Literature Review on Background Subtraction Model for Object Detection

Jaskirat Singh, Sandeep Kumar Sahoo

Literature Review on Background Subtraction Model for Object Detection

Jaskirat Singh
Dept. of Computing Science (Multimedia)
University of Alberta
Edmonton, Canada
jaskirat@ualberta.ca

Sandeep kumar Sahoo
Dept. of Computing Science (Multimedia)
University of Alberta
Edmonton, Canada
sksahoo@ualberta.ca

***Abstract*—Background subtraction is widely used technique in computer vision application for object detection and segmentation. Most of the work done on background subtraction considers static camera that identify moving objects by detecting areas in a video that change over time, which is not applicable in real world scenarios like moving camera mounted on the autonomous vehicle.it's a difficult problem due to the motion of both camera and the object.In this paper, we extend the concept of subtracting areas at rest to apply to video captured from a freely moving camera. we approach this problem by exploiting optical flow based background subtraction method.**

***Index Terms*—Image Processing, Video Processing, Computer Vision, Pose detection, Machine Learning, Action Recognition, CNN, Fish-eye lenses, Spatially-varying sensing, virtual reality, adaptive training, rehabilitation, virtual games**

I. Introduction

Recent advancements in the field of computer vision algorithms have lead to a grown interest in many areas of research such as tracking and detecting foreground objects from video. Human motion analysis is the current active research topic in computer vision. It has been implemented successfully in video surveillance, human-machine interface, and virtual reality systems, to recognize human motion, gestures, and analyzing human body structure. This can be achieved by using computer vision algorithms such as the background subtraction method and optical flow method. The main objective of this algorithm is to detect moving objects from the background determined by calculating the difference between two consecutive image frames. The optical flow method estimates the optical flow field and does clustering processing. The feature points are extracted from the current frame and then the clustering algorithm classifies the feature point. However, this approach is sensitive to noise makes it ineffective in real-world scenarios. The background subtraction method calculates the difference between the current image frame and the background frame to get the complete movement information and detect moving objects. Most algorithms are designed for indoor environments, and lacks accuracy in the outdoor environment in real-time which has various non-stationary objects with variations in lighting conditions.

II. Publication Review

A. *Motion detection using background contraints*

In this paper [1], Elnagar and Anup Basu introduced a technique for detecting moving object from a moving camera using background constraints. They considered both an object and camera are moving rigidly with respect to each other as well as the background. The author considered a camera system that mounted on a tilt device that allows rotation and translation. The displacement field at each point in the image is computed. From the above figure, the displacement vector is represented as the sum of translational component and a rotational component. For an active camera, a mapping function is defined to map pixels which corresponds to the same 3D point to the corresponding image plane positions. For each pair of images processed in the image sequence, the image[I(t)] at time t is mapped so as to correspond pixel by pixel with the image[I(t+1)] at time t+1.Let P be a 3D point that is projected on both the images: p(t+1) = [x(t+1),y(t+1)] = [x(t)+ut, y(t)+vt] = p(t) + V Any point lies in the image plane should satisfy the equation. Whereas any point that lies on an independently moving object is unlikely to satisfy the equation. Therefore, detecting moving objects becomes easy depending on this constraint, which can be used for both translation and rotation.

$$
\begin{aligned}
u = u_T + u_R &= \left[\frac{f_x U - xW}{Z}\right] \\
&+ \left[-\frac{xy}{f_y}A + \left(f_x + \frac{x^2}{f_x}\right)B - \frac{f_x}{f_y y}C\right] \\
v = v_T + u_R &= \left[\frac{f_y V - yW}{Z}\right] \\
&+ \left[\left(\frac{-y^2}{f_y} - f_y\right)A + \frac{yx}{f_x}B + \frac{f_y}{f_x x}C\right].
\end{aligned}
$$

Figure 1.

B. Pose recognition using randon transform

In this paper [11], Meghna Singh, Mrinal Mandal and Anup basu proposed a novel method for the recognition of pose estimation using randon transform. The proposed technique first involves background separation step to distinguish foreground object from background model. Then medial skeleton model is obtained from the resultant binary image by thinning the image. Finally the randon transform is applied to detect the orientation of the skeleton line which are not very localized and are rarely straight lines. So the proposed method threshold the Radon transform coefficient to extract the local maxima. Furthermore, the authors designed an efficient matching algorithm, Spatial maxima matching, to recognize the unknown pose by matching it to the known pose. Finally the score is computed between the total no of corresponding match and the no corresponding match. The higher the score, the closer the unknown pose is to the known pose. It is observed that an overall recognition rate of over 87 percent is achieved with leg and arm poses.The promising recognition rates achieved with the parameterized Radon transform emphasize the feasibility of the proposed method.

$$\Re_f\left[r,\alpha\right] = \sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} f_s\left(x,y\right)\delta\left(r - x\cos\alpha - y\sin\alpha\right)$$

Figure 2. Randon Transform

C. Representation, Analysis, and Recognition of 3D Humans:A Survey

In this paper [14], Stefano,Mohamed, Pavan and Anup basu provided an analysis of recent development of representation of human body in 3D data and organised in a taxonomy the representation methods based on the main characteristics of the 3D shape. They classified the methods based on the main characteristics of the 3D shape they capture. They divided the 3D representation into two sub categories: i) spatial, and ii) Temporal. the spatial categories captures properties of the surface and the volume of the face and body. The spatial modeling extracts information from the model's surface. The solutions for surface representation is that of separating them depending on whether they use extrinsic or intrinsic measurements. Methods used in topological representations extract topological information of the human body and face. Skeleton model is widely used to represent human topology. Various techniques like thinning algorithm are used to extract skeleton from the human body. skeletons were used for computing the 3D transformation between two poses of a shape and aligning objects. Methods used in representation based on keypoints by Zaharescu et al. is meshDOG a 3D feature detector that seeks the extrema of the Laplacian of a scale-space representation of any scalar function. The descriptor is able to capture the local geometric and/or photometric properties.

Human representation requires temporal modeling in order to model the temporal dynamics of the event. Trajectory based representation is used to extract features from each 3D frame and then comparing those sequences of those features. The matrix representation is extracted from the overall sequence or its parts, like covariance or Hankel matrices.Markov models are cases where the dynamics can be modeled as a Markovian process and transitions between states can be used to model complex temporal events.Recurrent neural netowrk(RNN) based model learns dynamic patterns from the temporal data.

D. A Multisensor Technique for Gesture Recognition Through Intelligent Skeletal Pose Analysis

In this paper [12], Nathaniel Rossol, Dr. Irene Cheng, and Dr. Anup Basu proposed a novel technique to improve hand pose estimation accuracy using smart-depth sensor technology for tracking hand gestures. Their technique addresses the occlusion issue using pose estimations from multiple leap motion sensors placed at different viewing angles. A classifier is built and trained in offline mode using a premeasured artificial hand. The model then can be used in real-time to intelligently select pose estimations while still running at a high sampling speed of over 120 Hz. Leap motion sensors provide very high 3-D positional accuracy for hands and fingertips at a very high sampling rate. Irrespective of the underlying hand pose recognition technique or sensor used, occlusion is the main problem when using a single vision-based sensor. Hence they proposed fusing data from multiple sensors placed at different viewing angles and analyzing the skeletal poses directly instead of examining the depth maps.

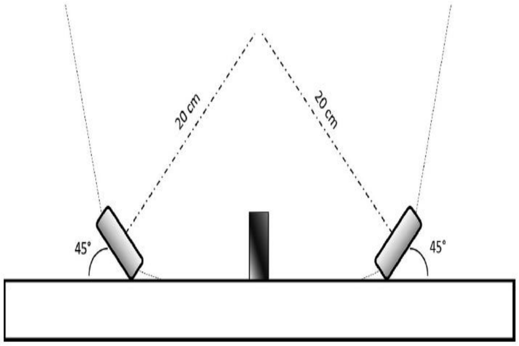

Figure 3. Two-sensor setup

Their multisensor skeletal pose estimation approach is composed of the following steps:

1) They used a trained support vector machine (SVM) model to intelligently determine the optimal pose estimation from an array of sensors. They built this model offline with a training dataset, which uses a feature vector composed of a subset of each sensor's output.
2) Then they converted each sensor's pose output into a global coordinate system so that the poses of all of

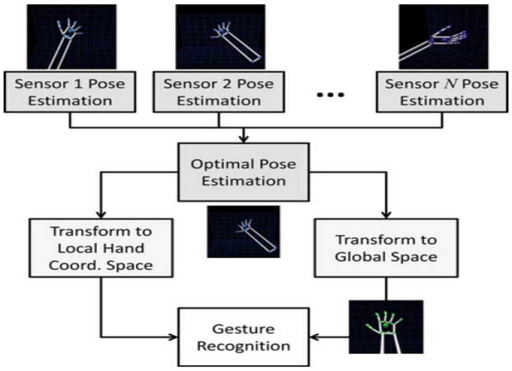

Figure 4. Flowchart of proposed Technique

the fingers are represented in a single unified space. Likewise, they also kept the positions of all of the fingers in a local hand coordinate system, which provides key information for the pose estimation model.

3) Finally, the local hand pose information and the global hand pose information are input into our gesture recognition model so that dynamic and static hand gestures can be tracked.

Ground truth values for hand poses were determined by using an artificial hand model that was manually placed into fixed and known poses. They focused mainly on tracking 3 hand poses (open hand, pinch gesture, tap gesture) that can replace certain commands commonly used on touch-screen interfaces.

The experimental results indicated that their multiple sensor approach reduced the total pose estimation error by 31.5% compared to a single sensor. The average pose estimation error (in mm) is calculated from the sum of the Euclidean distances of each fingertip from its ground truth position divided by the number of fingers. Using this metric, the worst-case and best-case performances were found to be 21.45 and 9.5 mm, their approach has an optimal accuracy of 90.8% on average.

E. Variable Resolution Teleconferencing

In this paper [2], Dr. Anup Basu, Allan Sullivan, and Kevin Wiebe proposed the usage of Variable Resolution(VR) for image compression, and a prototype teleconferencing system based on VR is introduced.

Videoconferencing is a prime application for VR compression for several reasons:

1) Videoconferencing requires fast compression of images.
2) Image quality is not a high priority.
3) The typical ‘talking head’ scene provides for an ideal fovea location.
4) VR provides constant compression, suitable for transmission of images over channels with fixed bandwidth.

The video conferencing system provides transmission of greyscale images from an image server to a display or viewer process. The server process is responsible for capturing, compressing, and transmitting the image. The display process accepts images, uncompresses and displays them on a screen. In addition to VR, an intraframe difference encoding routine provides additional compression. The difference between pixels in successive frames is calculated and, only changed pixels are transmitted. Frame rates on the order of 1.5 - 2 frames per second were achieved between SPARC stations across an Ethernet. With the compression values obtained (up to 98% with interframe encoding), the network can easily handle much higher frame rates. No special boards are required as the system is based on software. The video conferencing system can cheaply and easily be run on an entire network with no additional cost. Work is currently underway to integrate the motion video with audio and drawing programs. The ultimate goal is to produce a true multimedia system that can function in a generic environment, with little or no additional hardware. Such a system will be cheap, easy to upgrade and maintain, and will be portable across many systems.

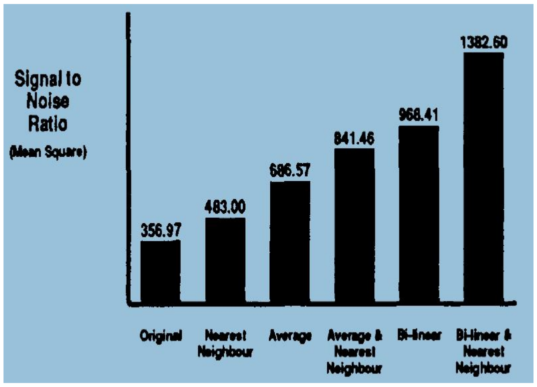

Figure 5. Comparision of SNR Ratio between various Methods

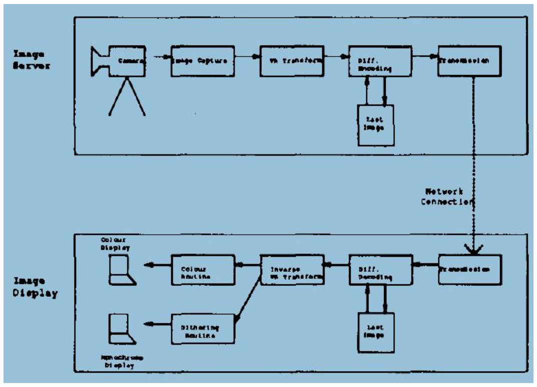

Figure 6. Flowchart of Video Conferencing System Prototype

F. Videoconferencing using Spatially Varying Sensing with Multiple and Moving Foveae

In this paper [4], Dr. Anup Basu and Kevin James Wiebe proposed a new method for videoconferencing using the concept of spatially varying sensing. Various techniques are discussed for combining information obtained from multiple points-of-interest (foveae) in an image. The Variable Resolution(VR) transform has two parameters that affect the resulting image: compression and alpha which controls the distortion at the edges of the image w.r.t. the fovea. A high alpha value gives a sharply defined fovea with a poorly defined periphery; a small alpha value makes the fovea and periphery closer in resolution. The main issue with the basic VR transform is that the images are not rectangular so some areas of the image must be clipped off, or the image will have unused pixels in the corners. The problem is magnified with high alpha values and when the fovea is not located in the center of an image. Two ways are proposed to fix this issue.

1) Modified Variable Resolution(MVR): Use multiple scaling factors, each scaling factor dependent on the angle theta in polar coordinates. This method maintains the isotropic properties of the original formulae but is relatively complex.
2) Cartesian Variable Resolution (CVR): Simplify the formulae by isolating the vertical and horizontal components. In this method, isotropic accuracy is reduced a little but Computational complexity is significantly decreased.

Sometimes there is more than one area of interest to the observer then a decision must be made to either reduce the resolution around each fovea to compensate or retain additional information for each additional fovea, reducing the compression ratio. The quality of an image then depends on the relative position of multiple moving foveae. Two distinct approaches are cooperative and competitive foveae.

cooperative foveae calculates the location of a point in the transformed image w.r.t. each fovea separately. The true location is then found by weighting the two estimated points according to the distance of the original point from the fovea. A higher weight is given to the location calculated using the closer fovea. A unique property of cooperative foveae is the existence of “ghost foveae” between them. If only two foveae exist, then the area of the highest quality in the scene will not only be at these foveae but also along the line connecting the foveae.

In competitive foveae, all foveae compete to calculate the position of a point in the transformed image. The fovea which is closest to any point in the original image will be the one that determines its transformed position.

The videophone component provide transmission of grey scale images from an image server to a display or viewer process. The server process is responsible for capturing, compressing and transmitting the image. The display process accepts images, uncompresses and displays them on a screen. Without using any additional hardware, it provided very high frame rates with very high compression upto (98%) using intaframe encoding. The rate at which it can compress images, especially on machines with limited processing speed, and the high quality present in the foveal region, make it ideal for the multimedia market.

G. Enhancing videoconferencing using spatially varying sensing

Human vision can be characterized as a variable resolution system-the region around the fovea (point of attention) is observed with great detail, whereas the periphery is viewed in lesser detail. In this paper, [3], Dr. Anup Basu, and Kevin James Wiebe presented that spatially varying sensing can indeed be useful in videoconferencing. A system that can incorporate multiple and moving foveae. There are various possible ways of implementing multiple foveae and combining information from them. Some of these alternative strategies are discussed and results are compared. They also show the advantage of using spatially varying sensing as a preprocessor to JPEG. The methods described here can be useful in designing teleconferencing systems and image databases. This paper is an extension of the above two papers [4] and [2]and they have generalized Multiple Foveae and Enhanced JPEG. The concept of cooperative foveae is generalized by creating weights depending on a certain power of distances to foveae. Ghost foveae disappear as the parameter grows, and the quality of the image in the foveae improves at the expense of quality between the foveae. If the power is high, it performs in a similar way as competitive foveae. Spatially varying sensing can be used very effectively to enhance the performance of JPEG at high compression ratios. After compression using JPEG with quality = 1, 98.55% compression with signal to noise ratio (SNR) root mean square (RMS) of 18.25 is achieved. The result of JPEG enhanced by a VR preprocessor achieves 98.56% compression with the SNR (RMS) of 47.23. An alternate way of combining VR with JPEG would be to change the quality parameter in JPEG compression depending on the distances from the foveae.

H. Interactive Multimedia for Adaptive Online Education

In this paper [6], Dr. Irene Cheng, Dr. Anup Basu1, and Randy Goebel presented a broader view of multimedia education, focusing on future applications. The authors envisioned providing publicly accessible education anywhere, at any time, and to anyone. To realize this vision, they proposed the Computer Reinforced Online Multimedia Education (CROME) framework, which integrates the main components of education, including learning, teaching, and testing, as well as adaptive testing and student modeling. Interactive multimedia items such as Drag and Drop, Logical-Mathematical, Language, and Educational items can help improve learning performance by enhancing user satisfaction and engagement. Multimedia content helps to improve concept representation, which is not possible in conventional MCQ and fill-in-the-blank formats. They referred to a curriculum-specific question as an “item”, following each example item can be used either in learning or testing depending on the log-on status: a practice or testing session respectively. CROME framework is designed for multimedia education and uses a combination of web-development tools in order to optimize the constrained resources and to provide user satisfaction. The development kits include Java 2D/3D applets, Javascript, Flash, J2ME, PHP, and MySQL. The choice aims at platform and browser independence. Similar to the multiple-choice format, a generic template is designed for each category of items, which share certain similarities. Multiple questions can be generated by altering the content inserted into the template. 3D items are used in Computer Adaptive Testing, a student's performance can be evaluated more accurately by considering partial scores. A parameter-based strategy is a more general approach for assigning initial difficulties to items. They used Math questions as examples to illustrate the concept. They incorporate a Process Analyzer in the CROME framework. The objective is two-fold: to assist students to improve their problem-solving skills using step-by-step hints and instructions, and to assist teachers in monitoring student performance so that proper help can be provided in time. Every response given by a student is recorded for performance analysis and student modeling. The Process Analyzer comes with a graphical user interface that provides an engaging and motivating learning environment. They implemented multimedia item authoring templates which are plug-ins integrated into the basic CROME framework. There are two main design challenges: First, the uniqueness of each innovative item type, and second, the need to provide different processing pipelines for online and offline item creation. These challenges are overcome by separating the interface implementation into easy-to-manage logical components. Although there is research in the literature focusing on certain target areas in education, their framework is unique in the way it incorporates automatic difficulty level estimation, item generation, process analysis, and testing beyond subject knowledge.

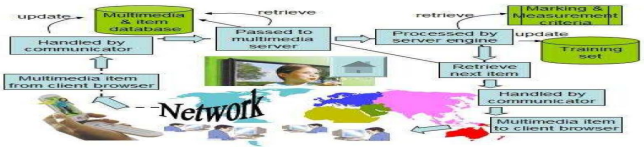
update
Multimedia & item database
retrieve
Handled by communicator
Passed to multimedia server
retrieve
Marking & Measurement criteria
Processed by server engine
update
Training set
Multimedia item from client browser
Retrieve next item
Handled by communicator
Network
Multimedia item to client browser

Figure 7. Flow Diagram

I. *Alternative models for fish-eye lenses*

The human visual system can be characterized as a variable-resolution system. Foveal information is processed at very high spatial resolution whereas peripheral information is processed at low spatial resolution. Various transforms have been proposed to model spatially varying resolution. Unfortunately, special sensors need to be designed to acquire images according to existing transforms and they are computationally expensive. Two models (FET and PFET) of the fish-eye transform are presented in [5] and their validity is demonstrated by fitting them to a real fish-eye lens. The variable resolution (VR) transform proposed in this paper, namely the Fish-Eye Transform (FET) is based on the characteristics of fish-eye lenses, making it as easy to implement using off-the-shelf lenses. Previously log-polar transform was used and its main drawback is that it needs specialized hardware to produce high-resolution images in real-time. They considered an alternative image transform, namely the Fish Eye Transform(FET) which is based on a simplification of the complex logarithmic mapping described by Schwartz. One problem of applying Schwartz's method directly is that image continuity across the vertical meridian is lost, and cannot be recovered easily. They used a simplified VR projection method and computed cortical polar coordinates and cortical Cartesian coordinates. The image continuity is preserved but the disadvantage of the simplified mapping function is that it produces strong anisotropic distortions in peripheral regions for large values of the distortion parameter lambda. Two modeling alternatives are evaluated, one using the FET, the other using a polynomial function (PFET) to approximate the distortion. An average of several estimates of the center of distortion is evaluated in order to obtain a better estimate. Using the polynomial function described by the PFET model and a set of N data pairs (T, p t), the objective of the least-squares method is to minimize an error function x. The problem now is deciding what degree of a polynomial represents accurately the recorded data. A standard way of solving this problem is to use the variance of the fitted model with the observed data. One increases the degree of the polynomial as long as there is a significant decrease in the variance u2. The PFET appears to have a better performance in compensating distortions than FET. Inverse mapping to restore the original image can be obtained using the inverse function of the simplified FET model. There is no simple way to obtain the inverse transform for the polynomial model. In future work, They wished to consider the problem of designing fish-eye lenses that fit a given parametric model and also intend to study the problem of fast and reliable obstacle avoidance using fish-eye lenses.

J. *Nose shape estimation and tracking for model-based coding*

In this paper, [10], Lijun Yin and Anup Basu proposed an efficient method to detect and track the nostril shape and nose-side shape which limits the search region and uses the shape-adaptive model as a remedy for the shrinking effect based on individual deformable templates. Individual templates are designed for the nostril and nose-side. There are 3 main steps in the method.

Firstly, the feature regions are limited to certain areas by using two-stage region-growing methods (i.e. global region growing and local region growing). In global regions growing, a large growing threshold is selected in order to explore more skin areas. In local region growing, the region information is estimated in the individual feature areas. The initial seed pixel of the skin is selected in the local area, and a smaller growing threshold can be selected so that a small skin area surrounding the facial organ is detected, and produces as much feature information as possible.

Secondly, the pre-defined templates are applied to extract the shape of the nostril and nose-side as they are important for composing facial expressions. A geometric template is applied to the nostril region, which is a twisted pair curve with a leaflike shape to detect the nostril shape and the nose side has a shape like a vertical parabola.

After the facial feature regions and shapes on various organs are estimated, a 3D wireframe model can be matched onto the individual face to track the motion of facial expressions.

A camera is mounted on an active platform (pan/tilt) to take an active video sequence, which shows a talking person with an unconstrained background and the camera rotation is less than 5 degrees. Nostril and nose-side are among the key features that are detected. The overall performance gives an indication of the capability of the system to detect most nose features accurately (Only 18 / 270 frames showed position error in the nose area beyond 3 pixels)

K. A Framework for Adaptive Training and Games in Virtual Reality Rehabilitation Environments

In this paper [13], Nathaniel Rossol, Irene Cheng, Walter F. Bischof and Anup Basu proposed a design for virtual rehabilitation system that uses virtual reality training and games to engage patients in effective instructions on the use of wheel chair and introduced a novel framework based on Bayesian networks for self-adjusting adaptive training in virtual rehabilitation environments. In order to allow clinicians to easily design and modify the environments for the patient, an XML scene node based framework was designed and implemented as a 3D-modelling plug-in tool. Using this tool, clinicians can change the layout of training rooms, adjust the position, scale, and rotation of objects, and even assign physical properties such as object mass.The second challenge of the implemented framework was to have the system automatically assess the patient skills and then it intelligently adapts the training simulation accordingly. The designed approach uses Bayesian networks to attempt to determine patient skill levels. The authors performed one experiment in order to determine if the training system was effective enough in teaching the participants to navigate around the obstacles in an environment faster than those who did not receive virtual training. The time taken to complete the obstacle course for the trained participants is mush faster, 81.5 seconds compared to 104.5 seconds for the non trained group.

L. Generating Realistic facial expressions with wrinkles for model based coding

In this paper [15], Lijun Yin and Anup basu presented a texture updating method for detecting and tracking facial textures in active wrinkle areas and mouth-eye areas. They used deformable template matching method to find the fiducial points on a face and developed mesh matching algorithm to track facial expression. A wire-frame model is first adapted onto the face images to track the movement of the expression. Then the texture of interest(TOI) is estimated based on fiducial points detected, such as corners of facial organs. Finally, the facial texture are synthesized using these texture of interest. The fiducial point tracking and model adaptation are performed automatically.

The adaptation procedure consists of two stages: 1)facial fiducial point estimation that produces a correct matching of a face model with the face that leads to correct matching of facial expressions and thus result in a correct selection of textures of interest.Using the information extracted from head motion and facial features (i.e., fiducial points, eye contour, mouth contour, and face silhouette), a 3D wireframe model can be fitted to the moving face. They developed a so-called “coarse-to-fine” adaptation algorithm using the extended dynamic mesh to implement the model adaptation procedure.

M. Perceptually Guided Fast Compression of 3D Motion Capture Data

In this paper [9], A. Firouzmanesh, I. Cheng, and A. Basu proposed a compression algorithm for motion capture, transmission for the interactive online 3D environment, and real-time synthesis on mobile data. The proposed method for the compression technique used here is wavelet coding as it is one of the efficient approaches for encoding multimedia data. The main idea of the compression technique is to select different numbers of coefficients for different channels of data based on the importance of each channel. One of the factors to evaluate animation quality is the viewer’s region of interest. In order to locate the high attention regions, the author used the Interactivity-Stimulus-Attention Model that explains how interactivity stimulates immersion of cognitive resources. the processing steps involves normalizing the bone values.Then calculating the variation of each channel using the below formula. Combine the effects of relative bone length and variation to obtain a weighted quantity. Finally compute the largest coefficient of each channel and Use run-length to encode the wavelet coefficients.

$$v_i = \frac{\sum_{p=m_i*W}^{m_i*W+W-1} \left| c_{i,p+1} - c_{i,p} \right|}{W}$$

Figure 8. Coefficient

N. Segmentation of Arterial Walls in Intravascular Ultrasound Cross-Sectional Images Using Extremal Region Selection

In this paper [7], Mehdi Faraji, Irene Cheng, Iris Naudin and Anup Basu introduced region detection strategy for segmenting the acquired IVUS image frames and experimented with the feature extraction method called External Regions of Extremum levels can segment the luminal and media-adventitia borders in IVUS frames. This proposed strategy involves below steps:

i)remove the typical artifacts of IVUS frames, such as ring-down effects and calibration squares ii)ROI are extracted and are filtered based on their types

iii)Perform region selection preocedure and segment the two regions labelling them as luminal and media

iv)Finally, both regions are traced by using contours.

The author applied their proposed technique on the publicly available dataset containing 326 IVUS B-mode images. The distance between the resultant extracted features with the original lumen and media are 0.22mm and 0.45mm respectively. The outcome of the experiment showed that the method was able to segment the region of interest with < 0.3 mm HD to the gold standards. Using this method, the segmentation results are pretty accurate and can be applied to analyse the internal structure of human arteries from Intravascular Ultrasound images taken with a 20 MHz catheter probe.

O. Hough transform for feature detection in panoramic images

Mark Fiala, Anup Basu [8] introduced an approach to detect features in panaromic non-SVP(single viewpoint) images using a modified Hough transform. The author proposed a new hough transform parameter space for detecting lines in images formed from non-SVP catadioptric panoramic image sensors. The proposed method locates horizontal line features. By mapping edge pixels to a new two-dimensional (2D) parameter space for each mirror lobe, the existence and location of horizontal lines can be found. The motivation for recognizing horizontal and vertical line segments is for its use in 3D model creation for man-made environments where the majority of line edge features are either horizontal or vertical. The double-lobed catadioptric optical arrangement presented in this paper allows two viewpoints to be captured by one image capture device (camera). Each point in this new parameter space corresponds to a plane in space instead of line. The family of planes which are viewed by a panoramic catadioptric sensor as a curved line have 2D and thus this Panoramic Hough transform space is 2D. Experiment was carried out which involves the below steps:

- For each point in the camera image convert to polar coordinates relative to Ucenter, Vcenter for pixels in the range Rmin–Rmax.

-Add the edge magnitude of each point in the image to all possible mapped points in the Hough image.

-Locate peaks in this Panoramic Hough image and declare the existence of a plane corresponding to Rfall.

The results demonstrated the robust performance of this transform and indicated one likely application being panoramic stereoimagery.

References

[1] ANUP BASU ASHRAF ELNAGARf. Motion detection using background constraints. In ., pages 795–799, 2007.

[2] A. Basu, A. Sullivan, and K. Wiebe. Variable resolution teleconferencing. In *Proceedings of IEEE Systems Man and Cybernetics Conference - SMC*, volume 4, pages 170–175 vol.4, 1993.

[3] A. Basu and K. Wiebe. Enhancing videoconferencing using spatially varying sensing. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 28(2):137–148, 1998.

[4] A. Basu and K. J. Wiebe. Videoconferencing using spatially varying sensing with multiple and moving foveae. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 2 - Conference B: Computer Vision Image Processing. (Cat. No.94CH3440-5)*, pages 30–34 vol.3, 1994.

[5] Anup Basu and Sergio Licardie. Alternative models for fish-eye lenses. *Pattern Recognition Letters*, 16(4):433–441, 1995.

[6] I. Cheng, A. Basu, and R. Goebel. Interactive multimedia for adaptive online education. *IEEE MultiMedia*, 16(1):16–25, 2009.

[7] Mehdi Faraji, Irene Cheng, Iris Naudin, and Anup Basu. Segmentation of arterial walls in intravascular ultrasound cross-sectional images using extremal region selection. In ., pages 795–799, 2007.

[8] Mark Fiala and Anup Basu. Hough transform for feature detection in panoramic images. *Pattern Recognition Letters*, 23(14):1863–1874, 2002.

[9] A. Firouzmanesh, I. Cheng, and A. Basu. Perceptually guided fast compression of 3d motion capture data. In ., pages 30–34 vol.3, 1994.

[10] Lijun Yin and A. Basu. Nose shape estimation and tracking for model-based coding. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, volume 3, pages 1477–1480 vol.3, 2001.

[11] Mrinal Mandal Meghna Singh and Anup basu. Pose recognition using randon transform. In ., pages 795–799, 2007.

[12] N. Rossol, I. Cheng, and A. Basu. A multisensor technique for gesture recognition through intelligent skeletal pose analysis. *IEEE Transactions on Human-Machine Systems*, 46(3):350–359, 2016.

[13] Nathaniel Rossol, Irene Cheng, Walter F. Bischof, and Anup Basu. A framework for adaptive training and games in virtual reality rehabilitation environments. In ., VRCAI '11, page 343–346, New York, NY, USA, 2011. Association for Computing Machinery.

[14] PAVAN TURAGA STEFANO BERRETTI, MOHAMED DAOUDI and Anup Basu. Representation, analysis, and recognition of 3d humans: A survey. In ., pages 795–799, 2007.

[15] Lijun Yin and Anup Basu. Generating realistic facial expressions with wrinkles for model-based coding. *Computer Vision and Image Understanding*, 84(2):201–240, 2001.