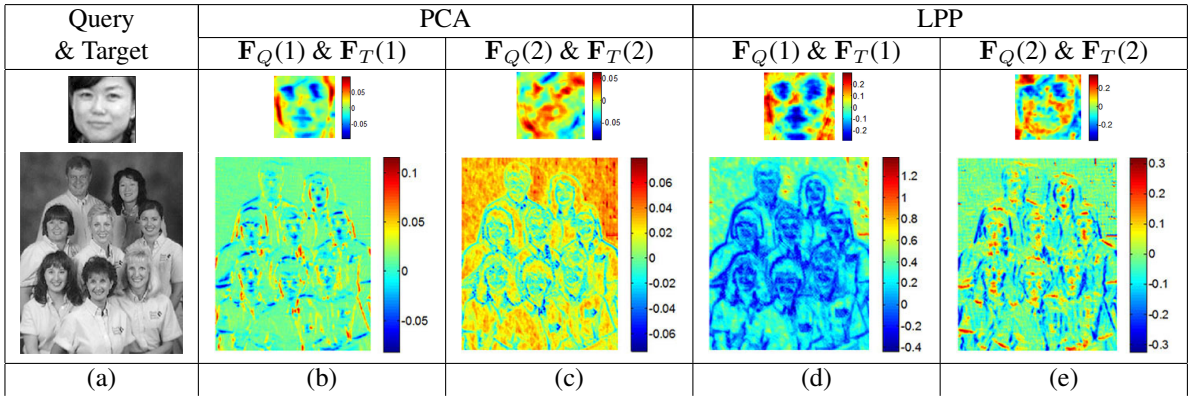


Fig. 2. Salient features shown after dimensionality reduction of LARK descriptors: (a) query & target images, (b)-(c) salient query (target) features $\mathbf{F}_Q$ ($\mathbf{F}_T$) learnt by projecting descriptors $\mathbf{H}_Q$ ($\mathbf{H}_T$) along two dominant principal components, (d)-(e) same LARK descriptors projected along two dominant eigenvectors of LPP (one can notice finer local details in these features)

2. LOCALITY PRESERVING GRAPH BASED DIMENSIONALITY REDUCTION

Consider the parameterized image surface $\mathcal{S}(\mathbf{x}_i) = \{\mathbf{x}_i, z(\mathbf{x}_i)\}$, where $\mathbf{x}_i$ denotes the 2D coordinate vector $\mathbf{x}_i = [x_{i_1} x_{i_2}]^{\mathrm{T}}$, having intensity $z(\mathbf{x}_i)$. We represent the descriptors computed at location $\mathbf{x}_i$ as l-dimensional vector $\mathbf{h}_i \in \mathbb{R}^l$. The descriptors are computed densely from the query Q as well as target image T over a uniformly spaced grid of $\mathbf{x}_i$'s. We denote the number of descriptors computed this way from query Q as N_Q and the same from target T as N_T, respectively. The descriptor vectors $\mathbf{h}_i$ are placed column wise to define the descriptor matrix for query as $\mathbf{H}_Q = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_Q}]$, and the same for target as $\mathbf{H_T} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_T}]$. So we have $\mathbf{H}_Q \in \mathbb{R}^{l \times N_Q}$ and $\mathbf{H}_T \in \mathbb{R}^{l \times N_T}$.

To distill the redundancy in densely computed descriptors, Seo *et al.* [3] have used PCA as a dimensionality reduction technique on the query descriptors $\mathbf{H}_Q$ to obtain salient features for object detection. In contrast, we learn a discriminatory subspace $\mathbf{v}$ from Q such that the query descriptors $\mathbf{h}_i$, when projected on $\mathbf{v}$, respect the local geometric pattern. In other words, if $\mathbf{h}_i$ and $\mathbf{h}_j$, extracted from Q, are closely spaced over the image manifold $\mathcal{S}$ then their projections $\mathbf{v}^{\mathrm{T}}\mathbf{h}_i$ and $\mathbf{v}^{\mathrm{T}}\mathbf{h}_j$ on a subspace $\mathbf{v}$ should be close as well. This objective is achieved by minimizing the following *locality preserving projection* (LPP) [1] cost function —

$$J_{\mathrm{LPP}} = \frac{1}{2}\sum_{ij}(\mathbf{v}^{\mathrm{T}}\mathbf{h}_i - \mathbf{v}^{\mathrm{T}}\mathbf{h}_j)^2\mathbf{K}_{ij}. \quad (1)$$

The cost function J_{LPP} with our proposed affinity measure $\mathbf{K}_{ij}$ (described in Sec. 3) incurs heavy penalty if *neighboring* descriptors $\mathbf{h}_i$ and $\mathbf{h}_j$ are mapped *far apart*. Upon simplification [1], expression (1) leads us to the final penalty as follows:

$$J_{\mathrm{LPP}} \quad = \mathbf{v}^{\mathrm{T}}\mathbf{H}_Q\mathbf{L}\mathbf{H}_Q^{\mathrm{T}}\mathbf{v}. \quad (2)$$

The matrix $\mathbf{D}$ is a diagonal matrix defined as $\mathbf{D}_{ii} = \sum_j \mathbf{K}_{ij}$, and $\mathbf{L} = \mathbf{D} - \mathbf{K}$ is known as the graph Laplacian. The constraint $\mathbf{v}^{\mathrm{T}}\mathbf{H_Q}\mathbf{D}\mathbf{H_Q}^{\mathrm{T}}\mathbf{v} = 1$ is imposed on $\mathbf{D}$ to prevent higher values of $\mathbf{D}_{ii}$ from assigning greater "importance" to descriptor $\mathbf{h}_i$. Minimizing J_{LPP} with respect to the aforementioned constraint leads us to the generalized eigenvalue problem. The projection vector $\mathbf{v}$ that minimizes (2) with respect to the constraint, is given by the minimum eigenvalue solution to the following generalized eigenvalue problem:

$$\mathbf{H}_Q\mathbf{L}\mathbf{H}_Q^{\mathrm{T}}\mathbf{v} = \lambda\mathbf{H}_Q\mathbf{D}\mathbf{H}_Q^{\mathrm{T}}\mathbf{v}. \quad (3)$$

The desired set of eigenvectors which builds our low dimensional LPP subspace comprises the trailing d eigenvectors computed as a solution of (3). We collect the set of d eigenvectors as columns of $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d] \in \mathbb{R}^{l \times d}$. Since the descriptors are densely computed they typically lie on a lower dimensional manifold making d a small integer such as 4 or 5. The descriptor matrices $\mathbf{H}_Q$ and $\mathbf{H}_T$ when projected on $\mathbf{V}$ lead to salient features for query Q and target T represented by $\mathbf{F}_Q$, $\mathbf{F}_T$ respectively as follows:

$$\mathbf{F_Q} = \mathbf{V}^{\mathrm{T}}\mathbf{H}_Q \in \mathbb{R}^{d \times N_Q}, \mathbf{F_T} = \mathbf{V}^{\mathrm{T}}\mathbf{H}_T \in \mathbb{R}^{d \times N_T}. \quad (4)$$

In contrast to the LPP, the objective of PCA is to preserve the global geometric structure of the data by projecting the descriptors along the direction of maximal variation, and thus pays no attention to preserving finer details of local pattern. Fig. 2 illustrates the difference showing the salient query features $\mathbf{F}_Q$ learnt with PCA versus LPP. The images in the figure demonstrate that LPP is able to preserve greater details in the salient features than PCA.

The image descriptors $\mathbf{h}_i$ in our proposed object detection framework are general and can be any (e.g., SIFT [12] or HOG [13]). However, we advocate the use of LARK descriptors [3, 14] because LARK is specifically designed to robustly encode the spatial layout of the image manifold.

3. LOCAL GEOMETRY REPRESENTATION WITH LOCAL GEODESIC COMPUTATION

In this Section, we present a geodesic view to construct the graph affinity matrix $\mathbf{K}$. The local geodesic distance between the two neighboring (i.e., closely spaced) pixels $\mathbf{x}_i$ and $\mathbf{x}_j$ on the image manifold $\mathcal{S}(\mathbf{x}_i)$ can be approximated [14] by the differential arc length ds_{ij} as follows:

$$\begin{aligned} ds_{ij}^2 &= dx_{i_1}^2 + dx_{i_2}^2 + dz^2, \\ &= dx_{i_1}^2 + dx_{i_2}^2 + (\frac{\partial z}{\partial x_{i_1}} dx_{i_1} + \frac{\partial z}{\partial x_{i_2}} dx_{i_2})^2. \end{aligned} \quad (5)$$

The second step in Eq. (5) follows from chain rule of differentials. We can discretize (5) by means of the following two approximations: $dx_{i_1} \approx \Delta x_{i_1 j_1} = x_{j_1} - x_{i_1}$, and $dx_{i_2} \approx \Delta x_{i_2 j_2} = x_{j_2} - x_{i_2}$, i.e., $\Delta x_{i_1 j_1}$ and $\Delta x_{i_2 j_2}$ representing displacements along the two image-axes. The discretized form of the differential arc length is given by the following expression:

$$\begin{aligned} ds_{ij}^2 &\approx \Delta x_{i_1 j_1}^2 + \Delta x_{i_2 j_2}^2 + \dots \\ &(\frac{\Delta z}{\Delta x_{i_1 j_1}} \cdot \Delta x_{i_1 j_1} + \frac{\Delta z}{\Delta x_{i_2 j_2}} \cdot \Delta x_{i_2 j_2})^2, \\ &= \Delta x_{ij}^{\mathrm{T}} \mathbf{C}_i \Delta x_{ij} + \Delta x_{ij}^{\mathrm{T}} \Delta x_{ij}, \\ &\approx \Delta x_{ij}^{\mathrm{T}} \mathbf{C}_i \Delta x_{ij}, \end{aligned} \quad (6)$$

where $\Delta x_{ij} = [\Delta x_{i_1 j_1} \ \ \Delta x_{i_2 j_2}]^{\mathrm{T}}$, and $\mathbf{C}_i$ is the local gradient covariance matrix (also called as steering matrix in [3]) computed at $\mathbf{x}_i$. The second step follows from the matrix-vector representation of previous step, and the last step is written by considering the fact that $\Delta x_{ij}^{\mathrm{T}} \Delta x_{ij}$ is data independent and trivial in a small local window [14].

However, straightforward computation of $\mathbf{C}_i$ based on raw image gradient at a single pixel may be too noisy. Therefore, we estimate $\mathbf{C}_i$ as an "average" gradient covariance matrix $\mathbf{C}_{\Omega_i}$ by collecting derivatives of the image signal $z(\mathbf{x}_i)$ over a patch Ω_i of pixels centered at pixel $\mathbf{x}_i$ (see Fig. 3(b)-3(c)) followed by an eigen-decomposition. Representing first derivatives as $\Delta_{i_1} z(m)$ and $\Delta_{i_2} z(m)$ along the two axes x_{i_1} and x_{i_2}, respectively, we write the final expression of $\mathbf{C}_{\Omega_i}$ as follows:

$$\begin{aligned} \mathbf{C}_{\Omega_i} &= \sum_{m \in \Omega_i} \begin{bmatrix} \Delta_{i_1} z(m)^2 & \Delta_{i_1} z(m) \cdot \Delta_{i_2} z(m) \\ \Delta_{i_1} z(m) \cdot \Delta_{i_2} z(m) & \Delta_{i_2} z(m)^2 \end{bmatrix}, \\ &= \nu_1 \mathbf{u_1} \mathbf{u_1}^{\mathrm{T}} + \nu_2 \mathbf{u_2} \mathbf{u_2}^{\mathrm{T}}, \\ &= (\sqrt{\nu_1 \nu_2} + \varepsilon)^{\alpha} (\frac{\sqrt{\nu_1} + \tau}{\sqrt{\nu_2} + \tau} \mathbf{u_1} \mathbf{u_1}^{\mathrm{T}} + \frac{\sqrt{\nu_1} + \tau}{\sqrt{\nu_2} + \tau} \mathbf{u_2} \mathbf{u_2}^{\mathrm{T}}), \end{aligned} \quad (7)$$

where, ν_1 and ν_2 are eigenvalues of $\mathbf{C}_{\Omega_i}$ corresponding to eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. Also in the derivation above, $\varepsilon, \tau, \alpha$ are regularization parameters to avoid numerical instabilities and kept constant throughout all the experiments in this paper at 10^{-7}, 1 and 0.1 respectively.

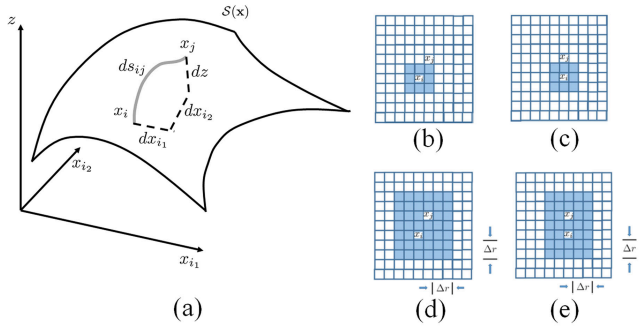


Fig. 3. (a) Geodesic distance ds_{ij} between the points $\mathbf{x}_i$ and $\mathbf{x}_j$ on the signal manifold $\mathcal{S}(\mathbf{x})$ is shown. Since patch Ω_i is centered at $\mathbf{x}_i$ as shown in (b) and (c), using definition $\mathbf{C}_{\Omega_i}$ (7) to compute ds_{ij} would make $ds_{ij} \neq ds_{ij}$. To make ds_{ij} symmetric, we make Ω common, as the circumscribing patch Ω_{ij} between $\mathbf{x}_i$ and $\mathbf{x}_j$ in (d)-(e). One can enlarge the common patch by adding an annular width of Δr pixels as required.

3.1. Building a Graph Laplacian with Geodesic Affinities

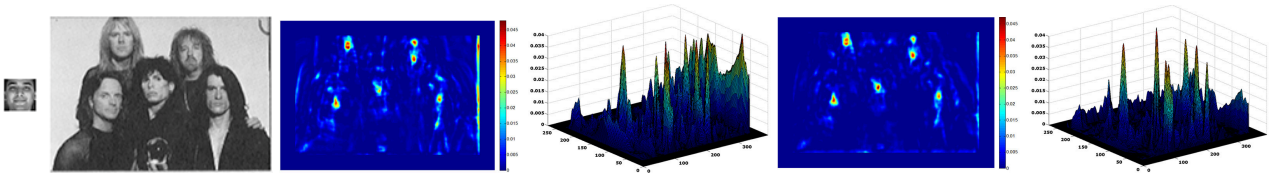
Next, we build a graph structure from query image using the query descriptors as the graph nodes. We define the affinities between neighboring descriptor locations $\mathbf{x}_i$ and $\mathbf{x}_i$ as follows,

$$\mathbf{K}_{ij} = \begin{cases} e^{-\frac{ds_{ij}^2}{2\sigma^2}} & \text{when } \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \gamma, \\ 0 & \text{otherwise}, \end{cases} \quad (8)$$

where σ is a smoothness parameter and γ a radius within which we limit our affinity computation. Unfortunately, computing ds_{ij} in (8) using $\mathbf{C}_{\Omega_j}$ (7) would render ds_{ij} (and hence $\mathbf{K}_{ij}$) non-symmetric. This is because the support Ω_i of $\mathbf{C}_{\Omega_i}$ is different from Ω_j of $\mathbf{C}_{\Omega_j}$ (see Fig. 3(b)-3(c)), and upon using Eq. (7) to compute ds_{ij} we have $ds_{ij} \neq ds_{ji}$. However, $\mathbf{K}$ is assumed symmetric in the derivation of LPP subspace (2). As a remedy to this technical predicament we make the support of $\mathbf{C}_{\Omega_j}$ in (8) common for both $\mathbf{x}_i$ and $\mathbf{x}_j$ as shown by a circumscribing rectangle in the Fig. 3(d)-3(e). The common support is denoted by Ω_{ij} and we denote the corresponding gradient covariance matrix as $\mathbf{C}_{\Omega_{ij}}$. One can enlarge the rectangular patch by adding an annular width of Δr pixels (equal to 2 shown in Fig. 3). The final expression of the affinity becomes the following:

$$\mathbf{K}_{ij} = \begin{cases} e^{-\frac{\Delta x_{ij}^{\mathrm{T}} \mathbf{C}_{\Omega_{ij}} \Delta x_{ij}}{2\sigma^2}} & \text{when } \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \gamma, \\ 0 & \text{otherwise}. \end{cases} \quad (9)$$

A straightforward solution of taking average of $\mathbf{K}_{ij}$ and $\mathbf{K}_{ji}$ to make $\mathbf{K}$ symmetric does not work in practice because such average oversmooths the dominant structure pattern over the image manifold. Next, we discuss the detection framework using salient features $\mathbf{F}_Q$ and $\mathbf{F}_T$.



(a) (b) (c) (d) (e)

Fig. 4. (a) Query & target image, (b) resemblance map $f(\rho)$ with PCA, (c) transformed correlation peaks $f(\rho)$ with PCA, (d) resemblance map $f(\rho)$ with LPP, (e) transformed correlation peaks $f(\rho)$ with LPP. The sharp peaks of $f(\rho)$ with less crowding is visible in LPP results.

4. DETECTION FRAMEWORK

To detect a given query Q in a target image T, we sweep Q over T and in each target position $\mathbf{x}_i$ (i.e., centered at pixel $\mathbf{x}_i$) we compute cosine similarity between $\mathbf{F}_Q$ and $\mathbf{F}_T$ as follows: $\rho(\mathbf{F}_Q, \mathbf{F}_{T_i}) = \text{trace}(\frac{\mathbf{F}_Q^{\text{T}}\mathbf{F}_{T_i}}{\|\mathbf{F}_Q\|_F \|\mathbf{F}_{T_i}\|_F})$. Since $\rho \in [-1, 1]$, to suppress small correlation values further down and to boost the correlation peaks up we transform the ρ according to Lawley-Hotelling trace statistic [15, 16] as follows: $f(\rho) = \frac{\rho^2}{1-\rho^2}$. Here, $f(\rho)$ gives us a confidence or resemblance map [3], the same size as the target image, and we threshold the map with τ to achieve 99% confidence level from the empirical probability density function (PDF) of $f(\rho)$.

5. EXPERIMENTAL RESULTS

We have evaluated our methodology on two data sets, namely, UIUC car data set [17] and MIT-CMU face data set [18]. The LARK kernel size is maintained at 9×9 giving rise to 81-dimensional descriptors. We reduce the dimensionality to a small integer of roughly 5 by the use of LPP.

UIUC data contain gray-scale car images at same scale, and at multiple scales in the test set. MIT-CMU faces is also a gray-scale image data set and we have evaluated our methodology on the same subset of images as done by Seo *et al.* [3]. The test set consists of 43 gray-scale images containing a total of 149 frontal faces, occurring at various scales, and 20 gray-scale images [3] having faces at unusually large ($> 60°$) angular orientation. For multi-scale and multi-oriented detection, we have scaled and rotated the query image and searched for the transformed query pattern in target.

The locality preserving projection turns out to be robust in detecting query pattern even in the presence of noise, low resolution, pose variation and viewpoint changes. The resemblance maps for the target image in Fig. 4(a) is shown in Fig. 4(b) (for PCA) and Fig. 4(d) (for LPP). The correlation peaks resulting from LPP tend to be much sharper than PCA. The apparent crowding of correlation peaks in PCA is absent in LPP. Fig. 5 presents some results on UIUC car data set. We have summarized the performances of our methodology in Table 1 in terms of equal-error detection rates (equal to recall rate when it becomes same as precision rate). The proposed methodology is also compared with training-based approaches in the table. We have consistently used a single query in both the data sets to arrive at these performances.

Fig. 5. Detection results on UIUC Car data set with a single query image (top left)

Table 1. Equal Error Detection Rates

Approaches		Data Sets		
		UIUC Car		MIT-CMU Faces
		Single Scale	Multi-Scale	
Training Free	Proposed	90.76	76.91	89.01
	Seo *et al.* [3]	85.39	74.51	86.58
With Training	Agarwal *et al.* [17]	77.01	44.00	-
	Mutch *et al.* [19]	99.94	90.06	-
	Kapoor *et al.* [20]	94.00	93.50	-
	Lampert *et al.* [21]	98.50	98.60	-
	Wu *et al.* [22]	97.60	-	-

6. CONCLUSION

In this paper we have proposed an embedding technique for local descriptors in the context of one-shot object detection. We have further connected the idea of LPP with a geodesic notion of affinity, thereby replacing the traditionally used heat kernel to define graph affinities. The proposed approach is general enough to be integrated with any choice of descriptors. Finally, we have demonstrated the efficacy of this approach with improved experimental results.

7. REFERENCES

[1] H. Xiaofei, Y. Shuicheng, H. Yuxiao, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 3, pp. 328–340, 2005.

[2] A. Berg, R. Farrell, A. Khosla, J. Krause, F. F. Li, J. Li, and S. Maji, "Fine-grained challenge 2013," *Available: https://sites.google.com/site/fgcomp2013/.*

[3] H. J. Seo and P. Milanfar, "Training-free, generic object detection using locally adaptive regression kernels," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1688–1704, 2010.

[4] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[5] C. F. Chen and Y. C. F. Wang, "Exploring self-similarities of bag-of-features for image classification," in *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011, pp. 1421–1424.

[6] K. Chatfield, J. Philbin, and A. Zisserman, "Efficient retrieval of deformable shape classes using local self-similarities," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 264–271.

[7] T. Deselaers and V. Ferrari, "Global and efficient self-similarity for object classification and detection," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1633–1640.

[8] A. Vedaldi and A. Zisserman, "Self-similar sketch," in *Computer Vision–ECCV 2012*, pp. 87–100. Springer, 2012.

[9] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[10] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman, "Multiple kernels for object detection," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 606–613.

[11] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *Image Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 349–366, 2007.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[14] H. J. Seo and P. Milanfar, "Face verification using the lark representation," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 4, pp. 1275–1286, 2011.

[15] M. M. Tatsuoka and P. R. Lohnes, *Multivariate analysis: Techniques for educational and psychological research*, Macmillan Publishing Co, Inc, 1988.

[16] T. Caliński, M. Krzyśko, and W. Wołyński, "A comparison of some tests for determining the number of nonzero canonical correlations," *Communications in Statistics: Simulation and Computation*, vol. 35, no. 3, pp. 727–749, 2006.

[17] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 11, pp. 1475–1490, 2004.

[18] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 1, pp. 23–38, 1998.

[19] J. Mutch and D. G. Lowe, "Multiclass object recognition with sparse, localized features," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. IEEE, 2006, vol. 1, pp. 11–18.

[20] A. Kapoor and J. Winn, "Located hidden random fields: Learning discriminative parts for object detection," in *Computer Vision–ECCV 2006*, pp. 302–315. Springer, 2006.

[21] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Beyond sliding windows: Object localization by efficient subwindow search," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[22] B. Wu and R. Nevatia, "Simultaneous object detection and segmentation by boosting local shape feature based classifier," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.