MOVING OBJECT DETECTION WITH BACKGROUND MODEL

N.V.Manu[1], M.Kirithikadevi[2]

[1,2]Dept of CSE, Chendhuran College of Engineering and Technology

Anna University, Chennai. (India)

ABSTRACT

Motion detection plays an important role in intelligent video surveillance. This paper introduces a particular background subtraction technique called ViBe. This technique updates the background model randomly and it has established model with fast, high precision and fast processing speed. ViBe algorithm provides the method of updated background model, but slowly eliminates ghost region. This paper presents an improved ViBe algorithm based on region motion classification. The algorithm considers the difference of the movement directions of feature points in foreground regions on adjacent frame, and defines a criterion function to evaluate the difference so that can quickly eliminate ghost regions. Experimental results show that the proposed algorithm quickly remove the ghost region and improve the detection accuracy.

Keywords: Background Subtraction Techniques, Computation Speed, Mainstream Techniques, Photogrammetric Applications.

1. INTRODUCTION

In photogrammetric applications, it is required to make measurements on digital images. A major difficulty is the distortion of many geometric scene properties due to imaging transformation in a perspective camera. The most influenced properties include the size of an object, the length ratio and parallelism between line segments, all of which are not preserved in a projected image. An ideal way to clear all the uncertainties is by recovering the real 3-D world from a 2-D projected image, which in most cases needs for more information than a single image can provide with unknown camera calibration. A semiautomatic measuring method using image data from a single camera has been proposed in [1], which uses minimal calibration information and the perspective geometry approach for menstruation. Motivated by the wide availability of video surveillance systems, we propose to extend single image metrology techniques. To video data acquired by stationary cameras, which is referred to as video metrology.

Video metrology enables improved automation, flexibility and accuracy. 1) Achieving automation is critical for handling large volumes of video data. Traditional photogrammetric systems involve some human interactions which are unpleasantly burdensome and often unreliable. For example, geometric features, such as orthogonal lines, parallel lines and vanishing points are usually manually labelled and even objects themselves need to be manually selected from surroundings. Video sequences enable an automatic solution: Objects can be first segmented using tracking or segmentation algorithms, then motion information can be used to estimate the minimal calibration of scenes. The need for human interactions is highly reduced. 2) Exploitation of motion information offers enhanced flexibility. Photogrammetric is effective in man-made environments characterized by plenty of solid parallel and orthogonal lines to infer the vanishing line/point. These strong cues are not always present. However, the tracked moving objects in video sequences provide away of estimating the vanishing line of ground planes and the vertical vanishing point. In most cases, menstruation from one single image is not reliable, especially when occlusion so instant disturbances occur. 3) Since video sequences provide additional temporal information, fusing multi frame measurements smoothes out the menstruation errors of individual frames and hence improves the accuracy of the final estimate of the measurement.

Video is the technology of electronically capturing, recording, processing, storing, transmitting, and reconstructing a sequence of still images representing scenes in motion. The capture process fixes the "natural" frame rate of the image sequence. Moving image sequence can be captured at the rate which is different from presentation rate; however this is usually only done for the sake of artistic effect, or for studying fast-pace or slow processes. In order to faithfully reproduce familiar movements of persons, animals, or natural processes, and to faithfully reproduce accompanying sound, the capture rate must be equal to, or at least very close to the presentation rate.

Recording is a process of capturing data or translating information to a format stored on a storage medium often referred to as a record. Historical records of events have been made for thousands of years in one form or another. Amongst the earliest are cave painting, runic alphabets and ideograms. Ways of recording text suitable for direct reading by humans includes writing it on paper. Other forms of data storage are easier for automatic retrieval, but humans need a tool to read them. Printing a text stored in a computer allows keeping a copy on the computer and having also a copy that is human-readable without a tool. Technology continues to provide and expand means for human beings to represent record and express their thoughts, feelings and experiences.

Motion capture, motion tracking, or mocap are terms used to describe the process of recording movement and translating that movement on to a digital model. It is used in military, entertainment, sports, and medical applications, and for validation of computer vision when it includes face and fingers or captures subtle expressions, it is often referred to as performance capture.

Camera movements can also be motion captured so that a virtual camera in the scene will pan, tilt, or dolly around the stage driven by a camera operator while the actor is performing, and the motion capture system can capture the camera and props as well as the actor's performance. This allows the computer-generated characters, images and sets to have the same perspective as the video images from the camera. A computer processes the data and displays the movements of the actor, providing the desired camera positions in terms of objects in the set.

II. LITERATURE SURVEY

We consider the problem of estimating and tracking the 3D configurations of complex articulated objects from monocular images, *e.g.* for applications requiring 3D human body pose or hand gesture analysis. There are two main schools of thought on this. Model-based approaches presuppose an explicitly known parametric body model, and estimate the pose by either: *(i)* directly inverting the kinematics, which requires known image positions for each body part (Taylor, 2000); or *(ii)* numerically optimizing some form of model-image correspondence metric over the pose variables, using a forward rendering model to predict the images, which is expensive and requires a good initialization, and the problem always has many local minima. There we used a method that recovers 3D human body pose from sequences of monocular silhouettes by direct nonlinear regression of joint-angles against histogram-of shape- context silhouette shape descriptors and dynamics based pose estimates. No 3D body model or labelling of image positions of body parts is required. Regressing the pose jointly on image observations and previous poses allows the intrinsic ambiguity of the pose-from-monocular observations problem to be overcome, thus producing stable, temporally consistent tracking. We use a kernel-based Relevance Vector Machine for the regression, thus selecting a sparse set of relevant training examples as exemplars. The method shows promising results on tracking unseen video sequences, giving an average RMS error of 4.1◦ per body joint- angle on real motion capture data. [1]

The problem of tracking curves in dense visual clutter is challenging. Kalman filtering is inadequate because it is based on Gaussian densities which, being unimodal, cannot represent simultaneous alternative hypotheses. The Condensation algorithm uses "factored sampling", previously applied to the interpretation of static images, in which the probability distribution of possible interpretations is represented by a randomly generated set. Condensation uses learned dynamical models, together with visual observations, to propagate the random set over time. The result is highly robust tracking of agile motion. Notwithstanding the use of stochastic methods, the algorithm runs in near real-time. [5]

Segment and recognize elementary actions such as run walk and draw on board, from a video sequence where one person performs a sequence of such actions. This is a fundamental problem in human action understanding and has a wide range of applications in e.g. surveillance, video retrieval and intelligent interface. It is nevertheless challenging due to the high variability of appearances, shapes and possible occlusions, and things are further complicated for continuous action recognition in our case where it is necessary to segment the input video sequence into continuous action segments. In addition, we show that this method can also be used to recognize the person who performs in this video sequence. By employing a Viterbi-like column generation algorithm, this approach allows us to explicitly encode segment-level properties into feature representation and still be solved efficiently. Experimental results on a variety of dataset demonstrate that our approach is flexible to cater for different scenarios, yet it is competitive comparing to the state-of-the-art methods. [11]

Then used two algorithms for on-line prediction based on a linear model. The algorithms are the well-known gradient descent (GD) algorithm and a new algorithm, which we call EG\. They both maintain a weight vector using simple updates. For the GD algorithm, the update is based on subtracting the gradient of the squared error made on a prediction. The EG\ algorithm uses the components of the gradient in the exponents of factors that are used in updating the weight vector multiplicatively. We present worst-case loss bounds for EG\ and compare them to previously known bounds for the GD algorithm. The bounds suggest that the losses of the algorithms are in general incomparable, but EG\ has a much smaller loss if only few components of the input are relevant for the predictions. We have performed experiments which show that our worst-case upper bounds are quite tight already on simple artificial data. [6]

Many problems in computer vision can be naturally formulated as parameter estimation problems: given an image or a video sequence x, we estimate the parameters of a model describing the scene or the object of interest. Examples include estimation of the configuration of an articulated body, the contraction of muscles in the face, or the orientation of a rigid object. Example-based estimation methods capitalize on the availability of a large set of examples for which the parameter values are known: they infer the parameter values for the input from the known values in similar examples. This does not require modelling global structure of the input/parameter relationship, which is only Part of this research performed while first two authors were at Mitsubishi Electric Research Labs, Cambridge, MA. Their support is gratefully acknowledged. An algorithm that uses new hashing-based search techniques to rapidly find relevant examples in a large database of image data, and estimates the parameters for the input using a local model learned from those examples. Experiments show that our estimation method, based on parameter-sensitive hashing and robust locally-weighted regression, is successful on the task of articulated pose estimation from static input. These experiments also demonstrate the usefulness of synthetically created data for learning and estimation. In addition to the use of local regression to refine the estimate, our work differs from that of others, in that it allows accurate estimation when examining only a fraction of a dataset. The running time of our algorithm is sub linear; in our experiments we observed a speedup of almost 2 orders of magnitude relative to the exhaustive exact nearest-neighbour search, reducing the time to estimate pose from an image from minutes to less than 2 seconds without adversely affecting the accuracy. We expect an optimized version of the system to run at real time speed. This has the potential of turning infeasible example-based estimation methods into attractive for such tasks. [10]

There is a simple, principled approach to detecting foreground objects in video sequences in real-time. Our method is based on an on-line discriminative learning technique that is able to cope with illumination changes due to discontinuous switching, or illumination drifts caused by slower processes such as varying time of the day. Starting from a discriminative learning principle, we derive a training algorithm that, for each pixel, computes a weighted linear combination of selected past observations with time decay. We present experimental results that show the proposed approach outperforms existing methods on both synthetic sequences and real video data. An online discriminative approach is proposed to address a key problem in video analysis foreground background separation. The proposed approach is derived from an online risk minimization framework, and is shown in experiments to outperform existing algorithms. The current work involves more about the temporal dynamics of the pixel processes, our future work will focus on exploiting the spatial properties of the sensor fields and the label fields to further improve the performance. [3]

III. PROPOSED APPROACH

First we implement the unsupervised learning approach to implement this process. In machine learning, unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning.

We conduct experiments on a variety of video sequence datasets. The results of the proposed algorithms are compared to standard background subtraction methods including adjacent frame difference (FD), mean-filter (MF) as well as to the recent variant of mixture of Gaussians (MoG) using recursive updates. Throughout the experiments, Gaussian kernels are used for the proposed approach. We conduct controlled experiments to evaluate the performance of sequences are constructed using a large quantity of background examples and a relative small amount of foreground examples.

We have presented an automatic method for robustly measuring object heights from video sequences. We first recover the minimal calibration of the scene based upon tracking moving objects then apply the single view metrology algorithm to each frame, and finally fuse the multi frame measurements using the (average deviation). This project examines the problem of moving object detection. More precisely, it addresses the difficult scenarios where background scene textures in the video might change over time.

IV. SYSTEM DESIGN

4.1 System Architecture

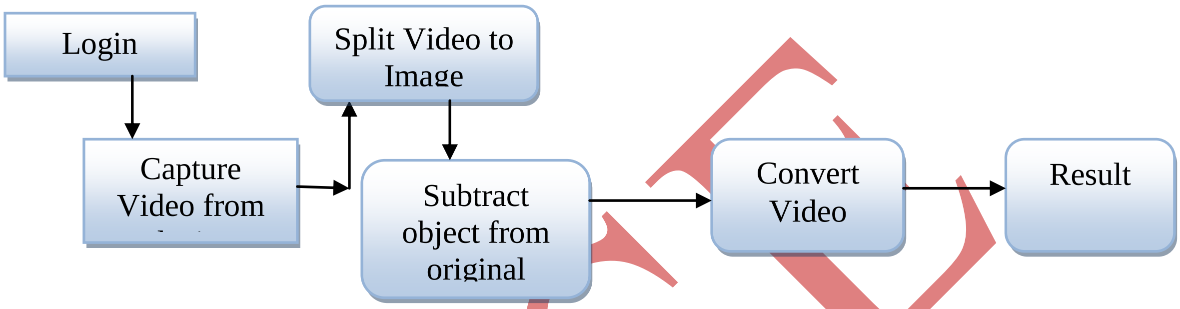

Fig1.System Flow Diagram

4.1.1 Input Design

Input design is one of the most important phases of the system design. Input design is the process where the input received in the system are planned and designed, so as to get necessary information from the user, eliminating the information that is not required. The aim of the input design is to ensure the maximum possible levels of accuracy and also ensures that the input is accessible that understood by the user.

The input design is the part of overall system design, which requires very careful attention. If the data going into the system is incorrect then the processing and output will magnify the errors.

The objectives considered during input design are:

- Nature of input processing.
- Flexibility and thoroughness of validation rules.
- Handling of properties within the input documents.
- Screen design to ensure accuracy and efficiency of the input relationship with files.
- Careful design of the input also involves attention to error handling, controls, batching and validation procedures.

Live Video Input: In this form we give live video capture from the web cam and give input to start our application to run.

Compare Image: In this we give image from the original video or from the current taken video as a input to compare the current position of the place.

Calculate image different: In this form we give the original image and changed image as an input to calculate the modified image height between the images.

4.1.2 Output Design

The output design is the most important and direct source of information to the user. The encoding time and file size for both the fractal as well as fast fractal technique are shown in output screen. The comparison of both techniques is done and the PSNR value is calculated. The reconstructed image is also displayed in the output screen. Output from the computer system is required to communicate the result of processing to the user and to provide permanent copy of these results for later consultation. While designing the output, the type of output format, frequency etc has been taken into consideration. Output designed to simply generate an output of the process whether it was successful or not.

Video Frame: In this output form we get the real and live video from the web cam in our video play from.

Mismatched Frames: In this process we get the output as a images those are not match with the original image are saved in this recorded folder image.

V. MODULES

1. User Interface Design
2. Video Capture
3. Frame Grabber
4. Standard Deviation Method
5. Image Comparison
6. Image Conversion
7. Object Detection
8. Frames into Video

5.1 User Interface Design

Interface design is involved in a wide range of projects from computer systems, to cars, to commercial planes; all of these projects involve much of the same basic human interaction yet also require some unique skills and knowledge. As a result, designers tend to specialize in certain types of projects and have skills centered around their expertise, whether that be software design, user research, web design, or industrial design.

5.2 Video Capture

Video is the technology of electronically capturing, recording, processing, storing, transmitting, and reconstructing a sequence of still images representing scenes in motion. The capture process fixes the "natural" frame rate of the image sequence. Moving image sequence can be captured at the rate which is different from presentation rate; when the video is capturing from the web camera we use the buffer storage to store the video from the output. From the buffer we get the specify data and use that to Frame grabber

5.3 Frame Grabber

This is a digital image taken by the host operating system or software running on the computer, but it can also be a capture made by a camera or a device intercepting the video output of the display. Screenshots, screen dumps, or screen captures can be used to demonstrate a program, a particular problem a user might be having or generally when display output needs to be shown to others or archived.

The term screen cast compares with the related term screenshot; whereas screenshot is a picture of a computer screen, a screen cast is essentially a movie of the changes over time that a user sees on a computer screen, enhanced with audio narration.

Our Step by step process:

- Get capture location from user
- Read data from that location
- Capture video from that location using java media framework
- Captured video convert into frames using Frame grabbing using(Frame grabber)

Frames stored into one location.

5.4 Standard Deviation Method

This algorithm used to compare the back ground image and capture image get from the live web camera. Standard deviation is a widely used measurement of variability or diversity used in statistics and probability theory. It shows how much variation or "dispersion" there is from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values.

Technically, the standard deviation of a statistical population, data set, or probability distribution is the square root of its variance. It is algebraically simpler though practically less robust than the average absolute deviation. A useful property of standard deviation is that, unlike variance, it is expressed in the same units as the data.

The standard deviation of a list of numbers is a measure of how much the numbers deviate from the average. If the standard deviation is small, the numbers are clustered close to the average. If the standard deviation is large, the numbers are scattered far from the average. The standard deviation of a list of numbers n1, n2, n3, and so forth is defined as the square root of the average of the following numbers: (n1 - a) ^2, (n2 - a) ^2, (n3 - a) ^2, and so forth.

5.5 Image Comparison

Image compression scheme, in which a part of coded data extracted in the coding process is hidden into the other parts of coded data of own image, especially into the block address data of the best matching block within restricted blocks. Compare two frames, any changes from that frame that frame move to recording position otherwise frames will be deleted.

5.6 Image Conversion

In this method we convert a colour image to gray scale. A massive number of image file formats are available for storing graphical data, and, consequently, there are a number of issues associated with converting from one image format to another, most notably loss of image detail.

Color in an image may be converted to a shade of gray by calculating the effective brightness or luminance of the color and using this value to create a shade of gray that matches the desired brightness. The following code snippet isn't really a good example of how it should be done in a production situation but it does show the principles involved clearly.

The effective luminance of a pixel is calculated with the following formula:

Y=0.3RED+0.59GREEN+0.11Blue

This luminance value can then be turned into a grayscale pixel using Color.

5.7 Object Detection

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance.

5.8 Frames into Video

In this module all mismatch frames from the static image are converting into a video. That video store in the server system as mpeg file format.

VI. CONCLUSION

We proposed a method for detecting moving objects with a background model that covers local and global changes in backgrounds using a spatio-temporal texture called ST-Patch features, which describe motion and appearance at the same time. Our proposed method performed better than the method applying normalized distance, which is an example of conventional features that describe appearance only, in three scenes. However, our proposed method only determines whether or not there is any moving object within a detection area, and cannot extract any information about moving objects. .

ACKNOWLEDGMENT

Our completion of this paper could not have been accomplished without the support of staff members those who are working with us. We are very much thankful to them. For the reference, we refer many articles and research papers of many authors and institutions those are available in online and offline. We offer our sincere appreciation for the learning opportunities provided by those authors and institutions. At last but not least, our heartfelt thanks to all our family members for caring us and for the huge support.

REFERENCES

[1] A. Agarwal and B. Triggs, "*Learning to track 3D human motion from silhouettes*," in Proc. Int. Conf. Mach. Learn., New York, 200, p. 2.

[2] C. Bregler and J. Malik, "*Tracking people with twists and exponential maps*," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Washington, DC, 1998, pp. 8–15.

[3] L. Cheng, S. Wang, D. Schuurmans, T. Caelli, and S. V. N. Vishwanathan, "*An online discriminative approach to background subtraction*," in Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance, 2006, p. 2.

[4] D. M. Gavrila, "*The visual analysis of human movement: A survey,*" Comput. Vis. Image Understand., vol. 73, no. 1, pp. 82–98, 1999.

[5] M. Isard and A. Blake, "*Condensation—Conditional density propagation for visual tracking*," Int. J. Comput. Vis., vol. 29, pp. 5–28, 1998.

[6] J. Kivinen and M. K. Warmuth, "*Exponentiated gradient versus gradient descent for linear predictors*," Inf. Comput., vol.132, no. 1, pp. 1–64, Jan. 1997.

[7] T. B. Moeslund, A. Hilton, and V. Krüger, "*A survey of advances in vision-based human motion capture and analysis*," Comput. Vis. Image Understand., vol. 104, no. 2, pp. 90–126, 2006.

[8] R. Poppe, "*Vision-based human motion analysis: An overview*," Comput. Vis. Image Understand., vol. 108, no. 1–2, pp. 4–18, 2007.

[9] L. Ren, G. Shakhnarovich, J. K. Hodgins, H. Pfister, and P. Viola, "*Learning silhouette features for control of human motion,*" ACM Trans. Graph., vol. 24, no. 4, pp. 1303–1331, 2005.

[10] G. Shakhnarovich, P. Viola, and T. Darrell, *"Fast pose estimation with parameter-sensitive hashing*," in Proc. IEEE Int. Conf. Comput. Vis., Washington, DC, 2003, pp. 750–757

[11] Q. Shi, L.Wang, L. Cheng, and A. J. Smola, "*Discriminative human action segmentation and recognition using semi-Markov model*," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2008, pp. 1–8.

AUTHORS PROFILE

N.V.Manu is currently a PG student in Computer Science Engineering from the Department of Computer Science and Engineering at Chendhuran College of Engineering and Technology, Pudukkottai. He received his Bachelor Degree in Computer Science Engineering from Shree Venkateshwara Hi-Tech Engineering College Othakuthirai, Erode District, and Tamilnadu. His Research areas include Data Mining and data warehoushing, Wireless Sensor Networks and distributed system.

M.Kirithikadevi is currently working as an Asst. Prof. from the Department of Computer Science and Engineering at Chendhuran College of Engineering and Technology, Pudukkottai. She received his Bachelor Degree from Bharathidasan University Trichy and Master Degree in Computer Science & Engineering from Anna University, Trichy. She Published 1 national Conference and 1 international journal. Her main research interests lie in the area of wireless sensor networks, distributed computing.