

SPRINGER BRIEFS IN ELECTRICAL AND
COMPUTER ENGINEERING · SIGNAL PROCESSING

Long Xu
Weisi Lin
C.-C. Jay Kuo

Visual Quality Assessment by Machine Learning



Springer

SpringerBriefs in Electrical and Computer Engineering

Signal Processing

Series editors

Woon-Seng Gan, Singapore, Singapore

C.-C. Jay Kuo, Los Angeles, USA

Thomas Fang Zheng, Beijing, China

Mauro Barni, Siena, Italy

More information about this series at <http://www.springer.com/series/11560>

Long Xu · Weisi Lin · C.-C. Jay Kuo

Visual Quality Assessment by Machine Learning



Springer

Long Xu
National Astronomical Observatories
Chinese Academy of Sciences
Beijing
China

C.-C. Jay Kuo
University of Southern California
Los Angeles, CA
USA

Weisi Lin
Nanyang Technological University
Singapore
Singapore

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISSN 2196-4076 ISSN 2196-4084 (electronic)
SpringerBriefs in Signal Processing
ISBN 978-981-287-467-2 ISBN 978-981-287-468-9 (eBook)
DOI 10.1007/978-981-287-468-9

Library of Congress Control Number: 2015935610

Springer Singapore Heidelberg New York Dordrecht London
© The Author(s) 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Science+Business Media Singapore Pte Ltd. is part of Springer Science+Business Media
(www.springer.com)

Preface

After visual signals (referring to image, video, graphics, and animation) are captured or generated, they undergo a variety of processings, including compression, enhancement, editing, retargeting, and transmission. These processes change the quality of visual signals. To measure the extent of such changes, visual quality assessment (VQA) has gained popularity as a hot research topic during the last decade. The psychological and physiological research results have been plugged into this research field to provide fundamental knowledge of the visual perception mechanism and theoretical support for developing VQA models. In addition, the newly developed computer vision, artificial intelligence, and machine learning techniques have been applied to this research field; they have cooperated with psychological/physiological principles to produce more powerful and general computational models of VQA.

On the basis of acquired knowledge about the human visual system (HVS) to visual perception, a variety of VQA approaches have been developed in seeking agreement with the perception mechanism of the HVS to visual signal stimulus. However, due to the sophisticated nature of the HVS, it is difficult to model the HVS response and perception to image/video features directly and explicitly in general, with the current understanding and knowledge on the HVS. Many model-based and signal-driven VQA systems have been developed with strong assumptions. Therefore, machine learning can be used to emulate the mechanisms of complicated models as a new trend of VQA development, without resorting to prior, unrealistic assumptions. There have been a variety of machine learning-based VQA approaches in the recent literature with increase in necessary databases publicly available. The learning-based VQA has become an emerging category of VQA, apart from the model-based and signal-driven ones.

The content of this book is arranged into six chapters. Chapter 1 is the introduction to VQA. The fundamental knowledge, history, and major approaches (including model-based, signal-driven, and learning oriented ones) of VQA are presented. The important relevant documents and major subjective database resources are also provided in this chapter to be a convenient reference for readers. Chapter 2 briefly introduces the basic concepts and methods of machine learning.

Chapter 3 states the basic and advanced image features. These features are applicable to both general machine learning tasks and for specific purposes. We also introduce the relevant issues concerning feature extraction and feature selection in this chapter. Chapter 4 gives the ML-based feature pooling strategies on VQA, where the traditional ML tools, the newly proposed pairwise rank learning approach, and an ensemble-based scheme accounting for feature pooling are presented in detail. In Chap. 5, a fusion scheme of VQA metrics is presented. This fusion scheme takes advantage of the combined metrics to overcome the shortcomings of each metric used individually. The final chapter concludes this book and gives the potential research prospects in the VQA field.

This book is mainly targeted at researchers, engineers, and students of computer science, information science, mathematics, and perception science who are interested in VQA, machine learning applications, and image/video processing.

October 2014

Long Xu
Weisi Lin
C.-C. Jay Kuo

Acknowledgments

We thank Dr. Issac Niwas Swamidoss for proofreading. The related research was partially supported by the National Natural Science Foundation of China under Grants 61202242, 61370113, Singapore National Research Foundation under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office, Media Development Authority, and Ministry of Education (MoE) AcRF Tire 1 Funding, Singapore, under Grant 2014-T1-002-068.

Contents

1	Introduction	1
1.1	Subjective Visual Quality Assessment	2
1.2	Image Quality Databases	3
1.3	Video Quality Databases	5
1.4	Classification of Objective Visual Quality Assessment	7
1.4.1	Availability of Reference	8
1.4.2	Methodology for Assessment	8
1.5	Further Discussion on Major Existing Objective Visual Quality Metrics	9
1.5.1	Model-Based Metrics	10
1.5.2	Signal-Driven Metrics	12
1.6	Evaluation Criteria for Assessing Metrics	13
1.7	Emerging Machine Learning-Based Visual Quality Assessment	14
1.8	Significance and Highlights of This Book	15
	References	17
2	Fundamental Knowledge of Machine Learning	23
2.1	Artificial Neural Networks	24
2.2	Support Vector Machine	24
2.2.1	Primal Form	26
2.2.2	Dual Form	26
2.3	Clustering	28
2.4	Bayesian Networks	29
2.5	Representation Learning	29
2.6	Sparse Dictionary Learning	29
2.7	AdaBoost	30
2.8	Deep Learning	31
2.8.1	Deep Neural Networks	31
2.8.2	Deep Belief Network	33
	References	34

3	Image Features and Feature Processing	37
3.1	Types of Image Features	38
3.2	Commonly Used Feature Detectors	38
3.2.1	Laplacian of Gaussian (LoG)	38
3.2.2	Difference of Gaussian (DoG)	39
3.2.3	Gabor Filter Coefficients	41
3.3	Commonly Used Feature Descriptors	42
3.3.1	Scale-Invariant Feature Transform (SIFT)	43
3.3.2	Speeded up Robust Features (SURF)	47
3.3.3	A Global Scene Feature: GIST	48
3.4	Feature Selection and Extraction	48
3.4.1	Fisher Scores	49
3.4.2	Generalized Fisher Scores	50
3.4.3	Laplacian Scores	51
3.4.4	mRMR Method	52
3.4.5	Max-Dependency and Two-Stage Method	53
3.4.6	Principal Component Analysis (PCA)	54
3.4.7	Singular Vector Decomposition (SVD)	56
3.5	Feature Learning	59
3.5.1	K-Means Clustering	60
3.5.2	Sparse Feature Learning	62
	References	63
4	Feature Pooling by Learning	67
4.1	Support Vector Machine Learning for Feature Pooling	68
4.2	SVD-Based Feature and Feature Pooling Using SVM	71
4.2.1	SVD-Based Feature Detection	71
4.2.2	Combining Features into a Perceptual Quality Score	73
4.2.3	Support Vector Regression (SVR)	73
4.3	Rank Learning for Feature Pooling	75
4.3.1	Categorical, Ordinal, and Interval Variables	75
4.3.2	Numerical Rating and Pairwise Comparison	76
4.3.3	Rank Learning Approaches	77
4.3.4	Pairwise Rank Learning Image Quality Assessment (PRLIQA)	79
	References	89
5	Metrics Fusion	93
5.1	Multi-method Fusion (MMF)	94
5.1.1	Support Vector Regression for Fusion	94
5.1.2	MMF Scores	96
5.1.3	Data Scaling and Cross-Validation	97

5.1.4	Training	97
5.1.5	Testing	97
5.2	Context-Dependent MMF	98
5.2.1	Context Definition	98
5.2.2	Automatic Context Determination.	99
5.2.3	Fused IQA Method Selection.	100
5.2.4	Evaluation.	101
5.3	Image Quality Assessment Using ParaBoosting Ensemble	102
5.3.1	Image Quality Scores (IQSs)	106
5.3.2	Evaluation of IQSs, Training Strategy, and ParaBoosting Ensemble (PBE)	114
5.3.3	Scorer Selection for PBE IQA	119
	References	121
6	Summary and Remarks for Future Research.	123
6.1	Summary.	123
6.2	Further Discussions on Future Possibilities	125
6.2.1	3D Quality Assessment.	126
6.2.2	Joint Audiovisual Quality Metrics	126
6.2.3	Computer-Generated Visual Signal Quality Assessment	128
6.3	Final Remarks	129
	References	129

Acronyms

AdaBoost	Adaptive boosting
AI	Artificial intelligence
AIQS	Advanced image quality score
AIQs	Advanced image quality scorers
ANN	Artificial neural network
BIQS	Basic image quality score
BIQs	Basic image quality scorers
BN	Bayesian network
CD-MMF	Content-dependent MMF
CNN	Convolutional (deep) neural network
CS	Computer science
CSF	Contrast sensitivity function
DBN	Deep belief network
DCT	Discrete cosine transform
DFT	Fourier transform
(D)MOS	(Difference) mean opinion score
DNN	Deep neural network
DoG	Difference of Gaussian
DWT	Discrete wavelet transform
FR	Full reference
GLCM	Gray-level co-occurrence matrix
GPU	Graphics processing unit
HDR	High dynamic range
HVS	Human visual system
IQA	Image quality assessment
IQSs	Image quality scorers
JND	Just noticeable difference
KLT	Karhunen–Loeve transform
LoG	Laplacian of Gaussian
MAP	Mean average precision
ML	Machine learning
MMF	Multi-method fusion
mRMR	Minimum redundancy maximum relevance

MSE	Mean square error
NDCG	Normalized discounted cumulative gain
NLP	Neuro-linguistic programming
NR	No reference
NSS	Natural scene statistics
PC	Pairwise comparison
PCA	Principal component analysis
PLCC	Pearson's linear correlation coefficient
PRLIQA	Pairwise rank learning image quality assessment
PSNR	Peak signal-to-noise ratio
PVQM	Perceptual visual quality metric
RBM	Restricted Boltzmann machine
RMSE	Root mean square error
RR	Reduced reference
SD	Standard definition
SIFT	Scale-invariant feature transform
SROCC	Spearman's rank order correlation coefficient
SURF	Speeded up robust feature
SVD	Singular value decomposition
SVM	Support vector machine
SVR	Support vector regression
VA	Visual attention
VQA	Visual quality assessment
WCSS	Within-cluster sum of squares

Chapter 1

Introduction

Abstract With the development of digital visual signal processing, efficient and reliable assessment of image quality becomes more and more important. Measuring the image quality is of fundamental importance for image processing applications, where the goal of image quality assessment (IQA) methods is to automatically evaluate the quality of images in agreement with human visual quality judgments. In the past decade, researchers have made great efforts to develop many IQAs to fulfill this goal. According to the availability of reference, these IQAs can be classified into full-reference, reduce-reference, and no-reference IQAs. For most of the applications, there is no reference available during visual signal processing, e.g., decoding of compressed visual signal in client's terminal without the original signal stored in server. The absence of reference raises a great challenge for visual quality assessment (VQA). To address this challenge, machine learning was widely used to approximate the response of the human visual system (HVS) to visual quality perception. This chapter will present the fundamental knowledge of VQA, overview of the state-of-the-art IQAs in the literature, resource of VQA, standards developed for subjective quality assessment, and summary of all related items.

Keywords Image quality assessment · Mean opinion score · Subjective quality assessment · Objective quality assessment · Human visual system

After images are captured, operations of compression, enhancement, editing, retargeting, and transmission are usually performed, resulting in image quality changes. To measure the extent of such changes, the mean square error (MSE) or one of its variants between an original image and a revised one is calculated. However, MSE/PSNR (peak signal-to-noise ratio) cannot reflect the actual human-perceived quality of images, and therefore a bunch of visual quality assessment (VQA) approaches from the aspects of the human visual system (HVS) and human visual psychology have been proposed during the last decade.

Since humans are the ultimate receivers of the majority of visual signals being processed, the most accurate way of assessing image/video quality is to ask humans for their opinions of the quality of an image or video, known as the subjective VQA. The subjective image quality scores gathered from all subjects are processed to be the mean opinion score (MOS), which is regarded as the ground truth of image quality.

However, subjective image quality is time-consuming and labor-consuming, so it is not applicable to real-time processing; in addition, it is not feasible to have subjects' intervention with in-loop and on-service processes, such as image/video communications, video streaming, video conferencing, and so on. For these applications, objective VQA without subjects' intervention is desirable. Many approaches for objective quality assessment have been proposed, while subjective VQA is usually used for benchmarking objective visual quality algorithms being built. The correlation between the MOSs provided by the subjective VQA and the predicted quality scores computed from the objective visual quality metric is used to measure the effectiveness of an approach for predicting the human perception of image/video quality.

In this chapter, the basic knowledge of subjective VQA is introduced in Sect. 1.1. Sections 1.2 and 1.3 give information about the mostly used image/video quality databases. Section 1.4 discusses the classification of objective VQA. Section 1.5 introduces major existing objective visual quality metrics. Section 1.6 discusses the performance measurement for assessing visual quality metrics. Section 1.7 highlights the emerging machine learning (ML) oriented VQA. The structure of this book is presented in the last section.

1.1 Subjective Visual Quality Assessment

On the recommendation of the International Telecommunication Union (ITU) [1–6], a number of viewers are asked to rate images or videos in subjective experiments, and the scores of viewers are processed to be the mean opinion score (MOS) or the difference mean opinion score (DMOS). In [6], comparative studies of objective video quality metrics have been carried out by Video Quality Expert Group (VQEG). Subjective experiments have been performed on a large number of test video sequences by laboratories and researchers. In [7, 8], quality degradation of video streaming was investigated to target a good perceptual quality of multimedia services. The publicly available databases of subjective scores and test material are reported in [9, 10] for quality degradation of compression and error-prone channels. Most researchers developed their objective metrics based on the available subjective quality databases.

Several international standards have been proposed to normalize and guide the process of subjective experiments. Some of them are briefly described as follows:

- (a) ITU-R BT.500-11 [1] proposed methods for subjective quality assessment of television pictures. It was widely used to suggest the viewing condition, instructions on how to perform subjective experiment, test materials, and the presentation of experimental results.
- (b) ITU-P.910 [3] described methods for digital video quality assessment with bit rates under 1.5 Mbps.
- (c) ITU-R BT.814-1 [4] recommended the brightness and contrast of display devices.