Salient Keypoint Selection for Object Representation

Prerana Mukherjee†, Siddharth Srivastava† and Brejesh Lall* †
* Bharti School of Telecommunication Technology and Management, Indian Institute of Technology, Delhi, India
† Department of Electrical Engineering, Indian Institute of Technology, Delhi, India
{eez127506, eez138300, brejesh}@ee.iitd.ac.in

***Abstract*—In this paper, we propose a keypoint selection scheme for SIFT and KAZE features and demonstrate their effectiveness in object characterization. The selection criterion rely on the detectability, distinctiveness and repeatability of the keypoints. These scores are combined to give a keypoint saliency score. The keypoints are ranked according to their saliency values and weak/irrelevant keypoints are filtered out based on a threshold value. These keypoints are further augmented with the keypoints obtained by applying SIFT to the texture map constructed using Gabor filter. The keypoint set represents the boundaries and object regions effectively. Experimental results validate the claims that the salient keypoints chosen by the proposed methodology are well suited for object representation.**

I. INTRODUCTION

Detection of appropriate keypoints has become an important problem which has lead to the development of ubiquitously popular techniques such as SIFT [1], SURF [2], ORB [3], FREAK [4] etc. Selecting relevant keypoints from a set of detected keypoints assists in reducing not only the computational complexity but also the error propagated due to irrelevant keypoints. In [5] the authors refer to the importance of a keypoint by obtaining a saliency score based on distinctivity, repeatability and detectability of the keypoints. In order to decide which of the features/keypoints are appropriate for classification (good features) we base our paper on these three criteria to choose the keypoints and rank them on the basis of their combined saliency scores. Distinctivity helps in obtaining keypoints which represent unique information within an image. Repeatability refers to detecting the same keypoint independently under different transformations. Incorporation of repeatability thus leads to choice of keypoints which are robust under various deformations. Detectability quantifies the strength of detecting the keypoint in various changed conditions.

A few techniques that have been proposed in literature to select keypoints are usually tuned to a specific application scenario. The authors in [6] propose a keypoint selection scheme for visual tracking. The algorithm is based on a significantly faster and efficient Suppression via Disk Covering. An important observation of their work was that the spatial distribution of the selected keypoint was pivotal in improvising visual tracking in videos. In another work [7], authors postulate a grid based keypoint selection methodology where they obtain a spatial distribution of the motion followed between the frames. The advantage of grid based system is that it is able to cover a larger region as compared to selecting keypoints based only on strongest response which may be concentrated in a smaller area. The work in [8] poses keypoint matching as an optical flow problem. The matches are determined by constructing a histogram over components of optical vectors between the corresponding keypoints. The bins with highest response and their neighbors are considered as putative matches which are further pruned by RANSAC algorithm. A graph based matching strategy followed by non maximal suppression within voxels has been studied in [9]. A strategy to select only those keypoints which have high probability of being matched has been described in [10]. This is achieved by training a random forest classifier over descriptors to be matched. The authors demonstrate that the most matchable keypoints are those which lie on regions with reasonably high Difference of Gaussian (DoG) responses. In our work, we show that stronger representation of such a property can be realized by a combination of SIFT and KAZE keypoints.

In our previous work [11], we have established the complementary nature of SIFT and KAZE features for object classification. It was observed with exhaustive experimental analysis that this combination worked substantially well and outperformed the state of the art techniques. In this paper, we propose a keypoint selection strategy based on importance of keypoints which is mathematically intuitive rather than a being heuristic approach. The proposed methodology is able to achieve the minimal set of salient keypoints representative for the object. The importance of a keypoint is obtained by empirically incorporating three key properties of keypoints viz. distinctiveness, detectability and repeatability to estimate its saliency. This approach is motivated by a similar work in [5], where the authors use binary descriptors. Instead, we adapt the technique for a combination of SIFT and KAZE descriptors for reasons given in the previous paragraph. To this end, we introduce keypoint ranking for a combination of SIFT and KAZE [12] features. KAZE features have strong response along the boundary of objects while SIFT captures shape, texture etc. similar to neuronal response of human vision system. Therefore, we hypothesize that a combination of SIFT and KAZE keypoints will be able to characterize object properties well and could boost the matching accuracy significantly. The characterization is further strengthened by incorporating texture computed by SIFT keypoint responses over a texture map on the original image using Gabor filter. In our evaluations we observe that the proposed keypoints are localized around the object boundaries and the regions

978-1-5090-2361-5/16/$31.00

inside the object (characterized by SIFT keypoints on the texture map) as opposed to keypoints distributed throughout the image.

In view of the above discussions the key contributions can be summarized as:

- To the best of our knowledge, this is the first work using KAZE with SIFT keypoints for salient keypoint selection aimed at object characterization and its subsequent use for object matching.
- Salient Keypoint selection of SIFT features on Gabor convolved image for representation of features inside object boundaries in context of object characterization is a novel approach.
- We adapt distinctiveness, detectability and repeatability scores for keypoints to euclidean space. Since the distances between the descriptors of the popular feature extraction techniques including those obtained with deep learning methods rely on euclidean distances, this adaptation would play a crucial role in expanding the proposed scheme to other methods.

Rest of the paper is organized as follows. Section II, we provide a brief overview of the techniques used while in Section III we discuss the methodology. The results are discussed in Section IV. Finally, we conclude the paper in Section V.

II. Background

In this work, we utilize a ranked combination of SIFT and KAZE keypoints along with keypoints computed from the texture map produced by Gabor filter. This combination gains from both the saliency induced by selection of keypoints, which are intrinsic to object characterization and texture features. The keypoints such chosen give a strong representation for the objects. SIFT and its variants [13] [14] have remained the strongest detectors and descriptors for a long time. SIFT is based on Gaussian Scale Space which smoothens an image irrespective of its content. It is well known that sharp edges or transitions are one of the key characteristics of objects [15]. Therefore, SIFT or any other detector based on uniform smoothening would tend to lose out on this crucial boundary information. In order to overcome this, we have augmented the strength of SIFT keypoints with KAZE features, which are based on non-linear anisotropic diffusion filtering [16]. The anisotropic diffusion filtering preserves the edge regions as compared to other regions. This property makes KAZE more responsive towards boundaries/edges in the image. Thus, the ability of KAZE to describe boundary and that of SIFT to represent region information, makes them a suitable candidate for fusion to identify suitable keypoints for object representation. The methodology to select keypoints is detailed in Section III. It would be appropriate to indicate here that the selection of keypoints is based on three distinguishing factors, namely, distinctiveness, detectability and repeatability of keypoints. Distinctiveness means how different the keypoint is from the rest of the keypoints in the image, detectability represents how robustly the keypoint can be detected under viewpoint/lighting changes and repeatability refers to the ability of keypoints to remain invariant to various transformations. We combine these three properties to obtain a score and rank the keypoints.

Additionally, we supplement the SIFT and KAZE keypoints from original image with the SIFT keypoints obtained from the texture map using Gabor filter. Further saliency map obtained using [17] is used to threshold out 'weak' keypoints. This saliency map makes use of non-linear scale space filtering which helps in retaining the edge information. It comprises of features computed at local, global and rarity level. Thus, the saliency map encompasses both spatial and frequency information and can capture multiple objects with varying saliency values in the same image. The computation of the saliency map is given as,

$$salmap = w_1 * local + w_2 * global + w_3 * rarity \quad (1)$$

where $local$ represents local features comprising of color, intensity, orientation, depth and motion maps. $global$ represents global features consisting of global contrast and spatial sparsity while $rarity$ represents rarity (Phase Spectrum of quaternion Fourier transform) features. The weights are calculated as, $w_i = \frac{\sigma_i}{\sum_{i=1}^{N} \sigma_i}$, where w_i is the weight of the i^{th} map and N are the total number of maps.

This step of merging the ranked keypoints with texture keypoints is necessary since in our experiments we noticed that SIFT keypoints not always necessarily lie on regions with textures (for example, keypoints on regions with uniform texture such as roof surfaces was very low). The texture map is obtained using Gabor Filter where the orientations used for filtering are obtained from the cumulative orientation of SIFT keypoints. The selection strategy for texture SIFT keypoints are based on saliency values (Section III-B).

III. Proposed Methodology

In this Section, we give a detailed overview of the keypoint selection scheme and ranking strategy. The workflow of the proposed method is shown in Figure 1. In the following subsections we describe the components in brief followed by a detailed description of the proposed methodology.

A. Keypoint Selection and Ranking

In order to show the invariance of the keypoint set against various transformations, we apply a set of transformations on the original images like rotation ($\pi/6, \pi/3, 2 * \pi/3$), scaling (0.5, 1.5, 2), cropping (20%, 50%), affine. The keypoint selection is performed by ranking the keypoints from SIFT and KAZE by defining a saliency score as below:

$$S_{KP(i)} = Dist(KP(i)) + Det(KP(i)) + Rep(KP(i)) \quad (2)$$

where $S_{KP(i)}$ is the saliency score, $Dist(KP(i))$ is the distinctivity, $Det(KP(i))$ is the detectability and $Rep(KP(i))$ is the repeatability of the i^{th} keypoint respectively and are given by the following equations.

$$KP(i) = \{(x_i, y_i), s_i\}, i = 1...N \quad (3)$$

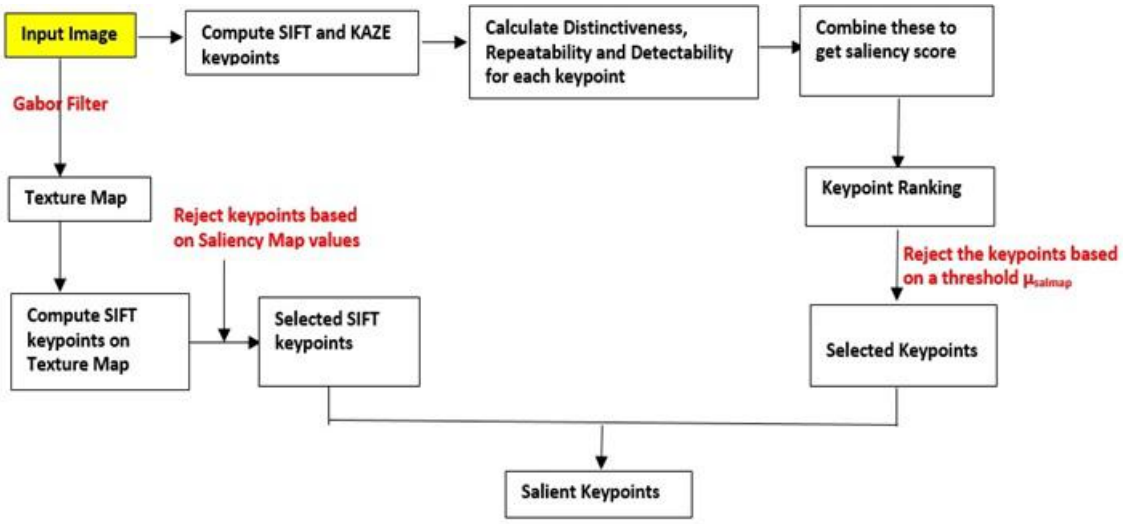

Fig. 1: Flow diagram for the proposed methodology

$$Dist(KP(i)) = \frac{1}{N-1} \sum_{(x_i,y_i)\epsilon KP(i), i \neq j} ED(d_i, d_j) \quad (4)$$

$$Rep(KP(i)) = \frac{1}{nTransf} \sum_{1}^{nTransf} ED(d_i, d_j^t) \quad (5)$$

$$Det(KP(i)) = \frac{1}{nTransf} \sum_{1}^{N} s_i \quad (6)$$

Equation 3 gives the description of i^{th} keypoint which gives the location (x_i, y_i) and response of the keypoint s_i. Equation 4 i.e. distinctiveness gives the summation of the euclidean distances between every pair of keypoint descriptors in the same image. Equation 5 i.e. repeatability gives euclidean distance (ED) between the keypoint descriptor in the original image to the keypoint descriptor mapped in the corresponding transform, t. Here, $nTransf$ is the number of transformations. Equation 6 i.e. detectability is the summation of the strengths of the keypoints in the original image and its respective transforms. All these scores are normalized to the range $[0, 1]$. Next, we select the KAZE and SIFT keypoints which have saliency score greater than the respective mean saliency scores and are given as follows,

$$SalientKP = KP(i) \quad s.t. \quad S_{KP(i)} \geq \mu_{salscore}, 1 \leq i \leq N \quad (7)$$

where N is the total count of keypoint from respective detector and $\mu_{salscore}$ is mean of the saliency scores.

$$\mu_{salscore} = mean(S_{KP(i)}), \qquad 1 \leq i \leq N \quad (8)$$

B. Texture Map based SIFT keypoints

The texture is computed using Gabor filter. Initially, SIFT keypoints are calculated on the original image. Then, the orientation histogram of the keypoints is constructed. The dominant orientations are found by binning the keypoint orientations into prespecified number of bins. The image is then convolved with Gabor filter (Equation 9) using these dominant orientations. Gabor Filters are the product of Gaussian with sinusoidal or cosinosoidal function. Next, the saliency map [17] is calculated for the original image. For each keypoint, if the saliency value is greater than the mean saliency then the keypoint is retained as illustrated in Equation 10.

$$G_{\theta,u,\sigma}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} . e^{2\pi\imath(ux\cos\theta + uy\sin\theta)} \quad (9)$$

where u denotes the frequency of the sinusoidal function, θ gives the orientation of the function, σ is the standard deviation of the Gaussian function.

$$TextureKP = KP(i) \quad s.t. \quad S_{KP(i)} \geq \mu_{salmap}, 1 \leq i \leq N \quad (10)$$

where TextureKP denotes the set of keypoints which are salient for representing the texture. μ_{salmap} denotes the mean of the saliency map. We finally combine the set of salient keypoints (SIFT and KAZE) with texture salient keypoints (SIFT) to form the set of salient keypoints. The algorithm for ranking the keypoints is given in Algorithm 1.

IV. Experimental Results and Discussions

We begin this section by introducing the data sets and evaluation strategy used. We then discuss the results in the subsequent subsections.

A. Data Sets and Evaluation

We evaluate our technique using two methods. First, we show that the set of salient keypoints chosen by our proposed algorithm is able to effectively characterize and represent the objects in the image. Second, we demonstrate the practical application of the proposed method for object matching. The

Algorithm 1 Algorithm for Ranking Salient Keypoints

procedure RANK–KP
  Compute SIFT ($SIFT_{KP}$) and KAZE ($KAZE_{KP}$) keypoints on input image I
  for each keypoint i
    $S_{KP(i)} = Dist(KP(i)) + Det(KP(i)) + Rep(KP(i))$
  end for
  $SalientKP = \{[SIFT_{KP} \quad KAZE_{KP}] \quad | \quad S_{KP(i)} \geq \mu_{salscore}\})$
  Distribute orientations of $SIFT_{KP}$ into c equal sized bins
  Compute texture map using Gabor Filter with c orientations
  $TextureKP = \{SIFT_{KP} \quad | \quad S_{KP(i)} \geq \mu_{salmap}\}$
  $RankedKP = [SalientKP \quad TextureKP]$
end procedure

first set of experiments are conducted on Caltech-101 dataset [18] which contain the corresponding annotation files for each object. The object matching experiments are performed on VGG affine dataset [19]. The experiments have been performed using MATLAB 2014a, OpenCV and vlfeat [20].

B. Object Representation

The Caltech-101 dataset consists of annotations for ground truth object boundaries and bounding boxes. For evaluation, we calculate SIFT and KAZE keypoints on all the images of this dataset. We also calculate the SIFT keypoints on the texture map obtained using Gabor Filter. The orientations for computing the texture map are obtained by first distributing the orientations of the SIFT keypoints obtained on original image into ten equally spaced bins. Then the center (average) of the minimum and maximum values of each bin is used as candidate orientations to the Gabor Filter. The keypoints thus obtained using texture map as well as the original SIFT and KAZE keypoints are filtered and ranked as per Algorithm 1. We also show that these combined set of keypoints represent the object properties i.e. boundary, appearance and texture better than the keypoints from various other detectors. Figure 2 shows an example of texture and saliency maps using the criteria discussed in Section III. Figure 3(a) shows the keypoints obtained using the proposed technique. As can be observed, the keypoints are concentrated around the object region.

We now discuss the empirical validation of the above discussion. Table I shows the percentage of keypoints lying within the object bounding box for various detectors. The results indicate that our proposed keypoint selection method, $RankedKP$ consisting of $SalientKP$ and $TextureKP$, significantly outperforms SIFT and SURF while slightly lagging behind KAZE on this evaluation metric. In order to better analyze this observation, we performed the same analysis around the contour of the object (Figure 2(d)). The results are shown in Table I(b). We observe that the proposed technique results in significantly high number of keypoints strictly within the object region (contour) especially as compared to KAZE keypoints. This shows that KAZE keypoints are not distributed evenly across the object, whereas the keypoints obtained by our proposed technique are spread in the object regions as well as the boundaries (For ex: Figure 2(d)).

TABLE I: Performance Analysis of various feature detectors for object representation.

Feature Detector	Keypoints inside Bounding Box (in %) (a)	Keypoints inside region (in %) (b)
SIFT	76	62
SURF	71	58
KAZE	87.62	69
RankedKP	84	82.7

C. Object Matching

Nine different transformations (scaling, rotation, affine, cropping etc.) were applied to the images from the VGG dataset for each class. The ranked feature set is calculated on the images belonging to the same class. They are matched and average Euclidean Distance score for all the matched keypoints is computed. The number of correct matches is computed by using homography matrices to check the mapping of the keypoints on the transformed image. The results for these experiments is illustrated in Table II. It was observed that the mean distance of keypoints by our ranking scheme was lesser by an order of magnitude as compared to SURF while being two orders lower than SIFT and KAZE. The lower distance demonstrates the ability of the chosen keypoints to represent the image information more robustly and also decreases the number of false positives as also indicated by Table II (b). It may seem that the results are biased due to the presence of affine transformations in the images of the VGG dataset and that constructed with fixed parameter set during calculation of ranked keypoints. But it is important to state that the reason behind using various transforms in our method is firstly to be able to capture maximum variations by unifying keypoints from different transforms. Secondly, since the transformations and parameters to perform them used in our method are constant it is very unlikely that the results would be biased in the dataset having six distinct homographies. Figure 3(b) shows the matching results obtained by our keypoint selection strategy on the VGG dataset. The number of keypoints gets reduced to result in a stronger representative keypoint set which gives higher matching accuracy.

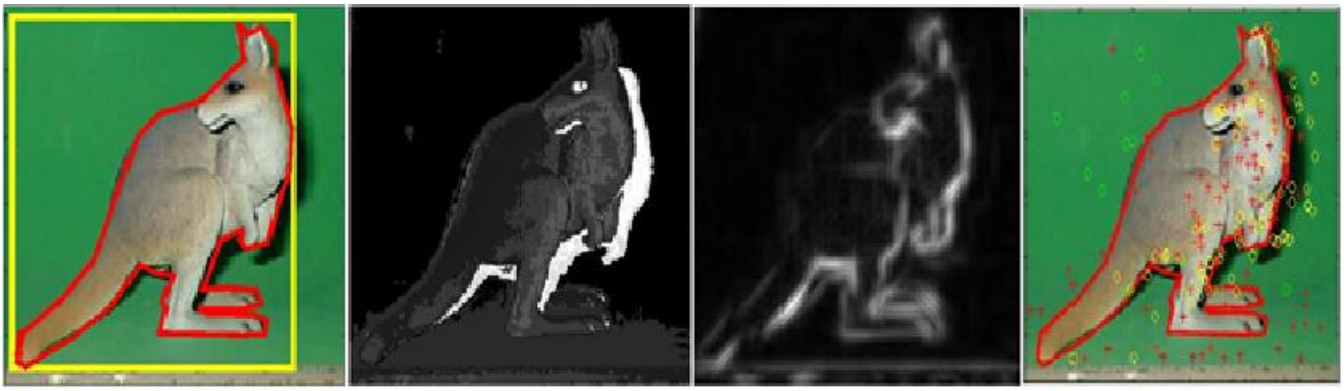

Fig. 2: Figure showing a) Object annotation b) Saliency Map c) Gabor filtered image (Texture Map) d) Ranked keypoints inside the object contour.

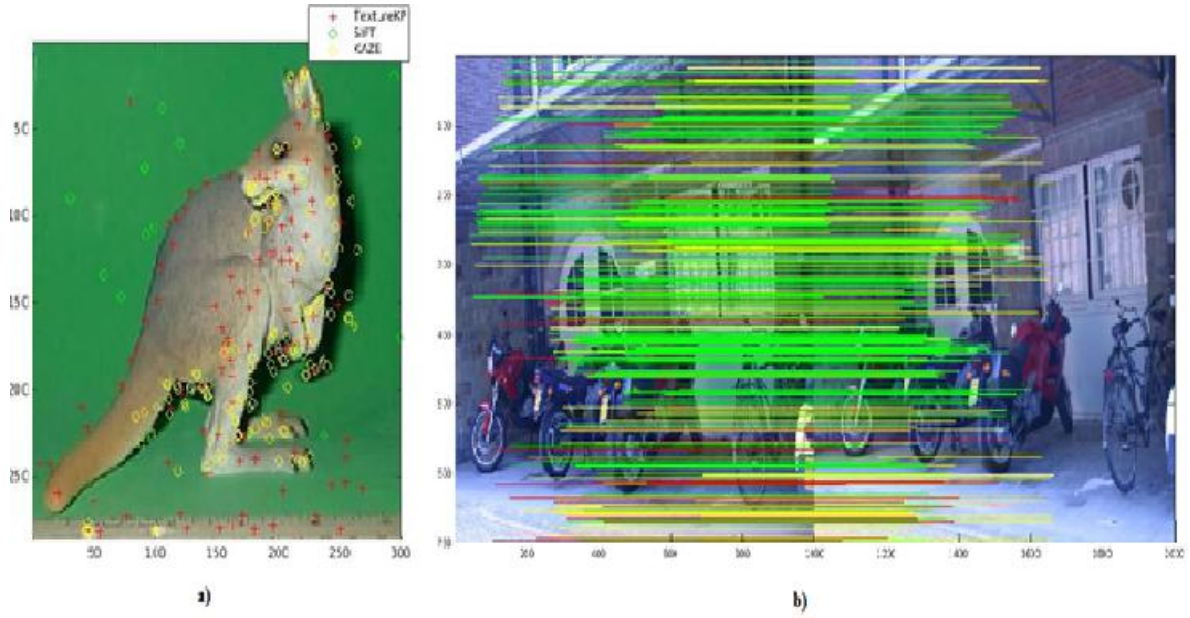

Fig. 3: a) Texture and Ranked (SIFT and KAZE) keypoints b) Correctly matched keypoints by the proposed selection strategy: red ($KAZE_{KP}$), yellow ($SIFT_{KP}$), green ($TextureKP$) on the bikes dataset (VGG).

TABLE II: Object Matching.

Feature Detector	Total KP matches (%) (a)	Correct KP Matches(%) (b)	Mean ED (c)
SIFT	67	90	0.0960
SURF	78	74	0.0752
KAZE	81	91	0.1297
RankedKP	89	96	0.0011

Figure 4 gives the average Euclidean distances between the matching scores obtained from different detectors. Our approach easily outperforms the considered detectors as it is able to give the keypoints which are around the object boundaries and also those which explicitly represents the texture of the object.

V. CONCLUSION

This paper proposes a novel keypoint selection scheme based on SIFT and KAZE. The technique incorporated texture information by finding SIFT keypoints on a texture map of the image obtained using Gabor Filter. As seen from the results, technique can characterize an object region more efficiently than other contemporary detectors. Moreover, the method is less prone to false positives along with demonstrating its effectiveness on a practical application of object matching. We expect that the proposed keypoint selection technique would help in extending the existing object matching and classification algorithms by explicitly providing a representation of the objects to be matched. The technique could also help in improvising the techniques for object localization, segmentation and many other domains. Hence, the hybrid keypoint selection

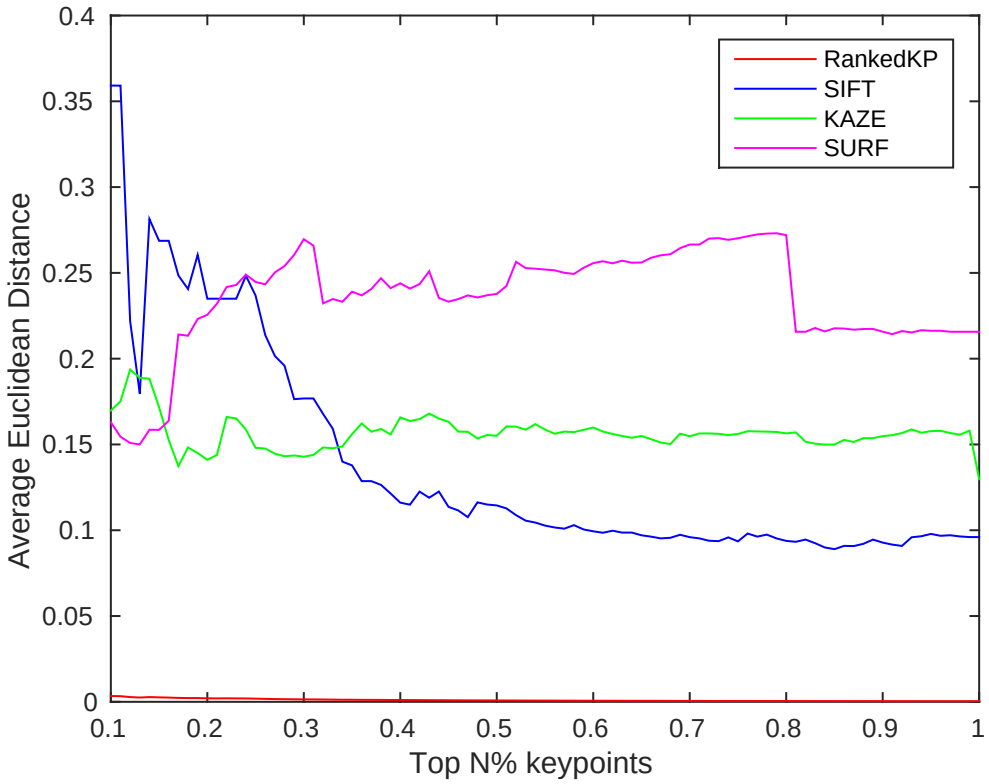

Fig. 4: Average ED vs top N% keypoints of the feature set

strategy holds promise to extend the existing state of the art in many application areas where objects are involved.

REFERENCES

[1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[2] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.

[4] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. Ieee, 2012, pp. 510–517.

[5] S. Buoncompagni, D. Maio, D. Maltoni, and S. Papi, "Saliency-based keypoint selection for fast object detection and matching," *Pattern Recognition Letters*, 2015.

[6] S. Gauglitz, L. Foschini, M. Turk, and T. Höllerer, "Efficiently selecting spatially distributed keypoints for visual tracking," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 1869–1872.

[7] V. Nannen and G. Oliver, "Grid-based spatial keypoint selection for real time visual odometry." in *ICPRAM*, 2013, pp. 586–589.

[8] Y. Liu, R. Feng, and H. Zhang, "Keypoint matching by outlier pruning with consensus constraint," 2015.

[9] S. Li and A. Calway, "Rgbd relocalisation using pairwise geometry and concise key point sets," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 6374–6379.

[10] W. Hartmann, M. Havlena, and K. Schindler, "Predicting matchability," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 9–16.

[11] S. Srivastava, P. Mukherjee, and B. Lall, "Characterizing objects with sika features for multiclass classification," *Applied Soft Computing*, 2015.

[12] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 214–227.

[13] K. E. Van De Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1582–1596, 2010.

[14] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–506.

[15] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 73–80.

[16] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 7, pp. 629–639, 1990.

[17] P. Mukherjee, B. Lall, and A. Shah, "Saliency map based improved segmentation," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1290–1294.

[18] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 4, pp. 594–611, 2006.

[19] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

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