ICIT 2015 The 7th International Conference on Information Technology
doi:10.15849/icit.2015.0015

Distributed 3D Object Recognition System Using Smartphones

Mustafa Ibrahim1, Omar El-gendy2 and Mohamed Farouk3
Center for Documentation of Cultural and Natural Heritage
Bibliotheca Alexandrina
Giza, Egypt
[1]Mostafa_ebrahim87@yahoo.com, [2]el-gendy@mcit.gov.eg, [3]mfarouk@mcit.gov.eg

***Abstract*—Object recognition and scene classification are generally considered one of the most important challenges in computer vision community, where, object recognition is a process of finding and identifying objects in a digital image or video sequence. One of the main problems in recognizing 3D object is extracting stable and consistent features vectors under different conditions, such as camera viewpoint, illumination and cluttered background. In addition, Processing and memory capacity of Smartphones still restrict the computational capacity of object recognition programs. In this paper, we propose a distributed 3D object recognition system to overcome computational capacity problem and improve scalability of objects that will simply be recognizable. The paper also proposes the use of k-Nearest Neighbors classifier with Speeded Up Robust Features algorithm to solve the problem of extracting stable and consistent features vectors. The system is remarkably capable of adapting to different network configurations and the wireless bandwidth, and improving the performance of recognizing multiple 3D objects using Smartphones devices.**

Keywords—Scale Invariant Feature Transform; Speeded Up Robust Features; k-Nearest Neighbors.

I. INTRODUCTION

One of the most significant developments in the last decade is the applications of 3D object recognition. Object recognition is a computer technology related to computer vision and image processing that deals with finding and identifying instances of semantic objects of a certain class in digital images or video sequence. The main factors that affect the accuracy of 3D object recognition systems are the variability in the illumination and the pose of the objects, in addition to time delay. The presence of these factors in recognizing 3D objects can lead to diminishing recognition reliability [1].

Obtaining 2D images from 3D scenes is the reverse process of reconstruction 3D Model from multiple 2D images. Thus, it can handle any 3D object as a sequence of 2D images.

Selecting the most consistent, stable and reliable feature extraction technique and appropriate classifier for classifying number of categories, which contain large number of features is a hard task that will be overcome by using Speeded Up Robust Features (SURF) algorithm in extracting features and K-Nearest Neighbor classifier in classifying objects.

Smartphones devices are very limited in executing high computational capacity programs such as objects recognition and image classifications programs, that's because of the high computational capacity required for this type of applications, in addition to low capacity of memory and processors of smartphones devices up to this day.

Offloading computations from smartphones to remote cloud resources has recently been rediscovered as a technique to enhance the performance of smartphone applications, while reducing the energy usage [2].

The rest of this paper is structured as follows: the next section presents the problem statement. Section 3 talks about some related works. Section 4 and 5 discuss the framework of the proposed recognition system. Section 6 explains the proposed distributed system. The comparative experiments and results are discussed in Section 7. Finally, the paper will be concluded in Section 8.

II. PROBLEM STATEMENT

Extracting stable and consistent features under different condition such as reflections, illumination and camera view point has been considered one of the main challenges in recognizing 3D objects. Besides that, Smartphones still restrict the computational capacity of object recognition programs, while,3D object recognition programs involve complex mathematical calculations and they require powerful memory and processor to cover their computational needs. Therefore, in this paper the proposed distributed 3D object recognition system can handle the computational capacity and scalability problems of smartphone resources. Moreover, using robust

doi:10.15849/icit.2015.0015

and fast algorithm such as SURF can solve the recognition problems with high accuracy.

III. Related Work

The most common way to tackle 3D detection is to represent a 3D object by a collection of independent 2D appearance models [3, 4, 5, 6, 7], one for each viewpoint. Several authors augmented the multi-view representation with weak 3D information by linking the features or parts across views [8, 9, 10, 11, 12]. This allows for a dense representation of the viewing sphere by morphing related near-by views [13], since these methods usually require a significant amount of training data.

Two general approaches have been taken to solve 3D recognition problem: pattern recognition approaches and feature-based geometric approaches. The first approach uses low-level image appearance information to locate an object, while the second constructs a model for the object to be recognized, and matches the model against the photograph.

The groundbreaking work of Schmid and Mohr showed that invariant local feature matching could be extended to general image recognition problems in which a feature was matched against a large database of images. They also used Harris corners to select interest points, rather than matching with a correlation window. The Harris corner detector is very sensitive to changes in image scale, so it does not provide a good basis for matching images of different sizes [14].

David G. Lowe extended the local feature approach to achieve scale invariance using Scale Invariant Feature Transform (SIFT) algorithm. This work also described a new local descriptor that provided more distinctive features while being less sensitive to local image distortions such as 3D viewpoint change. The high dimensionality of Lowe descriptors was a drawback of SIFT algorithm [14].

Speeded Up Robust Features (SURF) algorithm, on the other hand, is designed for much faster scale-space extraction. The detection of extrema is located on the determinant of Hessian approximated by Haar-wavelets. The descriptor is based on the polarity of the intensity changes. Sums of the gradient (oriented with the main orientation of the keypoints) and the absolute of gradient in horizontal and in vertical direction are computed [15].

However, some computation intensive applications cannot be run on smartphones since their computing power and battery life are still limited for such resources. Some of these applications including video encoding/decoding, image recognition, and 3D graphics rendering, could take a significant amount of time due to their computationally intensive nature. Processors on mobile devices are gradually getting faster year by year; however, without aid from special purpose hardware, they may not be fast enough for those computationally intensive applications [16].

Recently, it has been rediscovered that offloading computation using the available communication channels to remote cloud resources can help to reduce the pressure on the energy usage. Furthermore, offloading computation can result in significant speedups of the computation, since remote resources have much more compute power than smartphones [2].

IV. Speed up Rubost Features Algorithm

Feature extraction is one of the most important steps in image pattern recognition tasks. As happens with any pattern recognition algorithm, the performance of recognition algorithm strongly depends on the feature extraction method and the classification systems used to carry out recognition tasks [17].

Scale Invariant Feature Transform (SIFT) is an approach for detecting and extracting local features descriptors that are reasonably invariant to changes in rotation, scaling, lighting conditions and small changes in view point. SIFT features are also very resilient to the effects of "noise" in the image [17]. Generally, the high dimensionality of the descriptor is a drawback of SIFT algorithm at the matching step. For on-line applications relying only on a regular PC, each one of the three steps (detection, description, matching) has to be fast [18].

Speeded Up Robust Features (SURF) is a robust local feature detector, first presented by Herbert Bay et al. in 2006, that can be used in computer vision tasks like object recognition or 3D reconstruction. It is partly inspired by the SIFT descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT [17].

In SURF algorithm, the most valuable property of an interest point “Detector” is its repeatability. The repeatability expresses the reliability of a detector for finding the same physical interest points under different viewing conditions. Next, the neighborhood “Descriptor” of every interest point is represented by a feature vector. This descriptor has to be distinctive and at the same time robust to noise. The dimension of the descriptor has a direct impact on the time this takes, where, less dimensions are desirable for fast interest point matching [18].

In SURF feature vector. In order to bring in information about the polarity of the intensity changes, we also extract the sum of the absolute values of the responses,$|d_x|$and $|d_y|$. Hence, each sub-region has a four-dimensional descriptor vector v for its underlying intensity structure

$$v = (\sum d_x; \sum d_y; \sum |d_x| ; \sum |d_y|). \quad (1)$$

doi:10.15849/icit.2015.0015

Where d_x is the Haar wavelet response in horizontal direction, and d_y is the Haar wavelet response in vertical direction.

Concatenating this for all 4 x 4 sub-regions, this results in a descriptor vector of length 64.Figure 1 show the properties of the descriptor for three distinctively different image-intensity patterns within a sub-region [18].

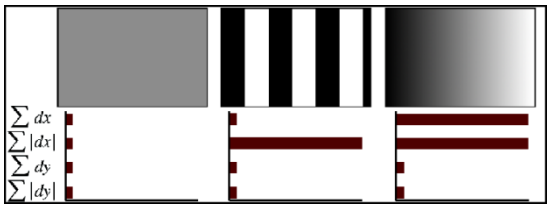


Fig. 1. The descriptor entries of a sub-region represent the nature of the underlying intensity pattern. Left: In case of a homogeneous region, all values are relatively low. Middle: In presence of frequencies in x direction, the value of$\sum|d_x|$ is high, but all others remain low. If the intensity is gradually increasing in x direction, both values$\sum d_x$ and$\sum|d_x|$are high [18].

Table (1) and (2) show that, SIFT has detected more number of features compared to SURF but it is suffered with speed. SIFT is slow and not good at illumination changes, while it is invariant to rotation, and scale changes. SURF is fast and has good performance as much as SIFT.

TABLE 1. COMPARISONS OF RESULTS OF SIFT AND SURF ALGORITHM [19]

Algorithm	Detected Feature Points		Matching feature point	Feature matching Time
	Image1	Image2		
SIFT	892	934	41	1.543 s
SURF	281	245	28	0.546 s

TABLE 2. COMPARISONS OF RESULTS OF SIFT,PCA-SIFT AND SURF ALGORITHM [20]

Algorithm	Time	Scale	Rotation	Blur	Illumination
SIFT	common	best	best	best	common
PCA-SIFT	good	common	good	common	good
SURF	best	good	common	good	best

V. K-NEAREST NEIGHBOR CLASSIFIER

K-Nearest Neighbor (KNN) is one of the most popular algorithms for pattern recognition, which has been proven to be a simple and powerful recognition algorithm. Many researchers have found that the KNN algorithm accomplishes very good performance in their experiments on different data sets [21].

K-Nearest Neighbor (KNN) is a supervised learning algorithm and it is a non-parametric method for classifying objects based on closest training examples in the feature space. In statistics, the term non-parametric covers techniques that do not rely on data belonging to any particular distribution [17].

The KNN classification algorithm predicts the test sample's category according to the K training samples which are the nearest neighbors to the test sample, and then judges it to that category which has the largest category probability [21].
The process of KNN algorithm to classify sample X is as follow [21]:

- Suppose there are *j* training categories $C_1, C_2, \ldots, C_j$ and the sum of the training samples is N after feature reduction, they become m-dimension feature vector.
- Make sample X to be the same feature vector of the form $(X_1, X_2, \ldots, X_m)$, as all training samples.
- Calculate the similarities between all training samples and X. Taking the i^{th} sample d_i $(d_{i1}, d_{i2}, \ldots, d_{im})$ as an example, the similarity $SIM(X, d_i)$ is as follows:

$$SIM(X, d_i) = \frac{\sum_{j=1}^{m} X_j . d_{ij}}{\sqrt{\left(\sum_{j=1}^{m} X_j\right)^2} . \sqrt{\left(\sum_{j=1}^{m} d_{ij}\right)^2}} \quad (2)$$

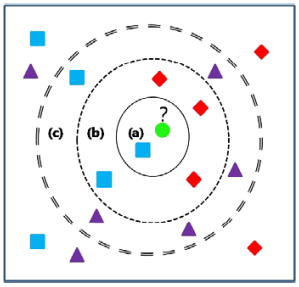
- Choose k samples which are larger from N similarities of $SIM(X, d_i)$, (i=1, 2,…, N), and treat them as a KNN collection of X. Then, calculate the probability of X that belongs to each category respectively with the following formula.

$$P(X, C_j) = \sum_{d} SIM(X, d_i) . y(d_i, C_j) \quad (3)$$

Where $y(d_i, C_j)$ is a category attribute function, which satisfied

$$y(d_i, C_j) = \begin{cases} 1, d_i \in C_j \\ 0, d_i \notin C_j \end{cases}$$

Judge sample X to be the category which has the largest $P(X, C_j)$ as shown in figure 2.



doi:10.15849/icit.2015.0015

Fig. 2. K-NN Classification. At the query point of the circle depending on the k value of 1, 5, or 10, the query point can be a rectangle at (a), a diamond at (b), and a triangle at (c)[22].

The accuracy of the k-NN algorithm can be severely degraded by the presence of noisy and irrelevant features, or if the feature scales are not consistent with their importance. So, extracting consistent and relevant features using SURF algorithm can help k-NN algorithm in classifying 3D objects with high accuracy.

VI. Distributed 3D Object Recognition System

This paper follows another line of research on building distributed 3D object recognition system. This distributed system is a software system in which, components located on networked computers communicate and coordinate their actions by passing messages [23].Therefore, the proposed system tries to employ the power of using distributed system in recognizing objects.

The proposed system for recognizing 3D objects consists of three steps. The first step is a sampling procedure that captures a finite set of candidate 3D locations in order to avoid the high computational cost of considering every potential location. The second step is extracting the most stable and consistent features of each captured image and merging them to create the marker of a captured object. The last step is matching the new image that was captured by a smartphone with stored markers list in a workstation and returns the result back to the smartphone.

Figure 3 illustrates a scenario for capturing multi scene of the same object from different viewpoints. For each 3D object, 18 (2*9) different vantage points have been selected to measure the 3D appearance of this object.

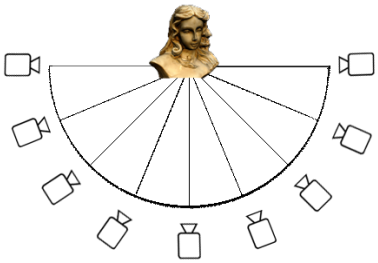


Fig. 3. Capture multi scenes of the same object from different multiple viewpoints

After extracting features of each object using SURF algorithm (local feature vectors for each image) and merging them, the obtained marker of each object has been ready to use in the matching process. Allocate all markers that have been created from extracting and merging features processes on a workstation to be used later in the features matching process.

Figure 4 shows the main architecture of the proposed distributed system. It consists of a workstation, wireless router and many smartphones. Although this is very simple distributed system architecture, it is a very effective system.

Fig. 4. The architecture of distributed 3D objects recognition system using smartphones.

The workstation will be connected to the router using Ethernet cable "Wired network", actually, wireless network can be used to connect the workstation and router instead of wired network, but it is better to connect them using a wired network because the wired network is faster and reliable than the wireless network.

In the proposed distributed system, the workstation can be configured as follows, install one of the editions of Windows Vista or Windows 7 on which IIS 7 (Internet Information Services) and above is supported before you proceed. Also be sure that you have administrative user rights on the computer.

After configuring the workstation, web services technologies can be used as a method of communication among many different electronic devices (smart phones and workstation) over a network.

Web services provide an infrastructure for maintaining a richer and more structured form of interoperability between clients and servers. In particular, web services allow complex applications to be developed by providing services that integrate several other services [23].

In The proposed distributed system, the developed web service is responsible for executing three procedures. The first procedure is loading and caching all markers of all objects that have been allocated on workstation web service (once any smartphone connects to the web service).The second procedure is receiving images that have been captured using smart phones and extracting their features using SURF algorithm. The last procedure is matching the extracted features with the loaded

doi:10.15849/icit.2015.0015

markers feature using KNN algorithm and sending result back to the smart phone as shown in figure 5.

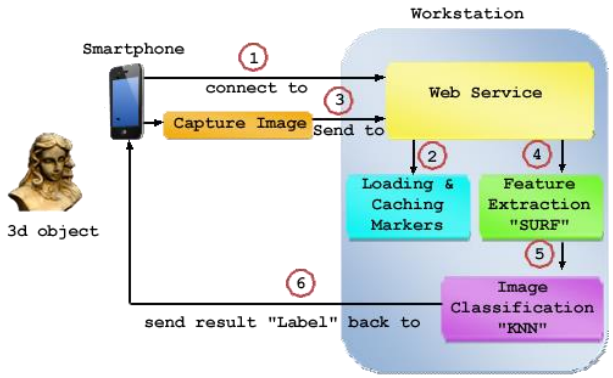


Fig. 5. Data flow diagram of the proposed distributed 3D object recognition system using smartphones.

VII. EXPERIMENTSAND RESULTS

In the experiments of the proposed system, we use some smartphone devices in testing system such as:

- Lenovo Tablet A3000 (OS: Android, v4.1, Processor: Quad-core 1.2 GHz Cortex-A7, RAM: 1 GB and Camera: 5 MP, 2592 x 1936 pixels, autofocus).
- HTC mobile phone Desire 816 (OS: Android, v4.4.2, Processor: Quad-core 1.6 GHz Cortex-A7, RAM: 1.5 GB and Camera: 13 MP, 4160 x 3120 pixels).
- Samsung mobile phone Galaxy Ace 3 (OS: Android, v4.2, Processor: Dual-core 1 GHz Cortex-A9,RAM: 1 GB and Camera: 5 MP, 2592 x 1944 pixels, autofocus).
- Samsung mobile phone Galaxy A5 Duos (OS: Android, v4.4.4, Processor: Quad-core 1.2 GHz Cortex-A53,RAM: 2 GB and Camera: 13 MP, 4128 x 3096 pixels, autofocus).

The configurations of the workstation which is responsible for executing all features extracting and matching procedures is HP Pro 3300 powered by the 2nd generation Intel® Core™ i5 processors running at 2.5 GHz, RAM: 4 GH, OS: Windows 7 professional and IIS 7 server. Both the smartphone devices and the workstation are connected within the same network segment via Wi-Fi 802.11g. D-LINK DSL-2640T router supports wireless speed up to 54 Mbps and interoperability with 802.11b wireless devices on the 2.4GHz frequency band.

The proposed system has been evaluated by real data provided by the Center for Documentation of Cultural and Natural Heritage (CULTNAT)as shown in figure 6.The experiment tested 440 different images of 11 objects whose sizes range from 120 KB to 200 KB and their dimensions are 352 x 288 pixel. Those images are captured from different distances and multiple view points.

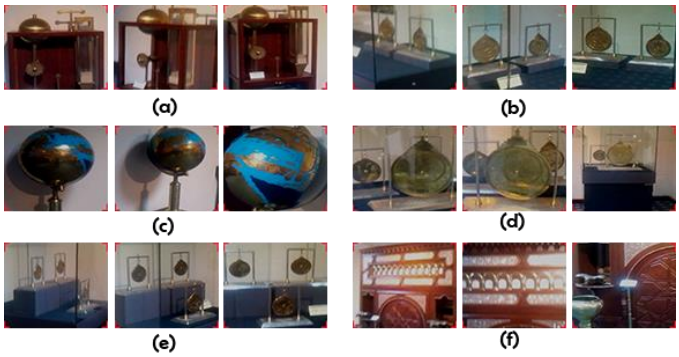
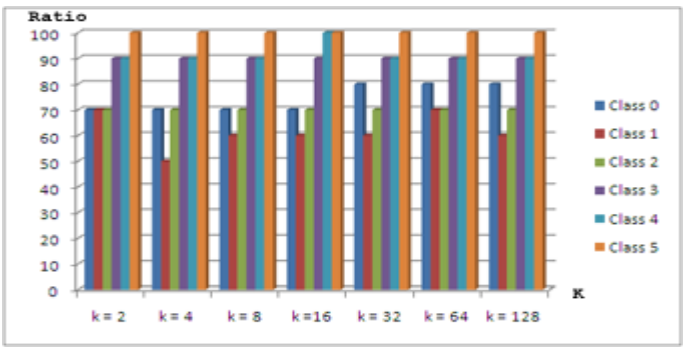


Fig. 6. Examples of different 3D objects have been captured from different distances and multiple viewpoints. In (a), Sand Clock model, In (b), The astrolabe of As-Sahli model, In (c) world map of the geography of the caliph ma'mun , In (d),The newest astrolabe model , In (e) The astrolabe of gafar al muktafi model, and In (f), Water clock Model

The version of OpenCV is 2.2.0 which is used in the workstation side. The main classes have been used are SURF Detector class for extracting features and Flann class for matching features.

During the experiment, no user applications on the smartphone devices other than the proposed system are launched.

Using the proposed system we conducted several experiments and here we represent some results of our experiments. Figure 7 and 8 appear the results of recognizing11 categories with different K values using K-Nearest Neighbor algorithm and 220 images as testing data. In this experiment we compare the results of recognizing different objects when k equals 2, 4, 8, 16, 32, 64 and 128. For readability we rename the used objects as class 0, class 1and so on, in addition to we separate the results in two charts. Figure 9 represents the average of success of this experiment. The best result was obtained when k = 64.



doi:10.15849/icit.2015.0015

Fig. 7. Chart of recognizing the first 6 classes using different k values (Where K=2, 4, 8, 16, 32, 64,128)

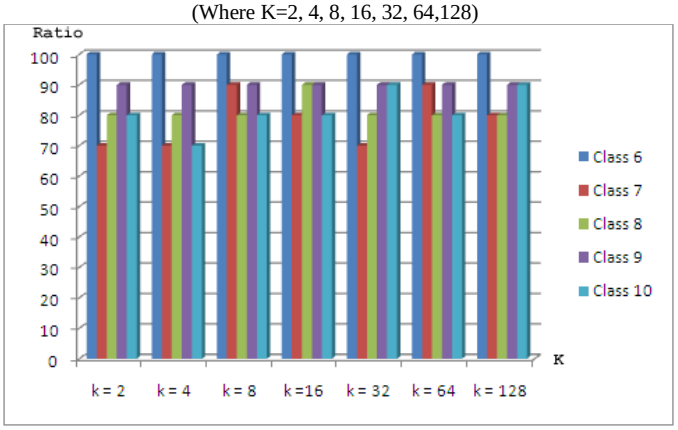


Fig. 8. Chart of recognizing the second 5 classes using different k values (Where K=2, 4, 8, 16, 32, 64,128)

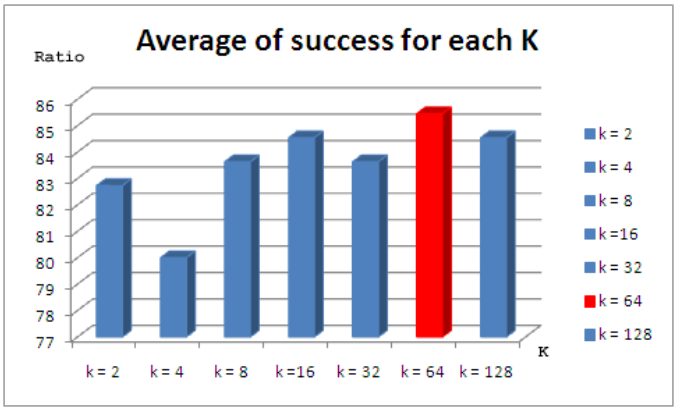


Fig. 9. The average of success of recognizing 11 objects using different k values (Where K=2, 4, 8, 16, 32, 64,128)

The previous experiments lead us to conduct other experiments with different values of k to help us in deciding which k must be used. Figure 10 and 11 represent the results of testing data when k=50, 100, 150, 200, 250 and 300. Figure 12 represents the average of success of this experiment. The best result was obtained when k =150.Increasing the value of k does not improve recognition performance but this consume more time than other experiments, so k=150 is the best one.

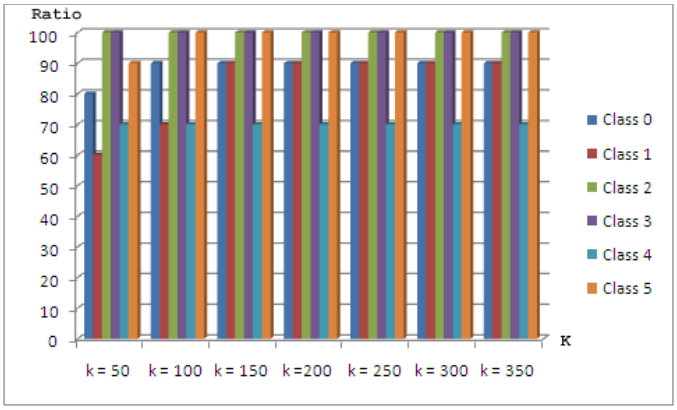


Fig. 10. Chart of recognizing the first 6 classes using different k values (Where K=50, 100, 150, 200, 250, 300,350)

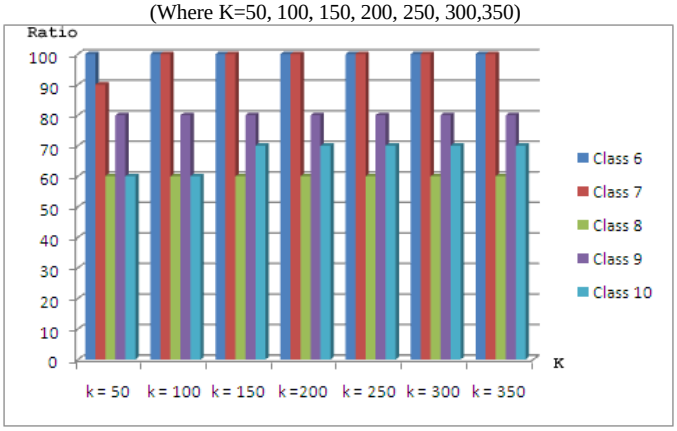


Fig. 11. Chart of recognizing the second 5 classes using different k values (Where K=50, 100, 150, 200, 250, 300,350)

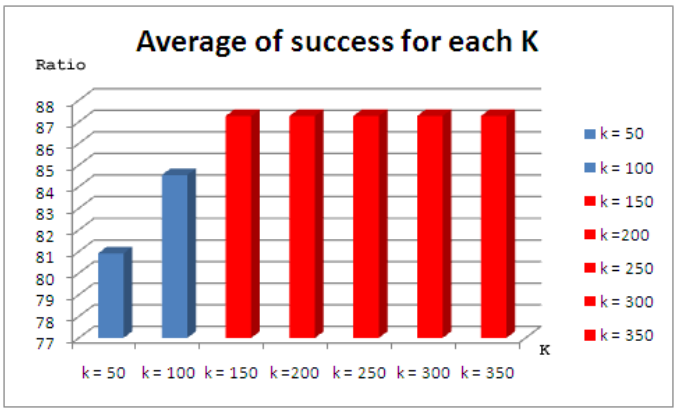


Fig. 12. The average of success of recognizing 11 objects using different k values (Where K=50, 100, 150, 200, 250, 300,350)

VIII. CONCLUSION

In this paper we proposed a distributed 3D object recognition system using smartphones. The system uses selective Speeded Up Robust Features algorithm to extract salient properties of appearance descriptors of local image patches. Furthermore, K-Nearest Neighbor classifier has been used as a simple and powerful recognition algorithm.

doi:10.15849/icit.2015.0015

Smartphones are only responsible for capturing images of 3D objects and sending them to the distributed workstation through a wireless network. All other processes of extracting and matching features are performed by the distributed workstation. Consequently, the proposed distributed system can handle computational capacity problem of smart phones and improve scalability of objects that will accurately be recognizable. 440 images have been used as a simple sample of testing data. Our experiments on a variety of 3D objects demonstrated the effectiveness of the proposed system.

IX. FUTURE WORK

A recent work shows that, extracting feature vectors can be accelerated using FREAK descriptors [24]. It will be interesting to try such methods to make our approach faster. Moreover, the accuracy could be improved by using support-vector-machines (SVM) classifier instead of K-nearest-neighbor classifier, as suggested by another recent work [22].

REFERENCES

[1] Krizaj Janez, Štruc Vitomir; Dobrisek, and Simon. "Robust 3D Face Recognition", journal of Electrical Engineering and Computer Science (Elektrotehniški Vestnik), Ljubljana, Slovenia, 2012.

[2] Roelof Kemp, Nicholas Palmer, Thilo Kielmann and Henri Bal, " a computation offloading framework for smartphones", Conference on Mobile Computing, Applications, and Services, Springer Berlin, Heidelberg, Germany, 2012.

[3] P. F.Felzenszwalb, R. B.Girshick, D.McAllester, and D.Ramanan, "Object detection with discriminatively trained part based models". IEEE Transactions on Pattern Analysis and Machine Intelligence,Chicago, 2010, 32, pp. 1627–1645.

[4] H.Schneiderman and T.Kanade, "A statistical method for 3d object detection applied to faces and Cars", Conference on Computer Vision and Pattern Recognition, San Francisco, 2000, pp. 1746–1759.

[5] A.Torralba, K. P.Murphy, and W. T.Freeman, "Sharing visual features for multi-class and multi-view object detection". IEEE Transactions on Pattern Analysis and Machine Intelligence,Chicago, 2007, 29, pp. 854–869.

[6] C .Gu, and X.Ren, "Discriminative mixture-of-templates for viewpoint classification". European Conference on Computer Vision, Crete, Greece, 2010, pp. 408–421.

[7] B.Pepik, M.Stark, P.Gehler, and B .Schiele, "Teaching 3d geometry to deformable part models", Computer Vision and Pattern Recognition, IEEE Conference, 2012.

[8] A.Kushal, C.Schmid, and J.Ponce, "Flexible object models for category-level 3d object recognition", Conference on Computer Vision and Pattern Recognition, San Francisco, 2007.

[9] A.Thomas, V.Ferrari, B.Leibe, T.Tuytelaars, B.Schiele, and L.V.Gool, "Toward multi-view object class detection", Conference on Computer Vision and Pattern Recognition, San Francisco, 2006.

[10] D.Hoiem, C.Rother, and J.Winn, "3D layoutcrf for multi-view object class recognition and Segmentation", Conference on Computer Vision and Pattern Recognition, San Francisco, 2007.

[11] Sun, M., Su, H., Savarese, S., and Fei-Fei, L, "A multi-view probabilistic model for 3d oblect Classes", Conference on Computer Vision and Pattern Recognition, San Francisco, 2009.

[12] N.Payet, and S.Todorovic, "Probabilistic pose recovery using learned hierarchical object models". International Conference on Computer Vision Barcelona, Spain, 2011.

[13] H.Su, M.Sun, L.Fei-Fei, and S .Savarese, "Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories", International Conference on Computer Vision,Kyoto, Japan, 2009.

[14] David G.Lowe, "Distinctive image features from scale-invariant keypoints", International journal of computer vision, 2004, pp. 91–110.

[15] D.Marimon, A.Bonnin, T.Adamek, and R.Gimeno, "DARTs: Efficient scale-space extraction of DAISY keypoints", Conference on Computer Vision and Pattern Recognition, San Francisco, 2010.

[16] Imai and Shigeru, "Task offloading between smartphones and distributed computational resources", Diss. Rensselaer Polytechnic Institute, 2012.

[17] M.Labib, M.Fakhr, and M.Ali, "Large scale linear coding for image classification", 1st ed, Germany, Deutschlan, 2014.

[18] Herbert Bay, Andreas Ess, TinneTuytelaars, and Luc Van Gool," SURF: Speeded up robust features", Computer vision and image understanding journal, 2008, vol. 110.

[19] P. M. Panchal1, S. R. Panchal2, and S. K Shah, "A Comparison of SIFT and SURF", International Journal of Innovative Research in Computer and Communication Engineering, Tamilnadu, India, 2013, vol. 1, Issue 2.

[20] Luo Juan and OubongGwun,"A Comparison of SIFT, PCA-SIFT and SURF", International Journal of Image Processing, Malaysia, 2009, vol. 3, Issue 4.

[21] N. Suguna, and Dr. K. Thanushkodi, "An Improved k-Nearest Neighbor Classification Using Genetic Algorithm", International Journal of Computer Science Issues, Mahebourg, 2010, Vol. 7, Issue 4, No 2.

[22] Kim, Jinho, Byung-Soo Kim, and Silvio Savarese. "Comparing image classification methods: K-nearest-neighbor and support-vector-machines" . Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, 2012.

[23] George Coulouris, Jean Dollimore, Tim Kindberg, and Gordon Blair, Distributed Systems Concepts and Design, 5th ed. Boston, Addison-Wesley, 2011.

[24] J.Krizaj, V.Struc , S.Dobrisek, D.Marcetić , and Ribarić, "SIFT vs. FREAK: Assessing the usefulness of two keypoint descriptors for 3D face verification", IEEE, Information and Communication Technology, Electronics and Microelectronics, 2014.