

ENHANCED AUGMENTED REALITY FRAMEWORK FOR SPORTS ENTERTAINMENT  
APPLICATIONS

A Dissertation  
Submitted to the Graduate Faculty  
of the  
North Dakota State University  
of Agriculture and Applied Science

By

Zahid Mahmood

In Partial Fulfillment of the Requirements  
for the Degree of  
DOCTOR OF PHILOSOPHY

Major Department:  
Electrical and Computer Engineering

November 2015

Fargo, North Dakota

North Dakota State University  
Graduate School

---

Title

Enhanced Augmented Reality Framework for Sports Entertainment Applications

By

Zahid Mahmood

The Supervisory Committee certifies that this *disquisition* complies with North Dakota State University's regulations and meets the accepted standards for the degree of

**DOCTOR OF PHILOSOPHY**

SUPERVISORY COMMITTEE:

Dr. Samee U. Khan

Chair

Dr. Jacob S. Glower

Dr. Sudarshan K. Srinivasan

Dr. Ying Huang

Approved:

11/23/2015

Date

Dr. Rajesh G. Kavasseri

Department Chair/Program Coordinator

## ABSTRACT

Augmented Reality (AR) superimposes virtual information on real-world data, such as displaying useful information on videos/images of a scene. This dissertation presents an Enhanced AR (EAR) framework for displaying useful information on images of a sports game. The challenge in such applications is robust object detection and recognition. This is even more challenging when there is strong sunlight. We address the phenomenon where a captured image is degraded by strong sunlight.

The developed framework consists of an image enhancement technique to improve the accuracy of subsequent player and face detection. The image enhancement is followed by player detection, face detection, recognition of players, and display of personal information of players. First, an algorithm based on Multi-Scale Retinex (MSR) is proposed for image enhancement. For the tasks of player and face detection, we use adaptive boosting algorithm with Haar-like features for both feature selection and classification. The player face recognition algorithm uses adaptive boosting with the LDA for feature selection and nearest neighbor classifier for classification. The framework can be deployed in any sports where a viewer captures images. Display of players-specific information enhances the end-user experience. Detailed experiments are performed on 2096 diverse images captured using a digital camera and smartphone. The images contain players in different poses, expressions, and illuminations. Player face recognition module requires players faces to be frontal or up to  $\pm 35^\circ$  of pose variation. The work demonstrates the great potential of computer vision based approaches for future development of AR applications.

## **ACKNOWLEDGEMENTS**

First and foremost thanks to ALLMIGHTY ALLAH Who has helped me throughout the course of my studies. All of my knowledge, strength, courage, health, and abilities are His Countless blessings upon me and there is no way to fulfill His right to thank Him.

Special thanks to Dr. Samee U. Khan, my adviser, for his help, guidance, and innovative ideas. I offer my sincere and deep hearted gratitude to my adviser who always encouraged me, and persistently conveyed the spirit and guidance required for the research. Without his kind guidance and continuous efforts, this disquisition would not have been possible.

Special thanks to my committee members, Dr. Jacob S. Glower, Dr. Sudarshan K. Srinivasan, and Dr. Ying Huang for their support, guidance, and helpful recommendations. Thanks to the Electrical and Computer Engineering staff members Jeffrey Erickson, Laura D. Dallman, and Priscilla Schlenker for all the unconditional help and favor.

I would like to thank my family. Their continuous support is always a source of motivation and encouragement for me. I especially like to pray and remember my great grandmother (late), father and mother, who are the only and every reason for whatever I am today and whatever I achieved in my life. I also would like to thank my loving brothers, sister, uncles, and cousins for their patience, time, and support. They have prayed and cooperated a lot with me that resulted in successful completion of this disquisition.

Finally, but not the least, I owe my heartiest thanks to COMSATS Institute of Information Technology for granting me Ph.D. scholarship, all my friends, especially to Dr. Tauseef Ali, and colleagues in the USA and Pakistan, who always helped me in the time of need.

## **DEDICATION**

I would like to dedicate this dissertation to my family, especially to my great grandmother (late), my brother Tahir Mahmood (late), and my parents for all the inexplicable love, support, prayers, and motivation.

# TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
DEDICATION.....	v
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
1. INTRODUCTION.....	1
1.1. Problem Statement.....	1
1.2. Motivation.....	2
1.3. Research Goals and Objectives.....	3
1.4. References.....	4
2. RELATED WORK.....	5
2.1. Image Enhancement.....	5
2.2. Face Detection and Recognition.....	6
2.3. The AR Applications.....	8
2.4. References.....	9
3. MULTI SCALE RETINEX FOR IMAGE ENHANCEMENT UNDER NON UNIFORM ILLUMINATIONS.....	12
3.1. Introduction.....	12
3.2. Retinex Theory.....	15
3.2.1. Single Scale Retinex (SSR).....	17
3.2.2. Multi Scale Retinex (MSR).....	18
3.3. Proposed Method.....	19
3.3.1. Applying Principal Component Analysis (PCA).....	19
3.3.2. Applying the MSR.....	20

3.3.3. Introducing the DFT .....	20
3.3.4. Enhanced Luminance (EL) Ratio Computation .....	21
3.3.5. Computing New RGB Values .....	21
3.3.6. Applying Contrast Stretching (CS) .....	21
3.4. Simulation Results and Discussion .....	22
3.4.1. Comparative Analysis .....	26
3.4.1.1. Mean .....	27
3.4.1.2. Standard Deviation .....	27
3.4.1.3. Entropy .....	27
3.4.2. Peak Signal to Noise Ratio (PSNR) .....	29
3.4.3. Absolute Mean Brightness Error (AMBE).....	31
3.4.4 Computational Complexity .....	32
3.5. Conclusions and Future Work.....	32
3.6. References .....	33
<b>4. AUTOMATIC PLAYER DETECTION, RECOGNITION, AND IDENTIFICATION FOR SPORTS ENTERTAINMENT APPLICATIONS .....</b>	<b>36</b>
4.1. Introduction .....	36
4.2. Player Detection using AdaBoost .....	39
4.3. Player Face Detection.....	40
4.4. Player Face Recognition.....	41
4.5. Player Statistics Retrieval.....	42
4.6. Simulation Results.....	42
4.6.1. Player and Face Detection.....	43
4.6.2. Players Face(s) Recognition .....	44
4.7. Conclusions and Future Work.....	46
4.8. References .....	47

5. EAR: ENHANCED AUGMENTED REALITY FRAMEWORK FOR SPORTS ENTERTAINMENT APPLICATIONS.....	50
5.1. Introduction .....	50
5.2. Simulation Results.....	51
5.2.1. Player Image Enhancement Module.....	51
5.2.2. Player Detection Module.....	51
5.2.3. Player Face Detection Module .....	53
5.2.4. Player Face Recognition Module .....	53
5.2.5. Further Analysis of Face Recognition Module .....	55
5.2.6. Computational Complexity .....	56
5.3. Conclusions and Future Work.....	57
5.4. References .....	58
6. CONCLUSIONS.....	61
6.1. The AR Applications.....	61
6.2. Final Remarks .....	61
6.3. Recommendations for Future Work.....	62

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1. Notations and their meanings.....	15
3.2. The entropy comparison of different image enhancement techniques.....	29
3.3. The PSNR comparison of different image enhancement techniques.....	30
3.4. Time cost comparison of the MSR with other image enhancement techniques.....	32
4.1. Notations and their meanings.....	39
4.2. Execution time of each module.....	44
4.3. Recognition performance with various size face samples.....	46
5.1. Accuracy of the developed EAR framework.....	53
5.2. Average minimum classification error rate.....	55
5.3. Summary of execution time of the developed EAR system.....	57

# LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1. The EAR framework.....	3
3.1. Block diagram of proposed image enhancement algorithm.....	14
3.2. Human visual system.....	16
3.3. Test Image 1: Image enhancement results.....	23
3.4. Test Image 2: Image enhancement results.....	24
3.5. Test Image 3: Image enhancement results.....	24
3.6. Test Image 4: Image enhancement results.....	25
3.7. Test Image 5: Image enhancement results.....	26
3.8. Comparative statistical analysis of the proposed technique with different image enhancement techniques.....	28
3.9. The AMBE comparison of the proposed technique with other image enhancement techniques.....	31
4.1. Framework of our proposed application.....	38
4.2. Example of positive (left) and negative (right) training samples.....	40
4.3. Example of rectangle features shown relative to the enclosing detection window.....	41
4.4. Face database of 150 different players.....	43
4.5. Sample images illustrating performance of players and face(s) detection module.....	44
4.6. Output of proposed system on sample images from database.....	45
4.7. Accuracy with number of training images used per subject.....	46
5.1. Complete output of the developed EAR system on sample images: (a) original degraded input player images, (b) player image enhancement, (c) player detection, and (d) player face recognition and display of statistics.....	52
5.2. Accuracy comparison by varying training images with and without player image enhancement.....	54

5.3. Performance of player face recognition module by varying player face image resolution..... 56

# 1. INTRODUCTION

## 1.1. Problem Statement

Augmented Reality (AR) superimposes virtual information on real-world data, such as displaying useful information on videos/images of a scene [1.1]. This dissertation presents an Enhanced AR (EAR) framework for displaying useful information on images (or videos) of a sports game. The challenge in such applications is robust object detection and recognition. This is even more challenging when there is strong sunlight [1.2]. We address the phenomenon where a captured image is degraded by strong sunlight. The developed framework consists of an image enhancement technique to improve the accuracy of subsequent player and face detection [1.3]. The image enhancement is followed by player detection, face detection, recognition of players, and display of personal information of players [1.4]-[1.5]. First, an algorithm based on Multi-Scale Retinex (MSR) is proposed for image enhancement. For the tasks of player and face detection, we use adaptive boosting algorithm with Haar-like features for both feature selection and classification. The player face recognition algorithm uses Boosted Linear Discriminant Analysis (BLDA) for feature selection and classic nearest neighbor classifier for classification. The system can be deployed in any sports where a viewer captures video/images. Display of players-specific information enhances the end-user experience. Detailed experiments are performed on 2096 diverse images captured using a digital camera and smartphone. The images contain players in different poses, expressions, and illuminations. Player face recognition module requires players faces to be frontal or up to  $\pm 35^{\circ}$  of pose variation. The work demonstrates the great potential of computer vision based approaches for development of AR applications.

## 1.2. Motivation

Today people live, work, and enjoy in a globalized society; hence understanding various activities, such as cultures or sports is always amusing. Paige *et al.* [1.6] defined artistic learning as a process of acquiring the knowledge, skills, and attitudes required for effective communication and interaction with people from other countries. Artistic learning, immersion, and simulation attract researchers from a variety of fields, including computer vision, machine learning, and image processing. Augmented Reality (AR), which enables the seamless connection of the digital and physical domains, is one of the newest technologies applied to various fields, such as sports entertainment applications. The AR has already begun to show promise in helping people learn more effectively and increase knowledge retention, relative to traditional 2D desktop interfaces [1.7].

The AR supports the understanding of complex phenomena by providing unique visual and interactive experiences that combines real and virtual information and help communicate abstract problems to learners. With the AR, designers can superimpose virtual graphics over real objects and allow users to interact with digital content through physical manipulation. One such scenario is when viewers are sitting in the stadium and they do not have access to TV that shows some important information and statistics of game and players.

The aforementioned discussion ratifies the apparent need and impetus for developing an AR framework that can be deployed in any sports entertainment to provide viewers accurate and interesting information of the players. Therefore, the research targets the development of an Enhanced Augmented Reality (EAR) framework to enhance end-users experience. As a matter of fact, this kind of research must continuously push the limits of what is possible in the AR, and indeed in this case, is leading edge in the development of the AR methodologies.

### 1.3. Research Goals and Objectives

The objective of our research is to develop methodologies for enhancing end-user experience by developing an application framework using image processing, computer vision, and machine learning techniques. The aforesaid situation demands for such methodologies those are able to cope with real life scenarios, such as bad lighting condition, occlusion, and face pose. Consequently, the developed methodology must have ability to perform image enhancement, player/face detection, and face recognition in real-time performance. Fig. 1.1 depicts the general overview of the EAR framework. Based on the aforesaid discussion, following are the specific objectives of our research work.

- To propose an image enhancement scheme to rectify the images degraded by sun light.
- To apply state-of-the-art machine learning algorithm to detect player(s) and face(s).
- To apply a robust computer vision algorithm to recognize player(s) face(s) and display the relevant information.



Fig. 1.1. The EAR framework.

#### 1.4. References

- [1.1] M. Yang and W. Liao, "Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction," *IEEE Transactions on Learning Technologies*, Vol. 7, No. 2, 2014, pp. 107-117.
- [1.2] J. Han and D. Farin, "A real-time augmented reality system for sports broadcast video enhancement," *MM'07*, September 23–28, 2007, Augsburg, Bavaria, Germany, pp. 1-4.
- [1.3] Z. Mahmood et al., "A color image enhancement technique using multiscale retinex," *11th International Conference on Frontiers of Information Technology (FIT)*, 2013, IEEE, pp. 119-124.
- [1.4] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, Vol. 57, 2004, pp. 137–154.
- [1.5] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Transactions on Neural Networks*, Vol. 17, No. 1, 2006, pp. 166- 178.
- [1.6] R. Paige, J. Jorstad, L Siaya, F. Klein, and J. Colby, "Culture learning in language education: A review of the literature," in *Culture as the core: Integrating culture into the Lang. education*, D. Lange and R. Paige, eds., Greenwich, CT, USA: Information Age, 2003 pp. 173–236.
- [1.7] M. Billinghamurst and A. Dunser, "Augmented reality in the classroom," *IEEE Comput.*, Vol. 45, No. 7, Jul. 2012, pp. 56–63.

## 2. RELATED WORK

In this chapter we discuss some of the work that is related to the research we have performed during Ph.D.

### 2.1. Image Enhancement

The objective of image enhancement is to improve the image quality in terms of human vision perception [2.1]. Often, images captured suffer in quality, such as whitish appearance, unclear details, and distorted colors [2.2]-[2.3].

In [2.1], authors present a Non-linear Transfer Function (NTF) based technique for color image enhancement. In the proposed method, the image enhancement is applied only on the luminance component of the HSV color image. Later, original H and S component image and enhanced luminance component image are converted back to the RGB image. The algorithm produces reasonable results for general outdoor images. Hongteng *et al.* [2.2] present a model for image contrast mapping. Their results demonstrate the feasibility in various applications, such as contrast enhancement, tone correction, and post-processing of de-hazed images. However, the algorithm does not explain elimination of severe dark contrast in captured images.

In [2.3], a technique is introduced for personal photograph enhancement utilizing Internet photo collections. Initially, the method constructs 3-D model followed by the augmentation of 2-D data with 3-D model. The work is computationally complex and does not address non-uniform illuminations. Authors in [2.4] introduce an image enhancement technique by applying power law transformations. However, the published work does not explain about correcting distorted colors in RGB images.

## 2.2. Face Detection and Recognition

Face detection locates the possible position of face(s) in an image, whereas face recognition aims to identify the query image from database of known and stored individuals.

Viola and Jones had developed the first real-time face detection system using AdaBoost [2.5]. They introduced the concept of integral image for quick Haar-like feature computation. Recently, Jeong *et al.* [2.6] proposed a Semi-Local Structure Patterns (SLSP), a novel feature extraction method based on local region-based differences. The SLSP successfully detected faces in non-uniform illumination variations, distortion, and sparse noise. Authors in [2.7] proposed a real-time and nonintrusive method based on the diffusion speed of a single image for efficient face detection. In particular, difference in surface properties between a live face and a fake one was efficiently revealed in the diffusion speed. The proposed approach accurately handled diverse malicious attacks regardless of the medium of the image by successfully detecting the face in an image.

In [2.8], authors proposed a new set of features, called qual Histogram of Oriented Gradients (qualHOG) for robust face detection. The proposed approach augmented face-indicative HOGs features with perceptual quality-aware spatial Natural Scene Statistics (NSS) features. Face detectors were trained on new features that provided statistically significant improvement in tolerance to image distortions over a strong baseline. Researchers in [2.9] discussed feasibility of a face constellation that enabled multiview face detection and localization. The proposed face constellation algorithm required only a single reference image of a face containing two manually indicated reference points for initialization. Detection results were reported for face detection with arbitrary pose. Recently, researchers from around the globe have addressed the face recognition problem in various domains.

Berg *et al.* [2.10] presented a method for face recognition that learnt a diverse set of highly discriminative features, known as Part-based One-vs-One Features (POOFs). Each of these features specialized in discrimination between two classes based on the appearance at a particular part. Usefulness of these features was demonstrated by presenting state-of-the-art results on bird species identification using the Caltech UCSD Birds (CUB) dataset and on Labeled Faces in the Wild (LFW) dataset. Authors reported an average accuracy of 73.30% on CUB and 93.13% on the LFW datasets.

Schwartz *et al.* [2.11] employed 70,000 feature descriptors using Multichannel Feature Weighting (MFW). The work was extended to the tree-based discriminative structures. Algorithm was tested on FERET and FRGC datasets with an accuracy of 92.75% and 94.63%, respectively. An approach in FR algorithms is Harmony Search Algorithm (HSA) that selected an optimal subset of features [2.12]. The proposed approach was compared with PCA where Euclidean distance and cosine similarity were used for classification. Authors reported 94% recognition accuracy on ORL database. Authors in [2.13] presented a novel video-based algorithm to recognize facial expressions. The approach consisted of face detection and face registration in video frames, while face recognition was achieved using the support vector machine. For anger, disgust, and surprise, algorithm yielded 82%, 88%, and 97% accuracy, respectively on FERA database.

Nguyen *et al.* [2.14] proposed a 3-D model based algorithm that efficiently distinguished between computer generated and natural faces by evaluating their dynamic behavior. The proposed algorithm had three major steps: **(1)** image/video normalization, **(2)** face model reconstruction using active shape model, and **(3)** facial characters identification. Experiments conducted on BUHMAP-DB and JAFFE databases reported 91.93% and 97.5%, accuracy.

### 2.3. The AR Applications

In the AR applications, virtual reality replaces the real world with a simulated environment. For example, a simulated image/video of a city may contain useful information about buildings, parking-lots, and restaurants [2.15].

In [2.16], an AR technique is presented to generate a visual enhancement for TV broadcasted court net sports. The proposed system generates redundant virtual scenes that promptly affect system performance. A graphics based AR technique proposed in [2.17] also do not produce satisfactory results due to loss of motion and texture of the players. In [2.18], researchers present a technique where they insert and manipulate 3-D virtual contents on to broadcasted tennis video. However, the proposed mechanism does not provide aesthetic viewing experience and struggles when tennis court net is invisible. Matsui *et al.* [2.19] develop a system that performs camera calibration and establishes mapping between the soccer field and virtual scene. However, the developed system has shortcoming of player pose restriction. User has to choose from three defined poses, such as walk, run, and stop.

In [2.20], researchers present a system for augmenting live broadcast sports with 3-D information. This work is successfully adopted by Fox Sports. However, authors do not address the issues if captured images/videos suffer in quality. Jungong *et al.* [2.21] present a real-time AR system by demonstrating experiments on volley ball and tennis during live sports. The work utilizes a probabilistic method based on Expectation Maximization (EM) procedure. The authors do not consider the situation where images/videos are distorted by non-uniform light. Researchers in [2.22], introduce a dual mode two way Bayesian inference approach by utilizing progressive observation modeling. However, results are shown only for team member detection.

## 2.4. References

- [2.1] D. Ghimire, J. Lee, “Nonlinear transfer function-based local approach for color image enhancement,” *IEEE Transactions on Consumer Electronics.*, Vol. 57, No. 2, 2011, pp. 858-865.
- [2.2] H. Xu, G. Zhai, and X. Yang, “Generalized equalization model for image enhancement,” *IEEE Transactions on Multimedia*, Vol. 16, No. 1, 2014, pp. 68-82.
- [2.3] C. Zhang, J. Gao, and O. Wang, “Personal photograph enhancement using internet photo collections,” *IEEE Transactions on Visualization and Computer Graphics.*, Vol. 20, No. 2, 2014, pp. 262-275.
- [2.4] G. Singh, and T. Kumar, “Color image enhancement by linear transformations solving out of gamut problem,” *International Journal of Computer Applications*, Vol. 67, No. 14, 2013, pp. 28-32.
- [2.5] P. Viola and M. Jones, “Robust real-time face detection,” *Int. J. Comput. Vis.*, Vol. 57, 2004, pp. 137–154.
- [2.6] K. Jeong, J. Choi, and G. Jang, “Semi-local structure patterns for robust face detection,” *IEEE Signal Processing Letters*, Vol. 22, No. 9, 2015, pp. 1400-1403.
- [2.7] W. Kim, S. Suh, and J. Han, “Face liveness detection from a single image via diffusion speed model,” *IEEE Transactions on Image Processing*, Vol. 24, N0. 8, 2015, pp. 2456-2465.
- [2.8] S. Gunasekar, J. Ghosh, and A. Bovik, “Face detection on distorted images augmented by perceptual quality-aware features,” *IEEE Transactions on Information Forensics and Security*, Vol. 9, No. 12, 2014, pp. 2119-2131.
- [2.9] S. Anvar, W. Yau, and E. Teoh, “Multiview face detection and registration requiring minimal manual intervention,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 10, 2013, pp. 2484-2497.

- [2.10] T. Berg and B. Peter, "POOF: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation," *Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 955-962.
- [2.11] W. Schwartz, H. Guo, J. Choi, and L. Davis, "Face identification using large feature sets," *IEEE Transactions on Image Processing.*, Vol. 21, No. 4, 2012, pp. 2245-2255.
- [2.12] R. Sawalha and I. Doush, "Face recognition using harmony search-based selected features," *International Journal of Hybrid Information Technology*, Vol. 5, No. 2, 2012, pp. 1-16.
- [2.13] S. Yang and B. Bhanu, "Understanding discrete facial expressions in video using an emotion avatar image," *IEEE Transactions on Systems and Cybernetics—Part B: Cybernetics.*, Vol. 42, No. 4, 2012, pp. 980-992.
- [2.14] Duc *et al.* "3D-model-based video analysis for computer generated faces identification," *IEEE Transactions on Information Forensics and Security*, Vol. 10, No. 8, 2015, pp. 1752-1763.
- [2.15] M. Yang and W. Liao, "Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction," *IEEE Transactions on Learning Technologies*, Vol. 7, No. 2, 2014, pp. 107-117.
- [2.16] J. Han and D. Farin, "A real-time augmented reality system for sports broadcast video enhancement," *MM'07*, 2007, Augsburg, Bavaria, Germany, pp. 1-4.
- [2.17] D. Liang *et al.* "Video2cartoon: generating 3-D cartoon from broadcast soccer video," In *Proceedings of ACM Multimedia*, 2005, pp.217-218.
- [2.18] X. Yu, X. Yan, T. Chi, and L. Cheong, "Inserting 3-D projected virtual Content into broadcast tennis Video," In *Proceedings of ACM Multimedia*, 2006, pp.619-622.
- [2.19] K. Matsui, M. Iwase, M. Agata, T. Tanaka, and N. Ohnishi, "Soccer image sequence computed by a virtual camera," In *Proceedings of CVPR*, 1998, pp. 860-865.

[2.20] Cavallaro *et al.*, “Augmenting live broadcast sports with 3D tracking information,” IEEE Multimedia, 2011, pp. 38-47.

[2.21] H. Jungong and D. Farin, “A real-time augmented-reality system for sports broadcast video enhancement,” 15th International Conference on Multimedia, ACM, 2007.

[2.22] Xing *et al.* “Multiple player tracking in sports video: A dual-mode two-way Bayesian inference approach with progressive observation modeling,” IEEE Transactions on Image Processing, 2011, pp. 1652-1667.

### 3. MULTI SCALE RETINEX FOR IMAGE ENHANCEMENT UNDER NON-UNIFORM ILLUMINATIONS<sup>1</sup>

This paper is submitted to IET Computer Vision and is in the second round of review. Authors of the paper are Zahid Mahmood, Tauseef Ali, Justin McIntosh, and Samee U. Khan.

#### 3.1. Introduction

Image processing transforms images through mathematical means to change characteristics, such as finding edges, segmentation, and enhancement [3.1]. With the recent advances and prevalence of imaging devices, millions of images are being created every day. Due to various factors, such as undesirable light sources and unfavorable weather conditions, contrast and visual appearance of the captured image may not always be pleasing. Therefore, image enhancement is often required for aesthetic and pragmatic purposes [3.2].

In [3.1], authors present a Non-linear Transfer Function (NTF) based technique for color image enhancement. In the proposed method, the image enhancement is applied only on the luminance component of the HSV color image. Later, original H and S component image and enhanced luminance component image are converted back to the RGB image. The algorithm produces reasonable results for general outdoor images.

Sun *et al.* [3.3] present a Luminance-Based (LB-MSR) image enhancement scheme by applying the PCA to obtain luminance channel. The luminance channel is enhanced using the MSR. Later, luminance channel is integrated with two chromatic channels followed by inverse

---

<sup>1</sup> Zahid Mahmood is the primary developer of the conclusions that are advanced here. Zahid Mahmood drafted and revised all versions of this chapter. Tauseef Ali and Justin McIntosh collected real images in sunlight to test the algorithm. Samee U. Khan supervised the research and vetted the mathematical and statistical analysis performed by Zahid Mahmood.

PCA transform to get output image. The algorithm produces reasonable results for general outdoor images, but fails to perform on images captured in strong sunlight. The proposed work lacks color restoration scheme. Researchers in [3.4] present an approach to enhance RGB images. Initially, RGB color space is converted to the Luminance Inphase Quadrature (YIQ) color space. Later, the MSR algorithm is introduced to enhance Luminance (Y) component followed by contrast enhancement. Only one image is tested during experiments and the work lacks color correction scheme.

In [3.5], researchers present a technique using the quantum-behaved particle swarm optimization with an adaptive strategy by proposing Beta ( $\beta$ ) transformation. The scheme also struggles to yield suitable enhancement on RGB images affected by non-uniform illuminations. Researchers in [3.6] introduce an image enhancement technique as an extension of the scalar-diffusion shock-filter coupling model. The proposed method utilizes single vectors of the gradient magnitude and the second derivatives. However, the algorithm is complex and consumes more time, than few of the techniques compared therein. A novel filter is proposed in [3.7] for edge preserving decomposition in an image. The filtered image contains local means and salient edges. The algorithm effectively enhances local details. However, the published work does not report experiments on low/high contrast images.

This paper presents a technique to enhance degraded colored (RGB) images by non-uniform illuminations using the MSR theory. The proposed technique has six major steps including application of: **(1)** the PCA, **(2)** the MSR, **(3)** the DFT, **(4)** enhanced luminance ratio calculation, **(5)** new RGB values computation, and **(6)** application of the contrast stretching. Extensive simulations performed on 994 diverse images reveal the superiority of the developed technique. The results of the proposed technique show that the visual appearance, brightness, and

perceptual quality of the images are increased. The proposed technique outperforms state-of-the-art techniques in terms of image quality (PSNR and AMBE) and execution time. Moreover, the statistical (mean, median, standard deviation, and entropy) comparison also reveals the superiority of the proposed technique. Fig. 3.1 shows the complete block diagram of the developed technique. Table 3.1 shows the symbols and their meaning used in this chapter. The main contributions of the paper are:

- We extend our earlier work presented in [3.8] and perform detailed and useful comparison of statistics, such as mean, median, standard deviation, entropy, the PSNR, and the AMBE of the proposed technique with state-of-the-art techniques.
- Our approach enhances real-life challenging images that are degraded by very strong sunlight under severe dark contrast having blackish appearance.
- The proposed MSR algorithm produces bright and perceptually high quality output images in less computation time than other image enhancement techniques. Particularly, the enhanced images of our proposed algorithm do not suffer from color distortions.

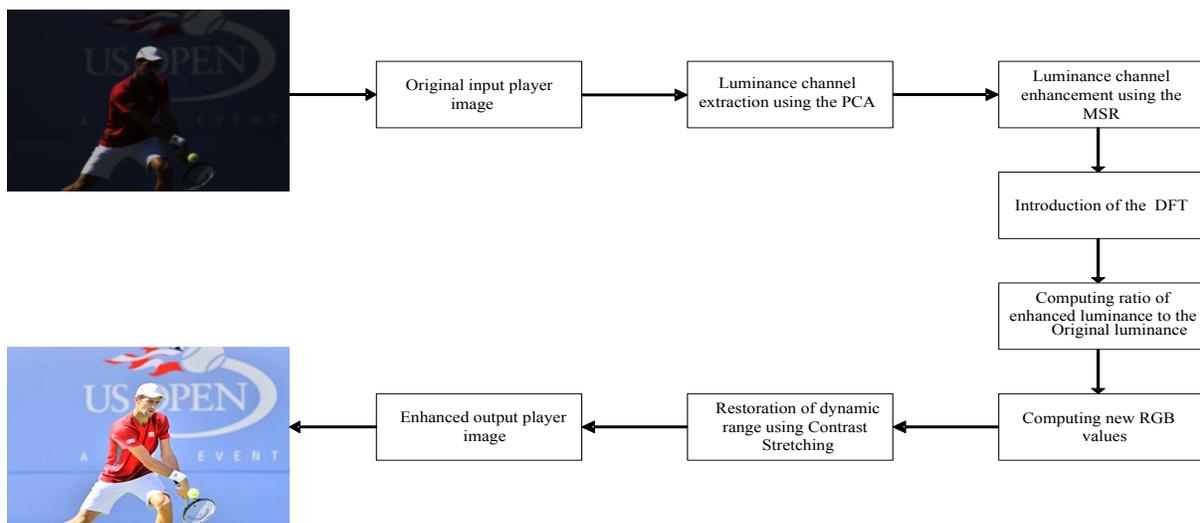


Fig. 3.1. Block diagram of proposed image enhancement algorithm.

Table 3.1. Notations and their meanings.

<b>Notation</b>	<b>Definition or Description</b>
AMBE	Absolute Mean Brightness Error
DRC	Dynamic Range Compression
HVS	Human Visual System
MSE	Mean Squared Error
MSR	Multi Scale Retinex
LB_MSR	Luminance Based Multi Scale Retinex
YIQ_MSR	Luminance (Y) Inphase (I) Quadrature (Q) MSR
NTF	Non Linear Transformation
PSNR	Peak Signal-to-Noise Ratio
SSR	Single Scale Retinex
$\sigma(x) = (SD)$	Standard Deviation
$c / C$	$c$ = Gaussian surround scale constant used in Equation (3.8) $C$ = Scales used in the MSR used in Equation (3.9)
$F(u, v) = DFT$	Discrete Fourier Transform of the signal
$H(x)$	Entropy
$(\omega_n)$	Weights associated with the MSR

### 3.2. Retinex Theory

The Retinex model is based on the assumption that the Human Visual System (HVS) operates with three retinal-cortical systems, each one processes low, middle, and high frequencies of the electromagnetic spectrum. The HVS consists of eyes, retina, and visual cortex that process the photons reaching the eyes. This processing is done independently. Each system

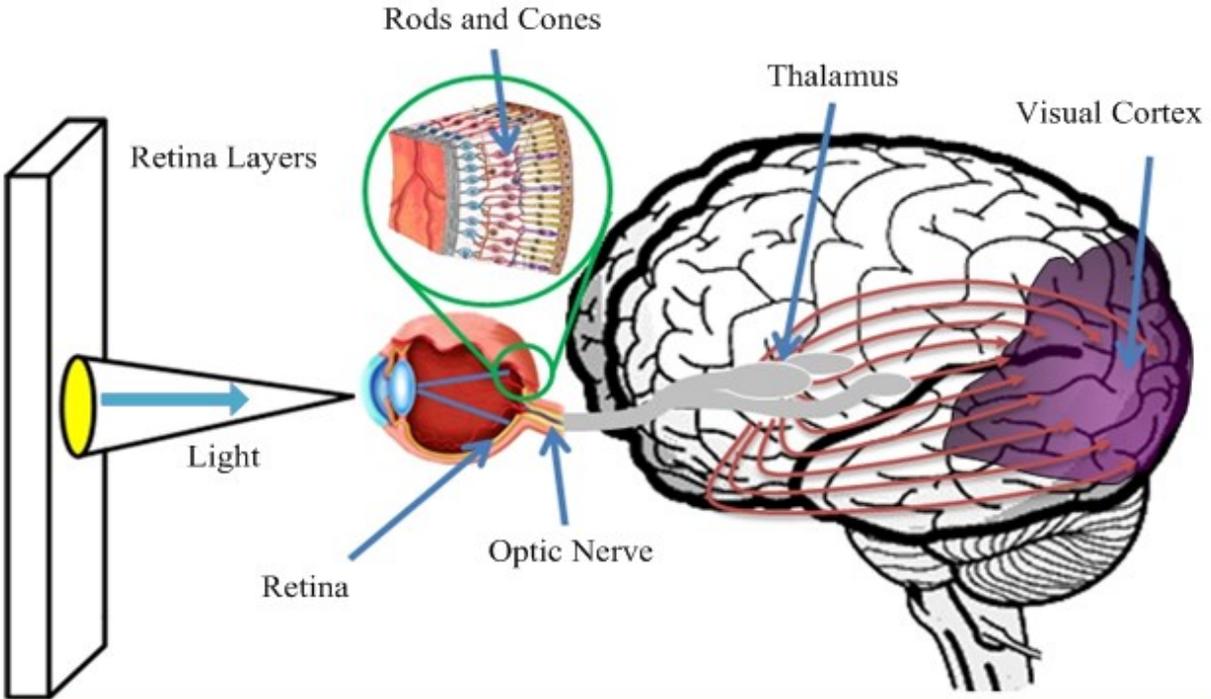


Fig. 3.2. Human visual system.

produces one lightness value that determines the perception of color in the HVS. The HVS produces the unique lightness values and each lightness value is related to RGB planes. Fig. 3.2 shows the general behavior of the HVS. In colored RGB images, the lightness is represented by the triplet  $(L_R, L_G, L_B)$  of lightness values in three chromatic channels. The main puzzle during human perception is the discrepancy found between the physical reflectance of objects and the colors perceived by the HVS. Color processing is completed in the cortex area.

Retinex theory is used to enhance images under varying illumination conditions. It has wide range of applications, especially in medical imaging, such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI), and X-Rays. The MSR can also be used for security and safety purposes like landing of planes in the foggy environment [3.9]. Moreover, the MSR can be helpful to improve the efficiency of algorithms in multimedia monitoring/surveillance and scenes [3.10], vehicle detection [3.11], and face detection [3.12]. Different implementations of the

retinex based algorithms, for example, [3.13], [3.14] have been published since its introduction. The MSR is composed of SSR, so in subsequent sections, details of the SSR and the MSR are explained.

### 3.2.1. Single Scale Retinex (SSR)

Recently, the SSR has been widely used by various researchers. Some researchers successfully apply Retinex theory in challenging situations, such as ultrasound liver images [3.15]. Mathematically, the SSR can be expressed as:

$$R_i(x, y) = \log I_i(x, y) - \log[F(x, y) * I_i(x, y)] \quad (3.1)$$

where  $I_i(x, y)$  is the image distribution in the  $i^{\text{th}}$  color band and  $F(x, y)$  is the normalized surround function satisfying  $\iint F(x, y) dx dy = 1$ . An image distribution  $I_i(x, y)$  is the product of scenes reflectance and illumination. Therefore,

$$I_i(x, y) = S_i(x, y) * r_i(x, y) \quad (3.2)$$

where  $S_i(x, y)$  is the spatial distribution of illumination and  $r_i(x, y)$  is the distribution of scene reflectances. The convolution with surround function works as averaging in the neighbourhood and is given by:

$$R_i(x, y) = \log \frac{(S_i(x, y) * r_i(x, y))}{(S_i(x, y) r_i(x, y))} \quad (3.3)$$

Mostly illumination has slow spatial variation, that means

$$S_i(x, y) \approx \overline{S_i(x, y)} \quad (3.4)$$

where  $\overline{S_i(x, y)}$  is average spatial distribution of the luminance, thereby indicates color constancy.

Therefore,

$$R_i(x, y) \approx \log \frac{r_i(x, y)}{r_i(x, y)} \quad (3.5)$$

Several surround functions can be used. The inverse square spatial surround function is given by:

$$F(x, y) = 1/r^2 \quad (3.6)$$

The technique proposed in [3.16] introduces artefacts in images and in some cases induces blurriness in colored images. Another group [3.17] apply exponential formula with absolute parameters given by:

$$F(x, y) = Kexp^{-r/c} \quad (3.7)$$

The technique in [3.17] shows better results but induces high contrast in images. In our simulations, we use a Gaussian function given by:

$$F(x, y) = Kexp\left(-\frac{x^2+y^2}{c^2}\right) \quad (3.8)$$

where  $K$  is selected, such that  $\iint F(x, y)dx dy = 1$ ,  $c$  is the Gaussian surround space scales constant, and  $(x, y)$  are spatial coordinates. During simulations, Gaussian function reveals the characteristics of being more regional having nice Dynamic Range Compression (DRC) over a large range of space constant. Selection of space constant is connected with visual angel in the direct observation. In [3.18], published work shows that there is a trade-off between the DRC and color rendition. The middle of the range from 50 to 100 pixels has an acceptable compromise.

### **3.2.2. Multi Scale Retinex (MSR)**

The MSR has been used in conjunction with other algorithms to enhance images. Because of the trade-off between the DRC and color rendition, an appropriate value of scale  $c$  of  $F(x, y)$  in Equation (3.8) must be chosen. For a compromise between the DRC and color rendition, a combination of different weighted scales of SSR that is actually MSR is found to be a good choice. Mathematically, the MSR is given by:

$$MSR = \sum_{n=1}^N \omega_n R_{ni} \quad (3.9)$$

where  $N$  is the number of the scales and  $R_{ni}$  is  $i^{\text{th}}$  component of the  $n^{\text{th}}$  scale from Eq. (3.1). An obvious question about the MSR is that what should be the number of scales, scale values, and associated weights ( $\omega_n$ ). Our simulations show that three scales are sufficient. We use scale values of:  $C_1 = 5$ ,  $C_2 = 85$ , and  $C_3 = 255$  to enhance small, medium, and large scales. The weights are changed to adjust color rendition and dynamic range compression. The MSR based images have significant dynamic range compression in the boundary between the lighted and dark parts, and reasonable color rendition in the whole image scale.

### 3.3. Proposed Method

#### 3.3.1. Applying Principal Component Analysis (PCA)

The PCA is a kind of vector space transformation and is frequently used to reduce multidimensional data sets to lower dimensions for analysis. The PCA transformation is a rotation of the original axis to new orientations that are orthogonal to each other and therefore there is no correlation between variables. In remote sensing, the PCA is usually used for image enhancement. Moreover, the PCA has better result in analysing data especially in data reduction for image processing. Using the PCA, the input RGB image is decorrelated into luminance and chrominance channels. The former component represents the achromatic channel, which carries luminance information and is processed further. While, the latter component represents the chrominance information and is left untreated.

As described earlier, the PCA provides orthogonality among the channels. Therefore, colors remain stable despite the variation in luminance.

### 3.3.2. Applying the MSR

Luminance channel is enhanced using three scales of the MSR. The MSR is a type of retinex algorithm that integrates different SSR outputs to generate a resultant output colored image having good dynamic range compression and color constancy. In our work, we use weight values  $(\omega_n) = \frac{1}{3}$  with three weight scales in Eq. (3.9).

### 3.3.3. Introducing the DFT

The DFT is introduced to speed up the enhancement process. The MSR accomplishes the image enhancement utilizing convolution theorem. The convolution operation is time consuming. If real-time implementation of color image enhancement is desired using the MSR, then it could be a problem since convolution is done in time domain and requires huge time to process an image. To reduce the processing time of the MSR and to optimize the algorithm, the convolution theorem in Fourier Transform is utilized. Therefore, the convolution in time domain is converted to product operation in frequency domain that results in significant decrease in computation time. Mathematical expression for the DFT and its inverse transform for 2-D signal of size  $M \times N$  is given below:

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp \left[ -j2\pi \left( \frac{ux}{M} + \frac{vy}{N} \right) \right] \quad (3.10)$$

$$f(x, y) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(u, v) \exp \left[ j2\pi \left( \frac{ux}{M} + \frac{vy}{N} \right) \right] \quad (3.11)$$

where  $F(u, v)$  is the frequency domain expression of the signal  $f(x, y)$ . Combining the DFT term in Eq. (3.10) and Eq. (3.11) with the MSR given in Eq. (3.9), the new form of the MSR is:

$$MSR(x, y) = \sum_{n=1}^N \omega_n \{ \log I(x, y) - \log [F^{-1}(I'(u, v)F'_n(u, v))] \} \quad (3.12)$$

where  $I'(u, v)$  and  $F'_n(u, v)$  are the DFT solutions of original image  $I(x, y)$  and the  $n^{\text{th}}$  surround function  $F_n(x, y)$ , respectively. The DFT is performed using the Fast Fourier Transform (FFT).

As luminance component is 2-D signal, the 1-D FFT is applied separately into columns and rows. Therefore, the DFT has greatly decreased the processing time of the MSR.

#### **3.3.4. Enhanced Luminance (EL) Ratio Computation**

In the fourth step, the ratio of the EL from step three to the Original Luminance (OL) is calculated. New RGB values for the color image are calculated using:

$$LR = \frac{EL}{OL} \quad (3.13)$$

where  $LR$  = Luminance Ratio,  $EL$  = Enhanced Luminance, and  $OL$  = Original Luminance. In this way, the enhanced luminance channel of the image is acquired.

#### **3.3.5. Computing New RGB Values**

In this step, new RGB values of the image are computed by the following equations.

$$(R_n, G_n, B_n) = \begin{cases} LR \times R_0 \\ LR \times G_0 \\ LR \times B_0 \end{cases} \quad (3.14)$$

where  $R_n$ ,  $G_n$ , and  $B_n$  are the new Red, Green, and Blue values, respectively. The  $R_0$ ,  $G_0$ , and  $B_0$  are original Red, Green, and Blue values, respectively.

#### **3.3.6. Applying Contrast Stretching (CS)**

Finally, the CS is performed on each RGB channels separately to obtain the enhanced output image. Two types of contrast stretching algorithms are the basic Contrast Stretching (CS) and the Ends-in-Search (EiS) [3.8]. In the histogram of a low contrast image, the pixels are concentrated on the right, left, or in the middle. Images having high contrast contain regions of both dark as well as light. The CS is performed on an image to stretch a histogram to fill the

entire dynamic range of the image. For Gaussian or near-Gaussian distributions best results are obtained. Mathematical form of the CS is given by:

$$\begin{cases} new\ pixel = \frac{old\ pixel - low\ pixel}{high - low} \times 255 \\ Low\ Pixel = 8 \leq low \leq 32 \\ High\ Pixel = 255 - low \end{cases} \quad (3.15)$$

For  $3 \times 3$  sliding window, we set coordinates (1, 2) as high pixel, (2, 1) old pixel, (3, 2) low pixel, and (2, 2) new pixel. After the CS is applied on each of the RGB channels, finally, we integrate each of the RGB channels to get the enhanced output image.

### 3.4. Simulation Results and Discussion

We perform detailed simulations on our internal Ubuntu Cloud (UC) setup running on 96 cores Supermicro Super-Server SYS-7047GR-TRF machine with on board 128 GB of RAM with MATLAB 2015 as simulation tool. The proposed algorithm has been applied on 994 colored images having variation in luminance and contrast. The proposed method is compared with other methods in terms of statistical analysis, image quality, and execution time. For experimental setup, size of all the input images is cropped to 256x256 pixels.

Figure 3.3(a) shows the results of state-of-the-art image enhancement methods along with the proposed technique. The enhanced image using the proposed method has better visual appearance, fine overall details, balanced luminance, and contrast across the whole image with no change in original color of the image comparison with other methods. In all the histograms, *x-axis* shows intensity of pixels, while *y-axis* shows the number of pixels. Moreover, histogram of the input test image mostly lies on the left side while for enhanced images the histogram has the better distribution of pixels. For the proposed method histogram of the output image covers the

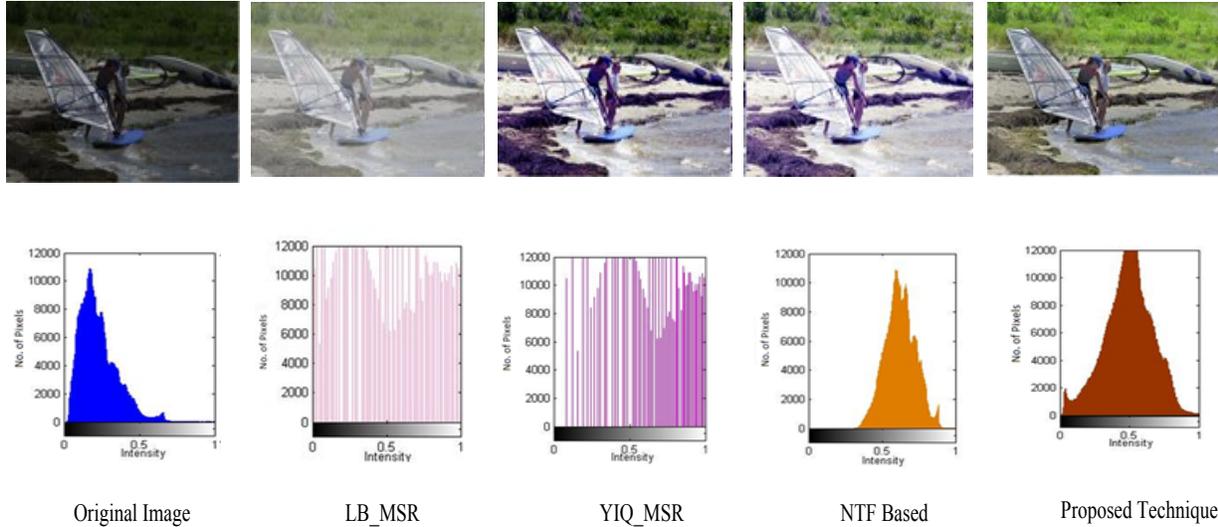


Fig. 3.3. Test Image 1: Image enhancement results.

whole region on  $x$ -axis. Furthermore, Fig. 3.3 shows that proposed algorithm has more clear details than the other image enhancement techniques. Application of the LB\_MSR [3.3] results in a whitish outlook. Outputs of the YIQ\_MSR [3.4] and the NTF based [3.1] do not have the satisfactory outlook. Aforementioned fact is also verified by the relevant histograms that cover whole ranges of  $x$ -axis with too many peaks and valleys. But, the result of the proposed technique has more natural outlook with nice image quality. Grass color, water, and blue color below the feet of the person in image are nicely restored to natural look.

Effectiveness of the proposed algorithm is further demonstrated in Fig. 3.4, where tower, sky color, and home details are revealed nicely than the other methods. The quality of the output image of proposed technique is much better than the compared techniques. Application of the LB\_MSR results in washed-out appearance and therefore results in elimination of fine details, such as windows in front of home. Similarly, application of the YIQ\_MSR and the NTF result in sparkling appearance of tower. Moreover, the land in front of the home has also unnatural outlook. However, the proposed algorithm reveals narrow details nicely, such as sky color, tower, windows, and grass outside of home.

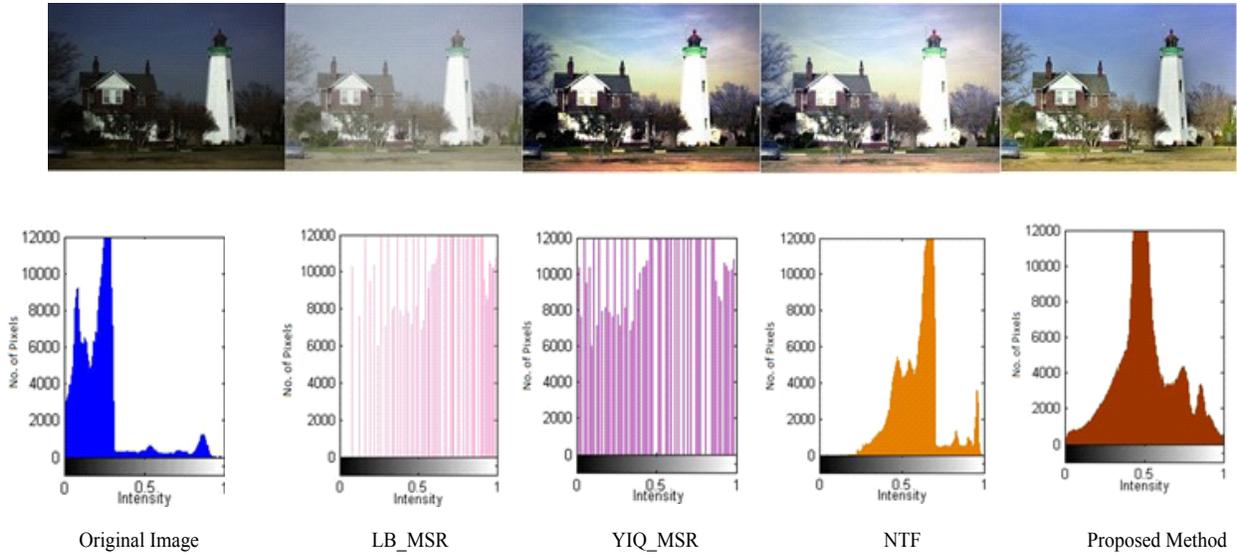


Fig. 3.4. Test Image 2: Image enhancement results.

In Fig. 3.5, trees and sky color outside the white building have more natural outlook in proposed algorithm output than the techniques compared therein. Clearly, the quality and visual appearance of the output image is restored to natural outlook. Moreover, in the output image of our proposed method, one can see that the contrast increases well and thereby preserving image details adequately. Our simulations on many of such outdoor test images have shown similar encouraging results.

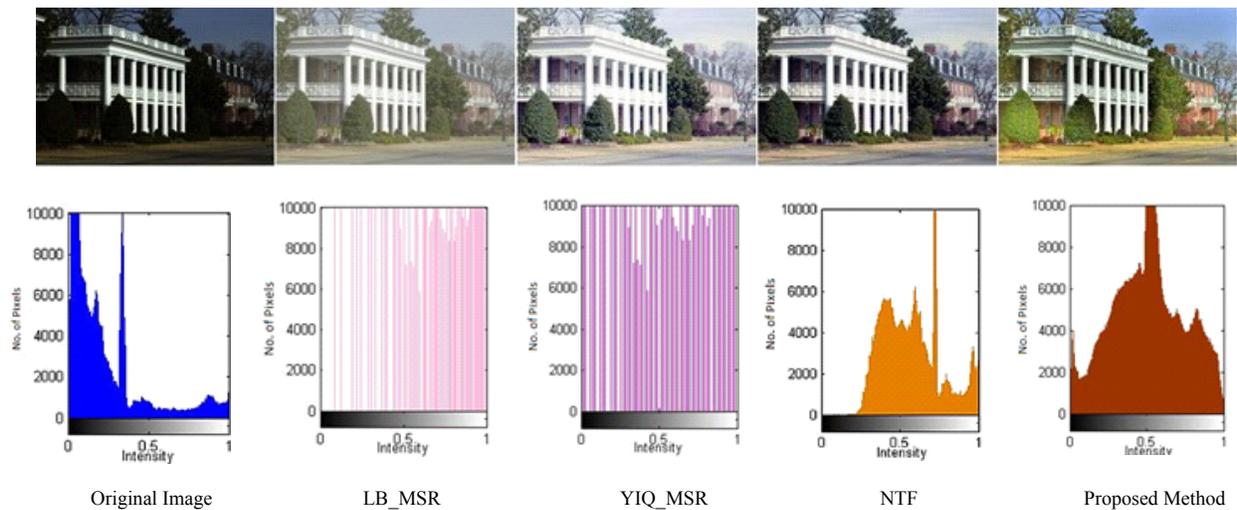


Fig. 3.5. Test Image 3: Image enhancement results.

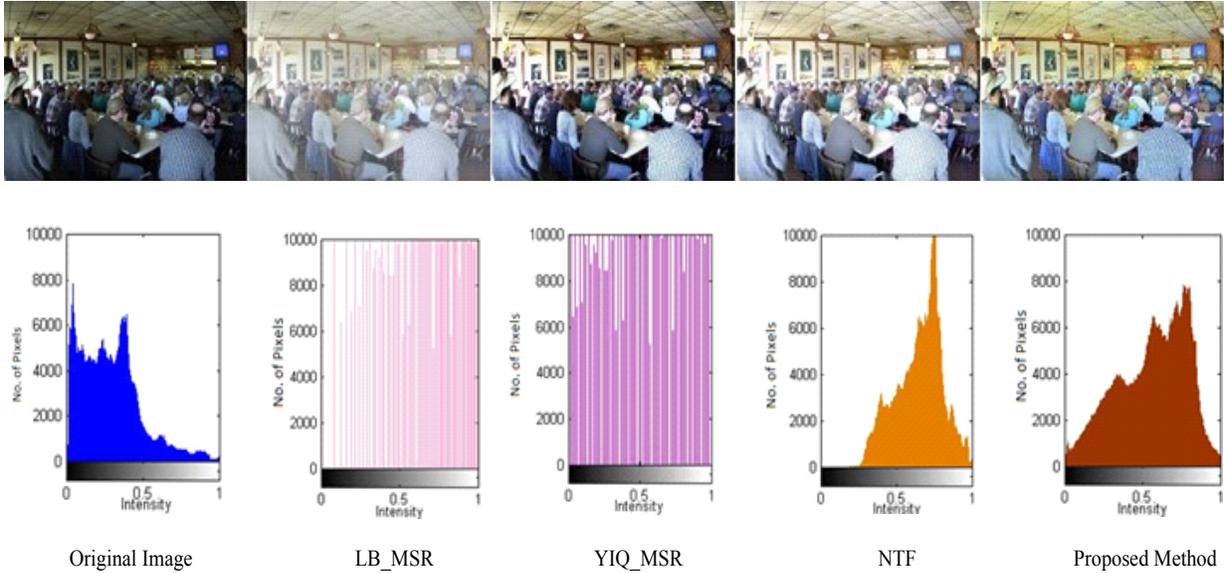


Fig. 3.6. Test Image 4: Image enhancement results.

Similarly, in Figure 3.6, audience and roof details have been enhanced nicely by the proposed algorithm. While, the LB\_MSR, the YIQ\_MSR, and the NTF based algorithms produce over bright, whitish, and blurry outputs. Moreover, the histograms of the outputs of LB\_MSR and YIQ\_MSR have irregular peaks and valleys, while for the NTF based; it lies mostly towards right side of  $x$ -axis. However, the histogram of the proposed algorithm covers the whole range of  $x$ -axis, thereby indicating better distribution of pixels.

To further validate the effectiveness of our developed scheme, we test our algorithm on images captured in very strong sun light, such as in baseball stadium and outdoor environment. Fig. 3.7 shows three input color images containing both dark and bright regions and result of the image enhancement by the LB\_MSR, the YIQ\_MSR, the NTF, and the proposed method. The input images have blackish/whitish appearance. Moreover, the original color of the image is changed after enhancing using image enhancement algorithms. For the YIQ\_MSR and the NTF based approach, the contrast has increased unnecessarily. However, the proposed algorithm on these images reveals nice details in the image. Specifically, in second row and right most image,



Fig. 3.7. Test Image 5: Image enhancement results.

the player face, white ball, boundary between the players, and grass color is nicely enhanced by proposed algorithm. Moreover, top and bottom row images in Fig. 3.7 are also restored to natural look by the proposed method.

**3.4.1. Comparative Analysis**

This section presents comparison of the proposed technique with other image enhancement methods. We present the detailed comparison in terms of useful statistics, such as mean, median, standard deviation, and entropy.

### 3.4.1.1. Mean

Let  $x$  denotes a discrete random variable representing discrete grey-levels (values of the R, G, B channels) in the range  $[0, L-1]$ , and  $p(x_i)$  denotes the normalized histogram component corresponding to the  $i^{\text{th}}$  value of  $x$ . The moment of  $x$  about its mean ( $m$ ) is defined as:

$$u_n(x) = \sum_{i=0}^{L-1} (x_i - m)^n p(x_i) \quad (3.16)$$

$$m = \sum_{i=0}^{L-1} x_i p(x_i) \quad (3.17)$$

where  $m$  is the mean value of  $x$  i.e., the average grey level and reflects the overall brightness of the image.

### 3.4.1.2. Standard Deviation

As can be seen from Eq. (3.16) and Eq. (3.17), when  $u_0 = 1$  and  $u_1 = 0$ . The second moment is given by:

$$u_2(x) = \sum_{i=0}^{L-1} (x_i - m)^2 p(x_i) \quad (3.18)$$

This expression is variance of  $x$  ( $\sigma^2(x)$ ). The Standard Deviation (SD) is defined as the square root of the variance.

$$(SD) = \sqrt{\sum_{i=0}^{L-1} (x_i - m)^2 p(x_i)} \quad (3.19)$$

The SD provides an estimable measure of the regional contrast variations. Visually optimized images are compactly assembled on a single mean value and have high SD. The SD spreads the signal excursions across the dynamic range to a maximum extent while limiting any spatial over/under shoots. Fig. 3.8 demonstrates the detailed statistical analysis comparison.

### 3.4.1.3. Entropy

Entropy is defined as the average amount of information contained in each image. Discrete entropy is the summation of the products of the probability of outcomes times' log

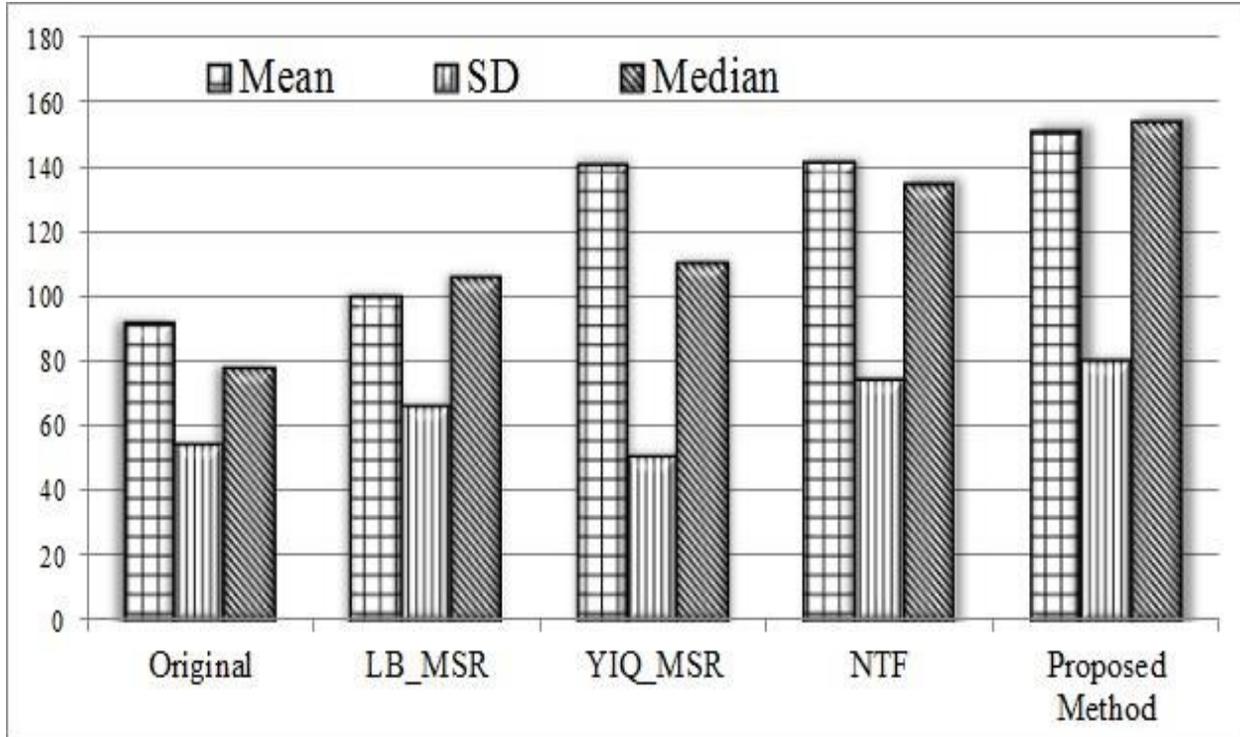


Fig. 3.8. Comparative statistical analysis of the proposed technique with different image enhancement techniques.

the inversed probability of the outcomes. Maximal entropy occurs when all the outcomes are equal. When the outcome is certainty, the minimal entropy is zero. For an image,  $n$  is the grey level,  $p(i)$  is the probability distribution taking all histogram counts. Discrete entropy is expressed as:

$$H(x) = \sum_{i=1}^k p(i) \log_2 \frac{1}{p(i)} - \sum_{i=1}^k p(i) \log_2 p(i) \quad (3.20)$$

In Table 3.2, we show the entropy values from Fig. 3.3 to Fig.3.7. The test images have low entropy values that imply little contrast. Whereas, entropy values of the proposed method are higher than the input degraded test images. The entropy values of the LB\_MSR, the YIQ\_MSR, and the NTF are too higher resulting in a washed out appearance in the final enhanced image. For proposed method, the output images have great contrast from one pixel to the next pixel that results in a natural, nice, and visually appealing image.

Table 3.2. The entropy comparison of different image enhancement techniques.

Figure	Original	LB_MSR	YIQ_MSR	NTF	Proposed Method
Figure 3.3	7.42	7.70	7.25	7.35	7.49
Figure 3.4	7.31	7.41	7.90	7.99	7.51
Figure 3.5	7.42	7.51	7.89	7.90	7.54
Figure 3.6	7.44	7.40	7.89	7.90	7.89
Figure 3.7	7.49	7.55	7.54	7.67	7.70

We successfully enhance 994 diverse images captured in very strong sun light in outdoor environments under non-uniform illuminations.

### 3.4.2. Peak Signal to Noise Ratio (PSNR)

The PSNR is an objective image quality metric. The PSNR is calculated from the difference in pixel values in an image. The PSNR is defined as follows:

$$PSNR_i(x, y) = 10 \log_{10} \left( \frac{255 \times 255}{\frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(x, y) - I'(x, y)]^2} \right) \text{dB} \quad (3.21)$$

where  $I(x, y)$  is original image and  $I'(x, y)$  is enhanced image. The  $M$  and  $N$  represent height and width of the image, respectively. The denominator in Eq. (3.21) represents the Mean Squared Error (MSE) of the original and enhanced images. Table 3.3 shows the PSNR comparison of the proposed technique with other image enhancement methods.

Table 3.3. The PSNR comparison of different image enhancement techniques.

Figure	LB_MSR(dB)	YIQ_MSR(dB)	NTF (dB)	Proposed Method (dB)
Figure 3.3	22.0019	30.1250	31.3491	43.6048
Figure 3.4	46.5942	27.8906	21.9879	46.5942
Figure 3.5	42.5634	43.0196	43.0198	72.2109
Figure 3.6	31.3481	36.4013	46.6264	46.6264
Figure 3.7	24.6704	23.7928	24.8692	36.8596

In Table 3.3, the PSNR values for the proposed method are comparatively better than the other image enhancement techniques. The PSNR values for Fig. 3.4 are same for the proposed method and the LB\_MSR method. Moreover, equal PSNR values are found for the proposed method and the NTF based scheme for Fig. 3.6. Therefore, we observe that the PSNR evaluation does not always correspond to subjective measure. The reason for the aforementioned behavior is that the PSNR eradicates image parts that have noise like appearance and uniformly processes image singularities, such as edges or textures. The fact is also obvious from Fig. 3.4 and Fig. 3.6 where the output images and histograms of each method are entirely different. Therefore, whenever the PSNR values of two images are same, it does not imply that they have the same subjective quality. Next, we evaluated the Absolute Mean Brightness Error (AMBE) and computational complexity as shown in next sections.

### 3.4.3. Absolute Mean Brightness Error (AMBE)

The AMBE is an objective measurement that is used to rate the performance to preserve the original brightness. The AMBE is the absolute difference between the mean of the input image and the enhanced output image and is defined as:

$$AMBE = |E(I(x, y)) - E(I'(x, y))| \quad (3.22)$$

where  $I(x, y)$  and  $I'(x, y)$  represent the original image and enhanced output image, respectively.  $E(.)$  denotes the expected value/statistical mean. We use the AMBE to estimate the excessive brightness change in the image. A lower value of the AMBE implies that the original brightness is preserved in a better way. For all of the cases in Fig. 3.9, the AMBE values of the MSR are much lower than the other image enhancement schemes. Therefore, the proposed MSR technique yields a better quality enhanced output image. The AMBE accurately senses the deviations in overall brightness and can be effectively used to observe the perceptually improved high quality output images.

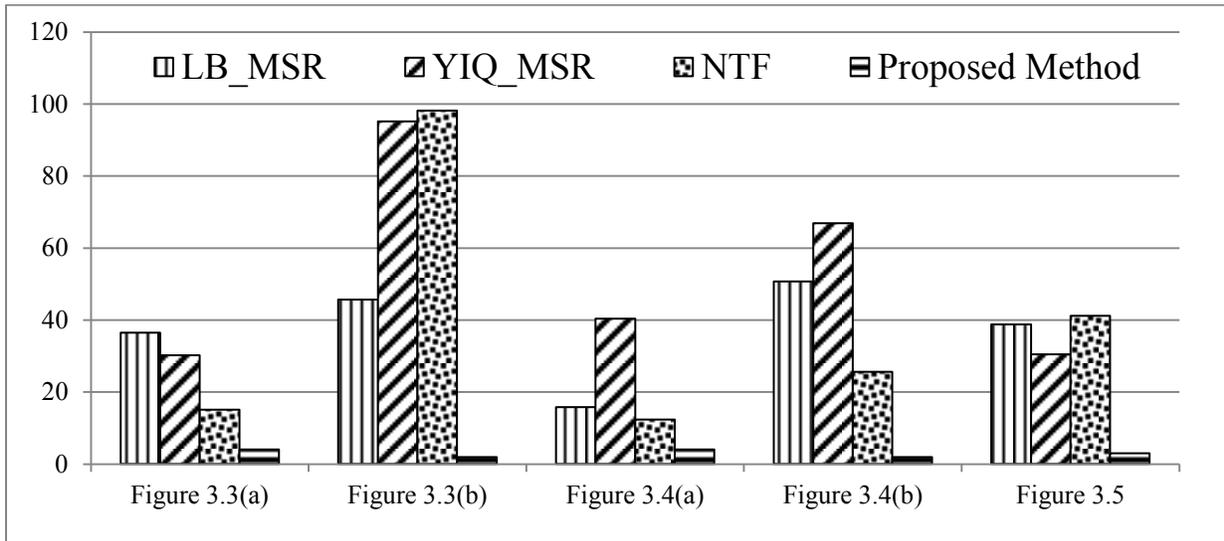


Fig. 3.9. The AMBE comparison of the proposed technique with other image enhancement techniques.

### 3.4.4 Computational Complexity

Computational complexity is evaluated in terms of time consumed in seconds to produce the enhanced output image. Each of the simulation is executed fifty times and results are summarized in Table 3.4. The proposed method is clear winner in terms of execution time followed by the MB\_MSR and YIQ\_MSR. In all the comparisons made, we observe the NTF based algorithm to be most computationally expensive.

Table 3.4. Time cost comparison of the MSR with other image enhancement techniques.

<b>Technique</b> <b>Figure</b>	<b>LB_MSR (s)</b>	<b>YIQ_MSR (s)</b>	<b>NTF (s)</b>	<b>Proposed Method (s)</b>
<b>Figure 3.3(a)</b>	1.6217	2.1740	4.5291	<b>1.4229</b>
<b>Figure 3.3(b)</b>	1.6221	2.1810	4.5268	<b>1.4230</b>
<b>Figure 3.4(a)</b>	1.6219	2.1860	4.5299	<b>1.4229</b>
<b>Figure 3.4(b)</b>	1.6207	2.1790	4.5291	<b>1.4227</b>
<b>Figure 3.5</b>	1.6216	2.1710	4.5297	<b>1.4234</b>
<b>Average</b>	1.6216	2.1780	4.5289	<b>1.4230</b>

### 3.5. Conclusions and Future Work

We presented a technique to enhance colored images using the MSR theory. The proposed technique has six major steps including application of the PCA, the MSR, the DFT, enhanced luminance ratio with new RGB values computation, and application of the CS. Extensive experiments were performed on 994 colored images captured in strong sunlight and under non-uniform lighting conditions. The high enhancement accuracy of the proposed algorithm demonstrates its effectiveness. Results of the proposed technique outperform the state-

of-the-art image enhancement methods in terms of perceptual quality, statistical comparison, and computational complexity.

In future, we aim to optimize the proposed algorithm to make it feasible to be deployed in embedded systems. The algorithm is potentially very useful for digital cameras, because a distorted image can be automatically enhanced in hardware without any manual intervention. An efficient implementation of the proposed method in embedded environment will be useful for modern flat display monitors and smart TVs to rectify degraded contents, such as images and videos. Moreover, several computer vision applications will benefit from the proposed algorithm by using it as a preprocessing module.

### **3.6. References**

- [3.1] D. Ghimire and J. Lee, “Nonlinear transfer function-based local approach for color image enhancement,” *IEEE Transactions on Consumer Electronics*, Vol. 57, No. 2, 2011, pp. 858-865.
- [3.2] R. Atta, and M. Ghanbari, “Low-contrast satellite images enhancement using discrete cosine transform pyramid and singular value decomposition,” *IET Image Process*, Vol. 7, No. 5, 2013, pp. 472–483.
- [3.3] B. Sun, W. Tao, and W. Chen, “Luminance based MSR for color image enhancement,” *International Congress on Image and Signal Processing*, 2008, pp. 358-362.
- [3.4] Z. Jiao, B. Xu, “An image enhancement approach using retinex and YIQ,” *International Conference on Information Technology and Computer Science (ITCS)*, 2009, pp. 476-479.
- [3.5] S. Xiaoping and W. Fang, “An image enhancement method using the quantum-behaved particle swarm optimization with an adaptive strategy,” *The Scientific World Journal, Mathematical Problems in Engineering*, 2013, pp. 1-14.

- [3.6] S. Bettahar, A. Boudghene, P. Lambert, and A. Benoit, "A PDE-based enhancement of color images in RGB space," *IEEE Transactions on Image Processing*, Vol. 21, No. 5, 2012, pp. 2500-2512.
- [3.7] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *IEEE Transactions on Image Processing*, Vol. 22, No. 1, 2012, pp. 70-79.
- [3.8] Z. Mahmood, T. Ali, and S. Khattak, "A color image enhancement technique using multi scale retinex," 11<sup>th</sup> Int. Conf. Frontiers of Information Technology (FIT), Pakistan, 2013, pp. 119-124.
- [3.9] B. Yu, B. Kim, and K. Lee, "Visibility enhancement based real-time retinex for diverse environments," 8th International Conference on Signal Image Technology and Internet Based Systems, 2012, pp. 72-79.
- [3.10] J. Tao and Y. Pingxian, "An enhancement algorithm for video surveillance image based on multi-scale retinex," *International Conference on Computational and Information Sciences*, 2011, pp. 22-25.
- [3.11] W. Kyung and Y. Ha, "Real-time multi-scale retinex to enhance night scene of vehicular camera," 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), 2011, pp. 1-4.
- [3.12] A. Poursaberi, N. Svetlana, and M. Gavrilova, "Modified multiscale vesselness filter for facial feature detection," 4<sup>th</sup> International Conference on Emerging Security Technologies, 2013, pp. 21-24.

- [3.13] X. Changyan, M. Staring, and Y. Wang, "Multiscale bi-Gaussian filter for adjacent curvilinear structures detection with application to vasculature images," *IEEE Transactions on Image Processing*, Vol. 22, No. 1, 2013, pp. 174-188.
- [3.14] J. Jang, Y. Bae, and R. Beom, "Contrast-enhanced fusion of multi sensor images using sub band-decomposed multi scale retinex," *IEEE Transactions on Image Processing*, Vol. 21, No. 8, 2012, pp. 3479-3490.
- [3.15] R. Zhang, Y. Huang, and Z. Zhao, "A ultrasound liver image enhancement algorithm based on multi-scale retinex theory," *5th International Conference on Bioinformatics and Biomedical Engineering, (ICBBE)*, 2011, pp. 1-3.
- [3.16] E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Natl. Sci*, 83, 1986, pp. 3078-3080.
- [3.17] L. Meylan and S. Susstrunk, "High dynamic range image rendering with a retinex-based adaptive filter," *IEEE Transactions on Image Processing*, Vol. 15, No. 9, 2006, pp. 2820-2830.
- [3.18] R. Chouhan, R. Jha, and P. Biswas, "Enhancement of dark and low-contrast images using dynamic stochastic resonance," *IET Image Processing*, Vol. 7, No. 2, 2013, pp. 174-184.

## **4. AUTOMATIC PLAYER DETECTION, RECOGNITION, AND IDENTIFICATION FOR SPORTS ENTERTAINMENT APPLICATIONS<sup>2</sup>**

This paper is submitted to Pattern Analysis and Applications. The authors of the paper are Zahid Mahmood, Tauseef Ali, Shahid Khattak, Laiq Hassan, and Samee U. Khan.

### **4.1. Introduction**

Sports' broadcasting is revolutionized due to the use of advanced video and image processing techniques. Continuous efforts are being made to enhance the contents of captured and transmitted images to provide the viewers more detailed and accurate information about game status and the players [4.1]. In recent times, computer generated visualizations are increasingly used in sports broadcasting to enhance the viewers experience like displaying simple data or current scores. Augmented Reality (AR) is a technique for overlaying virtual objects onto the real world. The AR of real sports games facilitates viewers to engage and immerse into the action. In the AR applications, virtual reality replaces the real world with a simulated one. Present AR techniques for sports video can be divided into two major categories. The first category is focusing on generating virtual scenes by means of multiple synchronous video sequences of a given sports game and second category aims at synthesizing virtual sports scenes from TV broadcasted video.

---

<sup>2</sup> Zahid Mahmood is the primary developer of the work presented in this chapter. Zahid Mahmood drafted and revised all versions of this chapter. Tauseef Ali, Shahid Khattak, and Laiq Hassan collected and provided real sports images from different sports arena to test the developed application. Samee U. Khan served as the supervisor of the proposed application and checked the maths and statistical analysis performed by Zahid Mahmood.

Until now, there is limited research on development of AR applications. The AR has been applied to few entertainment applications by using vision based tracking technique [4.2].

In [4.3] technique is presented to generate a visual enhancement for TV broadcasted court net sports. It had a major drawback as system generates a redundant virtual scene which abruptly affects system performance. In [4.4] virtual scenes are generated by means of multiple synchronous video sequences of a given sports game, but this technique also fails in real-time transmitted videos and images because of synchronization problems. A graphics based animation AR technique has been proposed in [4.5], but it also fails to produce the satisfactory results as it is unrealistic due to complete loss of texture and motion of the players. In [4.6] authors tried to insert and manipulate 3D virtual contents into broadcasted tennis video. This technique also struggles in situations like invisibility of one net.

In [4.7] the developed system performs a camera calibration algorithm to establish a mapping between the soccer court field in the image and that of the virtual scene. It had a major drawback of player pose restriction. User had to opt from only 3 defined poses such as walk, stop and run. The confinement to only three poses eliminates improvement in system performance. All these previous efforts have yet to reach the acceptable level of reality. Main contributions of our work are as follows.

- Our work is from the extension of the above in which we have developed a real AR based application on real-time transmission of baseball game images [4.8]. The developed system is free from problems in aforementioned techniques like synchronization, generation of virtual redundant scenes, and player pose.
- The developed application does not have problems, such as loss of texture and motion of players. All the viewers have to do is, just to capture an image using camera or smart

phone and after that detection and recognition of players along with the display of personal information occurs automatically.

- For the task of player and subsequent face detection, our method puts less restriction on the input images. The player face recognition algorithm requires frontal or near frontal faces. However it is quite robust to recognize low resolution faces. For example, a face size of 10x10 pixels can also be successfully recognized.

Fig. 4.1 presents a framework for displaying player statistics in still images that can easily be extended to real-time videos. In the step 1, we detect players using AdaBoost [4.9] and Haar-like features. In the step 2, we use the same algorithm to detect face in each player image. Step 3 matches the face with a database of player faces, while Step 4 retrieves statistics of players, such as name of the player, height, age, sports record, and nationality. Face recognition is performed using AdaBoost with LDA [4.10] as weak learner and NNC. Table 4.1 shows important notations with their meaning used in this chapter.

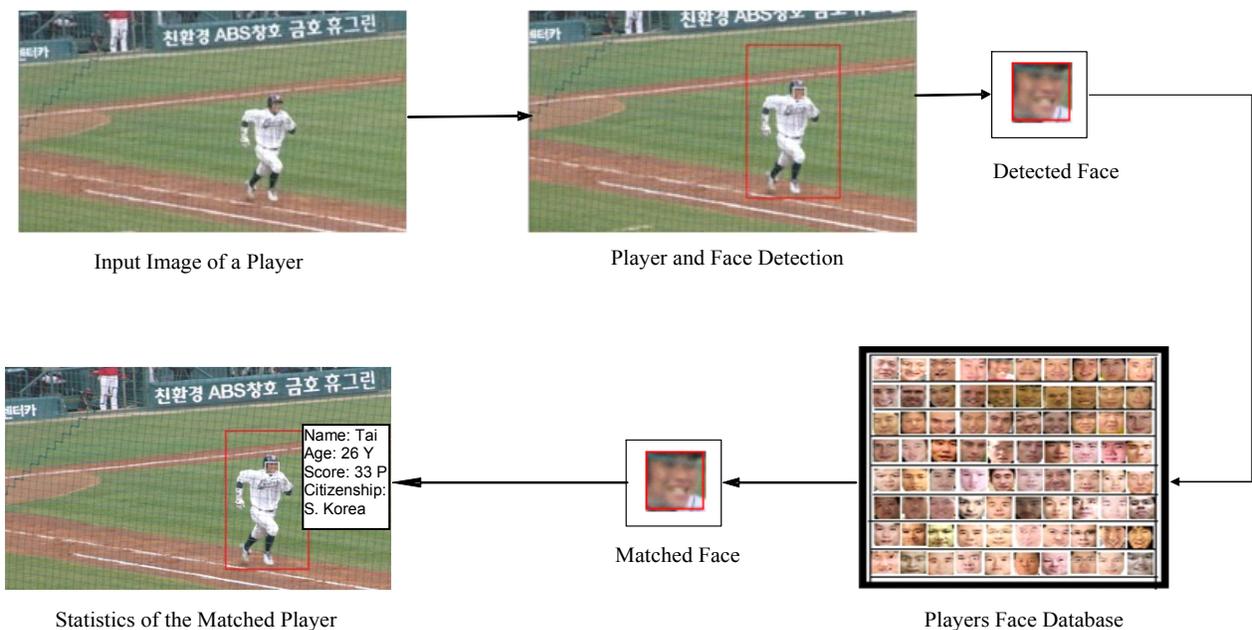


Fig. 4.1. Framework of our proposed application.

Table 4.1. Notations and their meanings.

Notation	Description
AdaBoost	Adaptive Boosting
AR	Augmented Reality
LDA	Linear Discriminant Analysis
BLDA	Boosted Linear Discriminant Analysis
NNC	Nearest Neighbor Classifier
$g(x)$	Classifier obtained to detect players in input image.

## 4.2. Player Detection using AdaBoost

AdaBoost algorithm constructs a strong classifier as a linear combination of weak classifiers. It improves the accuracy based on a series of weak classifiers employed at different stages. For player and subsequent face detection in our framework, we use the popular Viola and Jones approach [4.11]. A robust player classifier is obtained by supervised AdaBoost learning. Given a sample set of training data  $\{x_i, y_i\}$ , the AdaBoost algorithm selects a set of weak classifiers  $\{h_j(x)\}$  from a set of Haar-like rectangle features and combines them into a strong classifier. The strong classifier  $g(x)$  is defined as follows:

$$g(x) = \begin{cases} 1 & \sum_{k=1}^{k_{\max}} \alpha_k h_k(x) \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

where  $x$  is the input image,  $k$  is the number of weak classifiers,  $\theta$  is the decision threshold and  $g(x)$  is the final strong classifier.

A set of 630 positive samples and 1100 negative samples are used to train the classifier. The training data is selected from our own images database of baseball players. To speed up the computation time, all player images are first converted into gray-scale and size normalized to 20x40 pixels. Some examples of positive and negative data are shown in Fig. 4.2. Negative



Fig. 4.2. Example of positive (left) and negative (right) training samples.

samples are obtained from different public databases. A rough estimate of the size of the smallest and largest player can be obtained from the size of the input image and therefore the size of detection window is restricted to a certain range. The width of detection window is set to half of the height. This improves accuracy and searching time. An image captured in the sports arena contains players as well as other objects, such as audience and advertising boards. So the sole purpose of player detection is to confine the AR application processing area by eliminating the redundant information as mentioned above. Detecting player provides the viewers the exact position of their favorite players and increases the aesthetic experience as well.

### 4.3. Player Face Detection

Face detection employs boosting methodology, which is one of state-of-the-art approaches for object detection. Face detection algorithm based on feature subspace, maps face image to a certain feature space. The distinction between face and non-face is made according to the distribution law in feature subspace. In our work the objective of player face detection is to obtain spatial location of player faces. The simple features used are reminiscent of Haar basis functions. The Rectangle features are computed using Integral Image, which is sum of the pixels above and to the left of the point  $(x, y)$ . The Integral Image  $I(x,y)$  is defined in Eq. 4.2. Fig. 4.3 shows the two, three and four rectangle features.

$$I(x, y) = \sum_{x^* \leq x, y^* \leq y} i(x^*, y^*) \quad (4.2)$$

where  $I(x, y)$  is Integral Image at points  $(x, y)$  and  $i(x, y)$  is original image.

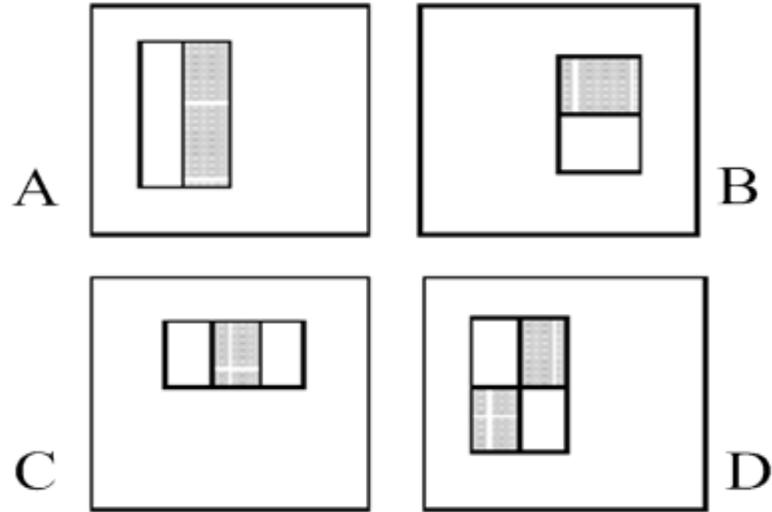


Fig. 4.3. Example of rectangle features shown relative to the enclosing detection window.

To detect player(s) face(s), use of Haar-like features and AdaBoost algorithm for both feature selection and classification results in less computation and execution time. The details of face detection approach are similar to player detection except for the use of different training data to build the classifier. After the player faces are detected, they are used as input to face recognition algorithm.

#### 4.4. Player Face Recognition

Currently, efforts are on for efficient recognition of face images. In our framework, the low resolution is a primary concern because the images are captured during sports in a big sports arena. Therefore, we selected the AdaBoost algorithm with Linear Discriminant Analysis (LDA) [4.10]. It is basically used to extract feature from the face image. The classification is performed using Nearest Neighbor Classifier (NNC). Additional advantage of the AdaBoost based approach is that it does not require large set of training images. We test the algorithm with a set of face images having a resolution of 5x5 pixels and a pose variation from frontal up to  $\pm 35^\circ$ . The

employed AdaBoost based approach outperforms standard PCA and LBP based approaches in terms of correct recognition rate considering both pose and low resolution [4.11, 4.12].

In the AdaBoost-LDA based face recognition algorithm, the performance of traditional LDA based approach is enhanced by incorporating it in the boosting framework. Each round of boosting generalizes a new LDA subspace particularly focusing on samples that are misclassified in the previous LDA subspace. The final feature extractor module is an ensemble of several specific LDA solutions. This kind of ensemble based approach takes advantage of both boosting and LDA and outperforms the LDA based systems in complex face recognition tasks [4.10].

#### **4.5. Player Statistics Retrieval**

The final step of the proposed AR application is to display the statistics of the player, once the input player's face is matched with a face in the database of players face images. We store players' information (statistics) in a database together with the identities of the players. For a detected input face image, the face recognition algorithm calculates a similarity metric called score of the input face to all face images in the database. The pair resulting in the maximum value of the score is considered as the correct match. The identity of the player corresponding to the maximum score is passed on to a module that retrieves the player's statistics. Finally, a rectangle is drawn nearby the player facial region and the statistics are displayed.

#### **4.6. Simulation Results**

Our proposed system is composed of three modules: player(s) detection, face(s) detection, and face(s) recognition. It is necessary to analyse the results of each module separately besides considering the performance of the whole system as a 'black box'. We use a dataset of 412 diverse images with different number of players in various lighting conditions and small



Fig. 4.4. Face database of 150 different players.

pose variation of face from frontal. To better evaluate the system feasibility for real world applications, we use two types of images: images captured with a smart phone and images captured with a good quality digital camera. There are a total of 150 different players in the database as shown in Fig 4.4. In Table 4.2, we give a quantitative summary of the performance of each module. Number of players per image essentially demonstrates the size of detected player. Fewer numbers of players per image implies image is captured from a small distance and player(s) in image has high resolution that helps to improve the performance of subsequent steps. In the subsequent sections, we discuss and analyze each module developed in our work.

**4.6.1. Player and Face Detection**

Fig. 4.5 shows some example images with detected players and faces by employed player detection methodology and feasibility of our approach for real-world scenarios. Player and face detection is achieved on images containing several players with variation in lighting conditions and small deviation from frontal pose. Players with full occlusion and extremely small sized players, such as 4x4 could not be detected by player detection module.

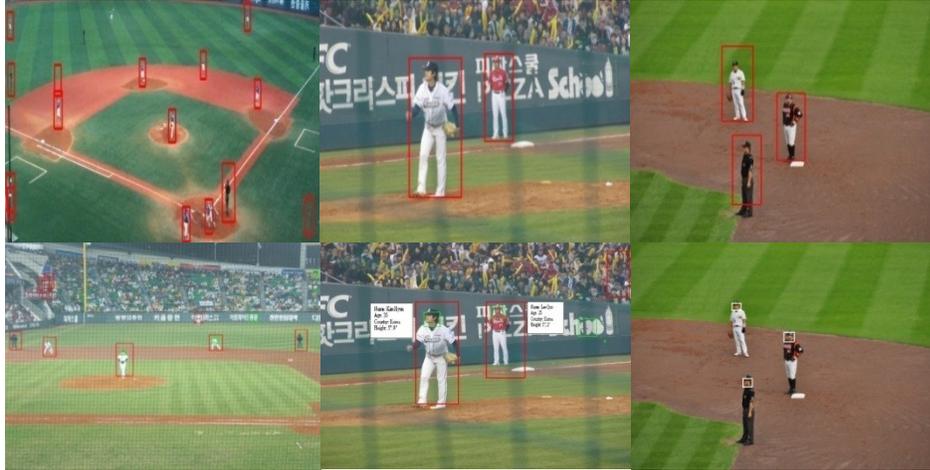


Fig. 4.5. Sample images illustrating performance of players and face(s) detection module.

#### 4.6.2. Players Face(s) Recognition

We partition the database randomly into two subsets: the training set  $Z$  and test set  $Q$ . Face Recognition algorithm is initially trained with  $Z$  training images and the resulting face recognizer is then applied to  $Q$  test images to observe classification error rate. Once the input face is matched to a face in the database, personal information of the matched player, such as

Table 4.2. Execution time of each module.

No of Players	Player Detection Time (milli seconds)	Player Face Detection Time (milli seconds)	Player Face Recognition Time (seconds)	System Execution Time (seconds)
1	490.87	190	2.21	2.89
2	490	190	2.21	2.89
3	540.11	240	2.23	3.01
4	598.33	263	2.25	3.10
5	1211.86	301.61	2.35	3.86
6	1211.86	301.61	2.35	3.86
7	1211.86	305.77	2.37	3.81
14	1221.62	Not Processed	Not Processed	Not Processed
17	1251.78	Not Processed	Not Processed	Not Processed

age, score, and nationality is retrieved from database of players' statistics. The developed application shows the feasibility of computer vision based approaches for enhanced sports broadcasting. Table 4.2 shows execution time of each module. Fig. 4.6 shows the final output of the proposed system on some sample images from our database of 412 images. Most of images contain several players at various poses and lighting conditions. In some cases only player detection and face detection succeeds, but recognition fails.

Fig. 4.7 shows accuracy of face recognition module when the numbers of training images are varied. In some cases only player detection succeeds and both subsequent face detection and recognition fails. The images in which player face is not visible are not processed further and in such cases no information about the player is displayed. Such kind of experiments is difficult to perform with players' database because of the difficulty to obtain many images of a single player

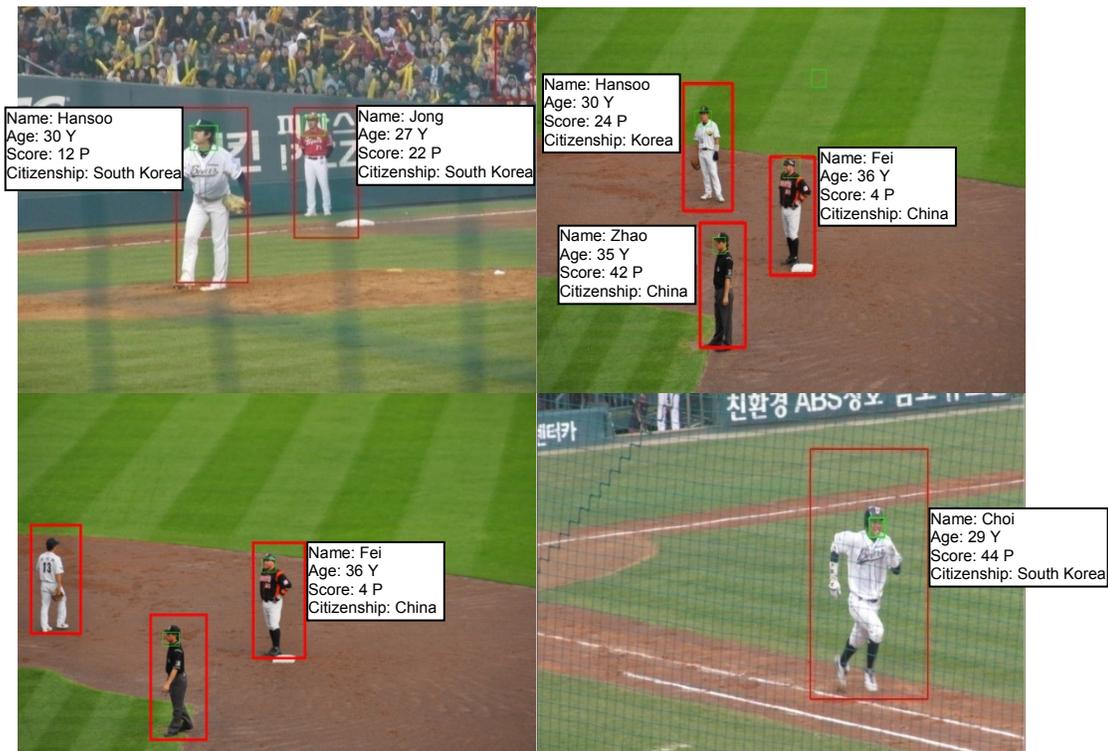


Fig. 4.6. Output of proposed system on sample images from database.

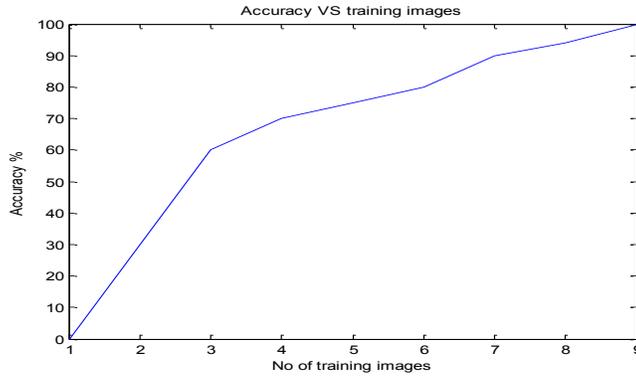


Figure 4.7. Accuracy with number of training images used per subjects.

during play particularly when we want to have more players in our database such as we consider 150 players in our experiments. It is clear as we increase the number of training images; accuracy is increased and reaches almost 100%.

Table 4.3 summarizes our results for 30 players in the database. The major obstacle to achieving higher accuracy from face recognition module is due to variation in pose. The face recognition module works well for as small as size of 5x5 with even 2 training samples. Players face databases ranges from size 231x251 to 10x10 pixels. Therefore, the system performs well in most of recognition tasks.

Table 4.3. Recognition performance with various size face samples.

Players	Error (%)					
	Face size: 20x20 pixels		Face size: 10x10 pixels		Face size: 5x5 pixels	
	(1Tr & 5Ts)	(2Tr & 5Ts)	(Tr & 5Ts)	(2Tr & 5Ts)	(Tr & 5Ts)	(2Tr & 5Ts)
30	15.5	5.5	16.4	11.75	18.5	11.22

#### 4.7. Conclusions and Future Work

We proposed the feasibility of computer vision based approach for application of enhanced end-user experience of sports. The primary goal of this work is to process images captured by an audience using digital camera or smart phone during play and display the relevant

statistics of each player depicted in the images. Statistics of interest may include age, score, and nationality. This work shows the feasibility of developing a real-time augmented reality application for enhanced users' experience and is based on computer vision and image processing techniques. One simple usage scenario of such system is when viewers are sitting in any public place or in a stadium and they do not have access to TV which shows important information of players and game status. Viewers using their smart phones or camera can start recording a video, enhancing video/image where detection and recognition of players and statistics display occur automatically with each frames in real-time.

In future, we aim to address the non-uniform illuminations in captured player images. Moreover, we aim to develop an image enhancement technique that will ultimately help in improving the face detection and recognition accuracy modules. Furthermore, we also aim to shift the developed AR application to the cloud computing environments [4.14].

#### **4.8. References**

- [4.1] Z. Mahmood, T. Ali, and S. Khattak, "Automatic player detection and recognition in images using AdaBoost," *Applied Sciences and Technology (IBCAST), 9<sup>th</sup> International Bhurban Conference*, 2012, pp. 64-69.
- [4.2] J. Looser, R. Grasset, and M. Billinghurst, "A 3D flexible and tangible magic lens in augmented reality," *Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007, pp. 51–54.
- [4.3] J. Han and D. Farin, "A real-time augmented-reality system for sports broadcast video enhancement," *Multi Media*, 2007, Augsburg, Bavaria, Germany.
- [4.4] N. Inamoto and H. Saito, "Free viewpoint video synthesis and presentation of sporting events for mixed reality entertainment," *In Proceedings of ACM ACE, Vol.74*, 2004, pp.42-50.

- [4.5] D. Liang, Y. Liu, Q. Huang, G. Zhu, S. Jiang, Z. Zhang, and W. Gao. "Video2cartoon: generating 3D cartoon from broadcast soccer video," In Proceedings of ACM Multimedia, 2005, pp.217-218.
- [4.6] X. Yu, X. Yan, T. Chi, and L. Cheong, "Inserting 3D projected virtual content into broadcast tennis video," In Proceedings of ACM Multimedia, 2006, pp.619-622.
- [4.7] K. Matsui, M. Iwase, M. Agata, T. Tanaka, and N. Ohnishi, "Soccer image sequence computed by a virtual camera," In Proceedings of CVPR, 1998, pp.860-865.
- [4.8] X. Yu, X. Yan, T. Chi, and L. Cheong, "Inserting 3D projected virtual content into broadcast tennis video," In Proceedings of ACM Multimedia, 2006, pp.619-622.
- [4.9] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, 2004, pp. 137–154.
- [4.10] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Transactions on Neural Networks*, Vol. 17, No. 1, 2006, pp. 166- 178.
- [4.11] Z. Mahmood, T. Ali, S. Khattak, and S. U. Khan, "A comparative study of baseline algorithms of face recognition," 13<sup>th</sup> International Conference on Frontiers of Information Technology (FIT), 2014, Pakistan, pp. 63-68.
- [4.12] Z. Mahmood, T. Ali, S. Khattak, and S. U. Khan, "The effects of pose and image resolution on automatic face recognition," *IET Biometrics*, 2015. (Accepted and to appear).  
Online version of this manuscript can be found at: [www.sameekhan.org/publications](http://www.sameekhan.org/publications).
- [4.13] J. Yu, Y. Rui, and B. Chen, "Exploiting click constraints and multi-view features for image re-ranking," *IEEE Transactions on Multimedia*, Vol. 16, No. 1, Jan 2014, pp. 159-168.

[4.14] Z. Mahmood, M. Jawad, M. U. S. Khan, T. Ali, and S. U. Khan, “MCB-AR: Mobile cloud based augmented reality framework for sports applications,” Under revision in IEEE Transactions on Cloud Computing, 2015.

# 5. EAR: ENHANCED AUGMENTED REALITY FRAMEWORK FOR SPORTS ENTERTAINMENT APPLICATIONS<sup>3</sup>

This paper is submitted to IEEE Transactions on Circuits and Systems for Video Technology. Authors of the paper are Zahid Mahmood, Tauseef Ali, and Samee U. Khan.

## 5.1. Introduction

In Chapter 2, we introduced the readers with image enhancement and AR. This chapter presents the details of the EAR framework. Technique described in Chapter 3 has been used as a preprocessing technique to improve the quality of the input image that helps in detection and recognition of players described in Chapter 4. The proposed EAR framework consists of five steps. In the first step, the input image, which may be degraded by strong sunlight, is enhanced by applying our developed Multi Scale Retinex (MSR) algorithm [5.13]. Applying the MSR algorithm results in higher detection (player and face) and recognition (face) accuracy.

In the second and third steps player and face detection are performed using Adaptive Boosting (AdaBoost) algorithm with Haar-like features for both feature selection and classification [5.12]. In the fourth step, we use a face recognition algorithm to match an input face with the faces in the database. For face recognition, we use Boosted Linear Discriminant Analysis (BLDA) for feature selection and Classic Nearest Neighbor Classifier (CNNC) for classification [5.11], [5.18]. In the fifth step, statistics of each matched player, such as name, age,

---

<sup>3</sup> Zahid Mahmood is the primary developer of the framework described in this chapter. Zahid Mahmood drafted and revised all versions of this chapter. Tauseef Ali collected and supplied original degraded sports images to test the developed framework. Samee U. Khan supervised the work and checked the statistical analysis in the chapter performed by Zahid Mahmood.

and nationality are retrieved based on the output of the player face recognition module. We test the system on images captured during soccer, cricket, and baseball games. However, the system can be deployed in any sports, where the viewers capture live video/images using digital cameras/smartphones while sitting in a stadium. Detailed experiments are performed on 2096 diverse images captured using a digital camera and smartphone during live sports. These images contain players in different sizes, pose, position, facial expressions, and illuminations. Accuracy of the developed EAR system is very high in most of the cases. However, accuracy of the face recognition module degrades when the face images are not frontal and the pose exceeds  $\pm 35^\circ$ . As the details of image enhancement, detection, and recognition have already been described. Therefore, to avoid duplications, we directly present the detailed simulation results below.

## **5.2. Simulation Results**

### ***5.2.1. Player Image Enhancement Module***

Fig. 5.1 shows some enhanced player images using the proposed image enhancement algorithm. In the left column of Fig. 5.1, the original images are shown that are degraded due to very low or strong sunlight. The enhanced images have better visual appeal and improved perceptual quality. The enhancement algorithm is very useful in such applications because it is expected that the images can be degraded by various factors, such as strong sunlight. In Table 5.1, we give a quantitative summary of the performance of each module before and after player image enhancement. Number of players per image gives an indication of the size of the detected players and generally decreases the detection and recognition accuracy.

### ***5.2.2. Player Detection Module***

The player detection module gives satisfactory results after the image enhancement. It successfully detects players in various poses, such as standing, running, throwing, waiving

hands, and diving as shown in Fig. 5.1. For difficult cases, such as audience an advertisement boards behind the players, the detection module is up to the mark. However, the case where only a small part of the player body is visible cannot be handled in most cases.



Fig. 5.1. Complete output of the developed EAR system on sample images: (a) original degraded input player images, (b) player image enhancement, (c) player detection, (d) player face recognition and display of statistics.

Table 5.1. Accuracy of the developed EAR framework.

No of Players per image	% Accuracy of the EAR Before and After Player Image Enhancement					
	Player and Face Detection (Before Enhancement)	Player and Face Detection (After Enhancement)	Face Recognition (Before Enhancement)	Face Recognition (After Enhancement)	Complete System (Before Enhancement)	Complete System (After Enhancement)
1	68.50	100	65.01	98.22	66.75	99.11
2	61.77	100	63.5	97.5	62.63	98.75
3	57.96	97.90	55.7	92.5	56.83	95.20
4	56.40	97.39	49.75	89.66	53.075	93.52
5	54.29	96.74	46.25	85.50	50.27	91.12
6	54.01	94.09	40.37	81.10	47.19	87.59
7	47.79	90.49	NP	NP	NP	NP
14	40.54	84.86	NP	NP	NP	NP

### 5.2.3. Player Face Detection Module

Player face detection module takes detected players as its input and detects the faces of the players up to face size of 5x5 pixels. Player face detection results can be seen in Fig. 5.1. We observed that player face detection module performs well after player image enhancement in all cases except when detected players are very small sized or when there is large variation in pose. However, in our current framework, we handle pose variation from frontal ( $0^{\circ}$ ) to  $\pm 35^{\circ}$ . The design of our framework easily allows for any future improvements, such as occlusion or large pose deviation because of its modular nature where a different implementation of one module does not affect the overall system.

### 5.2.4. Player Face Recognition Module

To perform the player face recognition experiments, we partition the database randomly into two subsets: the training set Z (8 images of each player) and test set Q (2 images of each player). Player face recognition algorithm is initially trained with Z training images and the resulting face recognizer is applied to Q test images to observe Classification Error Rate (CER). From Table 5.1, it is evident that the image enhancement module has a significant impact on the

face recognition module as well. Development and deployment of a robust face recognizer to large pose variations can be considered as future work for improving accuracy of the current system.

In Fig. 5.1 players with displayed statistics show success of all the modules of the system. In some cases only player detection and face detection succeeds. In some cases only player detection succeeds and both subsequent face detection and recognition fails. The images in which player face is not visible are not processed further and in such cases no information about the player is displayed. For extremely low resolution and images in which players are overlapping, face recognition algorithm does not work well. In such cases, player and face detection still exhibit considerable accuracy. Fig. 5.2 shows comparison and accuracy of the player face recognition module when the numbers of training images are varied before and after player image enhancement. Such kind of experiment is useful because in a certain situations, the number of training images per player available for training might be limited and the system should be able to perform well even if there are less training images available [5.6]-[5.10].

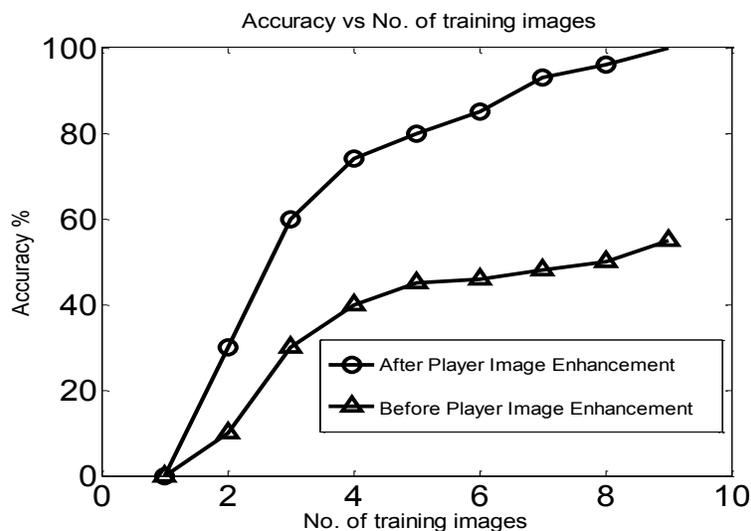


Fig. 5.2. Accuracy comparison by varying training images with and without player image enhancement.

### 5.2.5. Further Analysis of Face Recognition Module

We perform extensive experiments to ensure that the player face recognition algorithm is robust. One aspect of the learning based algorithms is the required number of training samples per subject to train the classifier. We take a subset from the whole player database. The subset contains 30 different players where each player has 10 different images. Different combinations of the training and testing samples are used to obtain classification error rate. In Table 5.2, we show results of the sensitivity of the classification error rate measured for various learning tasks arising from different database partitions. Each combination of the training and testing samples is executed fifty times to obtain minimum average Classification Error Rate (CER).

In Table 5.2, the e-30 (T\*) shows that iteration number is set to 30 and (T\*) shows the iteration number on which minimum error rate is obtained. As we increase the number of training samples, the CER decreases. For enhanced images, the CER decreases drastically and approaches to zero when there are nine training and one testing sample.

Table 5.2. Average minimum classification error rate.

Combination	Mean Error Rate %	
	(Before Enhancement)	(After Enhancement)
(2tr) & (8tst), e <sub>-30</sub> (T*)	26.00 (21)	15.90 (15)
(3tr) & (7tst), e <sub>-30</sub> (T*)	22.22 (29)	12.08 (17)
(4tr) & (6tst), e <sub>-30</sub> (T*)	14.44 (11)	4.14 (1)
(5tr) & (5tst), e <sub>-30</sub> (T*)	11.20 (30)	3.00 (10)
(6tr) & (4tst), e <sub>-30</sub> (T*)	6.87 (8)	3.20 (6)
(7tr) & (3tst), e <sub>-30</sub> (T*)	6.80 (8)	2.05 (5)
(8tr) & (2tst), e <sub>-30</sub> (T*)	5.30 (6)	1.21 (9)
(9tr) & (1tst), e <sub>-30</sub> (T*)	4.30 (26)	0.00 (1)

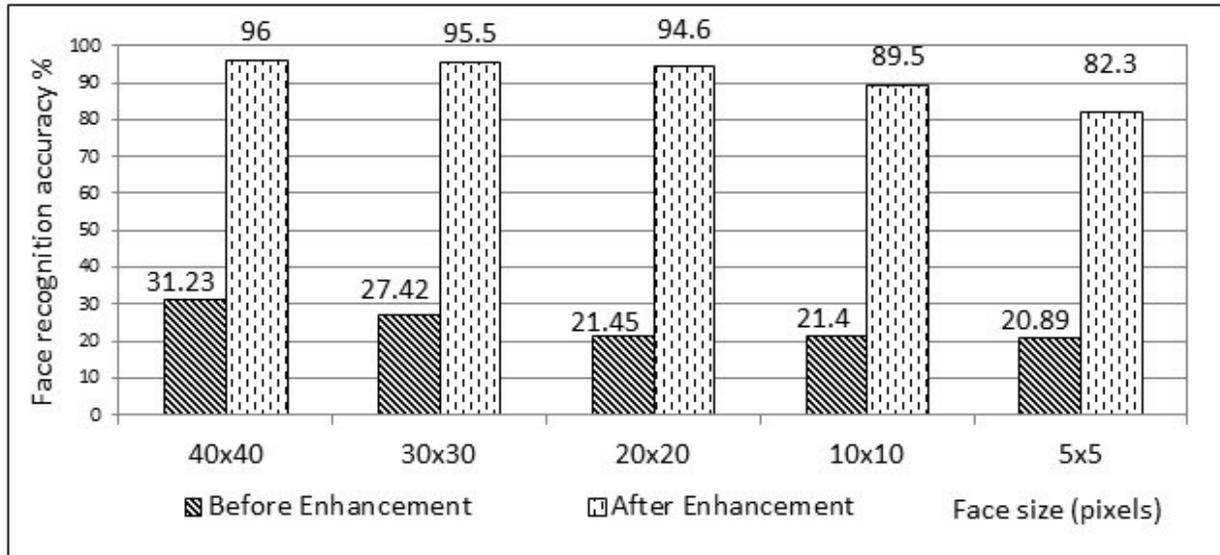


Fig. 5.3. Performance of player face recognition module by varying player face image resolution.

Fig. 5.3 summarizes our results based on the 2096 images in the database. It can be observed that player image enhancement significantly boosts the face recognition accuracy. We observe that the major obstacle in achieving higher accuracy from player face recognition module is due to the variation in face pose. The player face recognition module works practically well for even 5x5 pixels of the face region even with only 2 training samples. For player face size of 231x251, 200x200, 160x180, 100x140, 50x50, and 40x40 pixels, the face recognition accuracy is found to be same. Therefore, we only show the face recognition accuracy of 40x40 pixels in Fig. 5.3. The image enhancement algorithm is more useful for improving the accuracy of the face recognition algorithm when faces are of small size, such as 30x30, 20x20, 10x10, and 5x5 pixels.

### 5.2.6. Computational Complexity

We perform detailed experiments on our internal *Ubuntu Cloud* (UC) setup. Our setup runs on 96 cores Supermicro Server SYS-7047GR-TRF with on board 128 GB of RAM. Moreover, the setup is equipped with state-of-the-art cloud computing facilities.

Table 5.3. Summary of execution time of the developed EAR system.

No of Players	Execution Time (Seconds)				
	Player image enhancement	Player Detection	Player Face Detection	Player Face Recognition	Complete System
1	0.0210	0.0190	0.0170	0.7700	0.8270
2	0.0270	0.0199	0.0179	0.8770	0.9418
3	0.0271	0.0230	0.0189	0.8978	0.9668
4	0.0274	0.0249	0.0197	0.9811	1.0531
5	0.0274	0.0257	0.0211	1.0091	1.0833
6	0.0275	0.0271	0.0227	1.0999	1.0872
7	0.0275	Not Processed	Not Processed	Not Processed	Not Processed
14	0.0275	Not Processed	Not Processed	Not Processed	Not Processed
17	0.0279	Not Processed	Not Processed	Not Processed	Not Processed

Table 5.3 shows execution time of each module. As can be seen from Table 5.3, to completely process even six players in an image, the developed EAR framework consumes about 1.0872 seconds. The required processing time can be drastically reduced by parallel programming since all detected players can be processed independently from each other.

### 5.3. Conclusions and Future Work

We presented the feasibility of computer vision based approaches for developing enhanced augmented reality applications [5.1]-[5.3]. We developed an application/framework, which incorporates a robust enhancement approach in order to improve the accuracy of the detection and recognition. We observed that performance of most of the object (player and face) detector and face recognizer drops rapidly under non-uniform illuminations. The main contribution of the work is to develop augmented reality application for any sports that is enhanced using image enhancement scheme. Only those images are enhanced that are degraded by strong sunlight captured by an audience using digital camera/smartphone during play. This enhancement is tested and verified in the context of the presented application. We effectively

used machine learning and computer vision algorithms to detect and recognize players and display the relevant statistics of each player appeared in the input image. Statistics of interest may include name, height, and nationality. The work showed feasibility of developing an enhanced augmented reality application with high accuracy using intelligent combination of existing techniques [5.4], [5.5], and [5.17]. One simple use-case of the developed system is when viewers are sitting in a sports arena and they do not have access to TV that shows important information of players and game strategy. Viewers using their smartphones/camera can capture images/video where images are processed and information is displayed using the developed system. We observed that one of the major constraints in such applications is the robust face recognition across pose [5.14]-[5.16]. Applying a robust face recognition algorithm can significantly improve the system performance.

Future research can be carried out for solving problems, such as player pose and occlusion. Moreover, we aim to optimize the developed EAR framework to run in real-time for processing videos. Furthermore, we intend to shift the developed framework to mobile cloud computing environments.

#### **5.4. References**

[5.1] A. Dias, M. Ordaz, and J. Lopez, “From 2-D to augmented reality,” In Proc. CISTI, 2011, pp. 1–7.

[5.2] M. Yang and W. Liao, “Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction,” *IEEE Transactions on Learning Technologies*, Vol. 7, No. 2, 2014, pp. 107-117.

[5.3] M. Billingham and A. Dunser, “Augmented reality in the classroom,” *IEEE Comput.*, Vol. 45, No. 7, 2012, pp. 56–63.

- [5.4] N. Kuzmanovic, V. Mihic, T. Maruna, M. Vidakovic, and N. Teslic, "Hybrid broadcast broadband TV implementation in java based applications on digital TV devices," *IEEE Trans. Consumer Electroics* Vol. 58, No. 3, 2012, pp. 1056–1062.
- [5.5] J. Kim, J. Ha, B. Choi, Y. Jeong, and J. Hong, "Design and implementation for interactive augmented broadcasting system," *IEEE Transactions on broadcasting*, Vol. 60, No. 2, 2014, pp. 217-226.
- [5.6] R. Cavallaro, M. Hybinette, M. White, and T. Balch, "Augmenting live broadcast sports with 3-D tracking information," *IEEE Multimedia*, Vol. 18, No. 4, 2011, pp. 38–47.
- [5.7] R. Hervas, J. Bravo, and J. Fontecha, "An assistive navigation system based on augmented reality and context awareness for people with mild cognitive impairments," *IEEE Journal Of Biomedical And Health Informatics*, Vol. 18, No. 1, 2014, pp. 368-374.
- [5.8] Xiangzeng et al. "Tennis ball tracking using a two-layered data association approach," *IEEE Transactions on Multimedia*, Vol. 17, No. 2, 2015, pp. 145-156.
- [5.9] Liu, Jia, et al. "Automatic player detection, labeling and tracking in broadcast soccer video," *Pattern Recognition Letters*, 2009, pp. 103-113.
- [5.10] M. C. Hu et al. "Robust camera calibration and player tracking in broadcast basketball video," *IEEE Transactions on Multimedia*, Vol. 13, No. 2, 2011, pp. 266-279.
- [5.11] J. Lu, K. Plataniotis, A. Venetsanopoulos, and S. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Transactions on Neural Networks*, Vol. 17, No. 1, 2006, pp. 166- 178.
- [5.12] Kuo, S. Chorng, C. Lin, and C. Peng, "Using AdaBoost method for face detection and pedestrian-flow evaluation of digital signage," *International Symposium on Computer, Consumer and Control (IS3C)*, 2014, pp. 90-93.

- [5.13] Z. Mahmood et al., “A color image enhancement technique using multiscale retinex,” 11th International Conference on Frontiers of Information Technology (FIT), 2013, pp. 119-124.
- [5.14] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of cognitive neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86.
- [5.15] Ahonen, Timo, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, 2006, pp. 2037-2041.
- [5.16] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 1997, pp. 711-720.
- [5.17] G. Thomas, “Sports TV applications of computer vision,” *BBC Research and Development White Paper*, WHP 220, 2012, pp. 1–17.
- [5.18] Z. Mahmood, T. Ali, and S. U. Khan, “The effects of pose and image resolution on automatic face recognition,” *IET Biometrics*, 2015. (Accepted and to appear).

## 6. CONCLUSIONS

This chapter concludes the work and presents recommendations for future research. The dissertation focused on development of Enhanced Augmented Reality Application by emphasizing on: **(1)** image enhancement, **(2)** players detection, **(3)** players face detection, **(4)** players face recognition, and **(5)** players statistics display during various sports.

### 6.1. The AR Applications

While developing the Augmented Reality applications for sports, a major task is to identify the occluded face. This problem can be investigated by paying deep attention to small and regional facial details, such as the number and location of moles and scars in the face, shapes of different facial features, and the relative distances among different facial features. Currently, there is no automatic face recognition algorithm that can be used reliably to recognize occluded faces. One of the best strategies towards automation of facial comparison is to build a system, which is the combination of different individual classifiers, each performing comparison of a facial feature. For developing the future AR applications and to identify the occluded players' faces, eyes and eye brows regions should be given special attention to extract and classify important facial features.

### 6.2. Final Remarks

With the rapid improvement in biometric recognition technology, completely automated face recognition systems seem feasible in the future for several biometric modalities, such as face, fingerprint, and speech. For a system to be more useful in development of the AR applications for sports entertainments, in addition to the robust recognition algorithm, a state-of-the-art image enhancement algorithm is needed to rectify the images degraded due to non-

uniform illuminations under strong sun light. This was the focus of chapter 3 and chapter 5 considering general and practical aspects of image enhancement. The field of Augmented Reality applications is less mature and significant progress is yet to be done.

### **6.3. Recommendations for Future Work**

Till today, limited research has been done on the development of AR applications. In future, more formal study can be carried out to develop the AR application that can be reliably deployed in any sports. Based on the analysis presented in chapter 3 and chapter 5, modifications can be proposed to increase the accuracies of state-of-the-art object detection and recognition algorithms. Particularly, the recognition accuracies of most of the face recognition algorithms rapidly degrade under non-uniform illuminations.

Currently, the EAR system has been designed and evaluated on smartphone or digital camera images. We believe that in future, the EAR system will be potentially very useful for several purposes as highlighted in points below.

- **Coaches:** They want to know what are the weaknesses and strengths of their players. For example, in football a player can have high shooting power and low acceleration. So that they can provide personalized training to their players.
- **Directors:** They want to know which player is best for a given position in field, which player to hire for their clubs, etc.
- **TV and entertainment industry:** They want to provide more interesting information about the game, such as average speed of the ball, of a player, or was it an offside or not?
- **Miscellaneous:** There are other potential users, such as newspaper editors who want a summary of the match to publish, sports analyst, and gaming industry.

The aforementioned applications are based on accurate player detection and tracking. Once the locations of the players are found, all the useful statistics can be derived. Situations, such as variation in pose, low-resolution, and occlusion can be further investigated. The results in chapter 5 are based on using the image enhancement scheme to increase the face detection and recognition accuracy. It is interesting to investigate by how much the result and conclusions vary if another detection (player(s) and face(s)) and face recognition algorithm is used.

A useful extension of the EAR framework is to investigate it on the live videos. Similarly, the EAR framework can be shifted to mobile cloud computing environments to develop a mobile and cloud based AR application. Moreover, the recognition results of each facial feature can be weighted based on the recognition performance of a facial feature on a training data set. Furthermore, a more useful study towards automation of facial comparison is to develop automatic detectors for moles, scars, or tattoos, etc. To make this process easier, a semi-automatic segmentation can be performed, such as based on the eye locations. Alternatively, for improved recognition results, more manual labeling can be used for segmentation. For example, providing the coordinates of all of the players' facial features. This kind of manual labeling will not compromise the usefulness of state-of-the-art face recognition systems. This is because the goal is not necessarily a complete automation but also to assist the viewers in understanding any game status and strategy.