Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A BIBLIOGRAPHY OF OBJECT CLASS RECOGNITION AND OBJECT RECOGNITION BASED ON VISUAL ATTENTION

[1]VAHID ALIZADSEH SAHZABI, [2]KHAIRUDDIN BIN OMAR

[1]PhD Candidate, Pattern Recognition Research Group (CAIT), Faculty of Information Science and Technology, UKM, Malaysia

[2]Prof., Pattern Recognition Research Group (CAIT), Faculty of Information Science and Technology, UKM, Bangi, Malaysia

Email: [1]Vahid.alizadeh@ftsm.ukm.my, [2]KO@ftsm.ukm.my

ABSTRACT

Object class recognition has exhibited significant progress in recent years and is now an integral component of many machine vision applications. However, object class recognition using visual attention image segmentation is a novel idea, which has only been developed in the past decade. This paper presents a comprehensive survey on object class recognition and object recognition algorithms, in addition to their applications based on visual attention region selection methods that as recently published. Additionally, increased efforts have been directed to the development of a generic method for categorizing all objects in a domain including examples such as Winn's Method, used to recognize object classes at a glance. The Majority of object class recognition algorithms are highly dependent on shape matching results. The purpose of this review is to provide a comparison among the visual attention (bottom-up and top-down), object recognition (e.g., SIFT, SURF and PCA-SIFT) and object class recognition methods, aimed to researchers identifying the most appropriate method for a particular purpose. This survey is suitable for researchers in the pattern recognition field, providing familiarity with the existing algorithms for object classification from image acquisition steps to final output (i.e., image segmentation, object recognition and object classification). At the end of each part, the challenges, critical analysis table are provided and future directions of every method are suggested for developing new ideas end of this paper. Additionally, this approach allows researchers to find the definition of keywords and to obtain brief knowledge concerning how each method works and what obtained results are for various datasets.

Keywords: *Visual Object Recognition, Object Class Recognition, Object Classification, Object Categorization, Bottom-Up Visual Attention, Bottom-Up Visual Attention, Saliency Visual Attention*

1. INTRODUCTION

Object recognition is one of the major components of the process of learning visual categories and identifying new interesting objects in images. In essence, machine vision applications are based on the capability of object detection, scenes analysis and object classification. The human visual system, identifies objects using their shapes, sizes and colors [1]. Through all algorithms in object recognition and knowledge acquisition in machine learning, the most challenging issue is the confrontation with novel objects that the system has never seen (called unknown objects). On the one hand, there is no previous knowledge about those objects; on the other hand, object segmentation is difficult due to a lack of information concerning size, shape, color, etc. Object recognition methods mostly consist of two main stages, the training stage and the testing stage. In the training stage, the system is trained by labeled objects, while in the latter stage, the trained objects are detected using the gained information. In many real world applications, the problems associated with detecting novel objects are derived from i) the similarity in shape and

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

size, ii) the presence of unlabeled objects, iii) illumination, iv) viewpoints, v) occlusion and a cluttered background.

In the approach, solving the problems of object recognition and learning has made important strides during recent decades. A recent finding of object recognition techniques has shown the capability of accurate models to detect a particular object within various object classes. Today there are a significant number of standard datasets as a result of realistic background images and the categorizing of huge numbers of objects into thousands of categories. While a variety of approaches have been explored, each must make a few common choices. For example, how will the images and models be represented? Using that representation, how is a category learned? Given a new image, how is categorization or detection carried out?

The experiment of [2] survey the object recognition process utilized by the human brain cortex. They considered three exposures of natural images that differ in their candidate components during object recognition. They performed four types of experiments, including i) naming objects at different levels of specificity, ii) comparing of detection, categorization and identification performance with a two-alternative forced-choice design, iii) detection performance based on object category information, and iv)comparing performance in two tasks on a trial-by-trial basis for exposures. Based on their outcomes, the precise identification of an object can prove the object's presence in an image without requiring recognition of the object category.

[3]studied what people perceive upon first viewing a scene. The experiments were implemented with 22 native English speakers using 90 grayscale images to asking the participant to describe "what do you see in each image" within a short amount of time (from 27 to 500ms). Then, another group scored to each image description obtained from the participant. Individual scores were assigned to more than a hundred different attributes. According to the obtained results, the feature level of each description such as the shape or shade reports precedes the description at the semantic level; in contrast, objects are recognized during semantic level. This idea suggests the possibility of bias toward either scene recognition or object recognition.

In essence, vision applications are based on the capability of object detection, scenes analysis and classification. Object recognition and vision systems are utilized in many research areas, such as artificial intelligence, machine learning, information retrieval and data mining, e.g., video data mining, object detection in robots, texture recognition. A precise definition of an ''object,'' without taking into account the purpose and context, is ultimately impossible. However, it remains clear that it is possible to capture the appearance of the shapes of matter to which people often assign names [1].

As Figure. 1 illustrates, there are 3 main stages and 7 steps for all object class recognition applications, and the image acquisition can be performed by a camera or fixed image dataset. Stage1 is the image segmentation process. This approach uses variety of algorithms for which saliency based bottom-up visual attention is the most reliable method for region selection. This is because the top-down approach works based on prior knowledge; However, in this survey, only unsupervised learning methods have been discussed. Additionally, the bottom-up approach considers the salience regions, which is more similar to the process carried out in human behavior.

Stage 2 includes feature selection and object recognition, using a common object recognition method such as SIFT, PCA-SIFT, SURF, etc. Of all common recognition methods, SURF has the better accuracy and the best processing time. [4] compared three common object recognition methods (i.e., SIFT, PCA-SIFT and SURF) based on the aspects of processing time, scale, rotation, blurriness, illumination and affine transformation. In this way, they applied KNN (K nearest neighborhood) and RANSAC (random sample consensus) to three object recognition methods; Accordingly, KNN was used to find matches, while RANSAC was used to reject inconsistent matches. Based on the experiments, SURF was found to be the fastest, i.e., 100 times faster than the other methods, but it did not recognize the objects as well as SIFT or PCA-SIFT. Additionally, the comparison concluded that choosing the most appropriate method is highly dependent on the application and target of use.

Stage 3 is the final output of the process used to recognize object classes from extracted features, shape matching and other features that are employed in accordance with the method's presenters. In this way, the goal of this study is to

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

complete a survey of the proposed method and algorithm for the object class recognition process according to Stages 1-3 (i.e. image segmentation, feature extraction and object recognition, and object class recognition). The definition of "object classification" as a basic problem in the field of computer vision could be provided to avoid the confusion related to "object recognition". Shape-based object classification in range images aims to label the objects captured in range images based on the common patterns shared by the other objects in the same class. Its applications include autonomous robotic navigation and manipulation, as well as urban scene understanding. Shape has an important role in object class recognition, therefore the methods that work based on shape similarities have better performance than others. Some methods have used color or size and visual dictionaries combined with shape matching. This can increase the accuracy but is a type of supervised learning that requires prior knowledge. One of the most popular shape matching methods is using a shape skeleton for matching the shapes; [3][2][5]proposed a shape matcher based on a canonical skeleton that first simplifies the shape skeleton representation and then tries to obtain the most optimized shape to the original shape with the minimum error. For this purpose, two main steps must be considered, including the elimination of unstable external branches and the elimination of unstable internal branches. The results for both pose estimation and object recognition improvements of 12% and 4% respectively which as achieved by [5].

2. VISUAL ATTENTION BASED IMAGE SEGMENTATION

Of the various methods used to select regions of interest in images, visual attention is one of the most human-like in terms of behavior. Visual attention is motivated by the human visual system to pay attention to objects in images, in a manner that is highly similar to the zooming system used in cameras. However, a single image taken in an unconstrained environment is not sufficient to allow a computer algorithm, or even a human being, to decide where on object starts and another object ends. Figure 2 illustrates a recent publication from the WEB OF SCIENCE website.

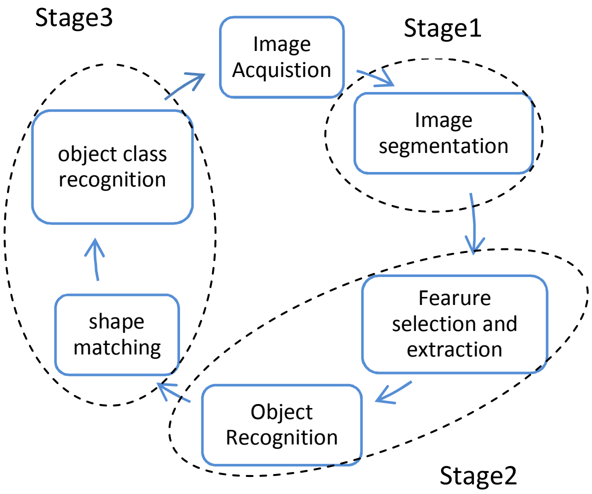

Figure 1: Object Class Recognition Process

From a psychological perspective, there are two types of visual attention systems, including top-down and bottom up visual attention. Bottom-up attention alerts us to salient items in the environment, but top-down attention modulates bottom-up signals when the desire to look for something specific is present [6]. That is, attention is the process of selecting and gating visual information based on either saliency in the image itself (bottom-up), or prior knowledge about the scenes, objects and their interrelations (top-down) [7, 8]. Attention influences the processing of visual information even in the earliest steps of processing in the primate visual cortex. The integration of bottom-up and top-down attention is performed in a saliency map which is related to a topographic representation of behavioral relevance and relative stimulus strength across a visual space. This map appears to be distributed across areas of the visual cortex, and is closely linked to the oculomotor system that controls eye movements and orients the gaze to locations in the visual scene that are characterized by a high salience [9]. The visual attention presenters have been typically used an object recognition algorithm, which is described in detail in the next part of this paper.

[1] proposed a saliency based bottom-up visual attention method for application in cluttered scenes. The method is bottom-up selective attention and uses grouping based on segmentation, which is a tow-dependent method for objects detection in images. Grouping and segmentation results are given inhomogeneous regions while bottom-up selective attention uses

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

saliency-based attention which relies on image contrast. The Selective visual attention algorithm assumes that, bottom-up attention will completely select the region surrounding particular object. Walther's method works based on three major components, colors (i.e. red, green, blue and yellow), orientation in four degrees (i.e., 0°, 45°, 90°and 135°) and pixel intensity. Next a competition is held among the candidate pixels based on feature combination values, finally yielding a take-all winner that expand (Fig. 3.a and Fig. 3.b).

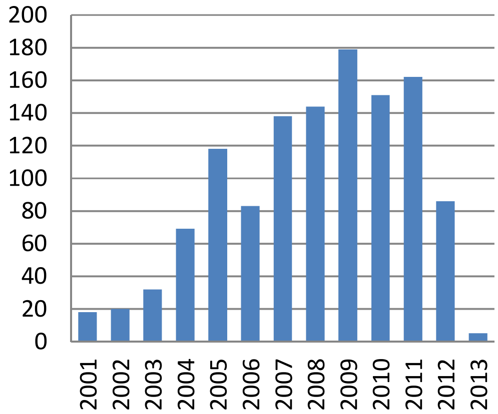

Figure 2: The Number of publications in the field of visual attention

{Milanese, 1994 #7} presented a combinational bottom-up and top-down method for extracting the region of interest in object recognition applications in cluttered scenes. [10] used a familiarity based combinational top-down and bottom-up visual attention method with the SIFT algorithm for object detection. Familiarity helps with attention toward known objects improving the performance in comparison with bottom-up saliency. In contrast, the method by [11] is able to detect regions of interest in images based on the concepts central to the human visual system. A top-down guided visual search module in the system identifies the most discriminate feature from the previously learned target object and then uses that feature to recognize the object. This enables significantly faster classification and is illustrated by identifying signboards in a road scene environment. The paper proposes an approach to extend the usage of features extracted from the attention model to the detection model thereby reducing the computational overhead default that exists in combining two different systems. This model decreases the computational complexity and increases the quality of the object search process

Figure. 3.a): The saliency toolbox output

Figure. 3. b): The saliency toolbox output whole object view

[12, 13] presented a primate visual system-based model for saliency-based visual attention, conducted in a focal manner to break down the problem of scene analysis. The approach develops understanding by computing conspicuous locations. The authors also reported a survey on feature combination map problems based on unrelated dynamic ranges and various visual modalities in the attentional saliency map. Ten years later [14] proposed a multiple object-based model according Itti's model, using visual attention for the recognition of multiple objects using a recognition system that observes the whole image. However, the role of the attentional module is only 20%. Itti and Walther used a saliency map to detect both objects and scenes, whereas Sun, Frintrop and others used alternative approaches. Hecht and Vecera [15]designed a mthod for

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

complex object recognition. The method was motivated by the Watson method and tried to cover the failure of Watson's method with respect to surface uniformity. Additionally, the proposed method illusrates that attentional selection only considers nonuniform objects when the surface changes. [16] framework is a video based attentional technique for detecting both multiple interesting objects and actions. The approach constructed both a temporal and spatial saliency map, which in large motion contrast temporal attention is over the spatial model and vice versa. [17] presented a region detection method according to the top-down attentional region of interest. Feature extraction in the method depends on the attention model and is concentrated on road scene environment. Whalter proposed a method for attending to proto-objects, capable of detecting individual objects in isolation within natural and complex scenes. [18] concentrated on salient color object recognition in natural scenes. Their human-like system approach extracts and analyzes color to find the most salient region.

[19] used the concept of selective tuning to propose a model for solving the problem of selection, routing the content within the visual attention-based model. In contrast, [20] discussed the feasibility, relationship and performance of tow subjective visual attention (i.e. a region of interest and visual fixation patterns). In their findings visual fixation patterns is more effective than the region of interest model. [21] presented an extraction method for the visual attentional region of interests'. They employed unsupervised techniques to develop content-based image retrieval solutions for grouping the most salient regions, i.e., regions that are perceptually similar.

[22] made the connection between concepts and ideas from different research areas such as -psychology, neuroscince and computer science-. Deco and [23]built a neurodynamical system according to the attentional control of the spatial resolution, used to analyze objects iteratively for object recognition. Their method then used a computational neuroscience approach to enhance the spatial resolution i.e., a 'what' and a 'where' stream. [24]elucidated the effects of stimulus-driven factors on the allocation of attention using stimulus salience by determining both the bottom-up and top-down attention. Additionally they identified the eye movement while observing complex scenes. [25]proposed a biological framework combined with object-based attention for selecting objects from coarse to fine locations according to the space-time content.

[26] proposed a method for recognizing objects in spatial 3D data. Their method was tested on a 3D laser scanner using a visual attentional system to detect the region of interest. [27]proposed a method based on the Reinforcement learning of visual classes to learn physical actions in interactive scenes using top-down and task-based visual attention.

Table 1: Critical analysis of visual attention

Year	Title of Method	Advantage	Incompetence
1998	The Saliency Method By Itti	i)Uses a saliency map to simulate bottom-up ii) Is suitable for detection of signs iii)Functions in real time iv)Has an accuracy of 70.4%	i)When noise increases the number of false detections is also ii) Fails to detect the targets salient that are unimplemented feature types
2005	The Saliency Toolbox	i) Is more like human visual attention ii) Has an accuracy of 95% for ROI iii) Is more accurate for object recognition iv)Is similar to Itti's method v)Uses bottom-up visual attention	i)Must train images several times to find all objects.
2010	The Color Saliency Model	i)Has an accuracy of 85.2% ii)Is faster than others approaches	i)Only works on color images ii) Performs poorly in colorful images
2011	The method by KP	i)Tries to use the most discriminative features ii)Has an accuracy of 98.2% with 6 features ii) Reduce the search time complexity	i)Relies on both the number of features and the size of an object ii)Is not similar to human visual attention

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. OBJECT RECOGNITION

Generally, there are two types of object recognition, the recognition of specific objects and the recognition of generic objects. In the case of specific object recognition the vision system detects specific and well-known objects such as the Egyptian Pyramids, Petronas Twin Towers. In contrast, in generic object recognition the recognition is considered based on the various appearances of an object that belong from different types of an object within a similar class. Specific object recognition relies on the matching and geometric verification paradigm. However, generic object classification employs several examples, which are different in appearance and shape; consequently the model can predict objects in new images by training images from a given category into learning a model which can predict objects in new images. As mentioned in the introduction, object class recognition is related to generic object recognition. Hence in this study, the object recognition methods are mostly reviewed which respect to generic object recognition in addition to the former approach, although only partially. Figure. 4 shows a number of object recognition publications from 2001, extracted from the WEB OF SCIENCE database.

[28] presented a connectionist model for detecting and learning multiple objects in images. Their model uses a set of confidence features, where the objects are recognized when the input feature overlaps with the trained features in a supervised mode. The model has the capability of detecting multiple objects, even when those objects occlude each other. [29] presented a learning method is based on use of multiple perspective of multiple objects in images. The proposed method is a factorial learning problem due to the number of objects varied based on the position differences of the objects in the images. The algorithm uses a mixture of Gaussian technique for background subtraction to extract the multiple views of the object positions; additionally to solve the factorial learning problem the model uses Greedy algorithm. [30] proposed a recognition and learning method for detecting objects regardless of differences in position, size and location. For this purpose, they carried out some pre-processing on images with classifiers that were employed to scan the whole image and search for objects. [31] proposed a method based on texture, layout, image context understanding for learning and recognizing multiple object classes. The method uses textons as the texture and layout filters in addition to a novel feature based on contrast; additionaly, share boosting was used for classification and feature selection due to its efficient classifiers. The results demonstrated the effectiveness of the method for highly textured objects (trees, grass), highly structured objects (cars, faces) and articulated objects (body, cows). [32] developed a Hough Transform based framework for object recognition. The framework solved the multiple extrema identification problem in Hough images, leading to the detection of multiple objects and dismissing the necessity of maximum suppression heuristics. However, the approach was weak when the interested objects placed close together or were occluded,although their method exhibits sufficent accuracy in straight line detection problems. [33] proposed a new supervised based framework that used multiple instance learning (MIL) for object recognition. The method is suitable for object/background discrimination in unseen images, and it uses MIL for multi-class classification.

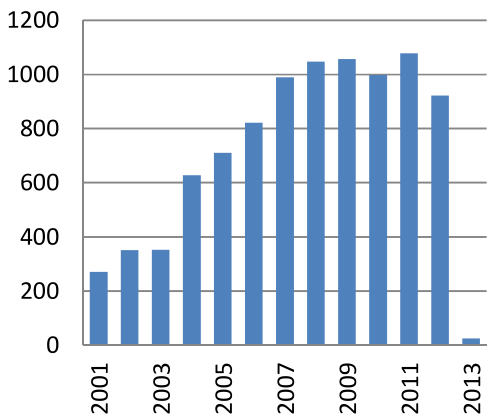

Figure 4: The Number Of Publications On Visual Attention By Year

[34] carried out a survey on generative and discriminative object recognition methods. Additionally they proposed a combinational generative and discriminative model for object recognition and localization based on local invariant features. The results obtained have shown that the generative object recognition model is appropriate for the classification of objects and has significant ability to localize the objects within an image. In contrast, discriminative model has the capability to perform rapid decision-making and focuses on highly informative features.

[35] introduced and developed the SIFT based on a set of local features. The features are invariant to image scaling, translation, and rotation and are

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

partially invariant to illumination changes and either affine or 3D projection. The algorithm uses some key techniques, which cover any local geometric deformation problem by using blurred image gradients in multiple orientation planes and at multiple scales. The use of low-residual-least-square analysis provides the final match between objects keys. Lowe found some distinctive image features from the SIFT output. The new method presented a robust object recognition algorithm from extracting distinctive invariant features to finding the maximum reliable matching keypoints between objects or scenes. The method includes four steps, given as: i) scale space extermum detection ii) keypoint localization iii) orientation assignment, and iv) keypoint description. The first stage uses the difference-of-Gaussian function to identify potential interest points. In the second step, the keypoints are extracted by measuring the location and scale, based on the keypoints stability. Then, one or more orientations are assigned to each keypoint based on the local image gradient direction. In the last step, the local image gradient is measured around each keypoint at the selected scale.

[36] proposed a method called PCA-SIFT shortly after SIFT was developed by Lowe. They used SIFT features to extract local image descriptors; however, instead of using the SIFT smoothed weight histogram, their algorithm applies principal component analysis (PCA) to the normalization of gradient patches. Their hybrid PCA and SIFT method improved the performance of the recognition algorithm up to 25%, and was found to be more flexible in cases of image deformation. The SURF algorithm is more robust and faster than SIFT, which is we elected to use SURF instead of SIFT in present study. By using an integral image for image convolution, the algorithm increased the speed by reducing the number of orientations in the image box. To match the keypoints, SURF calculates the minimum Euclidean distance between neighboring keypoints to find the nearest neighborhood keypoints. Figure. 6.a and Figure. 6.b illustrate successful and unsuccessful outputs of the SURF algorithm. In summary, the SURF method uses the Hessian matrix for the use of Gaussian or DOG. Additionally, to reduce the size of the image, it uses the integral image instead of the original image. Then, it finds some descriptors from point of interest in the images as well as feature vectors, and SURF performs some orientation assignments to calculate the orientation of each keypoint, making the scale invariant for each object transformation. Finally, all of the keypoints are matched between the source image and the target image.

[37] developed a new framework for detecting objects, found to be particularly useful for face detection. The method first extracts the image features using image an integral and then selects the features using the AdaBoost algorithm. The proposed framework considers several classifiers for reducing the time complexity and improving the detection accuracy. [38] presented a method for 3D scenes that recognizes the geometrically invariant parts of objects in the images. Generally, the model detects the parts of objects that are not mobile but are stable during movement of the whole object. [39] presented a moving object recognition algorithm for video sequences. The algorithm uses two cameras to find the distances between objects, i.e., the differences between frames illustrates the movement of objects against the background. The output of their algorithm is provided by a clustering algorithm which gives the position, number and size of the moving points.

[40] used Boosting technique to detect and learn generic objects. Their supervised learning method is based on weak initial hypotheses in cooperation with a boost leading to a final hypothesis for decision-making. According to the algorithm, the weak hypotheses are reproduced from four types of descriptors, given as the: gray values, the intensity moment, the moment invariant and SIFTs. The descriptors are obtained from local patches, based on certain spatial interest points and the weak hypotheses are the applied to every descriptors in one of the local patches. [41] considered the issue of image representation in generic object recognition. They used a hybrid method of autocorrolations with the bag-of-features approach to achieve some posterior probabilities to pre-classification of images ingeneric object recognition.

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 5 A) A Successful Output Of SURF

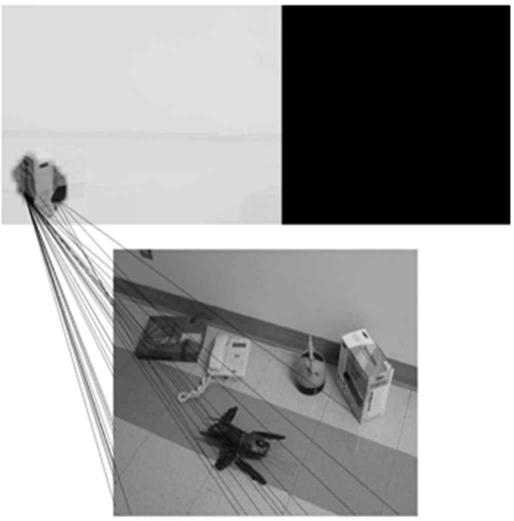

Figure 5.B) An Unsuccessful Object Recognition Using Surf At A Different Scale

[42] proposed a statistical framework for recognizing part based deformable rigid objects which is a expansion of the pictorial structure representation introduced by Fischler et al. The proposed framework consists of three major parts, including: i) an algorithm to find the best global matches for a set of pictorial structure models, ii) A statistical model to calculate multiple matches for a model, and iii) a former part that leads to a new statistical learning model from the labeled example images. This framework is general, in the sense that it is independent of the specific method used to represent the appearance of the parts, as well as the type of geometric relationships that exists between the parts. [43] introduced a method for recognizing objects and carrying out image segmentation by automatically considering occlusion, lighting changes and object deformation. The usefulness of using their method is the computing of multi level object segmentation and dense correspondence between the pixel of the original image and the target deformed image. A new method was introduced by [44] for object recognition and parts recognition known as semantic hierarchies. The semantic hierarchy algorithm works based on the hierarchical representation of objects appearance and parts to extract minimal features, using the hierarchy as a context for calculating the statistical analysis of extracted features. Part detection is obtained from the bottom-up top-down cycle.

ORB is one of the latest methods, proposed in last year by [45]. It is 14 times faster than SIFT and 300 times faster than SURF and is consisting of binary descriptors based on BRIEF features and oriented descriptors. ORB is a scale invariant method that works well in many situations. Table 2 provides a critical analysis of the various object recognition methods.

4. OBJECT CLASS RECGNITION

Object class recogniton is the process of categorizing and classifying object categories based on similiarities with other objects obtained from prior knowledge. In contrast, object recognition is used to recognize objects that have been seen before, where the recognition process will finds that object in the given images. In the process of recognizing object categories, human behavior is based on shape, color, size and other object characteristic. Figure. 6 illustrates the number of object class recognition publications based on the WEB OF SCIENCE webstite.

Object class recognition is similar to generic object recognition; however, object class recognition clarifies the object categories.The Classification process has a close relationship with object segmention in obtaining the shape of the target objects. Principally, the shape of an object plays an important role in object classification. In this way, the algorithms discussed in this paper are primarily simlar to the shape matching algorithm. This section is devided in two parts, i.e., similarity object class recognition and novel object class recognition; in both, shape matching is the main component of comparison. In the former algorithms, previous knowlegdge has been used for the class recognition process. However, in the latter algorithms there is no prior knowledge, and

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the objects are considered to be compeletely unseen. A visual dictionary of discriminative features or other object properties is a type of supervised learning method that is functional in the object class recognition process. Figure. 7 is an example of shape matching using shape skeletons [5].

Table 2: Critical analysis of object recognition

Year	Title of Method	Advantage	Incompetence
2004	SIFT	i)Works based on scale invariant features ii) Uses an orientation histogram of the sample points iii)Measures the Euclidian distance between two vectors	i)Is not very fast ii)Only considers the scale and rotation of the objects
2004	PCA-SIFT	i)Is similar to SIFT ii)Used PCA to extract distinctive descriptors iii)Measures the Euclidian distance between two vectors	i) Is very slow ii)The feature vector is small
2006	SURF	i)Uses an integral image to original one ii)Uses a fast hessian matrix iii)Is faster than the previous method	i)Uses nearest neighbor to match the keypoints
2011	ORB	i) An Alternative to SIFT and SURF ii)Uses BRIEF features iii) Very faster in omparison SURF and SIFT	i)Uses the nearest neighbor to match the keypoints

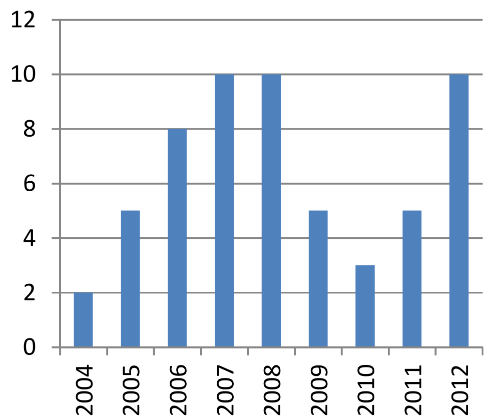

Figure. 6: The Number Of Publications On Object Class Recognition

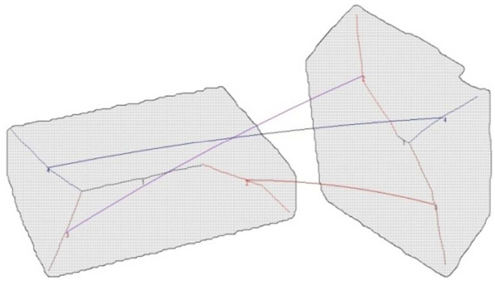

Figure. 7: A Shape Matching Result Based On A Shape Skeleton

[46] proposed a universal visual dictionary algorithm for recognizing object classes. Learning in the model is based on the supervised learning method in which the object classes learn from a set of training images. The model is suitable for application in semantic image retrieval and image understanding. The algorithm combines two main steps, including: extracting the visual words from a large visual dictionary and measurement of feature discrimination. Unlike the previous methods, which only consider some of the interest points or part of the universal visual dictionary, this method is focused on learning the discriminative features automatically instead of taking into account all of the pixels. This algorithm is powerful enough to classify both texture-rich (e.g. grass, sky and trees) and structure-rich (e.g. cars, bikes and planes) objects.

[47] used invariant features to develop an object class recognition method in clattered scenes. This method is a probability-base method that computes the likelihood of feature representations (such as appearance, shape, and occlusion), estimates the learning parameters and uses Bayesian manner to classify images. [48] proposed an object class detection algorithm using scale-invariant features from natural scenes. The method first clusters the local invariant descriptors, then trains a portion of the classifiers and finally selects the features to find discriminative descriptors. The differences between the mentioned method and that of the other existing methods are i) avoiding image normalization and ii) avoiding object labeling and background subtraction.

[49] developed and occluded an object detection method in multi view-points. The method is used for decision-making for in object detection and is based on descriptors such as shape, position or pixel values and the boundaries of occluded objects. The approach established a competitive approach between the descriptors in determining the final description and segmentation of the

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

objects. The method proposed by [50] is an extension of the existing single-view object classification method to a multi-view object representation. The obtained result has shown improvements in the method in comparison with that of the single-view representation.

The work by [51] explains a new clustering representation that uses local features and fast matching for large size dataset. This is a hierarchical agglomerative clustering algorithm that is combined with clustering for object class recognition. [52] separated the spares and local features in visual object classification. Then they illustrated the effectiveness of spars features, such as the number of features input into the model, lateral inhibition and feature selection.

[53] presented a method for a log-liner mixture model that extends the Gaussian mixture model (GMM) and promotes switching between the localized descriptors, such as the spatial information and incorrect assumptions.

A new supervised method proposes to recognize object class recognition with boosting presented by [54]. The method employs part base classifiers and designated according to the appearance at each object part; it also uses boosting to learn the model parameters. [55] introduced a hybrid multi-layer Adaboost algorithm for object class recognition based on heterogeneous features. The first layer extracts the features using PCA-SIFT and the shape context of the features, whereas the second layer calculates the spatial relationships between the features from the first layer, using Adaboost. Table 3 provides a critical analysis of the methods reviewed up to this point.

4.1. Similarity Based Object Classification

[56] provided a survey of similarity based and rule based object classification, including their psychological relationships. The obtained results demonstrate that people are primarily relying on similarity to determine object categories, despite the fact that they rely on rule based classification in cases when faced with novel objects. The current study was initially motivated by this idea. The approach is combination of rule-based and similarity-based (i.e., mostly on shapes) classification for seen and unseen (novel) objects.

In his article, [57] described a fragment-based method for object classification. This approach segments images by cropping objects, i.e., fragmenting and categorizing every fragment into classes according to object. Objects are inferred based on combining the detected fragments to develop the mutual information between objects. Experiments have shown the high accuracy of this approach in addition to low error rate in true classification.

[58] introduced a visual similarity-based method for object categorization according to human behaviors for finding the relationships between visually similar objects. The proposed method classifies objects based on their similarities, such as color, texture and shape, using Adaboost. Additionally, the method learns the objects that do not belong to any classes but exhibit similarity to several classes.

4.2. Novel Object Classification

[59] proposed a model based on the appearance and shape of objects for object classes. In the model, both appearance and shape are used to recognize object classes in cooperation with correlatons (adaptive vector quantized correlograms) and visual words. The method covers the problems of geometric transformation, occlusion and partially missing objects. In the same year, Winn presented a video-based combinational object class recognition method that classifies objects at a glance. The method uses visual feature extraction and efficient class models. In this way, the most challenging issue is the subtraction of objects from the background; this process is performed by using patch-based classifiers to discriminate the background (presented as a table in this work) and any objects. The proposed method is more efficient and robust because it covers problems such as the shadows, illumination and camera position. The proposed algorithm is a shape-based model for class recognition is according to a decision tree in which each node is associated with either appearance or shape [60].

[61] carried out a survey on learning and recognizing object class models. The work recently proposed is particularly considerate of the transfer of knowledge during the process of learning objects in comparison with human behavior. According to the groups used for types of knowledge transfer in the paper, three categories were established, transfer through prior parameters, transfer through shared features or parts, and transfer through contextual information. [62] introduced a method for learning novel classes based on one training image. The method is based on feature selection, which extracts the features from a feature filtering system. The

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

similarity of the extracted features and the novel feature specify the mutual feature classes for two objects.

[63] developed the idea of visual category recognition. They tried to measure the distance between a pair of images by learning the image-to-image distance. For this purpose the criterion is a patch-based feature vector for calculating such distances. [64] proposed an object classification model for any form of objects, named scalable object classification. Moreover, the method works for categorizing 3D object models and instance learning for novel objects. The proposed method is an extension of the TAX model used for unsupervised 2D object modeling and classifying, online learning and inferring according to the introduced visual dictionary words. The accuracy of the proposed method for an offline dataset is 94% and for a real time dataset, 88.4%.

[65] proposed a new framework for calssifying large image semantics via hierarchical classification. In the this method, salient objects are the intermediate of image semantics. The framework of image semantics is incorporated with the product of mixture-experts (PoM) method for decreasing high dimensionality and improving time complexity. Additionaly, the method covers the concept learning problem for object learning, concept model learning and the computational complexity for large image datasets. The framework has the capability of to specify and assess the results provided by users.

[66] presented a grid-based image classification method with a new probabilistic model i.e., 2D conditional random field. In this method, each image is segmented into blocks and a feature vector is incorporated with 2D neighborhood blocks. [67] proposed a multi NN (nearest neighborhood)-based classifier that categorizes the unlabeled objects. In this way, NN incorporates categorizers and feature descriptors.

Csurka's method emphasizes solving 3D object classification problems. The supervised method uses local invariant features and connects them to reproduce a compact summarization of appearance and geometric information [68]. In contrast, Csurka and Willamowski both have used the bag-of-keypoints in their methods. The Bag-of-keypoints is a set of vector quantization for affine invariant descriptors. Jointly these methods used SVM and Bayes classifiers for implementation and to comparison building [69, 70].

The method presented by [71] is an automatic UNV supervised framework that tries to classify and to learn buildings and vehicles using semantic analysis. The framework components include a content-based 3D mosaic (CB3M) representation, CB3M-based building detection and a probabilistic dynamic influence diagram.

Table 3: A Critical Analysis Of Object Class Recognition

Year	Title of Method	Advantage	Incompetence
2003	Fergus's Method	i)Uses scale invariant features and shape ii)Is aprobabilistic method	i)Only uses features that is not enough
2004	Darko's Method	i)Uses local invariant descriptors ii)Finds discriminative features	i)Does not normalize images ii)Does not label objects
2005	Hillel's Method	i)Uses boosting ii)Is an unsupervised learning process	i)Is not suitable for novel objects
2005	Zhang's Method	i)Uses Adaboost ii)Uses PCA-SIFT to feature extraction and shape context	i)Is slow
2008	Winn's Method	i)Uses visual dictionary to search object classes ii)Using patch-based classifiers to segment objects iii)Is invariant in terms of scale and rotation iv)Uses a decision tree that is based on the object shape	i)Uses pixel features ii)Uses only one parameter to recognize object classes

5. CONCLUSION AND FUTURE DIRECTIONS

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This paper is aimed to aid researchers in object recognition and object class recognition when choosing the most appropriate methods for image segmentation, object recognition, and object class recognition. As the first objective of this study, visual attention is to be in the early stage of development. Despite this, many methods have been proposed based on simulation of human visual system; even so, new algorithms are required to obtain the best accuracy in various environments. Top-down and bottom-up approaches which are both current areas of concentration for researchers; however, each possesses inadequacy in certain environments. In some cases a combinational algorithm of bottom-up and top-down approach is appropriate. Others researchers have used saliency mixed with bottom-up visual attention, although this approach is still evolving in application to contrast and illumination problems. The future direction of visual attention is to solve the problem of segmenting objects from images.

Object recognition was the second topic covered in this paper. Although object recognition has been a tremendous progress in recent years, there is long distance that must be covered bfore objects can be recognized accurately. The number of extracted features plays an important role in the accurate recognition of objects. Hence, an algorithm always must find sufficient number of features for detecting all objects in the dataset in the fastest way. SURF, SIFT, PCA-SIFT and other methods as mentioned in this paper, are proposed, although none works completely on any dataset. In the future, it is necessary to find more distinctive features for recognizing more objects in images.

Object class recognition is mostly performed using shape matching algorithms and other similarities, such as texture, color, etc. Winn's methods covered most types of object class recognition states (using universal dictionaries, novel objects), followed by [58] and Frome's methods. Another main component in object classification is the visual dictionary, which focuses on the discriminative features of objects in each category.

5.1. The Future Direction of Saliency Visual Attention and Challenges

Since 1988, even though many methods have proposed using top-down and bottom-up visual attention techniques in addition to those combined with salience objects in images, yet this area is still going forward. In recent years, Visual attention research has expanded joint with lots of features and descriptors, as mentioned earlier in this work. Numerous challenges exist for visual attention in region selection. One is the subtraction of object borders from regions of interest. Another is occlusion and differences in illumination on surfaces. Distinction in contrast or illumination leads to ambiguity in separating the visual attention area for object detection in segmented parts. The inadequacy of bottom-up and top-down visual attention when applied to various environments poses another obstacle. [72] proposed a switching method between top-down and bottom-up visual attention. The method tries to fill the gap between the method using top-down biased and bottom-up attentional selection and has been adopted for various robot processes, i.e., visual behavior. However the method is still limited by the constraints of those approaches.

The current subject of research includes the conceptual meaning of visual attention in addition to trying to find the processes used by the brain in attention regions and for analyzing objects in images, as detailed by [73], whose research is entitled "How do people come to recognize scenes at a glance?". Additionally, researchers are trying to develop a more accurate method for object recognition. [74] survey the current and future applications in the visual attention area of research. According to their experiments, the most important applications use prior knowledge conjointly with low-level visual features. Additionally, they compared 6 saliency maps of visual attention, such as that developed by [13]. However, these authors believe that the field of visual attention is only in its infancy.

5.2. The Future Direction of Object Recognition and Challenges

Object recognition experienced tremendous progress during the recent decades, and today, so many applications touch on the necessity of using object recognition algorithms. Most object recognition algorithms are based on features extracted from images. The challenge is to find the distinctive features to look up in images. Some methods implement pre-processing techniques to improve the accuracy of object recognition, thereby avoiding false positive detection. Nevertheless, it is a time consuming process and is not appropriate for real time applications. Other methods are more concentrated on speed, extracting several unique features in the initial steps and comparing those during the recognition process to identify any matching instance. This

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

may result in no matching features even though the target object still exists in the image. Hence, overcoming the lack of a sufficient number of unique features is another milestone in object recognition. In this way, the direction of object recognition is moving forward, finding more unique features capable of covering all of the objects in images. Even so, there remains long way to go within the field.

Recent methods have been considered to focus on the most unique features extracted from objects. Other have been presented to track and grasp the mobile objects in a video stream. Therefore, in the future, perhaps researchers' findings will yield more distinctive features in objects recognition.

Object class recognition is the next step after object recognition and classification. If an object cannot be recognized in specific, then object class recognition tries to find the category of that object. For this purpose, shape similarity is the first feature that would be compared, and the second feature could be size, color or frame. The problem occurs when the shape matching cannot yield any result. Then new ideas for the recognition of object classes have progressed toward the identification of more common features, used to discriminate between the objects in a category.

5.3. Future Direction of Object Class Recognition and Challenges

Object class recognition in the few past years has received more attention for the classification of objects into universal categories. Most of the proposed methods have used object recognition features to recognize the category of objects. Additionally, an object's appearance and shape are other parameters commonly used for object class recognition. The challenges in the recognition of object classes are i) confronting a novel object and ii) the ability to recognize objects classes that which are similar in shape but different in category. Other problems are the position of objects in images and occlusion, which can prevent obtaining the desired shape of an object. The use of a set of parameters for the placement of two objects in same category is the future direction of this research area. The new methods are similar to the previous methods, with the primary difference in the number of parameters used for classification. For future work, one idea is to combine the results of shape matching with human interaction or behavior detection during use of objects for which the replacement of object tracking by object recognition is the only necessity.

REFERENCES:

[1] Walther, D., et al., *Selective visual attention enables learning and recognition of multiple objects in cluttered scenes.* Computer Vision and Image Understanding, 2005. **100**(1-2): p. 41-63.

[2] Grill-Spector, K. and N. Kanwisher, *Visual recognition.* Psychological Science, 2005. **16**(2): p. 152.

[3] Fei-Fei, L., et al., *What do we perceive in a glance of a real-world scene?* Journal of Vision, 2007. **7**(1): p. 10.

[4] Juan, L. and O. Gwun, *A comparison of sift, pca-sift and surf.* International Journal of Image Processing (IJIP), 2010. **3**(5): p. 143-152.

[5] van Eede, M., et al., *Canonical skeletons for shape matching*, in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* 2006, IEEE Computer Society. p. 64-69.

[6] Connor, C.E., H.E. Egeth, and S. Yantis, *Visual attention: bottom-up versus top-down.* Current Biology, 2004. **14**(19): p. R850-R852.

[7] Desimone, R. and J. Duncan, *Neural mechanisms of selective visual attention.* Annual review of neuroscience, 1995. **18**(1): p. 193-222.

[8] Itti, L. and C. Koch, *Computational modeling of visual attention.* Nature reviews neuroscience, 2001. **2**(3): p. 194-203.

[9] Treue, S., *Visual attention: the where, what, how and why of saliency.* Current Opinion in Neurobiology, 2003. **13**(4): p. 428-432.

[10] Lee, S., et al., *Familiarity based unified visual attention model for fast and robust object recognition.* Pattern Recognition, 2010. **43**(3): p. 1116-1128.

[11] KP, S., J. Amudha, and Y.Kiran, *Feature Selection in Top-Down Visual Attention Model using WEKA.* International Journal of Computer Applications, 2011. **24**(4): p. 975-8887.

[12] Itti, L. and C. Koch, *A comparison of feature combination strategies for saliency-based visual attention systems.* SPIE human vision and electronic imaging IV (HVEI'99), 1999. **3644**: p. 373-382.

[13] Itti, L., C. Koch, and E. Niebur, *A model of saliency-based visual attention for rapid*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

scene analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1998. **20**(11): p. 1254-1259.

[14] Walther, D., et al., *Attentional selection for object recognition—a gentle way.* BIOLOGICALLY MOTIVATED COMPUTER VISION, 2010. **2525**: p. 251-267.

[15] Hecht, L.N. and S.P. Vecera, *Attentional selection of complex objects: Joint effects of surface uniformity and part structure.* Psychonomic bulletin & review, 2007. **14**(6): p. 1205-1211.

[16] Zhai, Y. and M. Shah. *Visual attention detection in video sequences using spatiotemporal cues*. in *Proceedings of the 14th annual ACM international conference on Multimedia*. 2006. ACM.

[17] Amudha, J., K. Soman, and Y. Kiran, *Feature Selection in Top-Down Visual Attention Model using WEKA.* International Journal of Computer Applications, 2011. **24**.

[18] Tian, M., S. Wan, and L. Yue, *A Color Saliency Model for Salient Objects Detection in Natural Scenes.* Advances in Multimedia Modeling, 2010: p. 240-250.

[19] Tsotsos, J.K., et al., *Modeling visual attention via selective tuning.* Artificial intelligence, 1995. **78**(1-2): p. 507-545.

[20] Engelke, U., H.J. Zepernick, and A. Maeder. *Visual attention modeling: Region-of-interest versus fixation patterns*. in *Picture Coding Symposium, 2009. PCS 2009*. 2009.

[21] Marques, O., et al. *Using visual attention to extract regions of interest in the context of image retrieval*. in *Proceedings of the 44th annual Southeast regional conference.* 2006. ACM.

[22] Frintrop, S., E. Rome, and H.I. Christensen, *Computational visual attention systems and their cognitive foundations: A survey.* ACM Transactions on Applied Perception (TAP), 2010. **7**(1): p. 6.

[23] Deco, G. and B. Schürmann, *A hierarchical neural system with attentional top–down enhancement of the spatial resolution for object recognition.* Vision research, 2000. **40**(20): p. 2845-2859.

[24] Parkhurst, D., K. Law, and E. Niebur, *Modeling the role of salience in the allocation of overt visual attention.* Vision research, 2002. **42**(1): p. 107-123.

[25] Sun, Y., *Hierarchical Object-Based Visual Attention for Machine Vision*. 2003, University of Edinburgh. College of Science and Engineering. School of Informatics.

[26] Frintrop, S., A. Nüchter, and H. Surmann, *Visual attention for object recognition in spatial 3D data.* Attention and Performance in Computational Vision, 2005: p. 168-182.

[27] Borji, A., M.N. Ahmadabadi, and B.N. Araabi. *Simultaneous learning of spatial visual attention and physical actions*. in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on* 2010. Taipei: IEEE.

[28] Basak, J., et al., *A connectionist model for category perception: theory and implementation.* Neural Networks, IEEE Transactions on, 1993. **4**(2): p. 257-269.

[29] Williams, C.K.I. and M.K. Titsias, *Learning about multiple objects in images: Factorial learning without factorial search.* Advances in neural information processing systems, 2003: p. 1415-1424.

[30] Schneiderman, H. and T. Kanade, *Object detection using the statistics of parts.* International journal of computer vision, 2004. **56**(3): p. 151-177.

[31] Shotton, J., et al., *Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context.* International journal of computer vision, 2009. **81**(1): p. 2-23.

[32] Barinova, O., V. Lempitsky, and P. Kohli. *On detection of multiple object instances using hough transforms*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* 2010. San Francisco, CA IEEE.

[33] Wang, D., B. Zhang, and J. Zhang. *Visual Object Recognition in Diverse Scenes with Multiple Instance Learning*. in *International Conference on Intelligent Robots and Systems*. 2006. IEEE.

[34] Ulusoy, I. and C.M. Bishop. *Generative versus discriminative methods for object recognition*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.

[35] Lowe, D.G. *Object recognition from local scale-invariant features*. in *Computer Vision, 1999. The Proceedings of the*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Seventh IEEE International Conference on. 1999. Ieee.

[36] Ke, Y. and R. Sukthankar, *PCA-SIFT: A more distinctive representation for local image descriptors*. Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004. **2**: p. 506-513.

[37] Viola, P. and M. Jones, *Robust real-time object detection*. International journal of computer vision, 2002. **57**(2): p. 137–154.

[38] Lazebnik, S., C. Schmid, and J. Ponce. *Semi-local affine parts for object recognition*. in *British Machine Vision Conference (BMVC '04)*. 2004. Citeseer.

[39] Costantini, G., D. Casali, and R. Perfetti. *Detection of moving objects in a binocular video sequence*. in *Cellular Neural Networks and Their Applications, 2006. CNNA '06. 10th International Workshop on*. 2006. Istnbul: IEEE.

[40] Opelt, A., et al., *Generic object recognition with boosting*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006: p. 416-431.

[41] Matsukawa, T. and T. Kurita, *Image representation for generic object recognition using higher-order local autocorrelation features on posterior probability images*. Pattern Recognition, 2011. **45**(2): p. 707-719.

[42] Felzenszwalb, P.F. and D.P. Huttenlocher, *Pictorial structures for object recognition*. International journal of computer vision, 2005. **61**(1): p. 55-79.

[43] Simon, I. and S.M. Seitz. *A probabilistic model for object recognition, segmentation, and non-rigid correspondence*. in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* 2007. IEEE.

[44] Epshtein, B. and S. Ullman. *Semantic hierarchies for recognizing objects and parts*. in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* 2007. Minneapolis: IEEE.

[45] Rublee, E., et al. *ORB: an efficient alternative to SIFT or SURF*. in *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011. IEEE.

[46] Winn, J., A. Criminisi, and T. Minka, *Object categorization by learned universal visual dictionary*. Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on 2005. **2**: p. 1800 - 1807.

[47] Fergus, R., P. Perona, and A. Zisserman. *Object class recognition by unsupervised scale-invariant learning*. in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* 2003. IEEE.

[48] Dorkó, G. and C. Schmid, *Object class recognition using discriminative local features*. INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE, 2005: p. 22.

[49] Hoiem, D., C. Rother, and J. Winn. *3d layoutcrf for multi-view object class recognition and segmentation*. in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* 2007. IEEE.

[50] Chiu, H.P., L.P. Kaelbling, and T. Lozano-Pérez. *Virtual training for multi-view object class recognition*. in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* 2007. Minneapolis, MN IEEE.

[51] Leibe, B., K. Mikolajczyk, and B. Schiele. *Efficient clustering and matching for object class recognition*. in *British Machine Vision Conference (BMVC'06)*. 2006.

[52] Mutch, J. and D.G. Lowe, *Object class recognition and localization using sparse features with limited receptive fields*. International journal of computer vision, 2008. **80**(1): p. 45-57.

[53] Weyand, T., T. Deselaers12, and H. Ney. *Log-linear mixtures for object class recognition*. 2009.

[54] Bar-Hillel, A., T. Hertz, and D. Weinshall. *Object class recognition by boosting a part-based model*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.

[55] Zhang, W., et al. *Object class recognition using multiple layer boosting with heterogeneous features*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* 2005. IEEE.

[56] Wood, T.J., *On the rules-to-episodes transition in classification: Generalization of similarity and rules with practice*. OPEN ACCESS DISSERTATIONS AND THESES, 1998: p. 2055.

[57] Ullman, S. and E. Sali, *Object classification using a fragment-based representation*. BIOLOGICALLY MOTIVATED

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COMPUTER VISION, 2000. **1811**: p. 116-129.

[58] Zhang, W., D. Samaras, and G.J. Zelinsky. *Classifying objects based on their visual similarity to target categories*. in *30th Annual Meeting of the Cognitive Science Society*. 2008. Cognitive Science Society.

[59] Winn, J. and J. Shotton, *The Layout Consistent Random Field for Recognizing and Segmenting Partially Occluded Objects (PDF)*, in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* 2006, IEEE. p. 37-44.

[60] Winn, J. and A. Criminisi. *Object class recognition at a glance*. in http://research.microsoft.com/vision/cambridge/recognition. 2006. Citeseer.

[61] Fei-Fei, L. *Knowledge transfer in learning to recognize visual objects classes*. in *Conference on Development and Learning*. 2006.

[62] Bart, E. and S. Ullman, *Cross-generalization: Learning novel classes from a single example by feature replacement.* Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on 2005. **1**: p. 672 - 679.

[63] Frome, A., et al., *Learning globally-consistent local distance functions for shape-based image retrieval and classification*, in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* 2007, IEEE.

[64] Kim, E. and G. Medioni. *Scalable Object Classification Using Range Images*. in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2011 International Conference on*. 2011. IEEE.

[65] Fan, J., et al., *Mining multilevel image semantics via hierarchical classification.* Multimedia, IEEE Transactions on, 2008. **10**(2): p. 167-187.

[66] Wen, M., et al., *2D Conditional Random Fields for Image Classification.* Intelligent Information Processing III, 2007: p. 383-390.

[67] He, D., S. Liang, and Y. Fang, *A multi-descriptor, multi-nearest neighbor approach for image classification.* Advanced Intelligent Computing Theories and Applications, 2010: p. 515-523.

[68] Savarese, S. and L. Fei-Fei. *3D generic object categorization, localization and pose estimation*. in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* 2007. Rio de Janeiro: IEEE.

[69] Willamowski, J., et al., *Categorizing nine visual classes using local appearance descriptors*. illumination, 2004. **17**: p. 21.

[70] Csurka, G., et al. *Visual categorization with bags of keypoints*. in *In ECCV International Workshop on Statistical Learning in Computer Vision (2004) Key: citeulike:581127*. 2004. Citeseer.

[71] Lyu, S. *Mercer kernels for object recognition with local features*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005.

[72] Xu, T., K. Kuhnlenz, and M. Buss, *Autonomous behavior-based switched top-down and bottom-up visual attention for mobile robots.* Robotics, IEEE Transactions on, 2010. **26**(5): p. 947-954.

[73] Walther, D.B., et al., *Simple line drawings suffice for functional MRI decoding of natural scene categories.* Proceedings of the National Academy of Sciences, 2011. **108**(23): p. 9661.

[74] Le Meur, O. and P. Le Callet. *What we see is most likely to be what matters: visual attention and applications*. in *Image Processing (ICIP), 2009 16th IEEE International Conference on* 2009. IEEE.