International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 2 | Issue 5 | ISSN : 2456-3307

A Comparison ECGM-Based Graph Matching and Enhnaced K-Means Clustering Multi-View Video Model

Dr. P. Sumitra, M. Senbagapriya

Department of Computer Science, Vivekananda College of Arts and Sciences for Women (Autonomous), Elayamapalayam, Tamilnadu, India

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 human using a simple binary decision task, a process faster than marking the correct person position, verifies a small subset of the data, composed of tracks with insufficient reliability. The proposed framework is generic and can handle additional trackers.In this thesis studied the most commonly use face edge detection techniques of Enhnaced 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

I. 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 distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems.

Purpose of Image processing

The purpose of image processing is divided into 5 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.

Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

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

1.1.2 TYPES OF DIGITAL IMAGES:

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

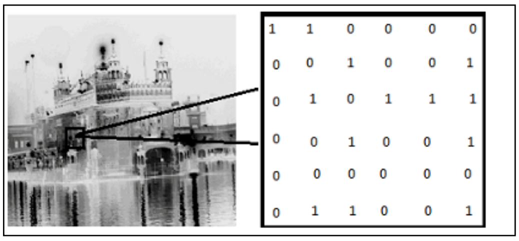

Figure 1.1binary Images

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

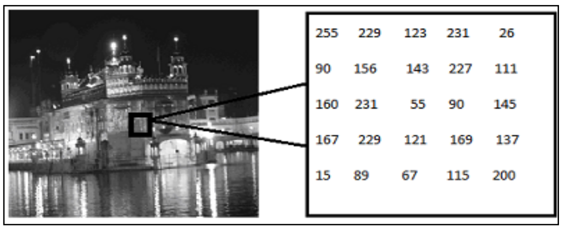

Figure 2. Grayscale Images

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 image. 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 3 values.

Indexed: Mostly all the colors images have a subset of more than sixteen million possible colors. For ease of storage and handling of file, the image has an related 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 give an index to the color in 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 color only.

1.3.3 DIGITAL IMAGE FILE TYPES

BMP:

Bmp stands for Bitmap. Every picture on a computer appear to be a BMP. In Windows XP the Paint program save its images automatically in bitmap format, however in Windows Vista images are saved now into JPEG format. 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 internet and large in file size.

Digital camera manufacturers obviously see the value in high quality images that eventually take up less space.

Benefits: Small size image, easily viewable from internet, Use millions of colors, and perfect for many type 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.

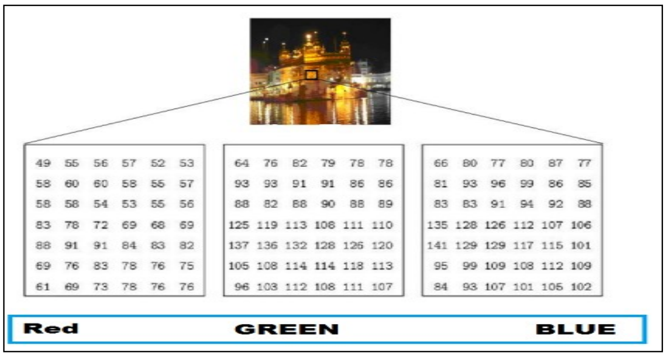

Figure 3. Color Images

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.

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 total number of 256 colors or it can be less. It is mostly used for loading the fast web pages. It also help to makes great banner and logo for different webpage. Different type of animated pictures are 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 very small file size, Easy to load, Benefit for Transparencies, and animations and Image maps

Downfalls: We can used 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 format, 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 logo's 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 use 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.

1.2 COLOR MODELS

For science communication, the two main colour spaces are RGB and CMYK.

1.2.1 RGB

The RGB colour model relates very closely to the way we perceive colour with the r, g and b receptors in our retinas. RGB uses additive colour mixing and is the basic colour model used in television or any other medium that projects colour with light. It is the basic colour model used in computers and for web graphics, but it cannot be used for print production. The secondary colours of RGB – cyan, magenta, and yellow – are formed by mixing two of the primary colours (red, green or blue) and excluding the third colour. Red and green combine to make yellow, green and blue to make cyan, and blue and red form magenta. The combination of red, green, and blue in full intensity makes white. In Photoshop using the "screen" mode for the different layers in an image will make the intensities mix together according to the additive colour mixing model. This is analogous to stacking slide images on top of each other and shining light through them.

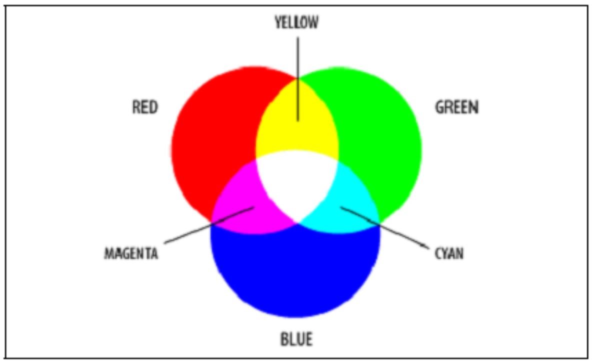

Figure 4. RGB Color Model

1.2.3 Gamut

The range, or gamut, of human color perception is quite large. The two color spaces discussed here span only a fraction of the colors we can see. Furthermore the two spaces do not have the same gamut, meaning that converting from one color space to the other may cause problems for colors in the outer regions of the gamuts.

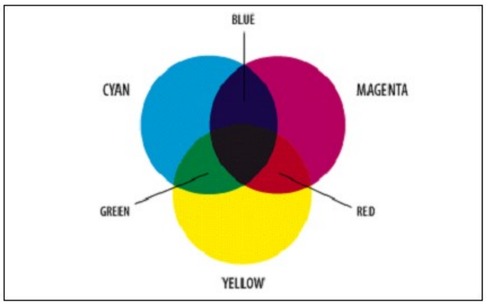

Figure 5. CMYK Color Model

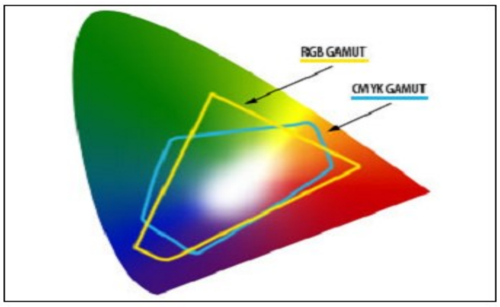

Figure 6. Different gamuts of the RGB and CMYK colour

1.3 CHARACTERISTICS OF IMAGE OPERATIONS

There is a variety of ways to classify and characterize image operations. The reason for doing so is to understand what type of results we might expect to achieve with a given type of operation or what might be the computational burden associated with a given operation.

1.3.1 Types of operations

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

1.3.2 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.

Operation	Characterization	Generic Complexity/Pixel
• *Point*	– the output value at a specific coordinate is dependent only on the input value at that same coordinate.	*constant*
• *Local*	– the output value at a specific coordinate is dependent on the input values in the *neighborhood* of that same coordinate.	P^2
• *Global*	– the output value at a specific coordinate is dependent on all the values in the input image.	N^2

Fig 1.7 Image Operators

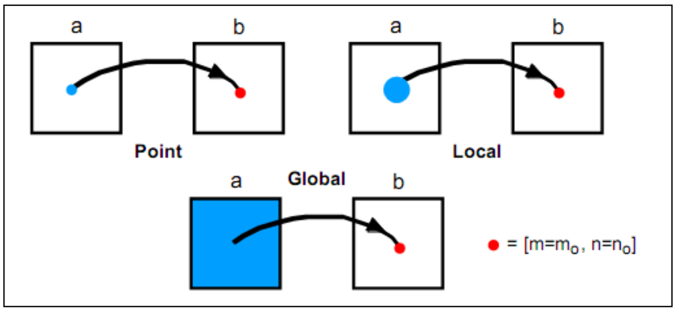

Figure 7. Illustration of various types of image operations

- 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. 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 sampling illustrated in Figure 3.

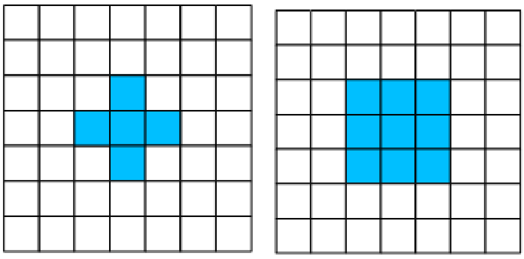

Figure 8. Rectangular sampling a and b

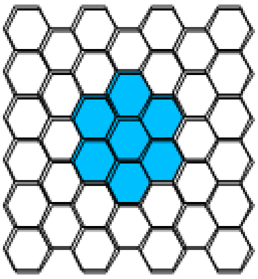

Figure 9. Hexagonal Sampling

2.1 ECGM-based graph matching method

- 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.

2.1.1 DATA INPUT

- Movie file selection using open file dialog control. AVI File is selected as input and saved in table.
- Movie file is selected from 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 'Bit maps' folder with movie id and frame id.
- Movie file is selected from table, frame id is selected from the retrieved bitmap frame id is selected and title sentence is added.
- Person names found in title is added into 'Face Names' table with movie id, frame id and name.

2.1.2 FACE ANNOTATION

- After the movie id selection, from the bitmap frames 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
- 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 database with X and Y location along with width and height of the area.

2.1.3 FACE CLUSTERING WITH ANNOTATION

- After the movie id selection, faces are clustered such that K Means clustering is applied with 'N' clusters is given as input.
- Based on the color difference in the bitmap pixels, the face similarity is calculated.

2.1.4 FORMING FACE RECOGNITION

- Based on the faces (multiple person) appeared in the bitmap frames, relationship between faces is formed.
- For example, one frame contains Face A, B and C other contains A and C. So A is more related with C and less related B.
- The edge weight is fixed based on the relationship/ common occurrence between faces.

2.1.5 FORMING PERSON ANNOTTAION GRAPH

- Based on the names (multiple person named) appeared in the bitmap frames title, relationship between names is formed.
- For example, one frame contains name A, B and C other contains A and C. So A is more related with C and less related B.
- The edge weight is fixed based on the relationship/ common occurrence between names.

2.1.6 PERSON RECOGNTION

- Matching is done based on common occurrences of faces and names.
- For example If frame 1 contains Face A, B and C with name X, Y and Z.
- Then frame 2 contains Face A and C with name X and Z, then it is sure that A and B have the names X and Z.
- After intersecting all the frames with Face/Person occurrence, try to match the person with faces.

3. ENHNACED K-MEANS CLUSTERING

- Frames from multiple movie files are consolidated (grouped) as if they are taken from 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 same character.

3.1.1 ADVANTAGES

- 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.

3.1.2 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.

3.1.3 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 notcarried out

II. CONCLUSION

A methodology for generating reference tracking data in long multi-camera videos, based on the consensus of a detector and several trackers. For multi-camera annotation, our methodology is the first to estimate the reliability of annotations, and to offer the possibility of balance accuracy and human-effort in the final annotation result. A novel probabilistic framework is to learn the error models of trackers, and how to apply them to estimate target position. Previous methods did not model the tracking error statistics of multiple trackers to increase the reliability of the estimated position. The valuation 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.

III. REFERENCES

[1]. A. Milan, K. chindler, and S. Roth, "Challenges of ground truth evalua-tion of multi-target tracking," in Proc. CVPRW, Jun. 2013, pp. 735-742.

[2]. H. Wu, A. C. Sankaranarayanan, and R. Chellappa, "Online empirical evaluation of tracking algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 8, pp. 1443-1458, Aug. 2010.

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

[4]. M. Kristan et al., "The visual object tracking VOT2014 challenge results," in Proc. ECCV Workshops, 2014, pp. 191-217.

[5]. J. Yuen, B. Russell, C. Liu, and A. Torralba, "LabelMe video: Building a video database with human annotations," in Proc. ICCV, Sep./Oct. 2009, pp. 1451-1458.

[6]. B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and Web-

based tool for image annotation," Int. J. Comput. Vis., vol. 77, nos. 1-3, pp. 157-173, 2008.

[7]. A. Dorado, J. Calic, and E. Izquierdo, "A rule-based video annotation system," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 5, pp. 622-633, May 2004.

[8]. M. Bertozzi, A. Broggi, R. Grisleri, A. Tibaldi, and M. D. Rose, "A tool for vision based pedestrian detection performance evaluation," in Proc. IEEE Intell. Vehicles Symp., Jun. 2004, pp. 784-789.

[9]. C. Spampinato, B. J. Boom, and J. He, "MTAP special issue on meth-ods and tools for ground truth collection in multimedia applications," Multimedia Tools Appl., vol. 70, no. 1, pp. 409-412, May 2014.

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