American-Eurasian Journal of Scientific Research 12 (3): 112-119, 2017
ISSN 1818-6785

DOI: 10.5829/idosi.aejsr.2017.112.119

A Survey on 3D Object Recognition

B. Bency Evelin and A.S. Radhamani

Department of Computer Science and Engineering, VV College of Engineering,
VV Nagar, Tisaiyanvilai Via, Tuticorin District, Tamil Nadu, India

Abstract: Object recognition is the process of labelling an object from an image. This is a key part in computer vision application. 3 D object recognition can help well in the field of robotic manipulation. There are various approaches determined to make the object recognition robust. The objective of this paper is to present various approaches of recognition and detection. It also illustrates the recognition and detection techniques used by different authors.

Key words: Object recognition · Object detection · Computer vision

INTRODUCTION

Object recognition is a technology in the field of computer vision for finding and identifying objects in an image or video sequence. Humans have the ability to recognize a massive amount of objects in images with little effort. The image of the objects may vary in different viewpoints, in many different sizes and scales or even when they are translated or rotated. The object detection plays an important role in most of computer vision and pattern recognition applications such as video classification, vehicle navigation, autonomous robot routing and surveillance. The object detection performs detection of objects and recognition of patterns in the frame of a video sequence [1]. In order to recognize an object, detection plays a major step. There are various approaches for detection. They are optical flow, background subtraction, frame differencing. The following sections give a brief review of the detection and recognition techniques and their applications.

Reviews: This section provides the detailed review of various object detection techniques, object recognition technique and the techniques used in various systems.

Review of Object Detection: The major aim of detection is to differentiate objects in the image from the background. The various methods of object detection are explained as follows.

Background Subtraction: To recognize an object from the video sequence, the video sequence has to be separated into foreground and background. The foreground and background contains objects of interest and the data that are not important for tracking respectively [2]. Before background subtraction, background modelling has to be done. If the background is constant, then the background is already modelled. Otherwise, background modelling has to be done. The background model is the reference model which is compared with the video sequence to detect the object by computing difference between the two. The background subtraction algorithm involves four steps pre processing, background modelling, foreground detection and data validation. The background modelling can be divided into recursive algorithm and non recursive algorithm [3].

Non Recursive Algorithm: A non recursive technique estimates the background using a sliding window approach. In this approach, a buffer of previous video frames is stored and the background image is estimated based on the temporal variation from each pixel within the buffer [4]. This technique does not depend on the history beyond those frames stored in the buffer. Some commonly used non recursive algorithms are frame differencing, median filter, linear predictive filter and non parametric model.

Recursive Algorithm: The recursive algorithm does not maintain a buffer for background estimation. On the other

Corresponding Author: B. Bency Evelin, Department of Computer Science and Engineering, VV College of Engineering, VV Nagar Tisaiyanvilai Via, Tuticorin District, Tamil Nadu, India.

hand, they recursively update a single background model based on each input frame. The input frame from distant past could have an effect on the current background model. This technique requires less storage than non recursive technique. But if any error exists in the background model, it remains for much longer period of time. Some of the known recursive techniques include approximated median filter, Kalman filter and mixture of Gaussians.

Frame Differencing: The frame differencing algorithm first converts the incoming frame into greyscale. Then, the current frame gets subtracted from background model. Then for each pixel, if the difference between the current frame and the background model is greater than a threshold, the pixel is considered as a part of foreground [1].

Optical Flow: The optical flow is the distribution of noticeable velocities movement of brightness patterns in an image [5]. It is used for computing the motion of pixels of an image sequence. It provides a point to point pixel correspondence. The correspondence problem determines the location of pixels of an image at time t+1 which are at time t [6].

Review of Object Recognition: The recognition of objects can be done even when they are partially blocked from view. This task is still a challenge for computer vision systems. Many approaches to the task have been implemented over multiple decades. There are various methods of object recognition. They are as follows,

Appearance Based Method: The appearance based method use sample images (called templates) of the objects to perform recognition. There are varying conditions which makes the object look different [7]. It includes changes in size, shape, lighting conditions and viewing directions. To achieve reliable recognition, a single template is not enough. However, it is not possible to represent all appearance of an object. Some of the appearance based method includes,

Edge Matching: The edge matching technique uses edge detection techniques to find edges. It detects the images in template and image and then compares the edge images to find the template.

Divide and Conquer Search: This search considers all positions as a set and determines lower bound on score at best position in cell. If the bound is too large, the cell is pruned otherwise, each cell is divided into sub cells and each sub cell is tried recursively. This technique is guaranteed to find all the matches that meet the criterion.

Grey Scale Matching: The edges are mostly robust to illumination changes. However, they throw away a lot of information. The pixel distance must be computed as a function of both pixel position and pixel intensity.

Gradient Matching: The gradient matching is another way to be robust to illumination changes without throwing away as much information. It is to compare image gradients. Matching is performed like matching greyscale images. The alternative of this technique is using correlation.

Large Modelbases: The large modelbases are one approach to efficiently searching the database for a specific image to use eigenvectors of the templates (called eigen faces). Modelbases are a collection of geometric models of the objects that should be recognised.

Feature Based Method: Feature based method is a search used to determine the reasonable matches between object features and image features. The major restriction is that a single position of the object must account for all of the possible matches. There are methods that extract features from the objects to be recognized and the images to be searched. It includes surface patches, corners and linear edges [7].

Interpretation Trees: The interpretation tree is a method to search for feasible matches. The search is through a tree. Each node in the tree represents a set of matches. Root node represents empty set. Each other node is the union of the matches in the parent node and one additional match.

Hypothesize and Test: Hypothesize is a correspondence between a collection of image features and a collection of object features. This is used for generating a hypothesis about the projection from the object coordinate frame to the image frame. This projection hypothesis is used to generate a rendering of the object. This step is usually known as back projection. The rendering to the image is compared and, if the two are sufficiently similar, the hypothesis is accepted. The hypothesis can b obtained by pose consistency, pose clustering, or by using invariants. Pose consistency is also called alignment because the object is being aligned to the image. The geometric constraints of pose consistency are the correspondence between the image features and the model features are not independent. The idea of pose clustering is that the each object leads to many correct sets of correspondences, each of which has (roughly) the same pose. Then voting is done on the pose. It uses an accumulator array to represent pose space for each object. In invariants, geometric properties that are invariant to camera transformations are used for matching the objects and images. This uses, geometric hashing, Scale Invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF) for detecting and describing the invariant points.

Bag of Words: In the Bag of Words method, an image is treated as a document. The words (that is, features) in the images need to be defined. It includes the following steps: feature detection, feature description and codebook generation. The BoW model is used in information retrieval from image for simplifying the representation of the given image.

Review of 3D Object Recognition

Globally Consistent Range Scan Alignment for Environment Mapping: Local frames of data and spatial relationships are maintained by the system is proposed by F. Lu *et al*. [8]. In this system, motion based detection technique such as Gaussian mixture and point to point matching is done. Matching is done through pair wise scanning. Range scan represent partial view of world. When each frame of sensor data is obtained, it is aligned to a previous frame. The systematic method propagates pose correction to all related frames. This approach is to maintain all local frames of data as well as relative spatial relationship between local frames. These spatial relationships are modelled as random variables and are derived from matching pair wise scans or from odometry. Then a procedure based on maximum likelihood criterion is formulated to optimally combine all spatial relations.

The advantage of this system is that, maintaining history of poses, allow all object frame to consistently register in global reference frame and spatial relations are derived and converted to object frame location. The disadvantage is that, continuous scanning is missing in the system and there is a problem in associating measurement with correct robot position.

Learning Hierarchical Object Maps of Non-stationary Environments with Mobile Robots: The algorithm for learning object models of non stationary objects is proposed by D. Anguelov *et al*. [9]. This system uses a similar form of frame differencing technique such as map differencing. For learning shape parameters EM (Expectation Maximization) algorithm is approximated. Bayesian model is used for estimating the total number of objects. The system proposes an algorithm that identifies classes of objects, in addition to learning plain object models. From multiple occurrences of same type of objects, this method provides the ability to learn shape models of individual objects. This learning helps the approach to generalize across different object models in addition to shape models until objects of same type are modelled.

The advantage is that, the algorithm used in the system outperforms previously developed non hierarchical algorithm in terms of predictive power and convergence properties. The disadvantage is that, the objects should not move during robotic mapping, the system does not learn attributes of object other than shape and the system does not provide relations between multiple objects and non rigid object structure.

Efficient RANSAC for Point-Cloud Shape Detection: An automatic algorithm to detect basic shapes in an unorganized point clouds are presented by R. Schnabel *et al*. [10]. In this system, a randomized shape detection algorithm is introduced which uses the differencing technique. RANSAC paradigm extracts shapes by randomly drawing minimal sets from point data and constructing corresponding shape primitive. The system detects plane, sphere cylinders, cones and tori based on random sampling. Due to the low quality of data or due to processing time constraints, this method is well suited when geometric data is automatically acquired. Here, the users avoid doing surface reconstruction method. This type of constraints is suitable for areas where high level model interaction is required, while measuring physical parameters or for interactive, semi automatic segmentation and post processing.

The advantage is that, the system is robust even in presence of high degree of noise and efficient for shape detection. The disadvantage is that, when there is large amount of data, it has greater impact on score evaluation. The system does not find shape proxies for every part of the service.

Object Recognition and Full Pose Registration from a Single Image for Robotic Manipulation: An approach for building metric 3D models of object using local descriptors from several images is presented by A. Collet *et al*. [11]. The detection method used by the system is the optical flow technique which uses Scale Invariant Feature Transform (SIFT) for detecting the invariant points which in turn can be used for recognizing through matching. The system uses a novel combination of Random Sample Consensus (RANSAC) and a Mean Shift Algorithm to register multiple instances of each object and for matching the local descriptors in the object. SIFT features is used to extract local descriptors. The system is separated into an offline object modelling stage and an on-line recognition and registration stage. The modelling stage of the system takes a sequence of images of an object from different viewpoints. The system does not maintain pose information. The segmentation of object in the training images is then done either manually or automatically. Next, SIFT features are extracted for each image and matched across the entire sequence.

The advantage is that, the system improves efficiency in recognizing multiple instance of same object. The disadvantage is that, the system lacks in pose estimation from multiple views and the object modelling stage is offline.

D-Clutter: Building Object Model Library from Unsupervised Segmentation of Cluttered Scenes:

An algorithm with a set of stereo images with one stereo pair per scene is presented by G. Somanath *et al*. [12]. The system uses the optical flow detection method and the recognition us done by feature based method such as hypothesizing through invariance and the invariance is determined by the SIFT detection and description. From the systems perspective, an object can be characterized in essence by its texture and its structure/shape. Images provide texture/colour information while 3D models retain both texture and shape/structure information, hence providing a more complete and concise representation. The system extracts the boundaries of different objects. It uses colour, texture and structure to refine previously determined object boundaries to achieve consistent segmentation.

The advantage is that, segmentation is consistent and the algorithm works significantly better for object discovery and image segmentation techniques. The disadvantage is that, as the object's library size increases, one to one comparison is impractical and searching a larger database is complex.

Unsupervised Learning of 3D Object Models from Partial Views: An algorithm for learning 3D object models from partial observations. Models learned are presented as point clouds by M. Ruhnke *et al*. [13]. The recognition of the system is done through hypothesize and test technique. This system presents an approach for unsupervised learning of object models. The approach used here can be applied to sequence of 3D scans and there is no problem even if the environment varies. In a scene, if there are more than one instances of same object from different perspectives, this approach merges the different individual views. Thus this approach can able to construct a model even from a single scan. Such a model contains all available structural information of an object.

The advantage of the system is that, it deal with partial views and can robustly learn accurate models from complex scene. The disadvantage is that, the system requires more views of object for better results.

Efficient Multi-View Object Recognition and Full Pose Estimation: An approach termed introspective multi view which combines single view averaging and full multi view approach is proposed in this system. Single view algorithm recognizes and register 3D model using local descriptor. Multi step optimization extends it to multiple views. Two standard approaches are popular for converting a single view vision algorithm to multiple views. The first, which is termed single view averaging, executes the single-view algorithm on each of the images and combines the resulting output, often using machine learning techniques. The approach that scales linearly with the number of images and has the ability to combine many single view algorithms at the same time is proposed by A. Collet *et al*. [14]. The second, which is termed full multi-view, combines multiple images by considering a network of cameras as one generalized camera which produces a single large aggregate image. The single-view algorithm is then applied to the generalized image. This approach is optimal that is, there is no loss of information. The detection and recognition approach used by this system is the optical flow and hypothesize and test respectively because the system uses SIFT for detection and description.

The advantage of the system is that, it combines the efficiency of single view algorithms and accuracy of generalized image algorithm. The disadvantage is that, the computational complexity grows exponentially with respect to the number of objects in a scene.

A Large-Scale Hierarchical Multi-View RGB-D Object Dataset: Random Sample Consensus (RANSAC) plane fitting is proposed by K. Lai *et al*. [15]. They also uses Adaptive Gaussian mixture model for background subtraction. Object detection uses background subtraction. Object detection is done through standard sliding window approach. The technique used for recognition is Feature based technique such as SIFT or Bag of Words (BoW) approach. Category and instance recognition is done. Through this technology the capabilities of robotics operation such as object recognition, manipulation, navigation and interaction can be increased. Both the depth based and vision based segmentation each do extremely well at segmenting objects under different conditions, here to create final object segmentation, the two are combined. The segmentation from depth is the starting point. Finally a filter was run on the segmentation mask to remove isolated pixels.

The main advantage of this paper is that, the system performs very good segmentation and combining visual and shape features in the system, improves accuracy. The major disadvantage is that, in the system there is a problem for recognizing small, dark, transparent and reflective objects. During RANSAC plane fitting, small, thin objects may get merged due to noise. Visual features are more useful than shape features for category level and instance level recognition.

Discovering Object Instances from Scenes of Daily Living: An approach to identify and segment objects from scenes that a person encounters in activity of daily living is proposed by H. Kang *et al*. [16]. This system uses optical flow for object detection and Bag of Words (BoW) for object recognition. This paper uses bottom up image segmentation. Here a procedure is developed that iteratively groups and refines segments. Image segmentation is typically noisy. With a single image, it is impossible to tell the segments that belong to a meaningful object. However, with multiple images, the segments that belong to the same object will show stronger correlation than the segments that belong to different objects or backgrounds. For specific instances, the correlations are measured explicitly as geometric consistency (scale, orientation) and appearance consistency (colour, texture and shape). This approach extracts object candidates as groups of mutually consistent segments by processing a noisy "soup" of segments.

The advantage is that, semi transparent texture less ambiguous object can be discovered by this system and this system significantly outperforms baseline system. The disadvantage is that, segment group's purity of the system should be over 80% and the measurement reflects how badly each object is confused with objects in interference image.

Toward Object Discovery and Modeling via 3D Scene Comparison: Recent simultaneous localization and mapping is adapted to align maps in this system. The system regularizes the surface patches to find whole objects in a complex scene. An algorithm that aligns the frame and then reconstructs each scene is proposed by E. Herbst *et al*. [17]. The algorithm aligns the frames of each RGB-D video using SLAM then reconstructs each scene as a dense set of surface elements. The measurement model tells that how likely each surface element is to have moved between two scenes. The detection technique used by the system is optical flow and nearest neighbour search which is a form of interpretation tree as the recognition technique.

The advantage is that, the system is robust to lack of texture and it is robust to noisy data. The disadvantage is that, object localization has to be improved and spatial regularization after combining multiple scenes has to be improved.

An Evaluation of the RGB-D SLAM System: The trajectory of handheld kinects is estimated by F. Endres *et al*. [18] and generates a dense 3D model of the environment. This paper evaluates the accuracy, robustness and processing time for 3 different feature descriptors (SIFT, SURF, ORB). This paper is a comparison paper for comparing various invariant feature detection techniques. SLAM (Simultaneous Localization and Mapping) for RGB-D camera is proposed in this paper. It extracts visual key points. RANSAC is used for location determination. Pose graph is optimized using non linear optimization. First, visual features are extracted from the incoming colour images. Then these features are matched against features from previous images. By evaluating the depth images at the locations of these feature points, a set of point-wise 3D correspondences between any two frames is obtained. Using RANSAC, the relative transformation between the frames is estimated based on these correspondences.

The advantage is that, the probabilistic occupancy estimation provides a means of coping with noise measurement and errors in pose estimation. The disadvantage is that, feature matching for many frames is costly to compute and the system produces high errors in 2 of 9 sequences.

A Robust Vision-Based Sensor Fusion Approach for Real-time Pose Estimation: A new Kalman based sensor fusion approach for pose estimation is proposed by A. Assa *et al*. [19]. The Kalman filter technique follows the background subtraction for object detection. It introduces an improved hybrid filter tuning technique to fuse data from multiple cameras and estimate object pose. The fusion algorithm has an adoption technique which tunes the process and measurement noise of filter.

The richness of outcome data is enhanced by combining multiple sensor data synergistically. This is done by employing sensor data fusion techniques. By fusing the information from multiple cameras, more robust and accurate estimation of the object pose can be obtained, though not all fusion attempts are necessarily successful .Image fusion was proposed to enhance the information in the image.

The advantage is that, the system offers higher accuracy and precision and is robust to many system parameter changes and the weighting scheme is introduced to eliminate less reliable data. The disadvantage is that, the system experiences more hardware overhead, this technique is computationally more expensive and is sensitive to noise measurement.

Table 1: Survey on Various Papers

S.No	Techniques	Description	Advantages	Disadvantages
1	A large-scale hierarchical multi-view RGB-D object dataset	• Uses RANSAC plane fitting. •Background subtraction -Adaptive Gaussian mixture model • Object detection - Sliding window approach.	• Performs very good segmentation. • Problematic for small, dark, transparent and reflective objects.	• Combining visual and shape features improve accuracy.
2	An evaluation of the RGB-D SLAM system	• Estimates the trajectory of handheld kinects. • Generates a dense 3D model of the environment. • RANSAC is used. Pose graph optimized using non linear optimization	• Can estimate errors in pose estimation.	• Feature matching for many frames is costly to compute. • They produce high errors in 2 of 9 sequences.
3	Efficient RANSAC for point-cloud shape detection	• Shape detection - an automatic algorithm. • Extracts shapes and constructs corresponding shape primitive. • Random sampling to detect plane, sphere cylinders, cones and tori	• Robust even in presence of high degree of noise. • Efficient for shape detection.	• When large data, it have greater impact on score evaluation. • Does not find shape proxies for every part of the service.
4	Globally consistent range scan alignment for environment mapping	•It maintains local frames of data and its spatial relationships. • Matching - pair wise scanning. • Sensor data obtained from each frame is aligned to a previous frame. •This systematic method propagates pose correction to all related frames.	• Maintaining history of poses, allows consistency • Spatial relations are derived and converted to object frame location.	• Continuous scanning is missing. • Problem in associating measurement with correct robot position.
5	Object recognition and full pose registration from a single image for robotic manipulation	• Build metrics of 3D object model using local descriptors from several images. • RANSAC and Mean Shift algorithm combined to register multiple instances of each object • SIFT - extract local descriptors.	• Improves the efficiency	• System lacks in pose estimation from multiple views. • Object modelling stage is offline.
6	Efficient multi-view object recognition and full pose estimation	• Single view averaging and full multi view approach is combined as introspective multi view. • Recognition - Single view algorithm. • Multi step optimization extends it to multiple views.	• Efficiency of single view and accuracy of generalized image algorithm is combined.	• The computational complexity grows exponentially.
7	D-clutter: Building object model library from unsupervised segmentation of cluttered scenes	• Extracts the boundaries of different objects. • Uses color, texture and structure to refine previously determined object boundaries.	• Segmentation is consistent. • Works better for object discovery and image segmentation.	• One to one comparison is impractical. • Searching larger database is complex.
8	Discovering object instances from scenes of daily living	• Identify and segment objects from scenes that a person encounters in activity of daily living. • This paper uses bottom up image segmentation. • Here a procedure is developed that iteratively groups and refines segments.	• Semi transparent texture less ambiguous object can be discovered. • Outperforms baseline system	• Segment group's purity should be over 80%. • Measurement reflects how badly each object is confused with objects in interference image.
9	Unsupervised learning of 3D object models from partial views	• 3D object models are learned from partial observations. • Unsupervised learning for object modelling.	• Robustly learn accurate models from complex scene.	• Requires more views of object for better results.
10	Toward object discovery and modeling via 3-D scene comparison	• Recent simultaneous localization and mapping is adapted to align maps. • Regularizes the surface patches to find whole objects from complex scenes. • Aligns the frame and reconstructs each scene.	• It is robust to lack of texture. • Robust to noisy data.	• Object localization has to be improved.
11	Learning hierarchical object maps of non-stationary environments with mobile robots	• EM algorithm - learning shape parameters. • Bayesian model - estimate total number of objects. • Identifies class of object and learns shape models of individual objects.	• Predictive power and convergence properties are better.	• Does not learn attributes of object other than shape. • Does not provide relations between ultiple objects and non rigid object structure.
12	A robust vision-based sensor fusion approach for real-time pose estimation	• Presents a new Kalman based sensor fusion approach for pose estimation. • Hybrid filter tuning technique - fuse data from multiple cameras and estimate object pose. • The fusion algorithm has an adoption technique which tunes the process.	• Offers higher accuracy and precision. • Robust for many system parameter changes. • Less reliable data is eliminated through weighting scheme.	• System experiences more hardware overhead. • Computationally more expensive. • Sensitive to noise measurement.

Applications of Object Recognition

Biometric Recognition: Biometric technology uses human physical or behavioural character to recognize any individual for security and authentication [20]. The identification of individuals can be made possible based on distinguishing features such as shape of hands (hand geometry), finger print, patterns of iris, etc. For biometric analysis, object recognition techniques such as template matching can be used.

Surveillance: The video surveillance system can be used in various fields for recognizing and tracking objects. Object recognition is required so that the suspected person or vehicle for example be tracked.

Industrial Inspection: The object recognition helps to recognize parts of machines. Thus every parts of the machine can be monitored continuously to avoid malfunctioning or damage.

Content-Based Image Retrieval (CBIR): When the retrieval is based on the image content it is referred as CBIR. A supervised learning system, called OntoPic, provides an automated keyword annotation for images and content–based image retrieval.

Robotic: The important issue in recent years in the field of object recognition is the research in autonomous robots. The most popular application is the humanoid robot soccer competition. The robot soccer players rely on their vision systems very heavily when they are in the unpredictable and dynamic environments [20]. The functions such as robot localization, robot tactic, barrier avoiding, etc., can be completed by collecting the environmental information as the terminal data. The vision system helps the robot to accomplish this task. This reduces the effort of computing. Thus the critical objects in the contest field can be recognized by the object features which are obtained easily from recognition technique.

Medical Analysis: The object recognition concept helps to detect tumour from MRI images. This can also be applied to detect diseases like skin cancer, etc.

Optical Character/ Digit/ Document Recognition: Characters in scanned documents can be recognized by recognition techniques.

Human Computer Interaction: Computer can be made able to interact with human in the real time environment by recognizing the human gestures which are stored in the system. The system can be any application on mobile phone, interactive games, etc.

Intelligent Vehicle Systems: The intelligent vehicle systems are very important to detect and recognize traffic signs and to detect and track vehicles [20]. In [21], such a system is developed. The detection phase of such system is developed to scan the scene and to quickly establish Regions of Interest (ROI). This phase use a colour based segmentation method for quick response. The AdaBoost training provides a set of Haar wavelet features which are used to detect the sign candidates within ROIs. Then, the Speeded Up Robust Features (SURF) is applied for the sign recognition. The local invariant features of a candidate sign is found by SURF and these features get matched with the template image features which exist in the data set. The template image which gives the maximum number of matches with the candidate sign is the object to be recognized.

CONCLUSION

Robust object recognition is a key step in application such as robotic manipulation. This paper had covered various object detection and recognition technique. Reviews of various solutions discussed in various papers by different authors were described in this paper. Applications of object recognition in various fields were described. Then a table is build to describe the system proposed by the author along with their advantages and disadvantages.

REFERENCES

1. Sukanya, C.M., Roopa Gokul and Vince Paul, 2016. "A Survey on Object Recognition Methods." International Journal of Science, Engineering and Computer Technology, 6(1): 48.
2. https://en.wikipedia.org/wiki/Background_subtraction.
3. Ragland, K. and P. Tharcis, 2014. “A Survey on Object Detection, Classification and Tracking Methods”. In International Journal of Engineering Research and Technology, 3(11) November-2014. IJERT.

4. Sen-Ching, S. Cheung and Chandrika Kamath, 2004. "Robust techniques for background subtraction in urban traffic video." Electronic Imaging 2004. International Society for Optics and Photonics, 2004.
5. Horn, Berthold K.P. and Brian G. Schunck, 1981. "Determining optical flow." Artificial Intelligence, 17(1-3): 185-203.
6. Fleet, David and Yair Weiss, 2006. "Optical flow estimation." Handbook of mathematical models in computer vision. Springer US, pp: 237-257.
7. https://en.wikipedia.org/wiki/Outline_of _object _recognition.
8. Lu, Feng and Evangelos Milios, 1997. "Globally consistent range scan alignment for environment mapping." Autonomous Robots, 4(4): 333-349.
9. Anguelov, Dragomir, *et al*., 2002. "Learning hierarchical object maps of non-stationary environments with mobile robots." Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc.
10. Schnabel, Ruwen, Roland Wahl and Reinhard Klein, 2007. "Efficient RANSAC for point-cloud shape detection." Computer Graphics Forum, 26(2). Blackwell Publishing Ltd, 2007.
11. Collet, Alvaro, *et al*., 2009. "Object recognition and full pose registration from a single image for robotic manipulation." Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009.
12. Somanath, Gowri, *et al*., 2009. "D-clutter: Building object model library from unsupervised segmentation of cluttered scenes." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
13. Ruhnke, Michael, *et al*., 2009. "Unsupervised learning of 3d object models from partial views." Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009.
14. Collet, Alvaro and Siddhartha S. Srinivasa, 2010. "Efficient multi-view object recognition and full pose estimation." Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, 2010.
15. Lai, Kevin, *et al*., 2011. "A large-scale hierarchical multi-view rgb-d object dataset." Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011.
16. Kang, Hongwen, Martial Hebert and Takeo Kanade, 2011. "Discovering object instances from scenes of daily living." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
17. Herbst, Evan, *et al*., 2011. "Toward object discovery and modeling via 3-d scene comparison." Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011.
18. Endres, Felix, *et al*., 2012. "An evaluation of the RGB-D SLAM system." Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012.
19. Assa, Akbar and Farrokh Janabi-Sharifi, 2014. "A robust vision-based sensor fusion approach for real-time pose estimation." IEEE transactions on Cybernetics, 44(2): 217-227.
20. Khurana, Khushboo and Reetu Awasthi, 2013. "Techniques for object recognition in images and multi-object detection." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(4): 1383-1388.
21. Murphy, Kevin, *et al*., 2006. "Object detection and localization using local and global features." Toward Category-Level Object Recognition. Springer Berlin Heidelberg, pp: 382-400.