HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 33–39 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-06

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

AUTOMATED LABELING PROCESS FOR UNKNOWN IMAGES IN AN OPEN-WORLD SCENARIO

DÁVID PAPP *[1] AND GÁBOR SZŰCS[1]

[1]Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Magyar Tudósok krt. 2., H-1117 Budapest, HUNGARY

Most of the recognition systems presume a controlled, well-defined research setting, where all possible classes that can appear during a test are known a priori. This environment is referred to as the “closed-world” model, while the “open-world” model implies that unknown classes can be incorporated into a recognition algorithm whilst being predicted. Therefore, recognition systems that operate in the real world have to deal with these unknown categories. Our objective was not only to detect data that originate from categories unseen during training, but to identify similarities between pieces of unknown data and then form new classes by automatically labeling them. Our Double Probability Model was extended by an image clustering algorithm, in which Kernel K-means was used. A new procedure, namely the Cluster Classification algorithm for the detection of unknowns and automated labeling, is proposed. These approaches facilitate the transition from open-set recognition to an open-world problem. The Fisher Vector (FV) was used for the mathematical representation of the images and then a Support Vector Machine introduced as a classifier. The measurement of similarity was based on the FV representations. Experiments were conducted on the Caltech101 and Caltech256 datasets of images and the Rand Index was evaluated over the unknown data. The results showed that our proposed Cluster Classification algorithm was able to yield almost the same Rand Index, even though the number of unknown categories increased.

Keywords: open-world problem, cluster classification, image classification, open-set recognition, image clustering

1. INTRODUCTION

In scenarios in the real world, the size of the available dataset continues to increase, therefore, any machine learning algorithm that operates in such an environment has to be capable of preventing growth. This is especially true in the case of image classification, because the growing dataset of tests can pose many difficulties, e.g. it is possible that some of the test images originate from categories that are unseen during training. Recognition systems should detect these unknown images and handle them in an appropriate way. In the rest of the paper the terms "unknown class or category" represent classes or categories that are unseen during training, and “unknown image” denotes images that originate from unknown classes or categories. One way of handling detected unknown images is to measure their similarities and identify new categories. Subsequently, these new categories can be added to the set of known classes. Based on this, three modules are required to solve such problems in the real world, namely a recognition system equipped with an unknown detector, a labeling process and an incremental learning process.

*Correspondence: pappd@tmit.bme.hu

Let us assume that there are K known classes $(C_1, C_2, \dots C_K)$ and U unknown classes in the test set at any given moment, where S_K and S_U denote the sets of known and unknown classes, respectively. Few distinguishable cases depend on the value of U:

1. $U = 0$,
2. $U = 1$,
3. $U > 1$.

Furthermore, a few more cases depend on the amount and type of available information concerning S_U:

(A) Training images,

(B) Set of attributes,

(C) Number of unknown categories (U),

(D) Nothing.

The cases that include 1 or A (e.g. 1A, 2A, 1B, 1C) produce the general multiclass classification because all categories are known a priori and positive-negative samples are available for each category during training. When

$U = 1$, the task is only to identify the unknown images because they originate from the same category, therefore, the similarity measurement is unnecessary. In this paper, the situation when $U > 1$ is considered.

As has been mentioned, 3A represents the traditional multiclass classification. 3B+3C is referred to as transfer learning or zero-shot learning [1], whereas according to the literature the case of 3C+3D is known as open set recognition [2,3] or the open-world problem [4]. The former refers to the detection of images that originate from unknown classes, and the latter includes the detection of unknown images and a labeling process to identify new classes, followed by the incremental learning of these new categories.

Our goal was to tackle the open-world problem as well as develop an algorithm that is able to detect the unknown images and then introduce new classes by automatically labeling the unknown data using unsupervised learning. Previously an algorithm referred to as the Double Probability Model (DPM) [5] was proposed, which is suitable as an unknown detector in an open-set environment.

There are several works that use a variant of Support Vector Machine (SVM) to solve the unknown detection problem, such as the Support Vector Data Description [6], One-class SVM [7, 8], Reject Option SVM (RO-SVM) [9] and the novel Weibull-calibrated SVM (W-SVM) [3]. The latter one was developed to operate under the Compact Abating Probability model, where the probability of class membership decreases (abates) as points move from known data towards unknown space. Scheirer et al. claim that W-SVM outperforms their previous solutions, namely the 1-vs-Set Machine Training algorithm [2] and the Pi-SVM [10]. On the other hand, it was shown that DPM outperforms W-SVM [5], therefore, in this paper the DPM was used for unknown detection. Bendale and Boult defined open world recognition and presented the Nearest Non-Outlier algorithm in [4], which adds object categories incrementally while detecting outliers and managing open space risk. They defined open world recognition in the form of three sequential steps: a multiclass open set recognition function with a novelty detector, a labeling process and an incremental learning algorithm. Although all of these steps should be automated, they presumed labels were obtained by human labeling. The main objective of our work and this paper is to propose an automated labeling process, the so-called Cluster Classification (CC).

In the next section, the DPM and image clustering methods are reviewed, subsequently, a baseline method is suggested for an open-world problem and finally our proposed algorithm, the CC, is presented. The third section contains experimental results and in the last section our conclusion is discussed.

2. Proposed open-world recognition system

2.1 Double Probability Model

The DPM [5] is based on the likelihood of a classifier and can be used with any kind of classifier that provides class membership probabilities for the images. As a result, after training the classifier, it is capable of making predictions with reliability values (scores) for each class, i.e. decision vectors. The range of the scores depends on the type of classifier (sometimes it is from 0 to 1 but it can be over any range. Only one condition is required, namely the larger score for class C_i should represent the higher likelihood of being a member of class C_i. In the training set or a validation set, the instances with corresponding scores are investigated in each class. The ground truth is known for this set, so the positive elements can be selected from each class. In order to calculate the conditional probability that a new instance belongs to class C_i according to its score, the cumulative distribution function (CDF) of positive scores should be determined, therefore, a reverse CDF of negative scores was created:

$$F_{P_i}(x) = p\left(C_i|\text{score} < x\right), \tag{1}$$

$$F_{N_i}(x) = p\left(\neg C_i|\text{score} > x\right), \tag{2}$$

where P_i and N_i denote the positive and negative elements, respectively. Note that the sum of these probabilities is not always equal to 1 (this is not a requirement).

A DPM was constructed based on the CDF and reverse CDF functions. During testing, the focus is on the likelihood of the occurrence of an unknown class compared with any of the known classes. Before the comparison, the probabilities of the known classes should be calculated. Scores (score_i for class C_i) for a new instance are obtained as outputs from the original classifier, and based on them the probability of class C_i occurring can be expressed as described in

$$P_{C_i} = F_{P_i}\left(\text{score}_i\right) \prod_{j=1, j\neq i}^{K} F_{N_j}\left(\text{score}_j\right). \tag{3}$$

An expression for the probability of class C_{K+1} is

$$P_{C_{K+1}} = \prod_{j=1}^{K} F_{N_j}\left(\text{score}_j\right). \tag{4}$$

If the probability of being a member of class C_{K+1} is higher than for any other (known) class, then the new instance will be a member of the unknown class. Otherwise the prediction is based on the original classifier, i.e. the class with the largest score will be selected. The decision with regard to the prediction of test instance j is formalized as

$$d_j = \begin{cases} C_{K+1} & | \quad P_{C_{K+1}} > \max_i \left\{P_{C_i}\right\} \\ \text{argmax}_j \left\{\text{score}_j\right\} & | \quad \text{otherwise} \end{cases} \tag{5}$$

At this point the algorithm is able to make a decision about test data if it originates from an unknown category. Also, should it originate from a known category, then based on the output of the classifier its known category can be determined.

2.2 Unknown image clustering

The image representations were created according to the Bag-of-Words [11, 12] model. Based on their visual content, each image was represented by a single high dimensional vector. In order to create these high-level descriptors, the local attributes of the images were investigated by calculating the low-level Scale Invariant Feature Transform (SIFT) [13] descriptor. Next, the Gaussian Mixture Model (GMM) [14–16] was used to define the visual code words and the Fisher Vectors [17, 18] to encode the low-level descriptors into high-level descriptors based on the visual code words. The Fisher Vectors were the final representations (image descriptors) of the images and were used as the input data for the clustering algorithm. After the final clusters of Fisher Vectors were formed, the image clusters could be produced by substituting the Fisher Vectors for the corresponding images.

The basis of our clustering approach is the well-known K-means clustering algorithm [19] which consists of two important inputs, namely the initial cluster centers and the number of clusters. The K-means clustering algorithm aims to minimize the sum of squared distances from all points to their cluster centers:

$$E = \min \left(\sum_{l=1}^{k} \sum_{x_i \in C_l} \|x_i - z_l\|^2 \right), \tag{6}$$

where k denotes the number of clusters, x_i represents a member of cluster C_l and z_l stands for the center of it.

However, the Fisher Vector consists of 65,791 dimensions, and the basic K-means clustering algorithm performs less efficiently when the clusters are non-linearly separable or the data contains arbitrarily shaped clusters of different densities. Therefore, an upgraded version of the K-means clustering algorithm was applied in the recognition system referred to as Kernel K-means [20–22]. The objective function of Kernel K-means is still to minimize the sum of squared distances, but it uses the kernel trick to transform the data points into infinite feature space $x_i \to \vartheta(x_i)$, as can be seen in

$$E = \min \left(\sum_{l=1}^{k} \sum_{x_i \in C_l} \left\| \vartheta(x_i) - \frac{\sum_{x_j \in C_l} \vartheta(x_j)}{N_l} \right\|^2 \right), \tag{7}$$

where N_l denotes the number of images in cluster C_l. The trick here is that explicit calculations in the feature space are never required, since transformed data points are only present as part of an inner product. Therefore, they can be substituted for their kernel representatives (the Gaussian kernel was implemented here).

In order to reduce the randomness of final clusters, the PlusPlus cluster center initialization algorithm was used before the iterative steps, which was proposed by D. Arthur and S. Vassilvitskii [23]. This approach aims to spread out the initial cluster centers and accelerate their convergence. The first cluster center is randomly selected from the data points, after that each subsequent cluster center is chosen from the data points with a probability proportional to its squared distance from the closest existing cluster center.

In the following sub-sections, the usage of the presented methods is discussed.

2.3 Baseline method

In this section, a baseline method of open world recognition is presented. First, at training time the classifier of the training data is trained with K known classes, then, at testing time classification of the test data ($K+U$ classes) is performed. The DPM is applied to the output of the classifier to detect unknown images U^{DPM}:

$$U^{\mathrm{DPM}} = \bigcup_{j=1}^{N_U} \{I_j | d_j = C_{K+1}\} \tag{8}$$

where I_j represents test instance j, N_U denotes the number of test instances in the test data, d_j stands for the decision of the DPM, and $\bigcup \{\dots\}$ is the operation of union.

Now, let us assume that information concerning U was provided (as in the case 3C), and U was used as the number of clusters. The Kernel K-means PlusPlus cluster center initialization algorithm (KK^{++}) was performed on U^{DPM} with $k = U$ clusters (which is the input parameter for the KK^{++}), and then the appropriate labels were assigned to the unknown images:

$$\begin{gathered} L_j = C_{K+i} | i = \mathrm{out}\left(\mathrm{KK}^{++}\right) \\ j = 1 \dots M, \quad i = 1 \dots U \end{gathered} \tag{9}$$

where M represents the number of unknown images; L_j and C_i denote the label of unknown image U_j^{DPM} and cluster identity, respectively. This concludes the baseline method for automated labeling. At this point the classifier can be retrained based on the previously known and new labels, and then the new test data classified.

2.4 Cluster Classification

In this section our proposed CC approach is presented, which is suitable for unknown detection and automated labeling. This algorithm contains extended training and testing phases. In training time, a classifier of the training data is trained with K known classes, then a pseudo-cluster is also created based on the K known categories. This means that the ground truth class labels are implemented rather than a clustering algorithm (to determine the final clusters), i.e. each category is a cluster. Subsequently, the images are substituted for their Fisher Vector

representations and the cluster centers calculated which will be used in the testing phase.

Let us assume T categories are found in the testing phase, and that $T > K$. The test data is classified into the K known categories and a DPM applied based on the decision vectors to detect the unknown images U^{DPM}. The next step is to form clusters using the Kernel K-means clustering algorithm starting from the K cluster centers that were calculated at training time from the pseudo-cluster. Afterwards, the remaining T-K cluster centers are determined following the PlusPlus initiation protocol. Furthermore, the training and test datasets were used together as the input data. Basically, with these modifications it was possible to guide the clustering algorithm, therefore, create more accurate clusters.

The following step of the testing phase is to classify the clusters $\{C_i\}$ by weighted majority voting of the members of the cluster. The vote is based on the class membership probabilities $(P_{C_i}; i = 1 \dots K+1)$ calculated in Eqs. 3–4. As was seen in Section 1, the definition of problem 3C assumes that the number of unknown categories exceeds 1. Nonetheless, the output of the DPM only yields $K+1$ alternatives instead of T. In spite of this, the classification of clusters that depend on $\{P_{C_i}\}$ can increase the number of alternatives to T as will be seen later. In Section 1, a differentiation was made between known and unknown images, and now this differentiation is broken down even more. The training data contains only known images, because each of them belongs to one of the set of known categories (S_K). From now on, the union of known images of the training data will be denoted by K^{GT} as can be seen in:

$$K^{\mathrm{GT}} = \bigcup_{j=1}^{N_K} \{I_j\} \tag{10}$$

where N_K stands for the number of images in the training data. On the other hand, the test data contains both known and unknown images. Furthermore, based on the output of DPM, the test data can be divided into two different subsets, namely predicted known images (K^{DPM}) and predicted unknown images (U^{DPM}), as can be seen in Eqs. 11 and 8, respectively.

$$K^{\mathrm{DPM}} = \bigcup_{j=1}^{N_U} \{I_j | d_j \neq C_{K+1}\} \tag{11}$$

The weight of the images can be calculated based on the cluster coherence. The coherence of a cluster can be determined by comparing the number of known images to the number of predicted unknown images inside that given cluster. It should be noted that known images inside the clusters either originate from K^{GT} or K^{DPM}, while the predicted unknown images are all part of U^{DPM}. If the number of known images exceeds the number of unknown images it is implied that a cluster exhibits "known coherence" (KC), and "unknown coherence" (UC) vice versa, as described in:

$$C^{\mathrm{coh}} = \begin{cases} \mathrm{KC} \mid \left\|\{K^{\mathrm{GT}} \cup K^{\mathrm{DPM}}\}\right\| \geq \left\|U^{\mathrm{DPM}}\right\| \\ \mathrm{UC} \mid \left\|\{K^{\mathrm{GT}} \cup K^{\mathrm{DPM}}\}\right\| < \left\|U^{\mathrm{DPM}}\right\| \end{cases} \tag{12}$$

where $\|X\|$ represents the number of elements in X, and the superscript coh indicates the coherence of cluster C.

The weights can be calculated as described in Eqs. 13 and 14. Intuitively, if an image is known and located inside cluster UC, then it is "punished" by assigning a lower weight to it; and vice versa, an unknown image is given a lower weight inside cluster KC. Moreover, the larger the difference between the numbers of known and unknown images implies a more severe punishment with regard to the value of weights.

$$w_j^{\mathrm{KC}} = \begin{cases} 1 + \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \notin U^{\mathrm{DPM}} \\ 1 - \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \in U^{\mathrm{DPM}} \end{cases} \tag{13}$$

$$w_j^{\mathrm{UC}} = \begin{cases} 1 + \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \in U^{\mathrm{DPM}} \\ 1 - \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \notin U^{\mathrm{DPM}} \end{cases} \tag{14}$$

Thereafter the final decision vector of cluster C_i can be calculated as:

$$V_i = \frac{1}{N_i} \sum_{j=1}^{N_i} w_j \times d_j \tag{15}$$

where N_i denotes the number of images in cluster C_i, w_j represents the weight and d_j stands for the decision vector ($\{P_{C_i}\}$) of image j. Note that d_j possesses $K+1$ elements ($+1$ from DPM), therefore, vector V_i also possesses $K+1$ elements. Consequently, the element with the maximum value of V_i determines the category of cluster C_i. The classification of cluster C_i is formalized in:

$$D_i = \begin{cases} \text{new class} & \mid \quad V_{K+1} = \max_j\{V_j\} \\ \mathrm{argmax}_i\{V_i\} & \mid \quad otherwise \end{cases} \tag{16}$$

The results of the classification of the clusters can be considered as a labeling proposal, i.e. label each image inside cluster C_i according to D_i. When decision D_i for cluster C_i is that it is part of a known category, then each image inside C_i gets labeled with the same category. On the other hand, when $D_i =$ a new class, a new category is created and each image in C_i gets labeled with the new category. Basically, the CC algorithm follows this labeling proposal.

3. Experimental Results

In order to measure the efficiency of the labeling process, experiments were conducted on the Caltech101 [24] and

Figure 1: Example images from the Caltech101 and Caltech256 datasets. The airplane, butterfly and windmill categories are represented by the left, middle and right columns, respectively.

Caltech256 [25] datasets. Example images from these datasets are shown in Fig. 1.

The former consists of 101 categories and 8,677 images, while the latter is composed of 30,607 images from 256 different classes. To create an open-world environment, 50 known and 50 unknown categories were randomly selected from the Caltech101 dataset, and 100 of both categories from the Caltech256 dataset. These random selections were repeated 5 times in order to calculate the average of the results of each experiment to obtain a more comprehensive overview of the efficiency of the CC algorithm with regard to these datasets. All of the known categories were available from the beginning of the tests, but the unknown categories were added incrementally over 10 steps, and in each step the Rand Index (RI),

$$\mathrm{RI} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \tag{17}$$

was evaluated over the unknown images, where TP, TN, FP, and FN denote the number of true positive, true negative, false positive and false negative decisions, respectively. The RI measures the similarity between the ground truth and predicted labels of the unknown images, in other words, the percentage of correct decisions.

Two methods were assessed and compared, namely the baseline method (DPM+KK) and the CC, which were discussed in Section 2.3 and 2.4, respectively. Both procedures used Fisher Vectors to mathematically represent the images encoded from 128 dimensional SIFT descriptors using a GMM consisting of 256 code words; a SVM equipped with a radial basis function (RBF) kernel was applied as a classifier. The results can be seen in Fig. 2 and Table 1.

The first diagram shows the results obtained from the Caltech101 dataset and the second from the Caltech256 dataset. The DPM with Kernel K-means and the CC are represented by dashed and solid lines, respectively. In both experiments, the CC algorithm yielded a higher RI, although during the first step the difference between the two methods was minimal. It can be seen that the RI of DPM+KK starts to decrease as the number of unknown categories increases, while the CC remains by and large unchanged.

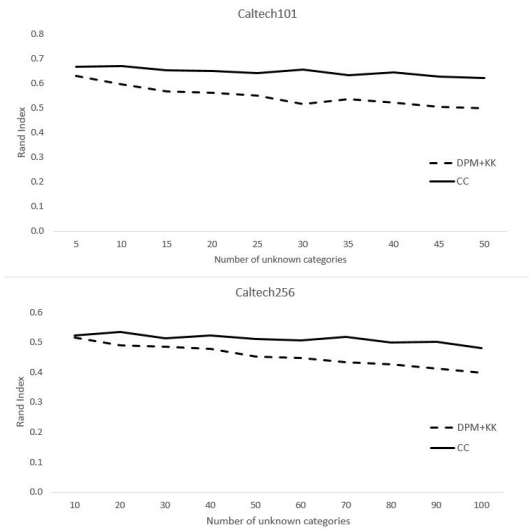

Figure 2: Averaged results of the 5-5 different test datasets that were randomly selected from the Caltech101 and Caltech256 datasets. The RI is plotted against the number of unknown categories. The diagrams compare the labeling performance of the DPM with Kernel K-means (DPM+KK) against the CC.

Table 1: Summary of the results obtained from the test data with the baseline (DPM+KK) and CC methods using the Caltech101 and Caltech256 datasets. The baseline column contains the RI values evaluated which depend on the number of unknown categories (un. cat.), and the CC column presents the improvements that result from CC as a percentage.

Caltech101			Caltech256		
un. cat.	base-line	CC (%)	un. cat.	base-line	CC (%)
5	0.629	6	10	0.514	1
10	0.594	13	20	0.489	9
15	0.567	15	30	0.484	6
20	0.561	16	40	0.478	9
25	0.550	17	50	0.452	13
30	0.514	28	60	0.448	13
35	0.536	18	70	0.433	19
40	0.522	23	80	0.426	17
45	0.505	24	90	0.412	22
50	0.497	25	100	0.397	21

4. Conclusion

In this paper, the problem of open world recognition was reviewed and the possible cases were differentiated based on our prior knowledge and actual information about the test data and, thus, the unknown space. The DPM and Kernel K-means algorithm were also reviewed in brief, followed by the presentation of two approaches, which perform multi-class classification, automatically detect unknown images and propose a labeling for them. The first method is a baseline technique where DPM was sequentially applied followed by Kernel K-means with a PlusPlus cluster center initialization algorithm. However, our proposed CC is a complex method of combining the unknown detector and clustering algorithm that seeks to determine the identity of formed clusters, while refining the decisions made by the classifier and unknown detector. The CC algorithm constructs a specific weight system to reward or punish images which were placed into a category that is presumably unsuitable for their estimated identity. Multiple experiments were conducted on two large datasets (Caltech101 and Caltech256), and the RI evaluated with regard to the unknown images. The results showed that the CC outperformed the baseline method, and was able to maintain almost the same RI, while the number of unknown categories increased.

Acknowledgement

The research was supported by the ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities.

REFERENCES

[1] Lampert, C. H.; Nickisch, H.; Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer, *2009 IEEE Conference on Computer Vision and Pattern Recognition,* 2009, pp. 951–958 ISBN: 978-1-4244-3992-8 DOI: 10.1109/CVPR.2009.5206594

[2] Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; Boult, T. E.: Toward open set recognition, *IEEE T. Pattern Anal.,* 2013 **35**(7), 1757–1772 DOI: 10.1109/TPAMI.2012.256

[3] Scheirer, W. J.; Jain, L. P.; Boult, T. E.: Probability models for open set recognition, *IEEE T. Pattern Anal.,* 2014 **36**(11), 2317–2324 DOI: 10.1109/TPAMI.2014.2321392

[4] Bendale, A.; Boult, T.: Towards open world recognition, *2015 IEEE Conference on Computer Vision and Pattern Recognition,* 2015, pp. 1893–1902 ISBN: 978-1-4673-6964-0 DOI: 10.1109/CVPR.2015.7298799

[5] Papp, D.; Szűcs, G.: Double probability model for open set problem at image classification, *INFORMATICA,* 2018 **29**(2), 353–369 DOI: 10.15388/Informatica.2018.171

[6] Tax, D. M.; Duin, R. P.: Support vector data description, *Machine Learning,* 2004 **54**(1), 45–66 DOI: 10.1023/B:MACH.0000008084.60811.49

[7] Cevikalp, H.; Triggs, B.: Efficient object detection using cascades of nearest convex model classifiers, *2012 IEEE Conference on Computer Vision and Pattern Recognition,* 2012, pp. 3138–3145 ISBN: 978-1-4673-1226-4 DOI: 10.1109/CVPR.2012.6248047

[8] Schölkopf, B.; Platt, J. C.; Shawe-Taylor, J.; Smola, A. J.; Williamson, R. C.: Estimating the support of a high-dimensional distribution, *Neural Comput.*, 2001 **13**(7), 1443–1471 DOI: 10.1162/089976601750264965

[9] Zhang, R.; Metaxas, D. N.: RO-SVM: Support vector machine with reject option for image categorization, In: Chantler, M.; Fisher, B.; Trucco, M.; (Eds.): Proceedings of the British Machine Conference (BMVA Press, UK) 2006, pp. 123.1-123.10. ISBN: 1-901725-32-4 DOI: 10.5244/C.20.123

[10] Jain, L. P.; Scheirer, W. J.; Boult, T. E.: Multi-class open set recognition using probability of inclusion, In: Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T.; (Eds.): Computer Vision – ECCV 2014., ECCV 2014. Lecture Notes in Computer Science, **8691** (Springer, Cham, Switzerland) 2014, pp. 393–409. ISBN: 978-3-319-10577-2 DOI: 10.1007/978-3-319-10578-9_26

[11] Fei-Fei, L.; Fergus, R.; Torralba, A.: Recognizing and learning object categories, *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Short course, 2007 http://people.csail.mit.edu/torralba/shortCourseRLOC/

[12] Lazebnik, S.; Schmid, C.; Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *2006 IEEE Conference on Computer Vision and Pattern Recognition,* 2006 **2**, pp. 2169–2178 ISBN: 0-7695-2597-0 DOI: 10.1109/CVPR.2006.68

[13] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision,* 2004 **60**(2), 91–110 DOI: 10.1023/B:VISI.0000029664.99615.94

[14] Reynolds, D. A.: Gaussian mixture models, In: Li, S. Z.; (Ed.): Encyclopedia of Biometric Recognition, 1st ed., (Springer, Boston, USA) 2009, pp. 659–663 ISBN: 978-0-387-73003-5 DOI: 10.1007/978-1-4899-7488-4_196

[15] Tomasi, C.: Estimating Gaussian mixture densities with EM: A tutorial (Tech. rep., Duke University) 2004 https://www2.cs.duke.edu/courses/spring04/cps196.1/handouts/EM/tomasiEM.pdf

[16] Browne, R. P.; McNicholas, P. D.; Sparling, M. D.: Model-based learning using a mixture of mixtures of Gaussian and uniform distributions, *IEEE T. Pattern Anal.,* 2012 **34**(4), 814–817 DOI: 10.1109/TPAMI.2011.199

[17] Perronnin, F.; Dance, C.: Fisher kernel on visual vocabularies for image categorization, *2007 IEEE Conference on Computer Vision and Pattern*

Recognition, 2007, pp. 1–8 ISBN: 1-4244-1179-3 DOI: 10.1109/CVPR.2007.383266

[18] Perronnin, F.; Sánchez, J.; Mensink, T.: Improving the Fisher kernel for large-scale image classification, In: Daniilidis, K; Maragos, P.; Paragios, N.; (Eds.): Computer Vision – ECCV 2010., ECCV 2010. Lecture Notes in Computer Science, **6314** (Springer, Berlin, Germany) 2010, pp. 143–156 ISBN: 978-3-642-15560-4 DOI: 10.1007/978-3-642-15561-1_11

[19] MacQueen, J.: Some methods for classification and analysis of multivariate observations, In: Le Cam, L. M.; Neyman, J.; (Eds.): Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, **1** (University of California Press, Berkeley, USA) 1967 pp. 281–297

[20] Chitta, R.; Jin, R.; Havens, T. C.; Jain, A. K.: Approximate kernel k-means: Solution to large scale kernel clustering, In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (ACM, New York, USA) 2011, pp. 895–903 ISBN: 978-1-4503-0813-7 DOI: 10.1145/2020408.2020558

[21] Dhillon, I. S.; Guan, Y.; Kulis, B.; Kernel k-means: spectral clustering, and normalized cuts, In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (ACM, New York, USA) 2004, pp. 551–556 ISBN: 1-58113-888-1 DOI: 10.1145/1014052.1014118

[22] Papp, D.; Szűcs, G.: MMKK++ algorithm for clustering heterogeneous images into an unknown number of clusters, *ELCVIA: Electronic Letters on Computer Vision and Image Analysis,* 2017 **16**(3), 30–45 DOI: 10.5565/rev/elcvia.1054

[23] Arthur, D.; Vassilvitskii, S.: k-means++: The advantages of careful seeding, In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, (Society for Industrial and Applied Mathematics, Philadelphia, USA) 2007, pp. 1027–1035 ISBN: 978-0-898716-24-5

[24] Fei-Fei, L.; Fergus, R.; Perona, P.: Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories, *Comput. Vis. Image Und.,* 2007 **106**(1), 59–70. DOI: 10.1016/j.cviu.2005.09.012

[25] Griffin, G.; Holub, A.; Perona, P.: The Caltech 256, *Caltech, Tech. Rep.*, 2012