Certificate

It is certified that the work contained in the thesis titled "DESIGN AND DEVELOPMENT OF SOME METHODS AND MODELS FOR CROWD ANALYSIS USING COMPUTER VISION AND DEEP LEARNING APPROACHES" has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive, Candidacy and State-of-the-art seminar.

Rajeev 4.7.22

Signature of Supervisor

Prof. Rajeev Srivastava

Professor, Department of Computer Science and Engineering

Indian Institute of Technology (BHU), Varanasi

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

Declaration by the Candidate

I, ***Santosh Kumar Tripathy*** certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of **Prof. Rajeev Srivastava** from ***July 2018 to July 2022***, at the ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not wilfully lifted up any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and included them in this thesis and cited as my own work.

Date: 4.7.22

Place: IIT (BHU), Varanasi

Santosh Kumar Tripathy
Signature of the Student

(Santosh Kumar Tripathy)

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Rajeev 4.7.22

Signature of Supervisor
(Prof. Rajeev Srivastava)
Professor, Department of Computer Science and Engineering
Indian Institute of Technology (BHU), Varanasi

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

Signature of Head of Department

(Prof. Sanjay Kumar Singh)

आचार्य व विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)

Copyright Transfer Certificate

Title of the Thesis: *Design and Development of Some Methods and Models for Crowd Analysis using Computer Vision and Deep Learning Approaches.*

Name of the Student: Santosh Kumar Tripathy

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University), Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the DOCTOR OF PHILOSOPHY.

Santosh Kumar Tripathy

Date: 04-07-2022 Signature of the Student

Place: IIT (BHU), Varanasi (Santosh Kumar Tripathy)

Acknowledgement

I would like to express my deepest gratitude to all those people who helped me directly or indirectly to complete this thesis work. Firstly, I would like to express my heartfelt gratitude to my supervisor, **Prof. Rajeev Srivastava**, for being a great mentor and the best adviser I could ever have. His advice, encouragement and critics are source of innovative ideas, inspiration are causes behind the successful completion of this Thesis work. The confidence shown on me by him was the biggest source of inspiration for me. It has been a privilege working with him from several years.

I would like to express my deepest appreciation to my research progress evaluation committee members Prof. K. K. Shukla of the Department of Computer Science and Engineering and Prof. Subir Das, Department of Mathematical Sciences IIT (BHU), for providing continuous support, encouragement, and advice.

I express my sincere thanks to all the Professors, Deans, office staff, supporting staff and PhD Research Scholars of Indian Institute of Technology (BHU) Varanasi India. I express my gratitude to Director, Registrars, Deans, Heads, and Student Alumni of the Indian Institute of Technology (BHU) Varanasi.

Special thanks to Dr. Subodh Srivastava (Assistant Professor, NIT, Patna) for his continuous support during my study. I'd like to extend my gratitude to Dr. Ankit Jaiswal (Assistant Professor, Bennet University), Dr. Roshan Singh (System Analyst IIT BHU Varanasi), Dr. Vibhav Prakash Singh (Assistant Professor MNNIT Allahabad), Dr. Jani Kuntesh Ketan (GEC, Gujrat), Dr. Gargi Srivastava (Assistant Professor, RGPIT Jais Amethi). I also wish to thank my lab members Saurav Arora, Divya Singh and Suryakant Singh for their consistent support and help during my research work.

I extend special thanks to the non-teaching staff in the Department, particularly, Mr. Manoj Singh, Mr. Ravi Bharti, Mr. Subham Pandey, Mr. Prakhar for their consistent support.

My parents, Smt. Rama Tripathy and Shri Trilochan Tripathy, who gave me the power and brain to work out on this research and their help at every level made me to see this success.

I owe thanks to my wife, Mrs. Sagarika Mishra and my son Sriyansh Aditya Tripathy for their continued and unfailing love, support and understanding during my pursuit of Ph.D. degree that made the completion of thesis possible. Words are insufficient to express my profound sense of gratitude to my loving brother, sister, sister-in-law, nephews and neice. Their encouragement and support gave me physical and moral strength throughout my career as well in the present research. Finally, I would like to wind up by paying my heartfelt thanks and prayers to the Almighty Lord Shiva, Mata Parvati, Sri Laxmi Narayan, Mata Saraswati, Sri Ganesh and Mata Santoshi for their unbound love and grace.

Santosh Kumar Tripathy

Santosh Kumar Tripathy

Abstract

In today's world, the exponential growth of the worldwide population has caused a considerable increase in crowd densities in places like rallies, public speeches, stadiums, mass transit, and tourist or pilgrim sites. Such places are vulnerable to crowd disasters. Crowd disasters are controlled by efficiently analyzing crowd scenes. The crowd analysis (CA) is beneficial for drawing better crowd management strategies, and public space design. It helps in developing an AI-based visual surveillance system to provide security and safety to the crowd. However, the CA is a tedious task that comprises several correlated tasks, out of which crowd counting and density estimation (CCDE), crowd congestion-level analysis (CCA), and crowd behavior analysis (CBA) are the minimum tasks required to control crowd disasters. Most existing solutions for crowd analysis are done manually in real world scenario, which are very complex and prone to error. Recently, deep learning techniques have significantly improved the performance of several computer vision tasks and contributed considerably to the development of AI-based solutions. Thus, the drawback of the manual process for the CA can be overcome by developing automated AI-based solutions using computer vision and deep learning solutions.

Further, CA using several single-task AI models will incur computational complexity overheads, which can be minimized by drawing better multitasking CA models. So, to control crowd disasters and provide security and safety to the crowd, efficient solutions using computer vision and deep learning approaches for significant tasks of CA such as CCDE, CCA, CBA, and multitasking CA, are required. However, as per the literature review reported later, the performance of several computer vision-based CA models is mainly affected by cluttered background, varying crowd densities, crowd

shape changes due to perspective distortion, illumination changes, and lack of availability of largescale CA datasets. Therefore, to address the issues mentioned above, the problem statement of the thesis is defined as the design and development of some methods and models for crowd analysis using computer vision and deep learning techniques.

This thesis mainly focuses on studying and analyzing the state-of-the-art CA techniques, finding their advantages and limitations, and proposing new methods and models to accomplish the objectives. This thesis aims to conduct a comprehensive literature review on four tasks of CA, i.e., CCDE, CCA, CBA, and multitasking CA. For each task, various methods and models concerning current research trends have been analyzed by mentioning their pros and cons and identifying possible research scopes. Various models using computer vision and deep learning approaches have been proposed in this thesis to fulfill the research scopes of each of the four tasks of CA.

The first contribution in the thesis is related to the task of Crowd Counting and Density Estimation (CCDE), where two models using deep learning techniques have been proposed. The first proposed model for CCDE is an **A**ttentive **M**ulti-**S**tream **CNN** (AMS-CNN) for video-based crowd counting. The main objective behind the AMS-CNN is to enhance the feature representation for crowd videos, minimize the effect of the cluttered background, and design attention mechanism for each stream to improve the counting performance. The second proposed model for CCDE is a novel cascaded deep architecture with weak supervision for video crowd counting. This model comprises two deep models named **L**ocal **D**ensity-map **R**egressor (LDR) and **G**lobal **C**rowd **C**ounting **R**egression (GCCR) modules. The LDR focuses on extracting multiscale spatial-temporal features using a multicolumn 3D Atrous (Dilated) CNN to tackle crowd shape changes due to perspective distortion. It also minimizes the effect of the cluttered background using a **H**ead **A**ttention **M**odule (HAM). The LDR considers the local distribution of

crowds and generated crowd density maps. On the other hand, the GCCR model is trained in a weakly supervised manner to exploit global crowd properties from the predicted density maps to obtain final crowd counting using a multi-layer perceptron neural network.

The second contribution to the thesis is related to the task of Crowd Congestion Level Analysis (CCA), where deep learning-based real-time **T**wo **I**nput **S**tream **M**ulti-**C**olumn **M**ulti-**S**tage **CNN** (TIS-MCMS-CNN) is proposed. In the proposed method, each of the two streams of TIS-MCMS-CNN has been built with three columns of multi-layers of CNN with different receptive fields to extract multiscale spatial and temporal features from two cues of video frames, i.e., frame and the flow magnitude of the frame. The extracted multiscale features are also known as scale-invariant features, which can handle crowd shape change due to perspective distortion. For experimental analysis, a dataset for the CCA is prepared using three publicly available benchmark crowd datasets. It is observed that the TIS-MCMS-CNN can process the frames in real-time.

The third contribution to the thesis is related to the task of Crowd Behavior Analysis (CBA). Under this contribution, two models have been proposed out of which the first model i.e., a **Mu**ltiscale **S**patial-**T**emporal 3D Atrous-Net with **P**CA-guided **O**C-**S**VM (MuST-POS) is designed using conventional machine learning and deep learning approaches to classify normal and panic crowd behaviors. On the other hand, the second model is a **T**wo-**S**tream **M**ultiscale **D**eep **A**rchitecture (TS-MDA) is developed using deep learning techniques for multiclass crowd behavior prediction. Both the models resolve the issue of human shape variation in the crowd videos, while the second model takes measured to minimize the effect of cluttered backgrounds in the crowd video.

The fourth contribution to the thesis is related to multitasking Crowd Analysis (CA), where an efficient multitasking deep model is proposed using a backbone structure

of multi-layer **D**epth-wise **S**eparable **CNN** (DSCNN) to predict crowd behaviors with crowd counting. The proposed model exploits spatial-temporal features with **F**low **A**ttention **B**locks (FABs) to provide optical flow attention to different scales features of the backbone network. The FABs focus on the moving pixels, thereby minimizing the effect of cluttered background. The multiscale flow attentive features are fused to handle crowd shape variation and perform multitask CA using a feed-forward network. In addition, a largescale multitasking CA dataset is also developed from the available benchmark crowd behavior datasets. Around 1,20,000 frames have been annotated to obtain ground truth crowd counting information.

The proposed models have been implemented, experimented and evaluated on several publicly available datasets. Experimental results and analysis show that the proposed models perform better than state-of-the-art methods reported in the literature. Extensive ablation studies for each of the proposed models have also been conducted to show the influence of several components of the proposed models.

The thesis concludes with an overall conclusion of the proposed research work, followed by a discussion of possible future research scopes in the areas of CA.

Table of content

List of Figures

List of Tables

List of Symbols

β_1	Decay rates of first moment
β_2	Decay rates of second moment
ξ	relaxation variable
ω	normal to the hyperplane of the feature space
$\varphi()$	maps the primitive feature space to a higher dimension
#	Cardinality
$v^t(x)$	the pixel corresponding to location x in the t^{th} resized frame
η	the learning rate
×	Not Applicable
✓	Applicable
.*	Elementwise multiplication
$\mathbb{R}^d$	Set of Real number of d dimension
$\emptyset$ and θ	Set of learnable parameters of deep models
λ	Lagrangian multiplier
$\cup$	Union
$\cap$	Intersection
Tanh	The activation function Tanh
L	Loss function
$Sum(.)$	The sum of all elements of the feature map(s)
$\odot$	Represents elementwise multiplication
nu	Hyperparameter of OC-SVM which defines the percentage of training samples will be treated as outliers.
$Concatenate()$	Concatenate the feature maps.
$CONCATE()$	Concatenate the feature maps.
	Concatenate the feature maps.
	Max Pooling
	Sum of all the elements of the given feature map(s)
	Convolution Operation
$argmin$	Returns the input for minimum output.

List of Abbreviations

CA	Crowd Analysis
CCDE	Crowd Counting and Density Estimation
CCA	Crowd Congestion-level Analysis
CBA	Crowd Behaviour Analysis
CBP	Crowd Behaviour Prediction
CNN	Convolutional Neural Network
Conv2D/2DCNN/Conv_2D	2D Convolutional Neural Network
Conv3D/3D CNN/Conv_3D	3D Convolutional Neural Network
FCN	Fully Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
ConvLSTM	Convolutional Long Short-Term Memory
GAN	Generative Adversarial Neural Network
AI	Artificial Intelligent
HOG	Histogram of Oriented Gradients
HOF	Histogram of Optical Flow
LBP	Local Binary Patterns
AMS-CNN	Attentive Multi-Stream CNN
DMR	Density Map-based Regression
SCR	Single Count-based Regression
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
OC-SVM	One Class Support Vector Machine
SVR	Support Vector Regression
GLCM	Gray Level Co-occurrence Matrix
SIFT	Scale-invariant Feature Transform
GPR	Gaussian Process Regression
GMM	Gaussian Mixture Model
MuST-POS	Multiscale Spatial -Temporal 3D Atrous Net with PCA guided OC-SVM
MuST-3AN	Multiscale Spatial -Temporal 3D Atrous Net

TS-MDA	Two-Stream Multiscale Deep Architecture
AMS-CNN	Attentive Multi-Stream CNN
ML	Machine Learning
OCC	One Class Classification
MCC	Multi Class Classification
HOT	Histogram of Oriented Tracklets
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
FPR	False Positive Rate
ER	Error Rate
NA	Not Applicable
Mean-ACC	Mean Accuracy
Acc	Accuracy
MOG	Mixture of Gaussian
FV	Frame Volume
RGB	Red Green Blue
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
ReLU	Rectified Linear Unit
AP	Average Pooling
3DAP	3D Average Pooling
MP	Max Pooling
3DMP	3D Max Pooling
FC	Fully Connected
DC	Densely Connected
GAP	Global Average Pooling
Dil_ConvLSTM2D	Dilated Convolutional LSTM 2D
Dil_Conv3D	Dilated 3D Convolution
TADM	Temporal Attentive Density Map
FADM	Foreground Attentive Density Map
SADM	Spatial Attentive Density Map