Unsupervised Learning of video image model for object recognition using Background subtraction with AND-OR Template

[1]Saravanan.S, [2]Malathi.D, [3]Shanmugam.K, [4]Dr.Vanathi.B
[1]PG scholar, [2,4]Professor,[3] Assistant professor
[1,2]SRM University,Chennai,India
[3,4]Valliammai Engineering College,India

***Abstract* - The framework for unsupervised learning of a hierarchical reconfigurable image template—the AND-OR Template (AOT) for visual objects has been studied and our proposed system consists of AOT template with back ground subtraction method in video analysis by using low rank detection algorithm. The background subtraction technique performs detecting moving objects from the current frame and the reference frame. The AOT includes hierarchical composition as "AND" nodes, deformation and articulation of**
parts as geometric "OR" nodes, and multiple ways of composition as structural "OR" nodes. The terminal nodes are hybrid image templates (HIT) that are fully generative to the pixels.

***Keywords* - AND-OR Template, background subtraction, hybrid image template (HIT), Image Content Analysis, low rank detection.**

I. INTRODUCTION

Image/Video Analytics is enabling a rapidly growing number of embedded video products such as smart cameras and intelligent Digital Video Recorders (DVRs) with automated capabilities that just a few years ago would have required human monitoring. Broadly, Image analytics is the extraction of meaningful and relevant information from the image. Image content Analytic(ICA)s refers the builds upon research in computer vision, pattern analysis and machine intelligence, and spans several industry segments including surveillance, retail and transportation.

Similar to human vision, which has a perceptual and cognitive aspect, image analytics uses computer vision algorithms which enable it to perceive or see, and machine intelligence to interpret, learn and draw inferences. The goal of image analytics is good understanding, which differs from motion detection. In addition to detecting motion, analytics qualifies the motion as an object, understands the context around the object, and is able to track the object through the frame.

Image analysis is the process of extracting meaningful information from images. Image analysis can include such tasks as finding shapes, detecting edges, counting objects, or measuring properties of an object.

Common image analysis algorithms include edge detection, shape detectors, color-based segmentation, and image Thresholding. By combining these common image processing techniques with region analysis functions, detailed statistics can be obtained from images to provide human analysts with additional quantitative and qualitative data. For example, the video contains 50images by using low rank detection method subtract the background and the foreground image

Background subtraction Technique

In [8], the author have to maximise speed and limiting the memory requirements, to more sophisticated approaches, proposed to achieve the highest possible accuracy under any possible circumstances. All approaches aim, however, at real-time performance, hence a lower bound on speed always exists.

- Running Gaussian average
- Mixture of Gaussians
- Kernel density estimation (KDE)
- Sequential KD approximation

Running Gaussian average

The method proposed to model the background independently at each ***(i,j)*** pixel location. The model is based on ideally fitting a Gaussian probability density function (pdf) on the last ***n*** pixel's values. In order to avoid fitting the pdf from scratch at each new frame time, t, a running (or on-line cumulative) average is computed instead as:

$$\mu_t = \alpha I_1 + (1-\alpha)\mu_{t-1} \qquad (1)$$

where *I* is the pixel's current value and μ_t the previous average, α is an empirical weight often chosen as a trade off between stability and quick update.

Mixture of Gaussians

Over time, different background objects are likely to appear at a same *(i,j)* pixel location. When this is due to a permanent change in the scene's geometry, all the models reviewed so far will, more or less promptly, adapt so as to reflect the value of the current background object. However, sometimes the changes in the background object are not permanent and appear at a rate faster than that of the background update. A typical example is that of an outdoor scene with trees partially covering a building: a same *(i,j)* pixel location will show values from tree leaves, tree branches, and the building itself. Other examples can be easily drawn from snowing, raining, or watching sea waves from a beach. In these cases, a single valued background is not an adequate model. A model of the background alone for a criterion as given in eqn (2) is required to provide discrimination between the foreground and background distributions. The probability of observing a certain pixel value, **x,** at time t by means of a mixture of Gaussians is given by

$$P_{(xi)} = \sum_{i=1}^{k} \omega_{i1t}\, \eta(x_t - \mu_{i,t}, \sum_{i,j}) \qquad (2)$$

where *K* is Gaussian distributions deemed to describe only one of the observable background or foreground objects.

Kernel Density Estimation

An approximation of the background *pdf* can be given by the histogram of the most recent values classified as background values. However, as the number of samples is necessarily limited, such an approximation suffers from significant drawbacks: the histogram, as a step function, might provide poor modelling of the time, unknown *pdf*, with the "tails" of the true pdf often missing. The background distribution by a non-parametric model based on Kernel Density Estimation (KDE) on the buffer of the last *n* background values. KDE guarantees a smoothed, continuous version of the histogram. In eqn (3), the background pdf is given as a sum of Gaussian kernels centered in the most recent *n* background values, xi:

$$P(x_i) = \frac{1}{n}\sum_{i=1}^{n} n\,(x_t - x_i \sum t \qquad (3)$$

Sequential Kernel Density approximation

Mean-shift vector techniques have recently been employed for various pattern recognition problems such as image segmentation and tracking. The mean shift vector is an effective gradient-ascent technique able to detect the main modes of the time pdf directly from the sample data with a minimum set of assumptions. However, it has a very high computational cost since it is an iterative technique and it requires a study of convergence over the whole data space. As such, it is not immediately applicable to modeling background pdfs at the pixel level.

II. RELATED WORK

Learning Hybrid Image Templates (HIT)

Zhangzhang Si and Song-Chun Zhu[3] proposed hybrid image template of information projection. Each template is composed of a number of image patches (typically 50, 100) whose geometric attributes (location, scale, orientation) may adapt in a local neighborhood to account for deformations and variations, and whose appearances are characterized, respectively, by four types of descriptors: local sketch (edge or bar), texture gradients (with orientation field), flatness regions (smooth surface and lighting), and colors.

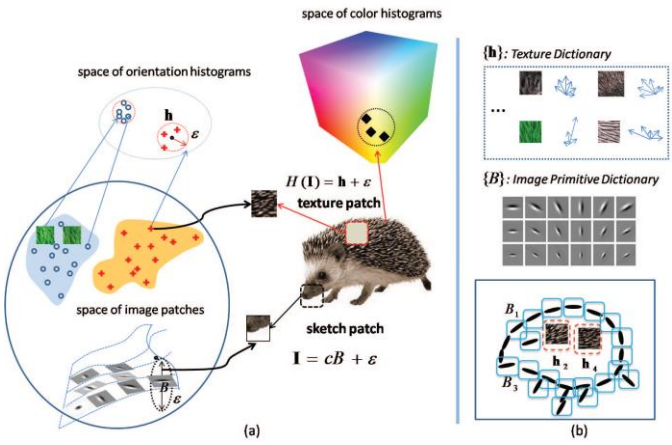

Figure 1.Analysing of image composition

Naturally, there are large variations in the representations of different classes, for example, teapots may have common shape outline, but do not have common texture or color, the animal hedgehog in Fig. 1 has distinct texture and shape, but its color is often less distinguishable from its background. So, the essence of our learning framework is to automatically select, in a principled way, informative patches from a large pool and compose them into a template with well-normalized probability model.

Object Detection with Discriminatively Trained Part-Based Models:

Pedro F. Felzenszwalb, Ross B. Girshick, David [4] proposed

Object detection system based on mixtures of multiscale deformable part models. Their system is able to represent highly variable object classes and achieves state-of-the-art results in the PASCAL object detection challenges. While deformable part models have become quite popular, their value had not been demonstrated on difficult benchmarks such as the PASCAL data sets. They system relies on new methods for discriminative training with partially labeled data. They combine a margin sensitive approach for data-mining hard negative examples with a formalism we call latent SVM.

Pictorial Structures for Object Recognition

Amir Sadovnik and Tsuhan Chen [2] proposed structure for object recognition. Sketch recognition and computer vision algorithms attempt to solve similar problems in different domains. For example, the tasks of object detection and localization in computer vision are closely related to the task of interpreting strokes as certain objects in drawings. In both, the goal is to take a 2D image and identify the existence and position of a previously learned object. However, many sketch recognition algorithms do not try to utilize the advancements made in the computer vision field on drawings. This is mainly because drawings are analyzed using ink data, and therefore there has been a lot of focus on developing algorithms which take advantage of this representation of the image.

Image Enhancement by Fusion in Courtourlet Transform

Melkamu H. Asmare, Vijanth S. Asirvadam [3] proposed by fusion countourlet transform.The composite image approach is proposed for enhancing still images. For image enhancement, one needs to improve the visual quality of an image with minimal image distortion. Contourlet Transform has better performance in representing the image salient features such as edges, lines, curves and contours than wavelet transform. It is therefore well-suited for multi- scale edge based color image enhancement.

Animal Detection Using Template Matching Algorithm

Traditional system requires a person who can view the system whole day. Animal detection by human eyes has been considered as the most reliable detection method if seen from the computational point of view. This is because the image structure in natural images is complex. In, it is found that a human observer is able to decide whether a briefly flashed animal scene contain an animal as fast as 150ms. In, median reaction time results indicate a speed accuracy of 92% for reaction time of 390ms and increase to 97% of correctness for 570ms. Though human detection is effective and achieve satisfactory level, human eyes can easily guttered causing decreasing of effectiveness. Furthermore, human eyes cannot work 24 hours a day to perform animal detection. These flaws can be curbed by applying computer vision in image processing for animal Detection.

Intelligent Video Surveillance System

In IVS, there are basically six components. These components are listed below.

Acquisition: This component is basically used for acquiring the images. There is a whole array of camera models to meet different monitoring needs. They are analogue and digital, and can be power-operated or not. Solar cameras are also being useful in many applications.

Transmission: The video captured by surveillance cam-eras must be sent to the recording, processing and viewing systems. This transmission can be done by cable (coaxial or fiber optic cables, stranded copper wire) or by air (infrared signals, radio transmission).

Compression: Digitized video represents a large quantity of data to be transmitted and archived. So, surveillance video must be compressed using codec, algorithms for reducing the amount of data by deleting redundancies, by image or between footage frames, as well as details that cannot be seen by a human eye.

Processing: Video management systems process video surveillance images, such as managing different video flows, and viewing, recording, analyzing and searching recorded footage. There are four major categories of video management systems, Digital Video Recorder (DVR), Hybrid Digital Video Recorder (HDVR), Network Video Recorder (NVR), IP video surveillance software.

Archiving: The video footage archiving period varies depending on surveillance needs, ranging from a few days to a few years. There are two types of archiving devices, internal and attached.

Display: Video surveillance can be viewed on different devices. In small facilities, the video can be viewed directly on the recorder, as the image is being recorded. Images are generally viewed remotely, on a computer, or on a mobile device such as a telephone or hand held device.

III. BACKGROUND SUBTRACTION AND AOT SYSTEM

Background subtraction is a computational vision process of extracting foreground objects in a particular scene. A foreground object can be described as an object of attention which helps in reducing the amount of data to be processed as well as provide important information to the task under consideration. Often, the foreground object can be thought of as a coherently moving object in a scene. We must emphasize the word coherent here because if a person is walking in front of moving leaves, the person forms the foreground object while leaves though having motion associated with them are considered background due to its repetitive behavior.

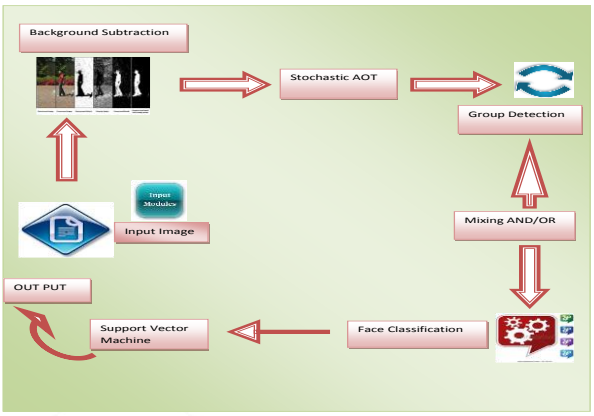

Figure 5: Architecture Diagram of Background Subtraction and AOT Model

In some cases, distance of the moving object also forms a basis for it to be considered a background, e.g. if in a scene one person is close to the camera while there is a person far away in background, in this case the nearby person is considered as foreground while the person far away is ignored due to its small size and the lack of information that it provides. Identifying moving objects from a video sequence is a fundamental and critical task in many computer-vision applications. A common approach is to perform background subtraction, which identifies moving objects from the portion of video frame that differs from the background model.

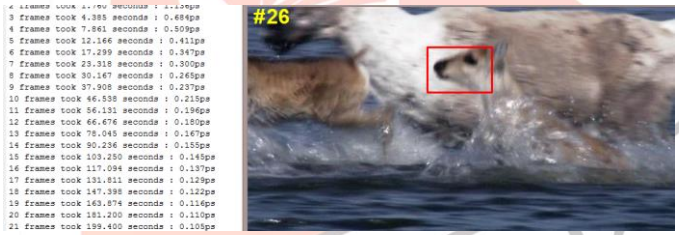

Figure 6: Object Tracking by frame difference (26th Frame)

Background subtraction is a class of techniques for segmenting out objects of interest in a scene for applications such as surveillance. There are many challenges in developing a good background subtraction algorithm. First, it must be robust against changes in illumination. Second, it should avoid detecting non-stationary background objects and shadows cast by moving objects. A good background model should also react quickly to changes in background and adapt itself to accommodate changes occurring in the background such as moving of a stationary chair from one place to another. It should also have a good foreground detection rate and the processing time for background subtraction should be real-time.

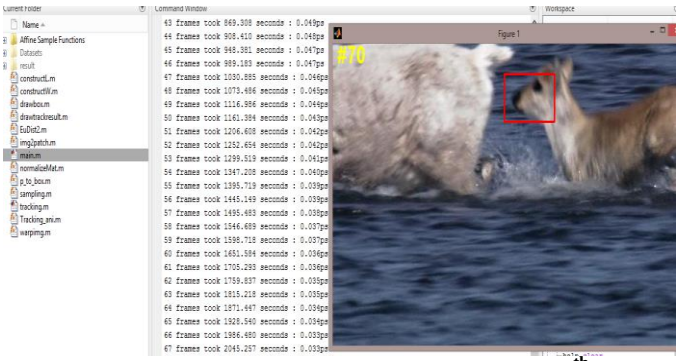

Figure 7: Object Tracking by frame difference (70th Frame)

The purpose of our work is to obtain a real-time system which works well in indoor workspace kind of environment and is independent of camera placements, reflection, illumination, shadows, opening of doors and other similar scenarios which lead to

errors in foreground extraction. The system should be robust to whatever it is presented with in its field of vision and should be able to cope with all the factors contributing to erroneousresults.

Figure 8: Object Tracking by frame difference (31st Frame)

Much work has been done towards obtaining the best possible background model which works in real time. Most primitive of these algorithms would be to use a static frame without any foreground object as a base background model and use a simple threshold based frame subtraction to obtain the foreground. This is not suited for real life situations where normally there is a lot of movement through cluttered areas, objects overlapping in the visual field, shadows, lighting changes, effects of moving elements in the scene (e.g. swaying trees), slow-moving objects, and objects being introduced or removed from the scene.
This Process can be done by

- Running Gaussian average
- Mixture of Gaussian
- Kernel Density Estimation
- Sequential KD approximation

AOT: Reconfigurable Object Templates

Compositional hierarchy used in object modelling with shared parts among categories. The OR nodes for structural variations are largely omitted or oversimplified.The method is that they do not build on a fully generative model to the level of pixels. Local geometry and appearance as a pooled histogram, but detailed object deformation is discarded during computation of histograms. It is capable of modelling a certain amount of object articulation, but the localization of object boundaries is imprecise because of the local histogram pooling in computing the HoG feature.

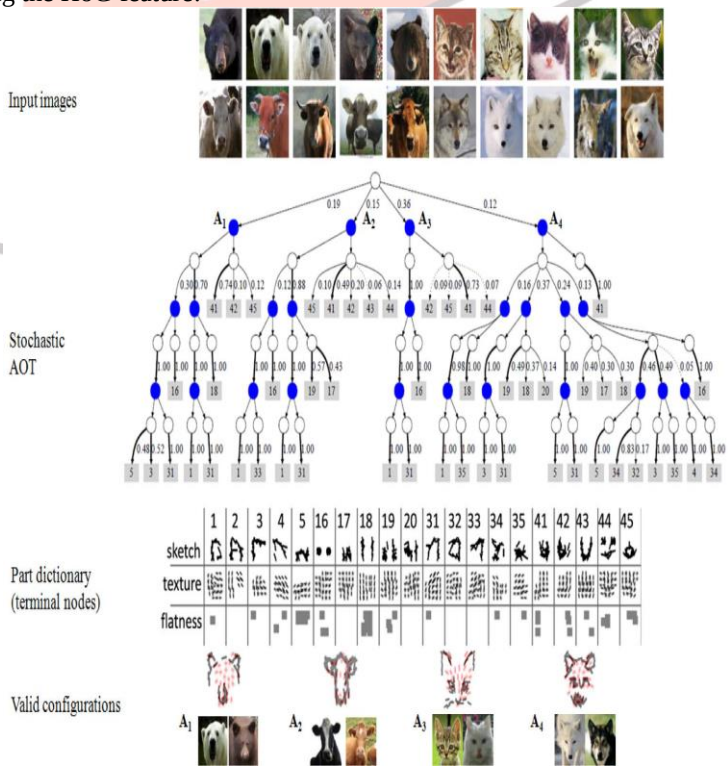

Figure 9 :AND-OR Template(AOT) learned by 200 animal faces of divided into four categories.Empty circle denoted by OR nodes.Shaded circle denoted by AND nodes which are combination of terminal nodes. Shaded rectangles are terminal nodes, which are Hybrid Image Template(HIT) for part appearances.

CONCLUSION

In this paper, we have proposed a background subtraction method of video analysis, which subtract foreground image and detailed ground from the video. We also propose a hierarchical reconfigurable object template, called the AOT model, which can capture rich structural variations and shape variations of objects. We have developed an unsupervised learning algorithm for learning AOTs from only images and no manual labelling. In our learning algorithm, the pervasive ambiguity of parts is overcome by 1) articulation of parts, 2) alternative compositions, both of which imply the importance of OR nodes. This is a major contribution of our work. We investigate the factors that Influence how well the learning algorithm can identify the underlying AOT, and we design a number of ways to evaluate the performance of the proposed learning algorithm through both synthesized examples and real-world images. The proposed AOT model achieves good performance on par with state-of-the-art systems in public benchmarks for object detection. Moreover, the AOT model has the advantage of significantly less training time and fewer parameters. In future work, we will allow the parts of visual objects to be no square regions and search for the optimal part segmentation with flexible shape decompositions.

REFERENCES

1. Zhangzhang Si and Song-Chun Zhu, " Learning Hybrid Image Templates (HIT) by Information Projection ", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 34, no. 7, JULY 2012.

2. Pedro F. Felzenszwalb, Ross B. Girshick,David McAllester, and Deva Ramanan, " Object Detection with Discriminatively Trained Part-Based Models ", International Journal of Computer Vision, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, September 2010.

3. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005.
   P.F.Felzenszwalb and D.P. Hutten ocher, "Pictorial Structures for Object Recognition," Int. Journal of Computer Vision, vol. 61, no. 1, pp. 55- 79, 2005.

4. T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 23, no. 6, pp. 681-685, June 2001.

5. Y.N. Wu, Z. Si, H. Gong, and S.-C. Zhu, "Learning Active Basis Model for Object Detection and Recognition," Int'l J. Computer Vision, vol. 90, no. 2, pp. 198-230, 2010.

6. Zhi-Hua Chenl, Xiao-Long Xiao, "Graph-Based Image Segmentation With Bag-Of-Pixels," Proceedings of the 2013 International Conference on Machine Learning and Cybernetics, Tianjin, 14-17 July, 2013

7. Massimo Piccardi, "Background subtraction techniques: a review" IEEE International Conference on Systems Man and Cybernetics, 2004

8. Mansi Parikh, Miral Patel, "Animal Detection Using Template Matching Algorithm". International Journal of Research in Modern Engineering and Emerging Technology, Vol. 1, Issue: 3, April 2013

9. C.Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," Proc. IEEE CVPR 1999, pp. 24&252, June 1999.