Contextual Analysis of Videos

Synthesis Lectures on Image, Video, and Multimedia Processing

Editor
Alan C. Bovik, ***University of Texas, Austin***

The Lectures on Image, Video and Multimedia Processing are intended to provide a unique and groundbreaking forum for the world's experts in the field to express their knowledge in unique and effective ways. It is our intention that the Series will contain Lectures of basic, intermediate, and advanced material depending on the topical matter and the authors' level of discourse. It is also intended that these Lectures depart from the usual dry textbook format and instead give the author the opportunity to speak more directly to the reader, and to unfold the subject matter from a more personal point of view. The success of this candid approach to technical writing will rest on our selection of exceptionally distinguished authors, who have been chosen for their noteworthy leadership in developing new ideas in image, video, and multimedia processing research, development, and education.
In terms of the subject matter for the series, there are few limitations that we will impose other than the Lectures be related to aspects of the imaging sciences that are relevant to furthering our understanding of the processes by which images, videos, and multimedia signals are formed, processed for various tasks, and perceived by human viewers. These categories are naturally quite broad, for two reasons: First, measuring, processing, and understanding perceptual signals involves broad categories of scientific inquiry, including optics, surface physics, visual psychophysics and neurophysiology, information theory, computer graphics, display and printing technology, artificial intelligence, neural networks, harmonic analysis, and so on. Secondly, the domain of application of these methods is limited only by the number of branches of science, engineering, and industry that utilize audio, visual, and other perceptual signals to convey information. We anticipate that the Lectures in this series will dramatically influence future thought on these subjects as the Twenty-First Century unfolds.

Contextual Analysis of Videos
Myo Thida, How-lung Eng, Dorothy Monekosso, and Paolo Remagnino
2013

Wavelet Image Compression
William A. Pearlman
2013

Remote Sensing Image Processing
Gustavo Camps-Valls, Devis Tuia, Luis Gomez-Chova, Sandra Jimenez, and Jesus Malo
2011

The Structure and Properties of Color Spaces and the Representation of Color Images
Eric Dubois
2009

Biomedical Image Analysis: Segmentation
Scott T. Acton and Nilanjan Ray
2009

Joint Source-Channel Video Transmission
Fan Zhai and Aggelos Katsaggelos
2007

Super Resolution of Images and Video
Aggelos K. Katsaggelos, Rafael Molina, and Javier Mateos
2007

Tensor Voting: A Perceptual Organization Approach to Computer Vision and Machine Learning
Philippos Mordohai and Gerard Medioni
2006

Light Field Sampling
Cha Zhang and Tsuhan Chen
2006

Real-Time Image and Video Processing: From Research to Reality
Nasser Kehtarnavaz and Mark Gamadia
2006

MPEG-4 Beyond Conventional Video Coding: Object Coding, Resilience, and Scalability
Mihaela van der Schaar, Deepak S Turaga, and Thomas Stockhammer
2006

Modern Image Quality Assessment
Zhou Wang and Alan C. Bovik
2006

Biomedical Image Analysis: Tracking
Scott T. Acton and Nilanjan Ray
2006

Recognition of Humans and Their Activities Using Video
Rama Chellappa, Amit K. Roy-Chowdhury, and S. Kevin Zhou
2005

Contextual Analysis of Videos

Myo Thida, How-lung Eng, Dorothy Monekosso, and Paolo Remagnino

ISBN: 978-3-031-01121-4 paperback
ISBN: 978-3-031-02249-4 ebook

DOI 10.1007/978-3-031-02249-4

A Publication in the Springer series
SYNTHESIS LECTURES ON IMAGE, VIDEO, AND MULTIMEDIA PROCESSING

Lecture #14
Series Editor: Alan C. Bovik, *University of Texas, Austin*
Series ISSN
Synthesis Lectures on Image, Video, and Multimedia Processing
Print 1559-8136 Electronic 1559-8144

Contextual Analysis of Videos

Myo Thida
Institute for Infocomm Research

How-lung Eng
Institute for Infocomm Research

Dorothy Monekosso
The University of the West England

Paolo Remagnino
Kingston University, UK

SYNTHESIS LECTURES ON IMAGE, VIDEO, AND MULTIMEDIA PROCESSING #14

ABSTRACT

Video context analysis is an active and vibrant research area, which provides means for extracting, analyzing and understanding behavior of a single target and multiple targets. Over the last few decades, computer vision researchers have been working to improve the accuracy and robustness of algorithms to analyze the context of a video automatically. In general, the research work in this area can be categorized into three major topics: 1) counting number of people in the scene 2) tracking individuals in a crowd and 3) understanding behavior of a single target or multiple targets in the scene.

This book focuses on tracking individual targets and detecting abnormal behavior of a crowd in a complex scene. Firstly, this book surveys the state-of-the-art methods for tracking multiple targets in a complex scene and describes the authors' approach for tracking multiple targets. The proposed approach is to formulate the problem of multi-target tracking as an optimization problem of finding dynamic optima (pedestrians) where these optima interact frequently. A novel particle swarm optimization (PSO) algorithm that uses a set of multiple swarms is presented. Through particles and swarms diversification, motion prediction is introduced into the standard PSO, constraining swarm members to the most likely region in the search space. The social interaction among swarm and the output from pedestrians-detector are also incorporated into the velocity-updating equation. This allows the proposed approach to track multiple targets in a crowded scene with severe occlusion and heavy interactions among targets.

The second part of this book discusses the problem of detecting and localizing abnormal activities in crowded scenes. We present a spatio-temporal *Laplacian Eigenmap* method for extracting different crowd activities from videos. This method learns the spatial and temporal variations of local motions in an embedded space and employs representatives of different activities to construct the model which characterizes the regular behavior of a crowd. This model of regular crowd behavior allows for the detection of abnormal crowd activities both in local and global context and the localization of regions which show abnormal behavior.

The last chapter suggests a number of research directions to be pursued for future work.

KEYWORDS

video context analysis, interactive Swarms, particle swarm optimisation, multi-target tracking, social behavior, crowded scenes, abnormality detection, visual surveillance, manifold embedding, crowd analysis, spatio-temporal *Laplacian Eigenmap*

Contents