Utility Person Detection and Multi-View Video Tracking Annotation Model

[1]M.Senbagapriya, [2]Dr. P.Sumitra

[1]*M.Phil Full Time Research Scholar, PG and Research Department of Computer Science Vivekananda College of Arts and Sciences for Women (Autonomous) Elayamapalayam.*

[2]*Assistant Professor, PG and Research Department of Computer Science. Vivekananda College of Arts and Sciences for Women (Autonomous) Elayamapalayam*

Abstract: - In this thesis a generic methodology for the semi-automatic generation of reliable position annotations for evaluating multi-camera people-trackers on large video data sets. Most of the annotation data are automatically computed, by estimating a consensus tracking result from multiple existing trackers and people detectors and classifying it as either reliable or not. A small subset of the data, composed of tracks with insufficient reliability, is verified by a human using a simple binary decision task, a process faster than marking the correct person position. The proposed framework is generic and can handle additional trackers. In this thesis studied the most commonly used face edge detection techniques of Enhanced Sobel Edge Annotation Algorithm (ESEAA). Higher-level edge detection techniques and appropriate programming tools only facilitate the process but do not make it a simple task.

Keywords: Image processing, Digital Image Processing, Analog Image Processing Two dimensional signals

----

1. Introduction

Image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distorti.00on during processing. Since images are defined over two dimensions (perhaps more),

digital image processing may be modeled in the form of multidimensional systems. Image processing is a method to convert an image into digital form and perform some operations on it, to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is an image, like video frame or photograph and output, may be image or characteristics associated with that image. Usually, image processing system includes treating images as two-dimensional signals while applying already set Signal processing methods for them. It is among rapidly growing technologies today, with its applications in various aspects of the business. Image Processing forms core research area within engineering and computer science disciplines too.

Image and Its Types

An image may be well-defined such as a two-dimensional function F(a,b).Where a and b are spatial (plane) coordinate, and the amplitude of F at any pair of coordinates (a, b) is called the intensity or gray level of the

image at that point. When a, b amplitude values of are all predetermined discrete quantity, we will call the image as digital image. A digital image is a collection of a finite number of elements, in which each element has a certain value and location. These items of the digital image are known as image elements, picture elements, pels, and pixels. Pixel is the word mostly used refers to the elements of a digital image.

Types of Digital Images:

Binary: In the binary image, the value of each pixel is either black or white. The image has only two possible values for each pixel either 0 or 1; we need one bit per pixel.

Grayscale: In a grayscale image, each pixel is a shade of gray, which has value normally 0 [black] to 255 [white]. This means that each pixel in this image can be shown by eight bits, the hat is exactly of one byte. Other grayscale ranges can be used, but usually, they are also the power of 2.

True Color or RGB: Each pixel in the RGB image has a particular color; that color in the image is described by the quantity of red, green and blue value in the image. If each of the components has a range from 0–255, this means that this gives a total of 2563 different possible colors values. That means such an image is "stack" of three matrices; that represent the red, green and blue values in the image for each pixel. This way we can say that for every pixel in the RGB image there are corresponding three values.
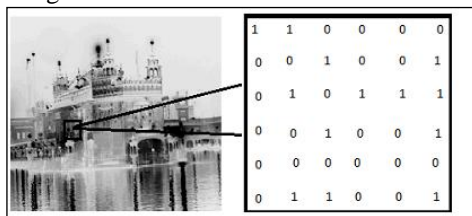

Fig 1.1 Binary Images

- Benefits: Small size image, easily viewable from the internet, Use millions of colors, and perfect for many types of images
- Downfalls: High compression loses quality of image, every time a JPG image is saved, it loses more and more
- quality of the picture.

GIF: GIF stands for Graphics Interchange Format. This format is best suited for text, drawing line screen shots, animations, and cartoons. Gif format is limited to a total number of 256 colors, or it can be less. It is mostly used for loading the fast web pages. It also helps to makes great banner and logo for the different webpage. Different type
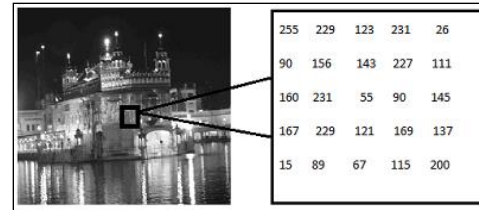

Fig 1.2 Grayscale Images

Indexed: Mostly all the colors images have a subset of more than sixteen million possible colors. For ease of storage and handling of the file, the image has a detailed color map, or we can say the colors palette, that is simply a list of all the colors which can be used in that image. Each pixel has a value associated with it, but it does not give its color as for as we see in an RGB image.

Instead, it provides an index to the color in the map. It is convenient for an image if it has 256 colors or less. The index values will require only one byte to store each. Some image file formats such as GIF which allow 256 colors only.

Digital Image File Types

BMP: Bmp stands for Bitmap. Every picture on a computer appears to be a BMP. In Windows XP the Paint program saves its images automatically in bitmap format, however in Windows Vista images are saved now into JPEG format. The bitmap is the basis platform for many other file types.

- Benefits: High-quality image, Easy to change and edit, No loss in image through process
- Downfalls: Difficulty while displayable on the internet and large in file size.

JPG, JPEG: JPEG stands for Joint Photographic Experts Group .jpeg format is mainly used for color photographs. It is not good with sharp edges, and it tends to blur the image a bit. This format became trendy with the innovation of the digital camera. Digital cameras mostly download photos to our computer as a Jpeg format. Digital camera manufacturers obviously see the value in high-quality images that eventually take up less space. of animated pictures is saved in GIF format. For example, the flashing banner would be saved as a Gif file format.

Benefits: It is supported mostly by all web browsers, it is minimal file size, Easy to load, Benefit for Transparencies, and animations and Image maps

Downfalls: We can use only basic colors, Complex pictures look horrible, No details of images are allowed.

PNG: PNG stands for Portable Networks Graphic. This is one of the best image formats. Still, it was not always well-suited with all web browsers and image software. This is the best image format to use for the website. It is also used for logos and screen shots.

TIFF: TIFF stands for Tagged Image File Format. This format has not been restructured since 1992 and is now owned by Adobe. It can store an image and data (tag) in the one file. This file is commonly used for scanning the data, faxing, word processing, etc. It is no common file format that can be used with our digital photos.

- Benefits: The image is perfect, Never loss any image.
- Downfalls: Due to massive file size there is difficulty in transferring of the file, not able to view on the internet, only some specialized program can view it.

Colour mixing model. This is analogous to stacking slide images on top of each other and shining light through them.

Types of operations

The types of operations that can be applied to digital images to transform an input image an[m,n] into an output image b[m,n] (or another representation) can be classified into three categories as shown in Fig 1.8.

Types of neighborhoods

Neighborhood operations play a key role in modern digital image processing. It is therefore important to understand how images can be sampled and how that relates to the various neighborhoods that can be used to process an image.

- Rectangular sampling – In most cases, images are sampled by laying a rectangular grid over an image as illustrated in Figure 1.10. This results in the type of sampling shown in Figure 1.10ab.
- Hexagonal sampling – An alternative sampling scheme is shown in Figure 1.11 and is termed hexagonal sampling.

Both sampling schemes have been studied extensively, and both represent a possible periodic tiling of the continuous image space. The image will restrict our attention, however, to only rectangular sampling as it remains, due to hardware and software considerations, the method of choice.

Local operations produce an output pixel value b[m=mo,n=no] based upon the pixel values in the neighborhood of a[m=mo,n=no]. Some of the most common neighborhoods are the 4-connected neighborhood and the 8-connected neighborhood in the case of rectangular sampling and the 6-connected neighborhood in the case of hexagonal.

Purpose of Image Processing

The purpose of image processing is divided into five groups. They are:

1. Visualization - Observe the objects that are not visible.
2. Image sharpening and restoration - To create a better image.
3. Image retrieval - Seek for the image of interest.
4. Measurement of pattern – Measures various objects in an image.
5. Image Recognition – Distinguish the objects in an image.

2. Literature Survey

2.1 A Comparative Study on Multi-Person Tracking Using Overlapping Cameras

M. C. Liem and D. M. Gavrila

This paper presents a comparative study for tracking multiple persons using cameras with overlapping views. The evaluated methods consist of two batch mode trackers (and one recursive tracker, which integrate appearance cues and temporal information differently. This paper [1] also added our improved version of the recursive tracker.

In, flow optimization is combined with POM. This is extended with an appearance model KSP does tracking by minimizing the flow trough a graph constructed by stacking POMs from a batch of sequential frames. Each POM location is a graph node, connected to its 9 neighbors in the next frame.

2.2 Efficiently Scaling Up Crowd Sourced Video Annotation

Carl Vondrick · Donald Patterson · Deva Ramanan

This paper[2] presents an extensive three-year study on economically annotating video with crowd sourced marketplaces. Our public framework has annotated thousands of real world videos, including massive data sets unprecedented for their size, complexity, and cost. To accomplish this, we designed a state-of-the-art video annotation user interface and demonstrated that, despite common intuition, many contemporary interfaces are sub-optimal. This extensive study has resulted in our release of VTIC (Video Annotation Tool from Irvine, California), an open platform for monetized, high quality, crowd source video labeling.

2.3 A Multi-View Annotation Tool for People Detection Evaluation

Ákos Utasi and Csaba Benedek

This paper [3] introduces a novel multi-view annotation tool for generating 3D ground truth data of the real location of people in the scene. Also, the height of the people can also be adjusted. To achieve precise ground truth data, the user is aided by the video frames of multiple synchronized and calibrated cameras. Finally, the 3D annotation data can b e easily converted to 2D image positions using the available calibration matrices. One key advantage of the proposed technique is that different methods can b e compared against each other, whether they estimate the real world ground position of people or the 2D position of the camera images.

2.4 Is My New Tracker Better Than Yours

Luka Cehovin, Matej Kristan, and Ales Leonardis

The problem of visual tracking evaluation is sporting an abundance of performance measures, which are used by

various authors, and largely suffers from lack of consensus about which measures should be preferred. This is hampering the cross-paper tracker comparison and faster advancement of the field. In this paper [4] provide an overview of the popular measures and performance visualizations and their critical theoretical and experimental analysis. We show that several measures are equivalent from the point of information they provide for tracker comparison and, crucially, that some are more brittle than the others.

2.5 Performance Evaluation Of Multi-Camera Visual Tracking

Lucio Marcenaro, Pietro Morerio Et Al

Main drawbacks in single-camera multi-target visual tracking can be partially removed by increasing the amount of information gathered on the scene, i.e., by adding cameras. By adopting such a multi-camera approach, multiple sensors cooperate for overall scene understanding. However, new issues arise such as data association and data fusion. This work [5] addresses the issue of evaluating the performance of a multi camera tracking algorithm based on Rao-Black realized Monte Carlo data association (RBMCDA) on real data. For this purpose, a new metric based on three performance indexes is developed.

2.6 Challenges of Ground Truth Evaluation of Multi-Target Tracking

Anton Milan, Konrad Schindler, and Stefan Roth

Evaluating multi-target tracking based on ground truth data is a surprisingly challenging task. Erroneous or ambiguous ground truth annotations, numerous evaluation protocols, and the lack of standardized benchmarks make a direct quantitative comparison of different tracking approaches rather difficult. The goal of this paper[6] is to raise awareness of common pitfalls related to objective ground truth evaluation. We investigate the influence of different annotations, evaluation software, and training procedures using several publicly available resources, and point out the limitations of current definitions of evaluation metrics.

2.7 The Visual Object Tracking VOT2013 challenge results

Matej Kristan et al

Visual tracking has attracted significant attention in the last few decades. The recent surge in the number of publications on tracking-related problems has made it almost impossible to follow the developments in the field. One of the reasons is that there is a lack of commonly accepted annotated data-sets and standardized evaluation protocols that would allow objective comparison of different tracking methods. To address this issue, the Visual Object Tracking (VOT) [7] workshop was organized in conjunction with ICCV2013. Researchers from academia, as well as an industry, were invited to participate in the first VOT2013 challenge which aimed at single-object visual trackers that do not apply pre-learned models of object appearance (model-free).

2.8 Active Frame Selection for Label Propagation in Videos

Sudheendra Vijayanarasimhan and Kristen Grauman

Manually segmenting and labeling objects in video sequences is quite tedious, yet such annotations are valuable for learning-based approaches to object and activity recognition. While automatic label propagation can help, existing methods simply propagate annotations from arbitrarily selected frame es (e.g., the first one) and so may fail to best leverage the human effort invested. The authors define an active frame selection problem: select k frames for manual labeling, such that an automatic pixel-level label propagation can proceed with minimal expected error [8].

2.9 A Semi-automatic Tool for Detection and Tracking Ground Truth Generation in Videos

I. Kavasidis, S. Palazzo, et al

A modular architecture which allows users to define the ground truth h by using an easy graphical user interface (GUI). The developed application integrates some computer vision techniques, with the purpose of enhancing the ground-truth generation process regarding both accuracy and human effort. In particular, Active Contour Models (ACM) are integrated to extract objects' contours automatically. Moreover, object detection algorithms and state-of-the-art edge detection techniques are employed to suggest to the user the most interesting shapes in the frame extract object contours automatically and easily, object detection aims at identifying possible interesting objects, and to do so the Gaussian Mixture Model algorithm (GMM) is employed

2.10 An innovative web-based collaborative platform for video annotation

Isaak Kavasidis, Simone Palazzo, Roberto Di Salvo, et al

Large scale labeled datasets are of key importance for the development of automatic video analysis tools as they, on the one hand, allow multi-class classifiers training and, on the other hand, support the algorithms' evaluation phase. This is widely recognized by the multimedia and computer vision communities, as witnessed by the growing number of the available datasets input video stream, our platform extracts video frames and provides a set of utilities to annotate each video frame and to follow objects across frames. It is a rich internet application, based on a standard client-server architecture: In the proposed tool, the tracking ground truth generation exploits the capabilities of multiple windows applications

to implement an easy-to-use and intuitive way to follow objects across consecutive frames

3. Problem Definition

3.1 Existing System

The existing system includes clustering person track using K-means, where the number of clusters is set to the number of distinguished speakers. Co-occurrence of names in script and face clusters in video constitutes the corresponding face graph and person graph. It modifies the traditional global matching framework by using ordinal graphs for robust representation and introducing an ECGM-based graph matching method.

For face and name graph construction, the existing system represents the person co-occurrence in rank ordinal level, which scores the strength of the relationships in a rank order from the weakest to strongest. The affinity graph used in the traditional global matching is interval measures of the co-occurrence relationship between characters.

3.1.1 Limitations of Existing System

- While continuous measures of the strength of relationship hold complete information, but it is highly sensitive to noises.
- Same person name for given face is tracked even if movies vary.
- Noise removal process is not discussed.
- The sequential statistics for the speakers is not carried out.

Data Input

- Movie files selection using open file dialog control. AVI File is selected as input and saved in the table.
- The movie file is selected from the table, and split into individual frames using AVI extractor "avifil32.dll" methods and saved as bitmaps. The bitmaps folder in the project is used to save all the frames. The record is saved into 'Bitmaps' folder with movie id and frame id.
- The movie file is selected from the table, frame id is selected from the retrieved bit map frame id is selected, and title sentence is added.
- Person names found in the title is added into 'Face Names' table with movie id, frame id, and name.

Face Annotation

- After the movie id selection, from the bit map frames, the face area is found out, and the details are saved in 'Faces' table with movie id, frame id and face data.
- From the frames, each one is selected.
- Converted into gray scale image

- The morphological filter is applied with erosion property (3X3) matrix is given as input for erosion process.
- Then Contour (Border) is found out. Then based on the given width/height ratio, places where images can be found out.
- Select face regions' Image data are saved into the database with X and Y location along with width and height of the area.

3.2 Proposed System

In addition to existing system implementation, the proposed system includes identifying face/names for different movies with different character names also. Enhanced Sobel Edge Annotation Algorithm (Enhnaccdfed K-Means clustering) is applied for clustering process. Moreover, also, different person names for given face is tracked even if movies vary. Median filter concept is applied to remove the noise either before/after the clustering process.

Proposed Annotation Algorithms

- Frames from multiple movie files are consolidated (grouped) as if they are taken from the single movie.
- For the given face (in the selected frames), the names appeared are grouped. For example, if Name Jack from Frame 1 of movie 1 and Name George from frame 1 of Movie 2, occurred for the same face, then both are considered as same actor name.
- Likewise, all the faces in all the frames are checked for combined name appearances in both movies taken for frame selection.
- Common names for same face data in two movie frames are compared for occurrences in multiple places and treated as the same character.

3.2.1 Advantages of the Proposed System

- Not sensitive to noises, since noise removal can be applied either before or after clustering process.
- Different character name for given face is tracked even if movies vary.
- Noise removal process is carried out.
- The sequential statistics for the speakers is also carried out.

3.3 Objective of the Proposed System

- Different type of the noise removal process is to be carried out for identifying the noise in the selected video for that clustering process is carried out in this noise removal.

- The previous systems are not studied such type of the noise removal and identifying images like process.
- The proposed system is carried out with different clustering for the noise removal.
- The proposed system also concentrates in detecting the face of the person who is found in the particular video.
- Total number of times the person image occur in that is been clearly calculated in this proposed system.

3.4 Need for Proposed System
While continuous measures of the strength of relationship hold complete information, but it is highly sensitive to noises.
- Same person name for given face is tracked even if movies vary.
- Noise removal process is not discussed.

The sequential statistics for the speakers is not carried out.

4. Methodology
4.1 General Background
Visual tracking has been studied for years, which has resulted in a rich and abundant scientific literature, summarized well in recent surveys Existing multi-view trackers are usually evaluated on relatively simple content such as people walking or standing without the presence of furniture and — for practical reasons — on short sequences, For real-life applications there is a considerable risk that existing solutions are over-tuned to specific datasets.

The evaluation of the accuracy and reliability of the proposed methodology 6-hour dataset, by a comprehensive visual inspection. Scalability of a semi-automatic methodology for annotation in multi-camera data sets of such length is addressed for the first time.

To make the new calibrated multi-camera dataset available to the community, and provide the annotation results.

4.3 Integration of Modules
The main page includes human reorganization, video dataset, annotation title, clustering process options in this form. Once the human reorganization is clicked camera options that are located or add camera is added in this form that includes details about the camera that is camera number or the id.

After that, the control will move to K-Means Clustering of Person faces menu in that menu the group is clustered, and the frame the selected pattern occurs is mentioned in that form. Then finally the total iteration taken for completing the process is carried out.
1. Recognizing Human face

2. Add pattern and Select snapshot
3. Video dataset
4. Split into frames
5. Add Annotation
6. Add and view person name
7. Persons face reorganization
8. Persons face in movies recognition
9. Person name wise clustering
10. K Means clustering of person name

4.3.1 Recognizing Human Face
In this module, the human face that is to be found out in the video should be noticed previously into the database that should be saved in the database. For storing that the camera should be added for taking the snaps of the human in particular position or the place where the camera located. The previously added camera details also view in the grid view control that is done in this module. The different electrical consumables that are on the camera are also noticed in this module to locate the human image taken in clear manner this process is called.

4.3.2 Add Pattern and Select Snapshot
In this module human face is chosen and then it is added to the database for that the specific area of the human face is selected and stored as the pattern. With that selected pattern image, the searching is carried out. For the selection of the snapshot that is taken in the camera will be noticed. The date and time the snap is taken are stored in the database.

The view is also provided for the pattern that is already stored that details are stored in the database and will be generated as the view using the grid view control. The selected pattern is done using the command click the specific camera is chosen, and with that camera id the snap waits. The view is generated using the grid view control and the snapshots stored will be shown in this grid view.

4.3.3 Video Dataset
In this module, the video is added that is in that video the person or the suspects face is recorded that video will be stored in the database for the search. The total video that is stored in the database can be viewed in this view dataset using the grid view control.

4.3.4 Split into Frames
This module the video or the image is applied to frames, and then each image is split into frames and stored in the database for the future reference. For that process command click is used to split the image into frames is been done in this module. For doing this picture box control is used and explore control is to explore the split image to the user.
4.3.5 Add Annotation
In this module, the movie id and the frame id is selected using the list box the previously added movie, and the

split frames are stored in this annotation module. The selected details are stored in the database. This module also has a view that is designed using the grid view control. The added annotation is also having in this module.

5.3.6 Add and View Person Name

This module has a specific segment that is personal details are added to the database that is the person to search for is added in this module. The view will be generated for the stored person particulars in this module. Select the movie id that is chosen from the list box the person who is appeared in the video is stored in this module. Moreover, additionally the character he got in that video is displayed in this module.

. This view is created using the grid view controls. This module also includes the separate view for displaying the person character name, and the movie id is shown in this module.

5.3.7 Persons Face Reorganization in Single Image

In this module, the image is selected and using the red green and blue check boxes the image is selected to find the faces, and that will be stored in the database. The size of the filter is also choosing using the list box control.

Before that, to load, the image into the frame command click is made to add the image in the frame for that single image is being used.

4.3.8 Persons Face Recognition Video Frames

In this module, the video is chosen to in from that video the frames are selected to recognize the face in that video. Once the video is selected, and the added frame in the image is processed after the process command click is done.

The face recognized in the single image is selected for that searching is done in this module and frames are identified and shown the occurrence of that frames that video. This also includes the filter size and the red, green, blue check boxes to choose the video frames. The selected image is stored in the data base after the save image click is processed.

4.3.9 Person Name Wise Clustering

This module movie id is selected, and person name is also given for the searching of the person that is the person who is in the video. That is the name fetched for the finding of occurrence, and the persons occurred as the group or cluster is also fetched from that video.

This also includes the person whose image occurred as the group and who is not with that group is listed in the message box. The frame selection details are also shown in this module. If the cluster is forming details and the other details can be updated using update command click.

4.3.10 K Means Clustering Of Person Name

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem.

The procedure follows a simple and easy way to classify a given data set through a certain number of clusters.K Means clustering is applied for clustering process. Moreover, also, different character names for given face is tracked even if movies vary. Median filter concept is applied to remove the noise either before/after the clustering process.

After the movie id selection, faces are clustered such that K Means clustering is applied to 'N' clusters is given as input. Based on the color difference in the bitmap pixels, the face similarity is calculated. Then to display the cluster information, Tooltip option is implemented in this module.

5. Testing and Implementation

After the source code has been completed, documented as related data structures. Completed the project has to undergo testing and validation where there are subtitle and definite attempt to get errors. The project developer treads lightly, designing and execution test that will demonstrate that the program works rather than uncovering errors, unfortunately, errors will be present, and if the project developer does not find them, the user will find out.

This project has undergone the following testing procedures to ensure its correctness.
1. Unit testing
2. User Acceptance Testing

Unit Testing

In unit testing, we have to test the programs making up the system. For this reason, Unit testing sometimes called as Program testing. The software units in a system are the modules and routines that are assembled and integrated to perform a specific function, Unit testing first on the modules independently of one another, to locate errors. This enables, to detect errors in coding and logic that are contained in the module alone. The testing was carried out during programming stage itself.

User Acceptance Testing

In these testing procedures, the project is given to the customer to test whether all requirements have been fulfilled and after the user is fully satisfied. The project is perfectly ready. If the user makes request for any change and if they found any errors those all errors has to be taken into consideration and to be correct it to make a project a perfect project.

System Implementation

When the initial design was done for the system, the client was consulted for the acceptance of the design so that further proceedings of the system development can be carried on. After the development of the system a demonstration was given to them about the working of the

system. The aim of the system illustration was to identify any malfunction of the system.

Testing: Once the coding process is begin and proceed in parallel, as each program module can be tested.

Installation: Installation is the process during which the current system is replaced by the new system. This includes conversion of existing data, software, and documentation and work procedures to those consistent with the new system.

Documentation: It is result from the installation process, user guides provides the information of how the use the system and its flow.

Training and support: Training plan is a strategy for training user so they quickly learn to the new system. The development of the training plan probably began earlier in the project.

The best-suited application package to develop the system is Visual C# .NET under windows XP' environment.

Experimental Analysis

The consensus tracker is based on simple probabilistic models of the tracking errors of individual trackers. The approach would be to ask a human to annotate some video frames, and then use this data to compute ground truth positions from which error statistics. Table 6.1 shows the person tracking ratio and error rate of individual Consensus tracker approach. Tracking errors are then modeled with distributions and obtained on those frames for which a position estimate.

Number of Frames	Consensus Tracker [%]	Error Rate [%]
100	52.32	49.51
200	59.76	58.62
300	65.33	65.34
400	73.56	64.36
500	77.25	75.26
600	79.82	77.85
700	81.35	80.26
800	82.55	80.12
900	84.25	83.45
1000	85.25	81.35

Table 7.1 Person Tracking Ratio of Consensus Tracker

The Fig. 7.1 illustrates Consensus Tracker approach, which involves estimating the uncertainty in the position estimate of each tracker and using this uncertainty estimate in a Bayesian fusion approach. Tracking errors

are then modeled with distributions and obtained on those frames for which a position estimate.
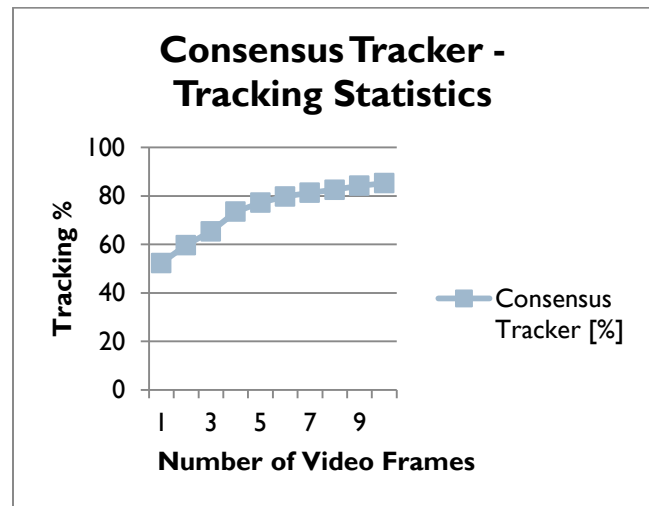

Fig 7.2 Consensus Tracker Statistics – Person Tracking Ratio

The Enhanced Sobel Edge Annotation Algorithm (ESEAA) is a novel methodology proposed for generating reference tracking data in long multi-camera videos, based on the consensus of a detector and several trackers. For multi-camera annotation in the proposed ESEAA methodology, it first estimates the reliability of annotations and offers the possibility of balance accuracy and human-effort in the final annotation result. Table 6.2 shows the person tracking ratio and error rate of ESEAA approach. In this approach, each frame is compared and performed clustering annotation for finding magnitude and direction of image gradients using Sobel operator.

The Fig. 7.2 illustrates the performance of person tracking the ratio of Enhanced Sobel Edge Annotation Algorithm (ESEAA) approach using multi-camera annotation. In this approach, Each Frame Compared and performed clustering annotation for finding magnitude and direction of image gradients using Sobel operator.
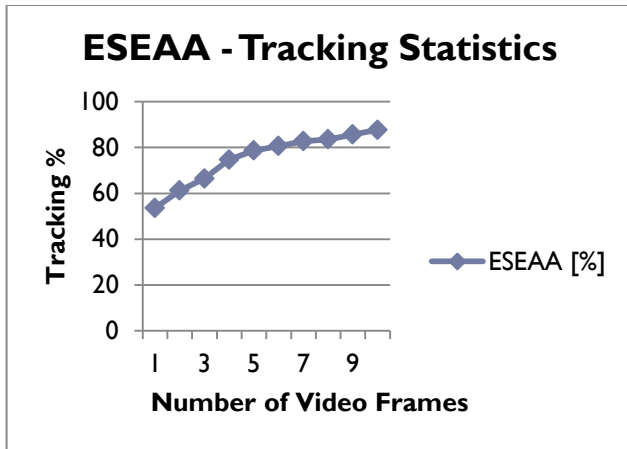
Fig 7.4 Enhanced Sobel Edge Annotation Algorithm – Person Tracking Ratio

The following Fig 7.3 showed the error rate of the existing Consensus Tracker and proposed ESEAA (Enhanced Sobel Edge Annotation Algorithm) approach. The error rate of both present and proposed models is compared, and the figure shows the average percentage of mistakes rate.
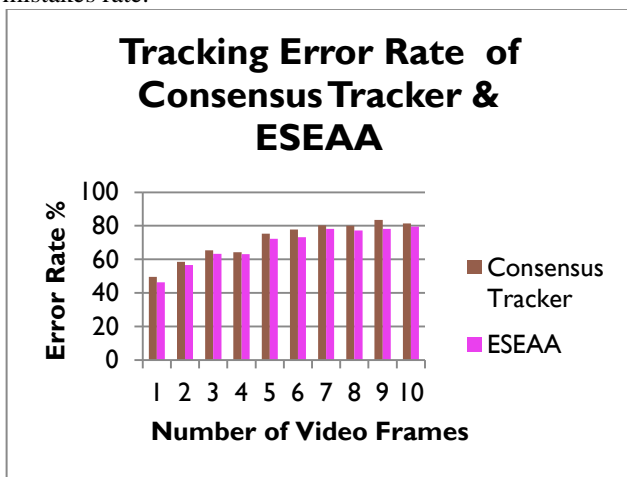

Fig 7.5 Comparison of Error Rate of ESEAA and Consensus Tracker Models

The following Table 7.3 showed the performance analysis results of the existing Consensus Tracker and proposed ESEAA (Enhanced Sobel Edge Annotation Algorithm) approach. The table describes the comparison result of both present and proposed models are compared and describes the average percentage of person tracking process.

Number of Frames	Person Tracking [%]	
	Consensus Tracker	ESEAA
100	52.32	53.62

200	59.76	61.25
300	65.33	66.45
400	73.56	74.68
500	77.25	78.69
600	79.82	80.65
700	81.35	82.75
800	82.55	83.54
900	84.25	85.68
1000	85.25	87.65

Fig 7.6 Comparison Analysis of ESEAA & Consensus Tracker Models

The following Fig 7.4 showed the performance analysis results of the existing Consensus Tracker and proposed ESEAA (Enhanced Sobel Edge Annotation Algorithm) approach. The table describes the comparison result of both existing and proposed models are compared and describes the average percentage of person tracking process.
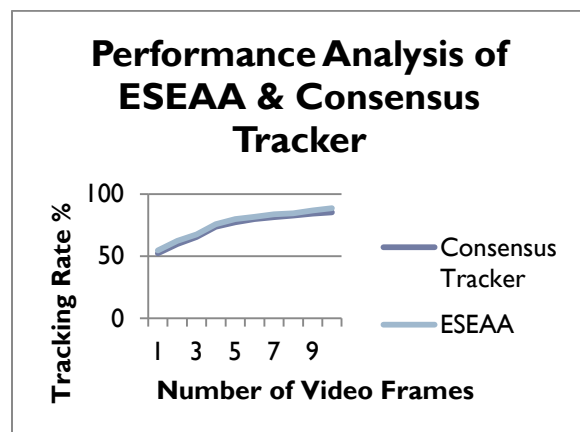

Fig 7.7 Performance Analysis of ESEAA & Consensus Tracker Models

6. Conclusion

The proposed system is developed using the pattern matching with the video that is captured by the camera that is matched with the annotated image. The annotated data set is compared with the video recorded in the camera. If there is any similar match found in the movie that will be sorted for the view. The sorting is done using the KNN classification algorithm to list the person who is detected in this video is classified in this form.

The proposed approach allows comparing the results of a given tracker to the consent of other trackers. The ultimate goal is to provide a graphical user interface which allows easy visualization of those portions of a data set with large discrepancies between tracking results. The

proposed system easy to allow the users to indicate the segments of video data in which their new tracker outperforms the compromise tracker.

7. Future Enhancement

The proposed methodology has some limitations in the aspect is the time and effort the user has to spend for the standard part of the semi automatic method of annotation. The visual inspection time can be reduced by optimizing the order in which frames with unreliable annotations are shown to the user.

The other ways of sampling annotations can be explored. Furthermore, by studying different values of trackless size, more efficient working points for visual inspection could be estimated in future work. Analyze and testing on more and larger video datasets is essential to explore when and how often trackers fail under diverse conditions and prioritizes the seeking of most bad annotations in the part of the future research work.

REFERENCES

[1] M. Liem and D. Gavrila, "A comparative study on multi-person tracking using overlapping cameras," in-proc. 9th Int. Comput. Vis. Syst., 2013, pp. 203–212.

[2] C. Vondrick, D. Patterson, and D. Ramanan, "Efficiently scaling up crowdsourced video annotation," Int. J. Comput. Vis., vol. 101, no. 1, pp. 184–204, Jan. 2013, doi: 10.1007/s11263-012-0564-1.

[3] A. Utasi and C. Benedek, "A multi-view annotation tool for people detection evaluation," in-proc. VIGTA, 2012, pp. 1–6.

[4] L.ˇ Cehovin, M. Kristan, and A. Leonardis, "Is my new tracker better than yours?" in proc. WACV, Mar. 2014, pp. 540–547.

[5] L. Marcenaro, P. Morerio, and C. S. Regazzoni, "Performance evaluation of multi-camera visual tracking," in Proc. AVSS, Sep. 2012, pp. 464–469.

[6] A. Milan, K. Schindler, and S. Roth, "Challenges of ground truth evaluation of multi-target tracking," in Proc. CVPRW, Jun. 2013, pp. 735–742.

[7] M. Kristan et al., "The visual object tracking VOT2013 challenge results," in Proc. ICCVW, Dec. 2013, pp. 98–111.

[8] S. Vijayanarasimhan and K. Grauman, "Active frame selection for label propagation in videos," in-proc. ECCV, 2012, pp. 496–509.

[9] I. Kavasidis, S. Palazzo, R. Di Salvo, D. Giordano, and C. Spampinato, "A semi-automatic tool for detection and tracking ground truth generation in videos," in Proc. VIGTA, 2012, pp. 1–5.

[10] I. Kavasidis, S. Palazzo, R. D. Salvo, D. Giordano, and C. Spampinato, "An innovative Web-based collaborative platform for video annotation," Multimedia Tools Appl., vol. 70, no. 1, pp. 413–432, May 2013.

[11] Ajin P Thomas, Sruthi P.S, Jerry Rachel Jacob, Vandana V Nair, Reeba R," Survey on Different Applications of Image Processing." International Journal of Computer Engineering In Research Trends.,vol.4,no.2,pp. 13-19,2017.

[12] Trisha Chakraborty, Nikita Nalawade, Abhishri Manjre, Akanksha Sarawgi, Pranali P Chaudhari," Review of Various Image Processing Techniques for Currency Note Authentication." International Journal of Computer Engineering In Research Trends.,vol.3,no.3,pp. 119-122,2016.

[13] Gunjan, Er. Madan Lal," Investigation of Various Image Steganography Techniques in Spatial Domain." International Journal of Computer Engineering In Research Trends., vol.3,no.6,pp. 347-351,2016.

[14] G.Prasanthi, A.Somasekhar," Anti-Theft Tracking and Controlling Of Vehicle According Us." International Journal of Computer Engineering In Research Trends., vol.2, no.12, pp. 898-903, 2015.