Object detection from Background Scene Using t-SNE -ORB Gradient Boost

Radhamadhab Dalai[1], Prof. Kishore Kumar Senapati[2]

[1]*Computer Science and Engineering, BIT Mesra, Ranchi*
Jharkhand, 835215, India

[2]*Computer Science and Engineering, BIT Mesra, Ranchi*
Jharkhand, 835215, India

Abstract

Object Classification using Gradient Boost is a robust mechanism used for computer vision problem domain. The acceleration of t-SNE—an embedding technique that is commonly used for the visualization of high-dimensional data in scatter plots—using two tree-based algorithms. In particular, the paper develops variants of the Barnes-Hut algorithm and of the dual-tree algorithm that approximate the gradient used for learning t-SNE embeddings in O (NlogN).Complex background adds challenge and error margin as well to the problem significantly lot algorithms for object detection are hard to comply with occlusion and pixel bending moment affect. In this paper a highly robust algorithm for gradient boost based t-SNE[16] for a less resolution image has been proposed and implemented using ORB detection with gradient boosting machine learning algortihm.The work has been compared with Adaboost and Surf based technology. The analysis of result shows 4.2% increase in performance of earlier model. The feature points extracted from ORB method are further processed to reduce the processing further. Only those points are selected which are triangularly farthest from centroid of it and only 1 point of feature selected. Thus the result is around 28%, much faster than earlier computation. The tree based GB has been implemented in this algorithm. With more number of feature points more classes need to be recognized and hence the computations performed is required an unreasonable amount of effort and time. So some nearby classes are assigned at same level using our algorithm to reduce the number of tree nodes. Overall performance of the proposed algorithm shows a significant increase in efficiency in computation time.

Keywords: Object detection, *machine vision,* t-SNE algorithm, *gradient boosting, Tree based GB algorithm*

1. Introduction

The Human visual system has the ability to process parts of image which are relevant, discarding the rest. This helps us to perceive objects even before identifying them. Object detection from very complicated background including multiple similar objects com-putationally detecting these relevant regions is a complex problem which takes cues from models in machine inteligence, Robotics and computer vision. It has gained a lot of attention in the recent years from the computer vision community owing to its use in object recognition [13], image segmentation [2], image re-targeting and cropping, image retrieval etc.

Works in Object detection are classified into three categories: Feature detection, feature processing, object classification (precomputation and processing), Object identification.

1. Feature-based object detection

In feature-based object detection, standardization of image features and registration (alignment) of reference points are important. The images may need to be transformed to another space for handling changes in illumination, size and orientation. One or more features are extracted and the objects of interest are modeled in terms of these features. Object detection and recognition then can be transformed into a graph matching problem. All types of features go through a grouping algorithm which finds matches using the various attributes for each feature type. A set of average feature attributes is created for each group. All the groups are checked for sufficient redundancy to ensure the feature occurs on multiple contours, meaning it is part of a significant outline. This means the number of features is now vastly reduced so a more intelligent grouping algorithm can be performed.

In feature-based object detection, standardization of image features and registration (alignment) of reference points are important. The images may need to be transformed to another space for handling changes in illumination, size and orientation. One or more features are extracted and the objects of interest are modeled in terms

of these features. Object detection and recognition then can be transformed into a graph matching problem.

2. Shape-based approaches

Shape-based object detection [4] is one of the hardest problems due to the difficulty of segmenting objects of interest in the images. In order to detect and determine the border of an object, an image may need to be preprocessed. The preprocessing algorithm or filter depends on the application. Different object types such as persons, flowers, and airplanes may require different algorithms. For more complex scenes, noise removal and transformations invariant to scale and rotation may be needed. Once the object is detected and located, its boundary can be found by edge detection and boundary-following algorithms. The detection and shape characterization of the objects becomes more difficult for complex scenes where there are many objects with occlusions and shading.

1. Color-based approaches

Unlike many other image features (e.g. shape) color is relatively constant under viewpoint changes and it is easy to be acquired. Although color is not always appropriate as the sole means of detecting and tracking objects, but the low computational cost of the algorithms proposed makes color a desirable feature to exploit when appropriate.

Massimo Bertozz et. al. [15] developed an algorithm to detect and track vehicles or pedestrians in real-time using color histogram based technique. They created a Gaussian Mixture Model to describe the color distribution within the sequence of images and to segment the image into background and objects. Object occlusion was handled using an occlusion buffer. This has been achieved tracking multiple faces in real time at full frame size and rate using color cues. This simple tracking method is based on tracking regions of similar normalized color from frame to frame. These regions are defined within the extent of the object to be tracked with fixed size and relative positions. Each region is characterized by a color vector computed by sub-sampling the pixels within the region, which represents the averaged color of pixels within this region. They even achieved some degree of robustness to occlusion by explicitly modeling the occlusion process.

2. Template-based object detection

If a template describing a specific object is available, object detection becomes a process of matching features between the template and the image sequence under analysis. Object detection with an exact match is generally computationally expensive and the quality of matching depends on the details and the degree of precision provided by the object template. There are two types of object template matching, fixed and deformable template matching. [3]

Adaboost (Adaptive boosting) [16] is a machine learning algorithm that addresses shape based technique of classification. It can be used with many different classifiers to improve the accuracy. Adaboost is adaptive in the sense that subsequent weak learners are tweaked [16]. Adaboost focuses on more previously misclassified samples.
Initially all samples are equal weights. Weight may change at each boosting round. It can be less susceptible to the over fitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better and the final model can be proven to converge to a strong learner. Steps of Adaboost classifiers are Bootstrapping, Bagging, Boosting, and Adaboost. A boost classifier is a classifier in the form.

2. Methodology

Input image are sometimes corrupted with noise and camera movement. So they need to be processed and noise-free image is restored from original noisy image. In our work the processing stage for noise and blurriness has been ignored and instead a standard median filter can be applied to remove impulse noise. The affect of reflection and brightness have been taken care of using *T-SNE* algorithm as underneath it had used ORB. The feature points are selected background color distribution may be multimodal, which could result in more than one color being included in the background model. This make possible to model periodic changes in the background, properly. If the current pixel value does not match any of the first D distributions, it is considered as a part of a foreground object.
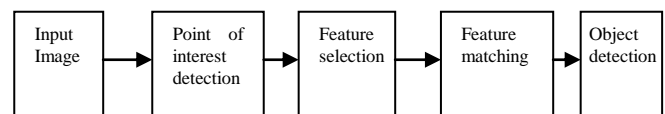

Fig 1: Flow of object detection technique T-SNE

2.1 Proposed algorithm

ORB is a combination of FAST feature detection and BRIEF Feature descriptor.

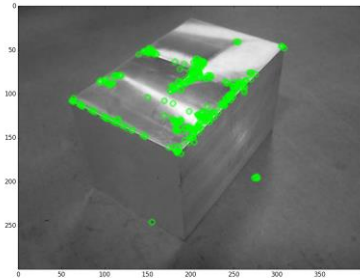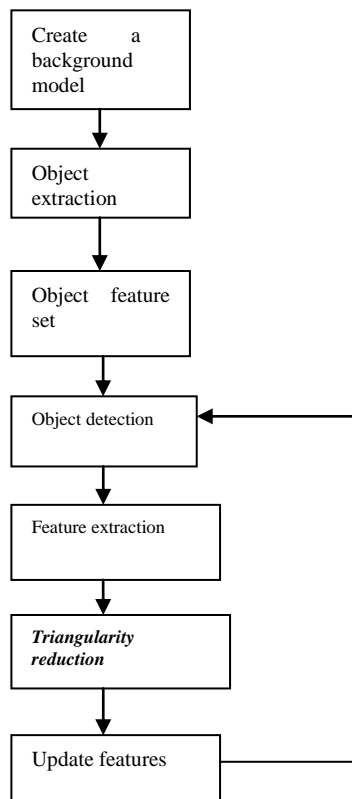***Fig 2: Input image of a solid image to the trained***

Fig 3: ORB based detected feature points

2.2 Detection algorithms

The proposed algorithm consists of following steps as shown in flow diagram mentioned below.

Fig 4: Flow diagram of proposed Algorithm

2.2.1 Creating a background model

Mixture of Gaussians is a widely used approach for background modeling to detect moving objects from static cameras. Numerous improvements of the original method developed by Stauffer and Grimson [1] have been proposed over the recent years and the purpose of this paper is to provide a survey and an original classification of these improvements. We also discuss relevant issues to reduce the computation time. Firstly, the original MOG are reminded and discussed following the challenges met in video sequences. Then, we categorize the different improvements found in the literature. We have classified them in term of strategies used to improve the original MOG and we have discussed them in term of the critical situations they claim to handle. Complex background creates negative impact on accuracy of detection and classification

2.2.2 Object extraction

Target object can be separated from background image using back ground segmentation approach and then can be analyzed using image segmentations algorithms [1, 2]. Background subtraction can roughly classify pixels of background and foreground, but the resultant segmentation result may be still quite noisy due to camera noises, illumination variations, and inappropriate threshold selections. Some post-filtering operations are subsequently performed to refine the segmentation result. To mitigate the distortion of the corresponding background model, the binary segmentation result is median filtered with a 3x3 mask, then is further refined with a morphological filter. Commonly, detectors use *maxima and minima points*, such as gradient peaks and corners; however, edges, ridges, and contours are also used as keypoints.There is no superior method for interest point detection for all applications. A simple taxonomy provided by edge-based region methods (EBR), maxima or intensity-based region methods (IBR), and segmentation methods to find shape-based regions (SBR) that may be blobs or features with high entropy.

2.2.3 Object featureset

There are many methods for object recognition, generally these can be split into algorithms which use global image features and

algorithms which use local image features. Global image features refer to properties of an image as a whole, such as color histogram, outline shape, and texture. The drawbacks of using global features for object recognition include sensitivity to clutter and occlusion, and difficulty in localizing an object in an image. Local image features describe small image regions around interest points, there are many different types of local images features such as edge and corner features; an advantage of local image features for object recognition is a greater robustness to occlusion and clutter compared to global features.The Scale Invariant Feature Transform (SIFT) developed by Lowe is a local image feature detector which finds stable scale invariant interest points in an image by taking the maxima and minima points in scale space, which is formed by taking the difference of successive Gaussian convolutions of the image. These points are then filtered to remove low contrast points or those located along an edge, as they are less likely to be reproducible. For the remaining points an orientation is assigned which is the direction of the local image gradient, and finally a rotation and scale invariant 128-dimensional description vector of the surrounding pixels is generated. As a result a SIFT feature can be characterized by its image position, scale, orientation, and description vector. SIFT features can be used to recognize an object by building a database of object features from training images. Features of a scene image are matched against a database of object features; a geometric consistency model is then applied to all matching feature pairs to remove inconsistent outliers and to determine the object location and orientation in the scene. This can be done using the SURF algorithm or the ORB algorithm .There are various ways to match a given image feature to a database of learned features. A simple method is to find any database feature that has a description vector with a Euclidean distance to the given feature's description vector below some threshold value.
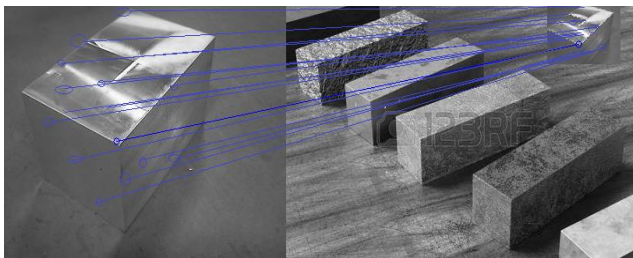
2.2.5 Object Detection

Fig 5: A ORB based feature matching and object detection using OpenCV

We describe a family of object detectors that provides state-of-the-art error rates on several important datasets including some solid basic blocks (iron, wood, and copper, concrete) as images and ORB based algorithm. The method builds on a number of recent advances. It uses the ORM machine learning framework and a rich visual feature set that incorporates Histogram of Oriented Gradient, Local Binary Pattern and Local Ternary Pattern descriptors. Partial Least Squares dimensionality reduction is included to speed the training of the basic classifier with no loss of accuracy, and to allow a two-stage quadratic classifier that further improves the results. We evaluate our methods and compare them to other recent ones on several datasets. Our basic root detectors outperform the single component part-based ones of methods to find shape-based regions (SBR) that may be blobs or features with high entropy.

2.2.5 Triangularity reduction

Each arrow line length denotes the distance from centroid of a particular class. So distance of feature points from centroid of a class intra-class is calculated, now the feature points are sorted according distance measure. Each three feature points are grouped in one block. Out of each block only one is selected for candidate feature point. The comparison is done within same class and each feature wise. The feature selected from key points based on ORB are taken as input in modified algorithm named as T-SNE. The key points are minimized as much possible as possible for less computation purpose with accuracy and precision to be kept same as earlier SURF based algorithm. The distance in graph diagram as shown below depicts the distance between each feature points (dotted) and solid line represents centorid of single class computed earlier. After further computation centroid is readjusted with modified or remained key feature points.
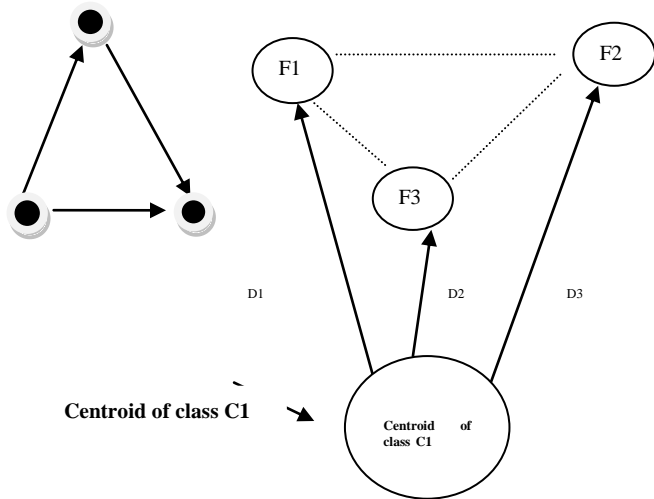
Fig 3.1(a): Feature points as represented as F1, F2, and F3.

Fig 3.1(b): Triangular showing inter distance between individual feature points

3.1 Classifiers algorithms

A classifier is a function that allocates a population's element value from one of the available categories. For instance, Spam Filtering is a popular application of Naïve Bayes algorithm. Spam filter here, is a classifier that assigns a label "Spam" or "Not Spam" to all the emails.KNN and SVM are another group of algorithms for classifying features set into a particular class of object set. Various types of KNN such as FAST-KNN, HKNN (Hyper plane distance nearest neighbor) can be used for classification problem.The algorithm for Boosting Trees evolved from the application of boosting methods to regression trees. The general idea is to compute a sequence of (very) simple trees, where each successive tree is built for the prediction residuals of the preceding tree. As described in the General Classification and Regression Trees Introductory Overview, this method will build binary trees, i.e., partition the data into two samples at each split node. Now suppose that you were to limit the complexities of the trees to 3 nodes only: a root node and two child nodes, i.e., a single split. Thus, at each step of the boosting (boosting trees algorithm), a simple (best) partitioning of the data is determined, and the deviations of the observed values from the respective means (residuals for each partition) are computed. The next 3-node tree will then be fitted to those residuals, to find another partition that will further reduce the residual (error) variance for the data, given the preceding sequence of trees.

It can be shown that such "additive weighted expansions" of trees can eventually produce an excellent fit of the predicted values to the observed values, even if the specific nature of the relationships between the predictor variables and the dependent variable of interest is very complex (nonlinear in nature). Hence, the method of gradient boosting - fitting a weighted additive expansion of simple trees - represents a very general and powerful machine learning algorithm.

3.2 Gradient boosting

Gradient boosting is one of the most powerful techniques for building predictive models. In this paper the gradient boosting machine learning algorithm has been used and the brief concept about its functioning has been mentioned below.

Gradient boosting involves three elements:

- A loss function to be optimized.
- A weak learner to make predictions.
- An additive model to add weak learners to minimize the loss function.

Algorithm

Input: training set $\{x_j, y_i\}_{i=1}^n$ a differentiable loss function number of iterations M.

Algorithm:

1. Initialize model with a constant value:

$$F_0(x) = \arg_\lambda \min \sum_{i=1}^n L(y_j, \lambda) \ldots\ldots\ldots\ldots (1)$$

2. For $m = 1$ to M:

   1. Compute so-called *pseudo-residuals*:
   $$r_{im} = -\left[\partial L(y, f(x)) / \partial F(x)\right]_{F_x = F_{m-1}(x)}$$

   2. Fit a base learner (e.g. tree) to pseudo-residuals, i.e. train it using the training set. $\{x_i, r_{im}\}_{i=1}^n$

   3. Compute multiplier by solving the following one-dimensional optimization problem:

   $$\gamma_m = \arg_\gamma \min \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \ldots (2)$$

   4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \ldots\ldots\ldots (3)$$

3. Output $F_m(x)$

t-SNE algorithm

t-Distributed stochastic neighbor embedding (t-SNE) minimizes the divergence between two distributions: a distribution that measures pairwise similarities of the input objects and a distribution that measures pairwise similarities of the corresponding low-dimensional points in the embedding. Assume we are given a data set of (high-dimensional) input objects D = (x1, x$_2$, ……. ,x$_N$) g and a function d(x$_i$, x$_j$) that computes a distance between a pair of objects, e.g., the Euclidean distance d(x$_i$, x$_j$) = kx$_i$ x$_j$ k. Our aim is to learn an s-dimensional embedding in which each object is represented by a point, $\mathcal{E}$ = {y$_1$, y$_2$,…..,y$_{N)}$ g with y$_i$ ϵ R^s (typical values for s are 2 or 3). To this end, t-SNE defines joint probabilities p$_{ij}$ that measure the pairwise similarity between objects x$_i$ and x$_j$ by summarizing conditional probability.

$$p_{j|i} = \frac{\exp(-d(x_i, x_j)^2 / 2\sigma_i^2)}{\sum\limits_{k=i} \exp(-d(x_i, x_j)^2 / 2\sigma_i^2)} \ldots\ldots\ldots\ldots (1)$$

It is straightforward to see that the evaluation of the joint distributions P is O(N^2), because both distributions involve a normalization term that sum over all N(N -1) pairs of unique objects. Since t-SNE scales quadratically in the number of objects N, its applicability is limited to data sets with only a few thousand input objects; beyond that, learning becomes too slow to be practical (and the memory requirements become too large).

4. Experiments and Results

The experimental environment is OpenCV 3.2.X and python 2.7. The feature points are calculated and stored and then using T-SNE algorithm the feature points are further adjusted. The processed and minimal feature points are stored and in training data image same feature points are extracted. Then they are matched and classified using enhanced gradient boot algorithm.GB algorithm has been implemented using Tree based GB algorithm. The enhanced algorithm is implemented adjusting the nodes of very similar features grouping into one node or in the same depth line in tree. Later this result is compared with earlier Surf and Adaboost and surf-SVM based approach. In more than above 80% case the result shows better accuracy.

Table 1: Average number of feature points found per image for detectors

Classifiers	SURF	SIFT	ORB
Key Points	506	56	34

Comparison with Adaboost and surf based approach with our proposed T-SNE and Gradient boost algorithm for following performance measures.

4.1 *Performance measures*

1) Accuracy: In the fields of science, engineering, industry, and statistics, the accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's actual (true) value.

$$Accuracy = \frac{TruePositive + TrueNegative}{Total\,number}$$

2) Precision: In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the find:

$$Precsion = \frac{TruePositive}{TruePositive + FalsePositive}$$

Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n.

3) Recall: Recall in information retrieval is the fraction of the documents that

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Table 2 Comparisons of perfomance measure for different Algortihms

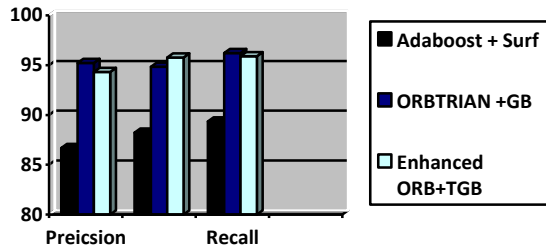| Classifier	Accuracy	Precision	Recall
Adaboost + Surf	88.22	91.67	95.30
ORB + Gradient Boost	94.83	95.21	97.2
tSNE + Enhanced GB	94.36	97.56	95.34

Fig 5: Graph diagram for comparison of results of 3 experiments of our proposed algorithm

Table 3: Comparison of classification techniques for Tree based Gradient Boost and Enhanced Tree based Gradient Boost algorithm

GB Algorithm	Tree based GB		t-SNE GB	
Feature points	100	200	100	200
Classification percentage (%)	72.12	76.56	89.40	88.48

4. Conclusions

In this paper, a simple and efficient method that utilizes object feature to detect objects in an image has been improved and experimented with various enhanced algorithms. Unlike recent methods, we obtain our saliency maps by estimating foreground regions in an image instead of using boundary priors. Our method combined with the optimization framework as enhanced Tree based Gradient Boost algorithm produces accurate and proper perform better than other methods in terms of MAE when tested on two widely used datasets. In future, we plan to investigate better cues rather than depending on contrast or boundary prior alone, better connectivity measures and better objectness proposal techniques that can perform well with backgrounds that are even more complex.

This paper proposed a kind of improved extraction algorithm based on ORB. Median filtering method is used to remove noise and using the ORB algorithm to detect feature points, then the hamming distance is adopted to feature points matching. Finally

using the t-SNE algorithm group and obtain the target coordinates in the scene to realize image correction and feature point minimization eventually achieve two images matching using enhanced gradient based boost algorithm. It shows that the calculating speed has been improved when getting more correct matching points. Median filter and edge filter algorithm can find more accurate matching points and improves sharpness of image but in terms of speed and accuracy of the perspective transformation proposed algorithm performs better than SURF and Adaboost based algorithm. The universal applicability of the improved algorithm is verified by two groups of different sampled images of solid steel block with various sets. Furthermore

References

[1] Jhing-Fa WangHan-Jen HsuJyun-Sian Li,"Intelligent Object Extraction Algorithm Based on Foreground/Background Classification", International Conference on Embedded and Ubiquitous Computing, EUC 2005 Workshops pp 101-110.

[2] Ali Saglam, Nurdan Akhan Baykan,"An Efficient Object Extraction with Graph-Based Image Segmentation", 2015, 4th International Conference on Advanced Computer Science Applications and Technologies (ACSAT).

[3] QuyIch Pham, Rudolf Jalovecký, Martin Polášek, "Using template matching for object recognition in infrared video sequences", 2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC).

[4] Alexander Toshev, Ameesh Makadia, Kostas Daniilidis, "Shape-based object recognition in videos using 3D synthetic object models", IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[5] Huiyan Zhao, Xiang Duan, Yongfeng Cao, "Color-Based Object Tracking Using Fast Survival-of-the-Fittest Method", Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007.

[6] Online resources for ORB algorithm: http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_orb/py_orb.html

[7] J.-H. Cho and S.-D. Kim,"Object detection using spatio-temporal thresholding in image sequences",ELECTRONICS LETTERS 2nd September 2004 Vol. 40 No. 18

[8] Chen Guodong, Zeyang Xia, Rongchuan Sun, Zhenhua Wang, Zhiwu Ren and Lining Sun,"A Learning Algorithm for Model based Object Detection",The 8th International Conference on Ubiquitous Robots and Ambient Intelligence,2011

[9] Jian Hou, Member, IEEE, Huijun Gao, Fellow, IEEE, Qi Xia, and Naiming Qi,"Feature Combination and the kNN Framework in Object Classification",IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 27, NO. 6, JUNE 2016

[10]   David G.Lowe "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol.60, no.2, 2004, pp.91-110

[11]   Lei Yu, Zhixin Yu and Yan Gong,"AnImproved ORB Algorithm of Extracting and Matching Features",International Journal of Signal Processing, Image

Processing and Pattern Recognition Vol. 8, No. 5 (2015), pp. 117-126

[12] Meng Fanqing , You Fucheng,"A tracking algorithm based on ORB", Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)

[13] Dr. Shubhangi D. C, Sabahat Fatima,"Feature Extraction of Images Using Two Different Descriptors and Their Differences", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, May2017

[14]   Geng Du*, Fei Su, Anni Cai,"Face recognition using SURF features" ,MIPPR 2009: Pattern Recognition and Computer Vision

[15] Massimo Bertozzi, Luca Castangia, Stefano Cattani, Antonio Prioletti, Pietro Versar,"360 degree detection and tracking algorithm of both pedestrian and vehicle using fisheye images",Intelligent Vehicles Symposium (IV), 2015 IEEE

[16] Liao Shaowen, Zhang Jianqing, Chen Yong,"Further Improvement of AdaBoost Algorithm",2015 Seventh International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)

[17] "t-distributed stochastic neighbor embedding", Online resource at https://en.wikipedia.org/wiki/T distributed_stochastic_neighbor_embedding.

Radhamadhab Dalai is a PhD scholar in the department of computer science and engineering in BIT Mesra, Ranchi, India .His interest areas of research are pattern recognition, image analysis, computer vision and robotics, machine learning

 Prof. Kishore Kumar Senapati is an professor in the the department of computer science and engineering in BIT Mesra, Ranchi,India.His interest areas of research are pattern recognition, computer algorithms, Bio-computing.