

Multi-class Assembly Parts Recognition using Composite Feature and Random Forest for Robot Programming by Demonstration

Yabiao Wang, Rong Xiong, Junnan Wang, Jiafan Zhang

Abstract— In robot programming by demonstration (PBD) for assembly tasks, one of the important purposes is to identify multi-class objects during demonstration. In this paper, we propose a composite feature representation method using color histogram, LBP, aspect ratio, circularity and Zernike moment, which is invariant to image translation, rotation and scale. Then Random Forest algorithm is employed to be trained as the classifier, by which the weight parameters of the composite feature are obtained simultaneously. Experimental results on 20 different kinds of objects demonstrate that our approach achieves high recognition accuracy with 99.33%. According to comparisons with other composite features and classification algorithms, the effectiveness with fewer collected samples and the efficiency using less model training time of our approach are verified. Our approach has been successfully applied in two PBD tasks – flashlight assembly and building blocks assembly.

I. INTRODUCTION

The robot programming by demonstration (PBD) enables a robot to perform a task by human demonstration [1]. Thus PBD gives workers an access to teach a robot directly rather than complex machine programming. And there is no need to re-design the programs when it comes to a new assembly task. Fig.1 shows the framework of the PBD system we designed, of which one key problem is to recognize the objects by human demonstration. Therefore, a multi-class objects recognition algorithm is required.

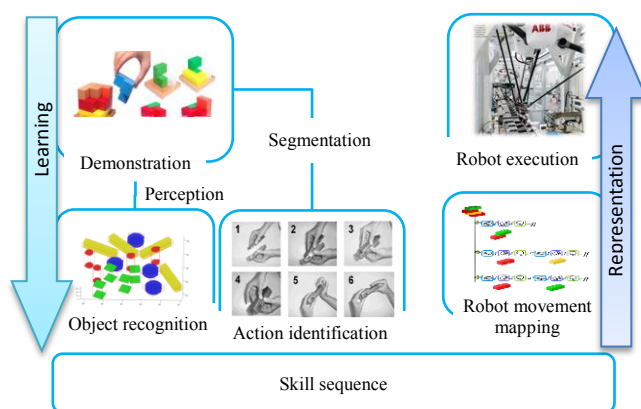


Fig. 1. Overall Framework of Programming by Demonstration

Multi-class object recognition is a complex problem in the field of pattern classification, especially for industrial parts.

Yabiao Wang, Rong Xiong and Junnan Wang are with State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, P.R. China. (Rong Xiong is the corresponding author with email: rxiong@iipc.zju.edu.cn)

Jiafan Zhang is with ABB Corporate Research, P.R. China.

For example, in the application of industrial parts assembly, there are various kinds of texture-less parts. It is a trivial and time-consuming work for workers to train the corresponding vision features when the parts manufactured change. Consequently, a multi-class recognition algorithm will greatly improve the intelligence of industrial robots and make industrial robots easy to use. Although some satisfactory achievements have been made for the recognition of specific object at present by extracting particular features and collecting vast amounts of training samples, e.g., traffic sign detection, face detection and recognition and fingerprint identification, there still exist some challenges for multi-class object recognition. Generally the challenge lies on two aspects. First, the recognition algorithm is not only robust to small differences of an object (e.g., the appearance of a cup is different under different perspectives), but also can identify different kinds of objects with similarity (e.g., tennis rackets and badminton rackets). Second, the complexity of identification model is generally increased with the categories of objects increasing.

Various methods for multi-class object recognition have been introduced. Both of the features representations and the classification methods are widely discussed. Bergevin et al[2] analyses the distribution of projective 2D curves and lines of 3D objects in different views, and they finish the work of recognition for different kinds of 3D objects. A variety of flowers are identified by Gehler et al[3]. Different features are selected by different models and features are given different weights in their research. Meng[4] proposes to analyze the relationship on different features and transform the features into semantic space. Then SVM is used to identify multi-class objects in semantic space. The research of Quelhas et al[5] shows that different kinds of scenarios can be classified by probabilistic latent semantic analysis (PLSA). The scenarios are represented by visual bag-of-visual-words in their paper.

With the research of multi-class object recognition, some new approaches have appeared currently. [6][7][8] adopted the idea to integrate a variety of features, because it is almost impossible to identify different kinds of objects by using a single feature description. Combination of different features is increasingly seen as an effective way by researchers. [9][10][11] proposed to use semantic representation for low-level visual features, where low-level visual features such as color histogram, gradient are extracted at first and then high-level semantic topics are sought out by some algorithms (e.g., *Probabilistic Latent Semantic Analysis* or *Latent Dirichlet Allocation*). With these processes, the performance of a recognition algorithm can be improved generally. [12][13][14][15] employed multi-instance multi-label approach, which is often used in scene classification. As an image often includes multiple scenes and thus it can't be described by a single vector. To complete

classification of different scenes, an image is often divided into several semantic regions and represented by different vectors. [16][17][18] used incremental learning, which fully draws lessons from human's cognitive process. When new samples are added to the trained classification model, the model learn from these new samples on the basis of previous knowledge to enhance its identify ability.

Although the research of multi-class object recognition has received much attention, there are still some problems. For example, feature extraction and model training are time-consuming in some research, and some algorithms need vast number of training samples. Though some approaches based on bag of words model are robust to rotation and noise, they discard the geometrical structure information essentially. As the information used is not sufficient in these methods, it is difficult to further improve the performance of identification.

In this paper, we propose a new composite representation of vision features for multi-class object recognition in industrial applications, in which feature representations including color histogram, local binary pattern, Zernike moment are employed, as well as some special features, such as aspect ratio and circularity are defined in this paper. Then Random Forest is used as the classifier to identify 20 kinds of objects. Our experiments on 20 kinds of objects demonstrate our representation has the advantages on translation, rotation, and scale invariance. Besides, our algorithm works effectively with fewer collected samples and less model training time than other algorithms such as deep learning based approaches, and using Random Forest which can gives the weights of features is much more accuracy than the approaches such as SVM, KNN, decision tree and etc.

The remainder of the paper is organized as follows. Section II introduces the framework of our algorithm. Section III describes our composite feature representation and classifier training method. Section IV presents some experimental results with a comparison with other methods. Conclusions and possible future topics of research are drawn in section V.

II. OVERVIEW OF OUR ALGORITHM

In this section, the algorithm framework and image segmentation in offline phase are described. Image segmentation focuses on accurate object segmentation. As there are varieties of objects in our experiment, it becomes tough to extract every object precisely from background.

A. Algorithm Framework

The flow of our algorithm (mainly on feature extraction and classifier training) is shown in Fig. 2. Objects are segmented from the background and the composite features are extracted. Principle Component Analysis (PCA) is applied to reduce the dimension of the features. Finally Random Forest is trained as the classifier.

The weights of each feature are obtained by learning training samples. As each feature makes different effect on recognition of different objects, thus the obtained weights enhance the generalization ability of our algorithm.

Random Forest and weights of each feature are acquired in offline phase. Segmentation and Feature extraction in online phase are identical with offline phase. Then Random Forest and weights of each feature are used to recognize multi-class objects in online phase.

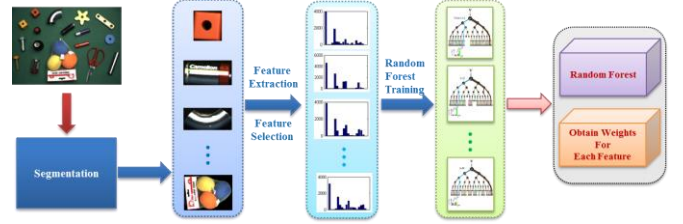


Fig. 2. Overall flow of our algorithm which is mainly on feature extraction and classifier training in offline phase. Random Forest is used to recognize multi-class objects in online working .

B. Image Segmentation

There are multiple objects in the images captured, and image segmentation is necessary to extract the objects from the background and separate the objects from each other.

The algorithm proposed by Otsu and Nobuyuki is utilized to obtain the coarse initial contour areas of every object, and Graph Cut[19] is applied to extract the accurate object contours based on coarse regions. Then composite features can be extracted effectively based on the accurate object contours.

III. COMPOSITE FEATURE REPRESENTATION AND CLASSIFIER TRAINING

A. Feature Extraction

There are variety of vision features in recent years, such as color histogram, SIFT, SURF, Harris Corner, Tamura feature etc. However, the color or the texture can be similar on different objects, so it is difficult to identify multi-class objects using single feature. In addition, some features such as SIFT are not able to be extracted under some perspectives. Therefore, a valid composite feature is proposed in this paper.

In this paper, color histogram, LBP, aspect ratio, circularity and Zernike moment are used to construct the composite feature. Color histogram is a coarse description of an object appearance and LBP is a fine description of an object texture. We also define aspect ratio and circularity to represent the shape. Zernike moment is illustrated to identify complex shapes among objects.

1) Color Histogram

The color histogram is calculated in the RGB space to capture the color distribution of the objects. To avoid the disturbance of the background, the color histogram is extracted within the contour area of the object rather than a rectangular area in our approach. For a single channel, the color histogram is calculated as equation (1)

$$p(r_k) = \frac{n_k}{S_{ROI}} \quad (1)$$

where r_k is kth intensity value and n_k is number of pixels with intensity r_k . S_{ROI} is the total number of pixels within the object contours. Finally, the color histogram of each channel is concatenated and normalized.

2) Local Binary Pattern (LBP)

LBP[20] is used to fine describe an object and it can be seen as a texture descriptor. Let (x_c, y_c) be a pixel coordinate, and there are P pixels in its neighborhood. Texture T is defined as the joint distribution of the gray levels of P pixels

$$T \approx t(s(g_0 - g_c), s(g_1 - g_c), \dots, s(g_{P-1} - g_c)) \quad (2)$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

A unique LBP value that characterizes the structure of this pixel is computed as

$$LBP(x_c, y_c) = \sum_{i=0}^{P-1} s(g_i - g_c) 2^i \quad (4)$$

Fig. 3 illustrates the processes.

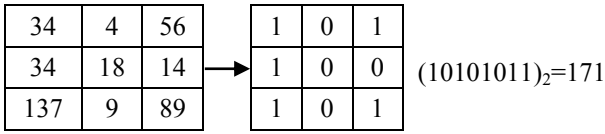


Fig. 3. Schematic diagram of LBP value

The neighborhood region of a pixel can be a square or circle. The intensity of neighbors which do not fall exactly in the center of pixels can be computed by bilinear interpolation.

There are many different ways to achieve rotation invariance of LBP. One way is to define a unique identifier to each rotation invariant pattern as follows

$$LBP_{P,R}^i = \min\{ROR(LBP_{P,R}, i) \mid i = 0, 1, \dots, P-1\} \quad (5)$$

where $ROR(x, i)$ is an operator that performs a circular bit-wise right shift on the P -bit number x i times.

The histogram of region of interest (ROI) based on LBP can be expressed as a texture feature descriptor to identify different texture distribution.

3) Aspect ratio and Circularity

Shape is an important character to identify objects, and aspect ratio and circularity is defined in this paper as an intuitive representation of shape.

2D points can be generated for arbitrary objects after segmentation. Minimum-area-rectangle (MAR) fitted on the given points can be computed by the algorithm proposed in [21], as shown in Fig. 4.

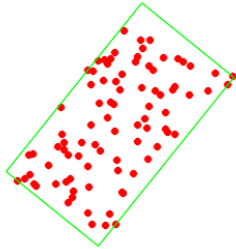


Fig. 4. The red circles are 2D points and the green rectangle is Minimum-area-rectangle(MAR). Generality MAR is a rotated rectangle of enclosing the input 2D point set.

The longer edge divides by the shorter edge is defined as aspect ratio in our approach. However, the aspect ratio of a

square is the same as a circle. For this reason, the circularity is introduced.

Similarity, an ellipse fitted a set of 2D points can be acquired. Circularity is given by

$$C = \frac{L_m}{L_n} \quad (6)$$

Where L_m, L_n are the major axis and minor axis of the fitted ellipse respectively.

It is worth noting that, the coordinates of 2D points for each object in image are transformed to the world coordinate using camera model. Therefore, these two features defined in this paper are robust to scale variance.

4) Zernike Moment[22]

It is necessary to utilize other descriptors to identify complex shapes as aspect ratio and circularity can only distinguish some simple shapes. There are some shape descriptors that is usually used, such as Fourier descriptor, moment description and shape context. Zernike moment is invariant to image scaling, translation and rotation and the orthogonality property can avoid information redundancy. Since the superiority of Zernike moment over regular moments is experimentally verified, Zernike moment is used in this paper.

In order to define Zernike Moments, we need to introduce the Zernike polynomials. Zernike polynomials are a set of complex polynomials which are orthogonal on the unit disk. The Zernike polynomials is given by

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jm\theta} \quad (7)$$

where

n the order of Zernike moment, $n=0, 1, 2, \dots$

m the repetition of Zernike moment, Positive and negative integers subject to $|m| < n$, $n - |m|$ even.

(ρ, θ) Coordinate in polar coordinate system.

$R_{nm}(\rho)$ is radial polynomial, which is defined as

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \cdot \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \cdot \rho^{n-2s} \quad (8)$$

Zernike moment can be represented as

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} I(x, y) V_{nm}^*(x, y) dx dy \quad (9)$$

where the star means complex conjugation.

Note that the pixel points need to be mapped to the range of unit disk to compute Zernike moment. 15 Zernike moments are used in this paper.

B. Feature Selection

The dimension of color histogram is usually high (765 in our case). To simplify models and enhance generalization, feature selection is required. Principle Component Analysis (PCA) is used to reduce the dimension of the Color histogram.

The total feature dimension is 933 in this paper. The dimensions of color histogram, Zernike moment and LBP histogram are 660, 15 and 256 respectively.

C. Random Forest Training

As Random Forest has advantages on high dimension features and small number samples, Random Forest is used as classifier in this paper. Our experiments show that Random Forest is not easy to over-fitting, and training time and used

training samples for random forest are less than other approaches such as deep learning based methods.

Random Forest[23] proposed by Breiman is a kind of classification and regression algorithm. It is a classifier consisting of a collection of tree-structured classifiers $\{h(X, \theta_k), k = 1, \dots\}$. Where $\{\theta_k\}$ are independent distributed random vectors and the final class is predicted by the majority of trees. Firstly a certain number of samples are randomly selected, then each tree grows with following procedure: at each node, choose the best attribute to split on with randomly selected sub-attributes. These two randomly selecting is helpful to reduce over-fitting for Random Forest. Fig. 5 shows the process of Random Forest generation.

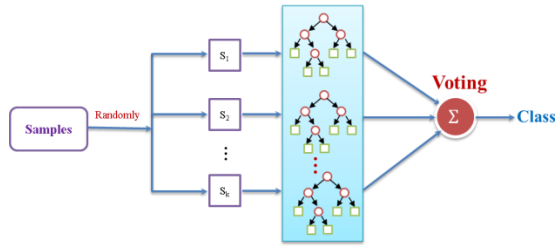







Fig. 5. The Generation process of Random Forest

IV. EXPERIMENTS

The objects used in our experiments are listed in Table I (There are a total of 20 kinds of objects). In order to test the performance of our algorithm, objects with the same color or the same shape are used (e.g. two Li batteries, some blocks), as well as some texture-less objects (e.g. pipe and metal screw) are included. In addition to the assembly parts (e.g. flashlight, building blocks), some objects such as scissors, toys are added in our dataset to test the generality of the algorithm. Note that different classes mean different objects in this paper.

TABLE I. 20 EXPERIMENTAL OBJECTS

Samples	Samples	Samples	Samples
 Flashlight_Head	 Metal_Screw	 Yellow_Cube	 Yellow_Flower_Block
 Flashlight_Tail	 Wooden_Screw	 Red_Cube	 Red_Flower_Block
 Flashlight_Middle	 Battery_Num_1	 Blue_Disk	 Toys
 Red_Li_Battery	 Elbow_Pipe	 Blue_Cuboid	 Three_Stick
 Blue_Li_Battery	 Pipe	 Scissor	 Six_Stick

A. Object Extraction

Otsu and Graph cut algorithm are used to extract every object in an image in this paper. Fig. 6 shows an example.

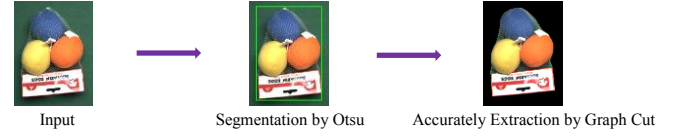


Fig. 6. The flow of object extraction

After foreground objects are extracted, feature extraction and selection are conducted to train Random Forest on offline phase.

B. Experiments on Recognition

Fig. 7 shows the scene of objects needed to recognize, and Fig. 8 gives the recognition results, which is illustrated in the ABB industrial robotic arm simulator. The simulated ABB industrial robot can grasp the objects to replay the assembly procedure learned from human.

It can be seen that the objects and pose of objects are the same as in real scene, which verifies the correctness of our algorithm.



Fig. 7. Building blocks and parts of flashlight in real scene.

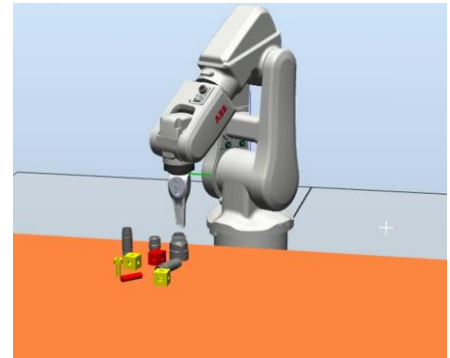


Fig. 8. The results of our approach are illustrated in a simulator of ABB Corporation. Compared with Fig. 7, we can find the results of recognition are all correct.

In addition to the recognition experiments on assembly parts, some results on other objects are presented in Fig. 9 and Fig. 10. The green fonts depending on recognition results are given by our algorithm, which are the names of objects (defined in this paper). The time of the whole recognition process on six objects in Fig. 9 is less than 1 second.



Fig. 9. Some objects in real scene.



Fig. 10. The segmentation and recognition result. The fonts in green are the names of objects, and are automatically labeled by the running result of our approach.

C. Comparative Experiments

1) Recognition accuracy on different composite features

In order to test whether the composite feature used in this paper is reasonable. The comparison experiments are made compared to other two composite features. One composite feature is color histogram with Zernike moment. The other is the composite feature used in our approach with grey level co-occurrence matrices (GLCM). The classification algorithm is Random Forest for these three composite features. 400 training samples are used in this paper (about 20 for every type of object). 120 samples are used for every test phase.

50 random trials are made in this paper. The random refers to 400 samples are randomly selected to train the classifiers, and the rest samples are used to verify the accuracy. Fig. 11 and Table II show the result.

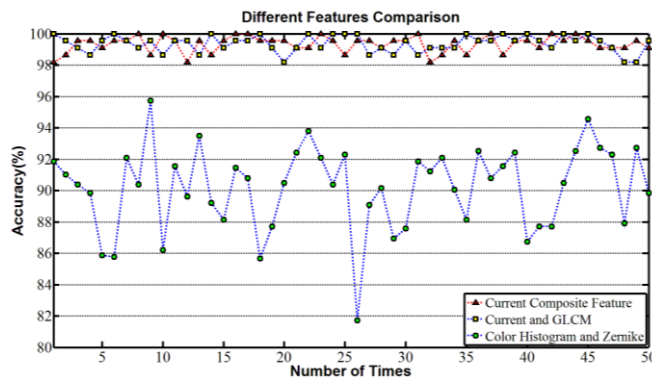


Fig. 11. Comparative accuracy rate of different composite features

Table II. The Average accuracy in 50 random tests

Different Composite Features	Accuracy
Current Composite Feature	99.33%
Current and GLCM	99.32%
Color Histogram and Zernike Moment	90.45%

We can see from table II the accuracy using color histogram and Zernike moment is low. This is because such a combination does not describe the texture details of different objects. In addition, the accuracy of the composite feature used in this paper with GLCM does not improve. So the composite feature of our approach is sufficient to identify different kinds of objects. Generally, the identification algorithm turns sensitive using more features, so we don't introduce other features in our approach.

2) Recognition accuracy on different classification algorithm.

Meanwhile, a comparison is made between Random Forest classifier and other classification algorithms. A large number of samples are not required as the proposed features are invariant to scaling, translation and rotation. The test way is similar to the random method mentioned in 1).

Besides the Random Forest used in for our approach, K-nearest neighbors, Softmax, Decision tree and SVM are used to compare with Random Forest (the training and verify samples are same for these classifiers each time). The final accuracy result is shown as Fig. 12 and Table III.

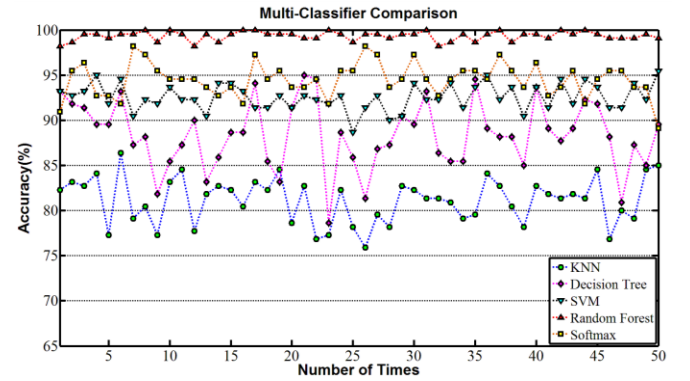


Fig. 12. Comparative accuracy rate of multiple classifiers

Table III. The Average accuracy in 50 random tests

Classification Algorithm	Accuracy
KNN	81.13%
Decision Tree	88.34%
SVM	92.55%
Random Forest	99.34%
Softmax	94.49%

The Random Forest achieves a higher accuracy than other classifiers, as shown in Fig. 12 and Table III. We can see that the Random Forest has advantages on high-dimensional features and small sample sizes, however, some classifiers like Softmax and SVM are prone to overfitting.

In addition, the accuracy of SVM, KNN and Softmax are greatly affected without the normalization of features (the accuracy of SVM and Softmax is less than 70% in this case),

but the accuracy of Random Forest is not sensitive to numerical value ranges of features. As Each dimension of the composite feature represents a node in the trees of random forest, so they are independent from each other when the trees split.

It is worth reminding that the training time of all classification algorithms is less than 3 minutes on our computer (Intel Core i5-3570@3.40GHz, RAM 8GB). As the training process is on offline phase, the time consumption is acceptable for our application.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a composite feature that describes the attribute of color, texture and shape. Random Forest is used to identify 20 different kinds of objects. Experimental results verify that our approach is more effective than other methods.

In the future, experiments on a wider range of objects will be conducted to improve our algorithm. Furthermore, since it is difficult to extract an object from a complex scene (e.g. Bin picking), depth information is expected to be introduced. Therefore, object recognition and pose estimation can be done simultaneously for robot grasping.

REFERENCES

- [1] Rozo, Leonel, Pablo Jiménez, and Carme Torras. "A robot learning from demonstration framework to perform force-based manipulation tasks." *Intelligent Service Robotics* 6.1 (2013): 33-51.
- [2] Bergevin, Robert, and Martin D. Levine. "Generic object recognition: Building and matching coarse descriptions from line drawings." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15.1 (1993): 19-36.
- [3] Gehler, Peter, and Sebastian Nowozin. "On feature combination for multiclass object classification." *Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009.*
- [4] Mength, H., et al. "Generic object recognition by combining distinct features in machine learning." (2005): 90-98.
- [5] Quelhas, Pedro, et al. "Modeling scenes with local descriptors and latent aspects." *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 1. IEEE, 2005.*
- [6] Shi, Meng, et al. "Handwritten numeral recognition using gradient and curvature of gray scale image." *Pattern Recognition* 35.10 (2002): 2051-2059.
- [7] Arivazhagan, S., and R. Ahila Priyadharshini. "Generic Visual Categorization using Composite Gabor and Moment Features." *Optik-International Journal for Light and Electron Optics* (2015).
- [8] Cao, Zhicheng, and Natalia A. Schmid. "Composite multi-lobe descriptor for cross spectral face recognition: matching active IR to visible light images." *SPIE Defense+ Security. International Society for Optics and Photonics, 2015.*
- [9] Sivic, Josef, et al. "Discovering objects and their location in images." *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 1. IEEE, 2005.*
- [10] Mousavian, Arsalan, Jana Kosecka, and Jyh-Ming Lien. "Semantically guided location recognition for outdoors scenes." *Robotics and Automation (ICRA), 2015 IEEE International Conference on. IEEE, 2015.*
- [11] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014.*
- [12] Zha, Zheng-Jun, et al. "Joint multi-label multi-instance learning for image classification." *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008.*
- [13] Zhang, Min-Ling, and Zhi-Hua Zhou. "A review on multi-label learning algorithms." *Knowledge and Data Engineering, IEEE Transactions on* 26.8 (2014): 1819-1837.
- [14] Chen, Zenghai, et al. "Multi-instance multi-label image classification: A neural approach." *Neurocomputing* 99 (2013): 298-306.
- [15] Song, Xiangfa, et al. "Sparse coding and classifier ensemble based multi-instance learning for image categorization." *Signal Processing* 93.1 (2013): 1-11.
- [16] Li, Li-Jia, and Li Fei-Fei. "Optimol: automatic online picture collection via incremental model learning." *International journal of computer vision* 88.2 (2010): 147-168.
- [17] Ross, David A., et al. "Incremental learning for robust visual tracking." *International Journal of Computer Vision* 77.1-3 (2008): 125-141.
- [18] Deng, Weihong, et al. "Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning." *Pattern Recognition* 47.12 (2014): 3738-3749.
- [19] Boykov, Yuri Y., and Marie-Pierre Jolly. "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images." *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. Vol. 1. IEEE, 2001.*
- [20] Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002): 971-987.
- [21] Schneider, Philip, and David H. Eberly. *Geometric tools for computer graphics*. Morgan Kaufmann, 2002.
- [22] Khotanzad, Alireza, and Yaw Hua Hong. "Invariant image recognition by Zernike moments." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12.5 (1990): 489-497.
- [23] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.