

Colour Object Detection in Video Sequence Using Open Computer Vision

M.Nithiyaa ¹, P.S.Sudarshana ²

Assistant Professor, Department of ECE, Knowledge Institute of Technology, Salem, Tamilnadu, India¹

Assistant Professor, Department of ECE, Knowledge Institute of Technology, Salem, Tamilnadu, India²

ABSTRACT: Object detection from a video is the vital and challenging fundamental mission of computer vision. It is a critical part in many applications like image search and scene understanding. It is still an open issue due to the variety and complexity of object classes and backgrounds. Colour based methods is the easiest way to detect an object from an image. For that the object and the background should have noteworthy colour dissimilarity in order to successfully segment the objects. An approach to automatically detect and track multiple colour objects in motion from a single camera is proposed. An object is represented as an assembly of the colour frames based on shape and dimension. Colour image is tracked by open CV (Computer Vision) source algorithm. This system can track objects with inter-object either static or non-static background. Evaluation results are based on a number of images and videos.

KEYWORDS: Object detection, open CV (Computer Vision)

I. INTRODUCTION

Detection and tracking of humans or objects is important for many applications, such as visual surveillance, human computer interaction, and driving assistance systems. For this task, we need to detect the objects of interest first (i.e., find the image regions corresponding to the objects) and then track them across different frames while maintaining the correct identities. In a video sequence an object is said to be in motion, if it is changing its location with respect to its background. The image is initially recorded by the camera and it tracks continuously the whole human body of a single person. The principle used in this method is using Gaussian transform and Background subtraction for converting the images into frames. Both range and textual information can be captured by the camera, and are perfectly aligned to each other. Though objects in the range image are still occluding each other, it is possible to segment these, by analysis of the distances between the persons present in the scene. The proposed work is to detect the object based on the dissimilarities of the colour in an image using the concept of background subtraction.

II. LITERATURE SURVEY

[1]The video is transferred to the computer connected in the network. The open CV software runs in Microsoft visual C++ studio and the face detection algorithm detect all the faces and it is finally recognized. The proposed method used in this is using open visual studio. In this, the output is continuous without any delay, when compared with other software the processing is very easy. But the output of the system is of three stages. The first step is original response of the detector, the second step is merged detector response which takes more amount of space and time. Even though its software processing is simple but it takes more time for processing.

[2] The Gaussian parameters are evaluated and updated using a simple heuristic to hypothesize which are most likely to be part of the "background process". Pixel values that do not match with the pixel's "background" are grouped using connected components. This is extremely useful when different regions have different lighting, because objects which appear in shaded regions do not generally exhibit as much noise as objects in lighted regions. Huge number of clusters makes their methods computationally expensive. Tracking and its recovery will be impossible. Non-target points can only translate in 2D. It cannot deal with the target object with changes in appearance. Computational cost is proportional to the number of non-target points. When any input devices are not responding continuously, then this method is not applicable because it will not respond properly.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

[3] The integration of 3D sensors, in particular a photonic mixture device, into current frameworks, in order to support the reliable detection and segmentation in dense situations. Reliable foreground segmentation and the computation of depth gradients. The proposed method is that the thresholding based on object size is not applicable, objects are almost entirely occluded would also be removed, and another method had to be found. The PMD camera is a systematic error due to the demodulation of the correlation function. This error will create a problem while executing the output with the image.

[4] This approach is automatically detect and track multiple objects, possibly partially occluded humans in a walking or standing pose from a single camera, which may be stationary or moving. Part detectors are learned by boosting a number of weak classifiers which are based on edge let features. In this method, the issue is the edge let which suffers from the narrow colour distribution in the colour histogram of the monochromatic target. Small changes in illumination - histogram will cause dramatic changes. In order to detect body parts at different scales, the input image is re-sampled to build a scale pyramid with a scale factor of 1.2 is needed.

[5] Inpainting algorithm fails in filling occlusion when the occluded region is large since there is loss of both structure and texture. The image is decomposed into texture and structure images using decomposition method based on sparseness of the image. The sparsed reconstruction of the decomposed results in an inpainted image is with all instant made structure. The structure synthesis is performed with texture on image. Finally the texture and structure images are combined together to get an image where the occlusion is filled. Even though this method is having more advantages, the performance is less due to mesh error during separating and combining the images. [6] An occlusion is the region between two overlapping objects with disparate manner. Detecting these occluded objects is crucial for many video sequence. The occlusion detection is decomposed into two independent sub-problems. The first is to detect foreground objects on a frame wise basis, by labelling each pixel in an image frame as either foreground and background. The second is to couple the object. Observation at different points in an sequence will yield the objects motion trajectory and occlusion.

III.METHODS DISCUSSION

3.1 Object Detection:

It is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class in digital images and videos. Initially the camera will record the video and the image will be lively recorded or the stored images will be tracked. A well-researched domain of object detection includes human detection and tracking. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance.

3.2 Object Tracking

It means track an object over a sequence of image. In a video sequence an object is said to be in motion, if it is changing its location with respect to its background. The motion tracking is actually the process of keeping tracks of that moving object in video sequence. Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging and video editing. Video tracking can be a time consuming process due to the amount of data that is contained in video.

3.3 Face detection

Face detection is a computer technology that determines the locations and sizes of human faces in digital images. It detects face and ignores anything else, such as buildings, trees and bodies. Face detection can be regarded as a more general case of face localization. In face localization, the task is to find the locations and sizes of a known number of faces (usually one). In face detection, face is processed and matched bitwise with the underlying face image in the database. Occlusion is anything that obstructs vision. Removing occlusion is very important for better perception and object recognition. Occlusion is a big challenge in many of the computer vision algorithms.. There are situations where unexpected and unappreciated people stand in front of historic places and monuments thus reducing the historic beauty of the picture.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

IV.METHODOLOGY

The Video camera sends stream of frames to the system. The video stream will be around 20 (frames/sec)-30(frames/sec). The video is transferred to the computer connected in the network. The OpenCV software that runs in Microsoft Visual C++ studio converts the video into image frames. The Object Detection Algorithm running in Visual C++ detects the objects. The entire process is done by Foreground subtraction from the background image. Then the objects are stored in the database. If the image is given as input in Visual C++, the images will be tracked using pixel wise tracking method. The complete information is stored in the memory hence the information can be retrieved at any time. Fig. 1 shows the block diagram of the proposed methodology.

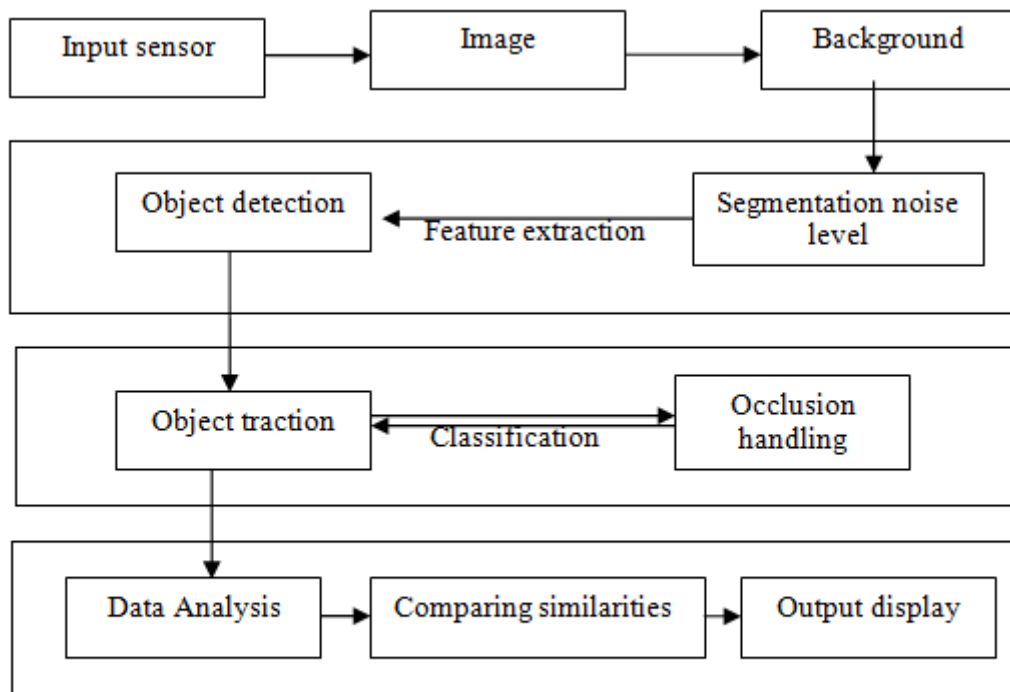


Fig. 1 Block diagram of the proposed method

V.OPEN COMPUTER VISION

OpenCV (Open Source Computer Vision) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. This library has more than 2500 optimized algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, etc. It is Developed by Intel Russia research center in 2000 and free for both commercial and non-commercial use. The Supported platforms are Windows, Android, +Maemo, FreeBSD, OpenBSD, iOS, Linux and OS X.

The user can get official releases from Source Forge or take the current snapshot under SVN from there. OpenCV uses CMake. OpenCV is written in C++ and its primary interface is in C++, but it still retains a less comprehensive though extensive older C interface. There are now full interfaces in Python, Java and MATLAB/OCTAVE (as of version 2.5). The API for these interfaces can be found in the online documentation. Wrappers in other languages such as C#, Perl, Ch, and Ruby have been developed to encourage adoption by a wider audience.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

VI. PROPOSED WORK

6.1 OBJECT REPRESENTATION

In a tracking scenario, an object can be defined as anything that is of interest for further analysis. For instance, boats on the sea, fish inside an aquarium, vehicles on a road, planes in the air, people walking on a road, or bubbles in the water are a set of objects that may be important to track in a specific domain. Objects can be represented by their shapes and appearances. In this section, object shape representations for tracking will describe first and then address the joint shape and appearance representations.

Object silhouette and contour:

Contour representation defines the boundary of an object. The region inside the contour is called the silhouette of the object. Silhouette and contour representations are suitable for tracking complex non rigid shapes. Object representation using silhouette method is shown in Fig.2.

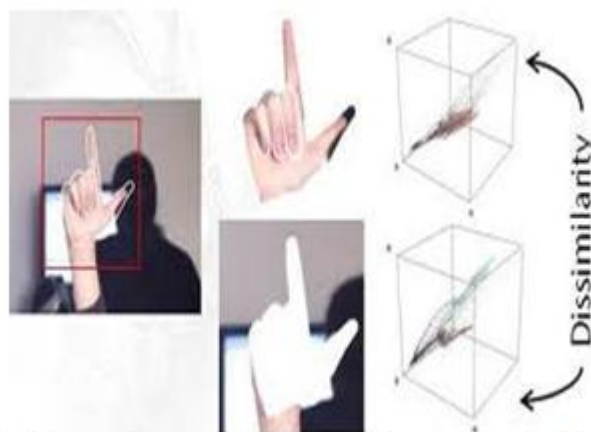


Fig. 2 object representation using silhouette method

6.2.OBJECT DETECTION

Every tracking method requires an object detection mechanism(Shown in Fig.3) either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections. This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracker's task to perform object correspondence from one frame to the next to generate the tracks.



Fig.3 Object Detection

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

Point Detectors

Point detectors are used to find interest points in images which have an expressive texture in their respective localities. Interest points have been long used in the context of motion, stereo, and tracking problems. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint.

Video Analysis

- System for semantic understanding of visual data.
- Intelligent video surveillance.

Typical Task

- Object Identification and Detection
- Tracking and Reorganizing
- Scene Monitoring
- Event Analysis



Fig.4 object tracking

Background Subtraction

Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are marked for further processing. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the background subtraction.(Shown in Fig.5 and 6)

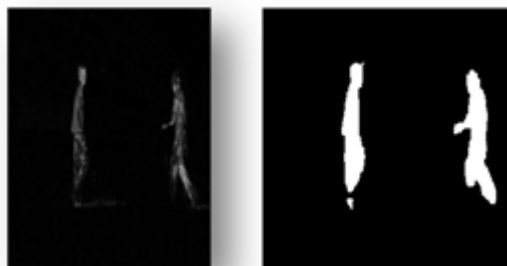


Fig.5 (a)Background Subtracted image (b)BW image

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015



Fig.6 (a)Video frame (b)Background Subtraction

6.3.OBJECT TRACKING

Object tracking (shown in Fig.7)is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behaviour. Therefore, the use of object tracking is pertinent in all tasks.



Fig.7 Object tracking using region based technique

6.4. HAAR TRANSFORM

The Haar wavelet is a sequence of rescaled "square-shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal function basis. The Haar sequence is recognised as the first known wavelet basis and extensively used as a teaching example.

Haar used these functions to give an example of an orthonormal system for the space of square-integrable functions on the unit interval $[0, 1]$. The study of wavelets, and even the term "wavelet", did not come until much later. As a special case of the Daubechies wavelet, the Haar wavelet is also known as D2.The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines. Fig.8 shows the Haar transform sample graph.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

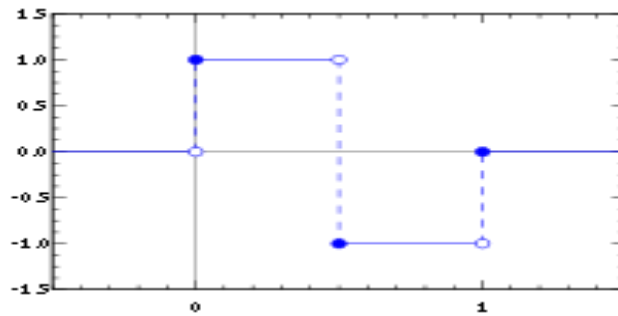


Fig.8 Haar Transform Sample Graph

The Haar wavelet's mother wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{-----(1)}$$

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{-----(2)}$$

Its scaling function $\phi(t)$ can be described as

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{-----(3)}$$

Haar-like features are used in object recognition and were used in the first real-time face detector. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. For example, let us say we have an image database with human faces. It is a common observation that among all faces the region of the eyes is darker than the region of the cheeks. Therefore a common haar feature for face detection is a set of two adjacent rectangles that lie above the eye and the cheek region. The position of these rectangles is defined relative to a detection window that acts like a bounding box to the target object.

6.5. USE OF HAAR TRANSFORM IN IMAGE

Modern cameras are capable of producing images with resolutions in the range of tens of megapixels. These images need to be compressed before storage and transfer. The Haar transform can be used for image compression. The basic idea is to transfer the image into a matrix in which each element of the matrix represents a pixel in the image. For example, a 256×256 matrix is saved for a 256×256 image. JPEG image compression involves cutting the original image into 8×8 sub-images. Each sub-image is an 8×8 matrix.

The 2-D Haar transform is required. The equation of the Haar

transform is $B_n = H_n A_n H_n^T$, where A_n is a n×n matrix and H_n is n-point Haar transform. The inverse Haar transform is

$$A_n = H_n^T B_n H_n$$

Each pixel's new value is set to a weighted average of that pixel's neighborhood. The original pixel's value receives the heaviest weight (having the highest Gaussian value) and neighboring pixels receive smaller weights as their distance to the original pixel increases. This results in a blur that preserves boundaries and edges better than other.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Special Issue 6, May 2015

VII. RESULTS AND DISCUSSIONS

The OpenCV library can be used to implement major algorithms. By employing OpenCV an object can be detected and tracked based on colours. By changing HSV values (hue, saturation, values), we can track the specific colour of the objects. There is a different HSV value for each and every colour, by fixing those values the objects can be tracked. Initially the values is set to zero then it is to be changed for particular object.

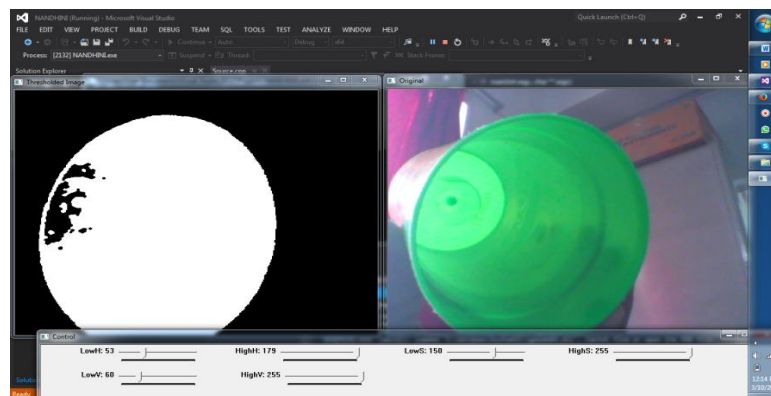


Fig.9 Detection of green coloured object

. The output will be in the form of zeros and for an instance, in a video sequence the consistence of the particular object is tracked. The errors which are appearing in the inputs are removed by Gaussian filters. Fig.9 shows the detection of colour(green) based object. As soon as the object is detected, the output screen is appeared with the black and white colour, white colour represents the colour which we had tracked and the black colour represents remaining colour, in order to differentiate with the output. The future work involves tracking of multiple colour objects in motion, which will be very useful in identification purpose.

REFERENCES

- [1] Vasuhi. S.and Vaidehi V. "Target Detection and Tracking for Video Surveillance", WSEAS Transactions on Signal Processing Volume 10, pp. 168-177, Apr 2014.
- [2] B.Heisele "Motion-based Object Detection and Tracking in Color Image Sequence," ACCV, 2000
- [3] Dejan Arsic, Bjorn Schuller, Benedikt Hornler and Gerhard Rigoll "Resolving Partial Occlusions In Crowded Environments Utilizing Range Data And Video Cameras", 16th International Conference on Digital Signal Processing, July 2009.
- [4] Daniel Olmeda, Arturo de la Escalera and Jos M Armingol "Detection and Tracking of Pedestrians in Infrared Images" International Conference on Signals, Circuits and Systems, 2009.
- [5] Bincy Antony M and Narayanankutty K A "Removing Occlusion In Images Using Sparse Processing And Texture Synthesis", International Journal of Computer Science , Engineering and Applications (IJCSA) Vol.2, No.3, pp 117-124, June 2012.
- [6] Kapil.A. Chavan and P.P.Halkarnikar "Occlusion detection in video sequences", IOSR Journal of Computer Engineering (IOSR -JCE), Volume 16, Issue 5, Ver. 5, Oct. 2014
- [7] D. Peters, D. Raskovic, and D. Thorsen, "An Energy Efficient Parallel Embedded System for Small Satellite Applications", *ISAST Transactions on Computers and Intelligent Systems*, vol.1(2), 2009.
- [8] Xu Xiaoli, Zuo Yunbo, "Integrated Design of Measurement and Control System for Photovoltaic Generation" Proceedings iCECE 2010, IEEE PRESS- 2010.
- [9] Urvashi Bakshi, Rohit Singhal, "A Survey on Face Detection Methods and Feature Extraction Techniques of Face Recognition" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 3, May-June 2014.
- [10] Wang Peizhen, "Research on Fault Analysis of Photovoltaic Array" Dissertation for the Degree of Doctor of Engineering. Hefei University of Technology, 2005.
- [11] Texas Instrument – Product and datasheet for the wireless transceiver cc1120: <http://www.ti.com/product/CC1120/>. Archived from the original on 5 May 2011.
- [12] Brendan Tran Morris and Mohan Manubhai Trivedi (AUGUST 2008) 'Ieee Transactions on Circuits and Systems for Video Technology, Vol. 18, No. 8'.
- [13] I-Hsien Chen and Sheng-Jyh Wang Hsinchu, Taiwan, *R.O.C. (ICVS 2006)Proceedings of the Fourth IEEE International Conference on Computer Vision* .