ON HARNESSING VIDEO CONTENT

by

GAURAV HARIT

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfillment of the requirements of the degree of

Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY, DELHI

INDIA

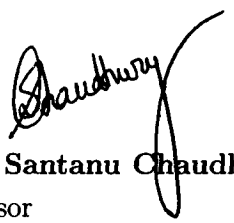
JANUARY 2007

. Video Content

Certificate

This is to certify that the thesis titled **"On Harnessing Video Content"** being submitted by Gaurav Harit to the Department of Electrical Engineering, Indian Institute of Technology, Delhi, for the award of the degree of Doctor of Philosophy, is a record of bona-fide research work carried out by him under my guidance and supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted to any other university or institute for the award of any degree or diploma.



Prof. Santanu Chaudhury
Professor
Department of Electrical Engineering
Indian Institute of Technology
New Delhi - 110 016

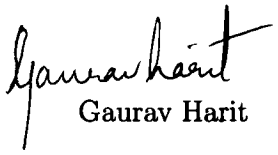
Acknowledgments

I am grateful..

- To my supervisor, Prof Santanu Chaudhury, for his guidance, encouragement, impetus, highly positive attitude, and for his deep insight into the problems and the ideas which have formed the core of this thesis. I owe him immense gratitude for allowing me to work at my own pace, for carefully reviewing and giving valuable feedback for all my papers so that they got accepted promptly, for supporting me financially, for funding all my conference travels, and for giving me full freedom in the Multimedia lab. His superior mentoring skills and efficient thinking have played a great role in pushing me through when I would reach a stagnation and keeping me actively involved. My acknowledgment in words will always stand pale to what I have inculcated while working under him.
- To my SRC members, Prof S. D. Joshi, Prof Prem Kalra, and Prof Subrat Kar, for their useful suggestions while monitoring the progress of my research.
- To Prof. J. K. Chatterjee, Head, Electrical Engineering Department, Prof A. N. Jha, Chairman DRC, and Dr. S Prakriya, for their prompt help whenever I approached them with a problem.
- To Mrs. Usha Bhola, technician in-charge of Multimedia Lab for arranging all what I needed, and for being a counsellor. To Mr. Fateh Singh, former technician in-charge of our lab for his care during the initial years of my PhD. To Sushilaji, helper in our lab, and to the mess staff of Kumaon hostel, where I resided for 5 years, for their cheerful greetings, good wishes, and every possible help.
- To Dr. Hiranmay Ghosh, for his useful suggestions on how to improve my work related to deploying video ontologies. To Geetika, for sharing her research experiences which counted a lot.
- To Venu Kesam, for making me most comfortable in his little sweet home, when I was having accommodation problems.
- To my friends, since my school days, my college times, and here at IIT Delhi,

who have inspired me to believe in myself. They will surely recognize themselves.

- To my relatives, who have always wanted that I'd be with them during all the holidays, but had to contend with just few hours.
- To my father Prof K. C. Harit, my mother Smt. Usha Harit, and my younger sister Pratibha, for stabilizing my mood swings, for boosting my tempo, and most of all for their prayers which have got me sail across the formidable ocean.



Gaurav Harit

New Delhi

January 8, 2007

Abstract

Video content is rich in information. Meaning from the video content can be extracted at different levels of abstraction. In this thesis, we develop techniques for processing and analyzing video at sub-object level, object level, and cognitive level. We extract spatio-temporal regions which depict homogeneity in color. A video shot segmented into tracks of homogeneous color regions is then processed in a grouping framework to identify the subjects. We then formulate a measure of prominence for the subjects and interpret the overall scene into general categories by taking into account the perceptual attributes of subjects as well as the context.

We also develop formalisms for representing video content knowledge in a formal ontological framework and deploying the knowledge for automatic semantic annotation using Bayesian network as a tool. The aggregate of techniques presented in this thesis is an attempt to harness video content by enabling effective conceptual access to the extracted semantic meta-data which characterizes video information.

Our contributions are briefly outlined as follows:

1. We have developed a novel clustering methodology, termed as Decoupled Semantics Clustering Tree (DSCT), which uses decoupled clustering in different feature sub-spaces. We have done extensive experiments that show that the proposed clustering methodology overcomes problems like model selection, cluster suitability, smoothing, etc in traditional clustering methods.

2. We have developed a perceptual grouping algorithm which makes use of a specified spatio-temporal grouping model and identifies the perceptual clusters or subjects in the scene. Gestalt principles have been used in the formulation of perceptual associations and a spatio-temporal coherence model, and thus our

grouping scheme is able to identify general classes of subjects which come out as perceptually distinct with strong associations between parts within the grouping. We have also shown how domain knowledge can be integrated with the grouping process to benefit detection of subjects, as well as to perform grouping with recognition.

3. We have developed a novel concept of perceptual prominence for subjects in the scene. Prominence is closely linked to how one cognitively interprets a scene. Our formalism for scene interpretation makes use of perceptual prominence and mise-en-scène features.

4. For representing video content knowledge in an ontology, we have proposed new constructs to the Web Ontology Language (OWL). We call our extensions as Multimedia-OWL (M-OWL). We propose a probabilistic reasoning framework to infer concepts in presence of uncertain relations between concepts and observable video features.

5. We develop technique for automatic creation of semantic-metadata using a Bayesian network which encapsulates the semantic content knowledge. Video annotations are exploited to provide conceptual access – browsing and querying in a video collection.

Contents