EAR: Enhanced Augmented Reality System for Sports Entertainment Applications

Zahid Mahmood[1], Tauseef Ali[2], Nazeer Muhammad[1*], Nargis Bibi[3], Imran Shahzad[1], and Shoaib Azmat[1]

[1]Electrical and Computer Engineering from COMSATS Institute of Information Technology, University Road, Abbotabad, 22010 - Pakistan.
[e-mail: zahid0987, imranshahzad, shoaibazmat@ciit.net.pk]
[2]Faculty of Computer Science, Mathematics, and Engineering, University of Twente, Netherlands.
[e-mail: t.ali@utwente.nl]
[1*] Department of Mathematics, COMSATS Institute of Information Technology, Wah Cantt, 47040 - Pakistan.
[e-mail: nazeermuhammad@ciitwah.edu.pk]
[3]Faculty of Computer Science, Fatima Jinnah Women University, Islamabad, Pakistan.
[e-mail: nargis@fjwu.edu.pk]
*Corresponding author: Nazeer Muhammad

Abstract

Augmented Reality (AR) overlays virtual information on real world data, such as displaying useful information on videos/images of a scene. This paper presents an Enhanced AR (EAR) system that displays useful statistical players' information on captured images of a sports game. We focus on the situation where the input image is degraded by strong sunlight. Proposed EAR system consists of an image enhancement technique to improve the accuracy of subsequent player and face detection. The image enhancement is followed by player and face detection, face recognition, and players' statistics display. First, an algorithm based on multi-scale retinex is proposed for image enhancement. Then, to detect players' and faces', we use adaptive boosting and Haar features for feature extraction and classification. The player face recognition algorithm uses boosted linear discriminant analysis to select features and nearest neighbor classifier for classification. The system can be adjusted to work in different types of sports where the input is an image and the desired output is display of information nearby the recognized players. Simulations are carried out on 2096 different images that contain players in diverse conditions. Proposed EAR system demonstrates the great potential of computer vision based approaches to develop AR applications.

1. Introduction

THE AR is a recent technology that enhances the user experience with additional computer generated information superimposed on the real world data [1]. The AR applications add a layer of information over the data and depict innovative digital constructs in real space. With the help of the AR technology, information about the surrounding real world of the viewer becomes interactive and digitally manipulable [2], [3]. Typically, the AR creates a meaningful image by superseding chunks of the real world images with synthetic images by augmenting useful information [4].

1.1   The AR and Computer Vision

In the AR applications, virtual entities can change the way the real world is perceived by the users' by adding, removing, or changing few aspects of the real world [7]. For example, a simulated image or video of a city may contain useful information about buildings, parking-lots, and restaurants [2]. Such applications require object detection and recognition as a preprocessing step. In these applications, computer vision technology is making significant progress in the last two decades. The integration of the AR and computer vision can result in enormous useful applications including the sports entertainment applications. Inducting latest AR and computer vision technology into sports entertainment applications provides novel opportunities and new challenges. Nowadays, for TV and satellite broadcasting companies, the AR technology is becoming an attractive modality [5]. Televised sports games use the AR knowledge to provide an enhanced experience to viewers. The ultimate transmitted images/videos contain enhancements that assist the viewers in understanding the games. TV is a one-way service and is restricted in terms of interactivity. In this paper, we develop an EAR system using a novel combination of existing image processing, computer vision, and machine learning techniques, which produces high accuracy (high recognition rate of players in a large and diverse database of player images). Specifically, the paper discusses a system that automatically enhances the input image, detects and identifies each player, and displays players' information besides each player.

1.2   Existing Work in the AR and Sports Applications

This section presents some of the recently published works in the AR and sports entertainment applications. In [1], the authors present an AR technique that recognizes both number characters and colors. To use this application reliably, an administrator first selects the jersey colors by capturing an image of each jersey before start of a game. This jersey colors information is then associated with team names, which is then stored in a cloud-based database. Finally, statistics and information about each team or players is aesthetically presented. The AR technique presented in [10] extracts players' shape and textures to generate visual enhancements. While in [11], authors utilize computer vision and computer graphics methods to attain 3-D locations of the play field, players', and the ball. In [12], researchers develop a system that inserts virtual contents on to aired tennis video. Matsui et al. [13] present a system that performs camera calibration, establishes mapping between the soccer fields, and evaluates players' pose, such as walk, run, or stop. In [14], researchers propose a hybrid broadcasting service framework by combining broadband Internet and conventional TV broadcasting models. The proposed work facilitates viewers by a touch-screen stimulated mechanism.

Rick et al. [15] present a system for augmenting live broadcast sports with 3-D information by inserting a graphical element into a live soccer scene. Jungong et al. [16] demonstrate an AR

system on volley ball and tennis during live sports. Their work utilizes a probabilistic method based on Expectation Maximization (EM) procedure. Xiangzeng et al. [17] propose a two-layered data association method to track tennis ball. At local and global layers various dynamic slice methods are applied to capture trajectories. To enhance the contents of sports events, an automatic cameraman scheme to detect players' during football match is presented in [18]. Jia et al. [19] present multiple player detection, labeling, and tracking scheme. The players' position and scale are determined by a boosting based detector. The player tracking is attained by markov chains of monte carlo method. The proposed algorithm is tested on FIFA World Cup 2006 videos. In [20], the developed scheme discusses baseball detection and tracking during a play.

Minchun et al. [21] present robust player tracking in broadcasted basketball video. In their work, players' positions and trajectory information is analyzed for situations, such as identifying open events, repossessing video clips, and planning implicit/explicit strategies. Slawomir et al. [22] design a novel segmentation system for football player detection in broadcasted videos. The system utilizes Histogram of Oriented Gradients (HoG) and Support Vector Machine (SVM). Developed system is evaluated under various lighting conditions and camera positions. Researchers in [23] develop a dual mode Bayesian inference approach to track multiple players in a sports video.

The aforementioned works are some of the nice examples where computer vision techniques are either used to perform automatic analysis of a given sports or enhance the users' experience of the broadcasted videos. We develop a similar application to enhance the users' experience and solve critical issues, such as bad lighting conditions, detection, and recognition from low-resolution facial images.

1.3   Contributions

We explore the potential of computer vision techniques to develop an EAR sports broadcasting applications with following major contributions.

- The developed EAR system is an enhanced version of our earlier developed application [26]. It addresses several drawbacks in our earlier solution. We incorporate a robust image enhancement technique, which is particularly important when captured images are degraded by non-uniform, strong sunlight, or bad illuminations. Enhancing players' image results in a nice visual appearance and significant increase in player and face identification accuracy. To the best of our knowledge, player image enhancement scheme has never been used in any AR sports system.

- Our system puts less restriction on the input images to detect players' and their faces'. Very small resolution of players, such as 30×30, 20×20, and 10×10 pixels can be successfully located. Our employed face recognition module is robust and efficiently recognizes low-resolution faces. Experimental results show that a face size of up to 5×5 pixels can be recognized in near real-time. However, player face recognition module processes faces from frontal (0^0) up to $\pm35^0$ pose variation.

- We present the fully automated and complete system for sports entertainment applications from player detection to display of the useful statistics beside the players. We believe that the proposed EAR framework can be easily adapted to develop future state-of-the-art sports applications. It is important to mention here

that the proposed EAR system has been evaluated under bad lighting conditions, face pose variations, and classification error rate as defined below.

Fig. 1. Proposed Enhanced Augmented Reality (EAR) system
(image captured using digital camera at Fargo sports arena, USA)

Definition 1: Bad lighting condition: A natural lighting condition that results in an image that contains non-distinct and undetectable objects and are not easy for most of the algorithms to detect or recognize objects. For example, a very bright image that is obtained due to strong sunlight, a very dark image because of low light, or a blurred image due to object/camera motion. This paper interchangeably uses non-uniform illuminations, strong sun light, or dark image to mean the bad lighting conditions. Moreover, the original images shown in this paper are naturally degraded without manual noise addition.

Definition 2: Players' face pose: We consider the face pose as orientation or angle of the face in specific direction. During play, players' face pose varies randomly in different directions. For the EAR system, we analyze face pose from frontal (0^0) to $\pm35^0$. It is important to state that the term face pose should not be mixed with the face expression, such as smiling, laughing, or crying in this paper.

Definition 3: Classification Error Rate (CER): The CER is the fraction of face test samples that are incorrectly classified.

Definition 4: System Accuracy: we consider the system accuracy as a mean of detection and recognition rates before and after image enhancement.

1.4   Overview of the EAR System

We develop a system that can be used to enhance users' experience. We propose a preprocessing technique to increase the quality of the input image that helps in detection and recognition of players. **Fig. 1** illustrates our five steps developed EAR system.

Step-1: The input image degraded by strong sunlight is enhanced by using our developed Multi Scale Retinex (MSR) algorithm [30]. The MSR has proven to be very effective in enhancing those images that are affected by severe dark contrast. In the EAR system: **(i)** the

MSR processes only luminance channel, which is extracted using the PCA and **(ii)** the Discrete Fourier Transform (DFT) accelerates the enhancement process followed by application of Contrast Stretching (CS). Applying the MSR algorithm results in higher detection (player and face) and recognition (face) accuracy.

Step-2 and Step-3: Player and face detection is achieved using AdaBoost algorithm combined

Table 1a. Abbreviations and notations

Notation	Description	Notation	Description
AdaBoost	Adaptive Boosting	$\hat{\epsilon}_t$	Pseudo-loss
BLDA	Boosted Linear Discriminant Analysis	$\mathcal{L}\,(\mathcal{R}_t, \widehat{D}_t, A_t)$	The LDA feature extractor
CNNC	Classic Nearest Neighbor Classifier	Υ_{t+1}	Mislabel distribution in the LDA subspace
DFT	Discrete Fourier Transform	$h_f(z)$	Face classifier for recognition
DRC	Dynamic Range Compression	w_t	Probability distribution
CMY	Cyan Magenta Yellow	$h\,(x)$	Classifier to detect players and players faces
HVS	Human Visual System	*MSR*	Multi Scale Retinex Output Image
ISOMAP	Isometric Mapping	SSS	Small Sample Size
HSI	Hue Saturation Intensity	*SSR*	Single Scale Retinex
OpenCV	Open Source Computer Vision	PCA	Principal Component Analysis
$H(x)$	Entropy	PSNR	Peak Signal-to-Noise Ratio

with Haar-like features for features extraction and classification [31]. The AdaBoost is a supervised learning algorithm that proved to be very effective for real-time applications. AdaBoost constructs a strong classifier by linearly combining many weak classifiers.

Step-4: A face recognition algorithm matches an input face with the stored faces in the database. Face recognition uses Boosted LDA (BLDA) [37] for feature selection and Classic Nearest Neighbor Classifier (CNNC) [39] to classify players' face(s). In the BLDA, performance of traditional LDA-based approach is improved by incorporating it in the boosting framework. Each round of boosting generalizes a new LDA subspace particularly focusing on examples, which are misclassified in previous LDA subspace. The final feature extractor module is an ensemble of several but very accurate specific LDA solutions.

Step-5: Statistics of each recognized player, for example, name, goals, and birth date etc. are retrieved based on the output of the player face recognition module.

We test the system on images captured during soccer, cricket, hockey, tennis, and baseball games. However, the system can be deployed in any sports, where a viewer captures images using digital cameras or smartphones while sitting in a stadium. We perform simulations on 2096 images captured during live sports that contain players in different dimensions, pose, position, expressions, and illuminations. Accuracy of the proposed EAR system is very high after enhancement. However, the accuracy decreases drastically when the players face images are not frontal and the pose exceeds $\pm 35^0$.

This paper is organized as follows. In Section 2, we briefly describe the player image enhancement algorithm. Section 3 presents the player and face detection methodology. The employed player face recognition algorithm is explained in Section 4. Section 5 illustrates the procedure for displaying the players' statistics. Results are analysed in Section 6. Finally, Section 7 concludes the paper and hints further research directions. To each readership, **Table 1a** summarizes the frequently used symbols, notations, and their meanings used in the paper.

2. Player Image Enhancement Using Multi Scale Retinex (MSR) Theory

With the fast advance of technologies and the prevalence of cameras, millions of images are

being created every day. Due to non-uniform illuminations, unfavorable weather, or failure of the imaging device itself, the captured image may not always be satisfactory. Therefore, image enhancement is often required for the aesthetic and pragmatic purposes. Recently, Retinex theory has been developed and used to enhance images under dark lighting conditions [27]-[30]. For the EAR system player image enhancement is achieved by applying the MSR [30]. **Algorithm 1** explains the pseudo code of the MSR algorithm. The proposed MSR algorithm has several distinct steps.

Algorithm 1. Pseudo code of the MSR Algorithm

Input: RGB colored image $I(x, y)$ with three channels R, G, and B; c_1, c_2, and c_3 three scales
Output: Enhanced colored image
Apply the PCA to obtain the luminance channel   ► on each color channel
begin
    foreach $c_i \in \{R, G, B\}$ **do**   ► for each RGB channel and scale
        foreach c_i
        $\left| \; SSR_i(x, y) = log\, I_i(x, y) - log[F(x, y) * I_i(x, y)] \right.$ ► achieve the MSR in three SSRs
        end
        $MSR(x, y) = \sum_{n=1}^{N=3} \omega_n \{log\, I(x, y) - log[F^{-1}(I'(u, v)F'_n(u, v))]\}$ ► $\omega_n = \frac{1}{3}$ for each c_i and $I'(u, v)$ and $F'_n(u, v)$ is the DFT
        $Luminance\ Ratio\ (LR) = \frac{MSR(x, y)}{Original\ Luminance}$
        $(R_n, G_n, B_n) = \begin{cases} LR \times R_0 \\ LR \times G_0 \\ LR \times B_0 \end{cases}$ ► new RGB values
        $New\ Pixel\ Value = \frac{Old\ Pixel\ Value - Low\ Pixel\ Value}{High\ Pixel\ Value - Low\ Pixel\ Value} \times 255$
        $Low\ Pixel\ Value = 5 \leq Low\ Pixel\ Value \leq 26$ ► contrast stretching applied on each RGB channel to obtain the output image
 $High\ Pixel\ Value = 255 - low\ Pixel\ Value$
    end
end

Initially, PCA is applied on the input player image to extract the luminance and chrominance channels [45]. Only luminance channel is processed further. The PCA was chosen for current work because, PCA retains optimum transformations for image data processing and has components that are positive as a whole with largest energy portion. The PCA also provides orthogonality among the luminance and chromatic channels. Therefore, colors remain stable despite the variation in luminance. It is important to state here that there exist several color spaces based on the tristimulus values, such as CMY, HIS, or YC_bC_r [46]. We empirically observe that the equations defining YC_bC_r, CMY, and HIS are formed in a way that rotates the entire nominal RGB color cube and scales it to fit within a larger color cube. Therefore, there are some points left within the YC_bC_r, CMY, and HIS color cube that cannot be represented in the corresponding RGB domain [46]. These conversions also require high computations to determine how to correctly interpret and display YC_bC_r, CMY, or HIS signals. As shown in Table 1b, the luminance channel obtained by using the PCA has almost similar values in the corresponding RGB domain.

In the second step, luminance channel is enhanced using three scales of the MSR. In the third step, Discrete Fourier Transform (DFT) is introduced to optimize and speed up the enhancement process. In the fourth step, enhanced luminance ratio is computed. In the fifth step, new RGB values of image are computed by multiplying the luminance ratio with initial RGB values. Finally, Contrast Stretching (CS) is applied on each RGB channels separately to obtain the enhanced output image. The CS is applied on 3×3 pixels sliding window with coordinates defined as: high pixel value = $(1, 2)$, old pixel value = $(2, 1)$, and low pixel value = $(3, 2)$. New calculated values are assigned to coordinates $(2, 2)$.

Table 1b. Comparison of color spaces (normalized values)

Color	RGB [46]	CMY [46]	YC_bC_r [46]	HSI [46]	PCA
Red	1,0,0	0,1,1	0.299, −0.16874,0.5	0^0,1,0.33	0.9999,0,0
Green	0,1,0	1,0,1	0.587, −0.33126, −0.41869	120^0,1,0.33	0,0.9999,0.0001
Blue	0,0,1	1,1,0	0.114,0.5, −0.08131	240^0,1,0.33	0,0,0.9999
Cyan	0,1,1	1,0,0	0.701,0.1687, −0.5	180^0,1,0.77	0,0.9999,0.9899
Magenta	1,0,1	0,1,0	0.413,0.3313,0.41869	300^0,1,0.66	0.9999,0,0.9999
Yellow	1,1,0	0,0,1	0.886, −0.5,0.08131	60^0,1,0.66	0.9999,0.9899,0.0001
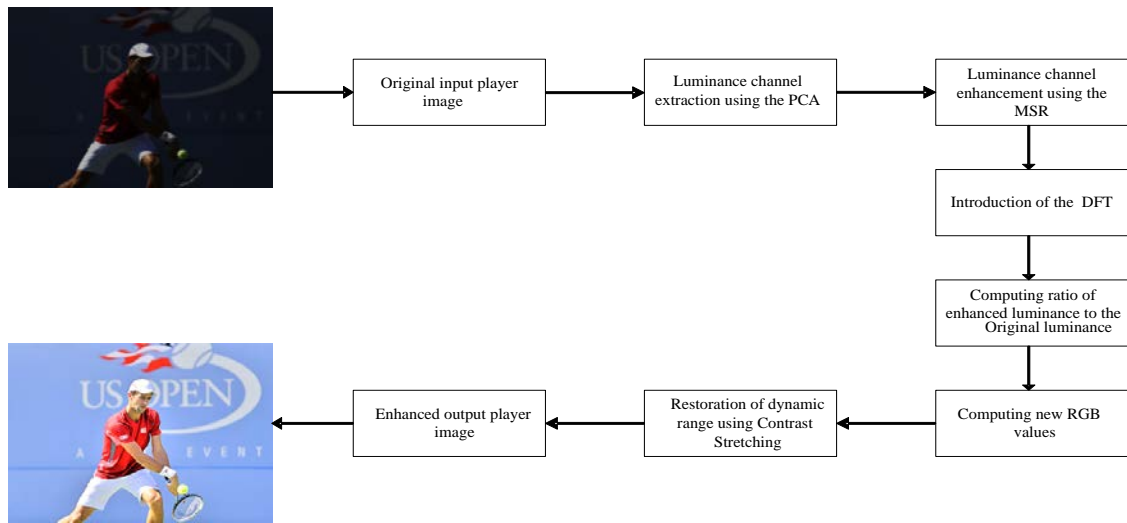
Fig. 2(a). Flow of proposed player image enhancement algorithm
(image captured using smartphone during US open, 2015)

The MSR is achieved using three scales ($N=3$) of the SSR with c values as $c_1 = 20$, $c_2 = 90$, and $c_3 = 255$ as shown in line 17 of Algorithm 1. We find a visually appealing output image while using weights (W_n) = (1/3) for each scale. However, we also observe that different weights can be used. The MSR yields the output image with enhanced luminance. After computing the LR in line 22 of **Algorithm 1**, new RGB values are computed followed by the application of Contrast Stretching (CS) to restore the dynamic range. While using the CS, new computed pixel value is assigned to (2,2) that we refer as a new pixel. In our experiments, we set *low* pixels value between 5 and 26 in line (25) for a suitable enhancement. For cases, where *low = high*, a washed out output image was obtained. The MSR based player image enhancement algorithm is extensively tested on images captured under non-uniform lighting conditions. **Fig. 2(a)** illustrates the block diagram of our developed image enhancement algorithm. The MSR algorithm performs exceptionally well on player images that contain severe dark contrast. Moreover, the MSR algorithm outperformed the Luminance Based Multi Scale Retinex (LB_MSR) [40], Luminance (Y) Inphase (I) Quadrature (Q) MSR (YIQ_MSR) [41], and Non Linear Transformation (NTF) [42] techniques in terms of visual appearance, brightness, computational cost, Entropy, and PSNR values. Once the input degraded image is enhanced, in next step, player detection is achieved in the enhanced image.

3. Player and Face Detection

We refer player and face detection as object detection in our study. To detect objects, we employ the Viola and Jones methodology that is considered as one of the state-of-the art approaches for object detection [31]. Standard implementation of the Viola and Jones face detector is available in various programming languages among which the one in OpenCV[1] is the most popular. This approach is based on Adaptive Boosting [8,9,25,26,38] (AdaBoost) and Haar- like features [32]. For a set of training images (x_i, y_i), the AdaBoost selects combination of several weak classifiers ($h_j(x)$) from Haar-like features and pools into a robust classifier that gives better classification accuracy [31]. The strong classifier $h(x)$ is expressed as:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where x is the input image, $h_t(x)$ is the tth weak classifier, $\alpha_t = \log\frac{1}{\beta_t}$. β_t is error ratio, and T are the total number of weak classifiers.

$$I(x, y) = \sum_{x^* \leq x, y^* \leq y} i(x^*, y^*) \tag{2}$$

where $I(x,y)$ is Integral Image at location (x,y) and $i(x,y)$ is original image. In general, the AdaBoost algorithm performs well for object detection. Only in some cases there are false detections, such as people from the crowd, advertisement boards, or any cautionary sign placed near the boundary line of the stadium. For our implementation, we collect a diverse set of 7000+7000=14,000 positive samples (player and face images) and 10000 negative samples (non-player images) to train the classifier. The positive training data are manually cropped from a subset of images of our collected images database. To decrease the computation time, the cropped positive images are changed to gray and resized to 18×36 pixels. While the negative images were set to 640×480 pixels. **Fig. 2(b)** shows the few training samples, whereas **Fig. 2(c)** shows Haar-like features that are applied to detect player's face. These features are obtained using Integral Image as described in Equation (2). Readers are referred to [31] for detailed description of AdaBoost and Haar-like features.

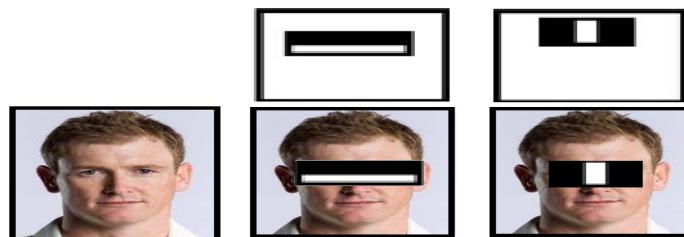**Fig. 2(b).** Positive (left) and negative (right) training samples

Fig. 2(c). Haar-like features used for player face detection [31]

---

[1] http://opencv.org/

4. Player Face Recognition

Currently, researchers from all over the world are focusing to develop an automated face recognition system that can be reliably used on any database to conclude the authenticity of any face image. Some well-known algorithms for face recognition are the PCA based approach (also called *Eigenfaces* [34]), Local Binary Patterns (LBP) [35], and Linear Discriminant Analysis (LDA) [33], [36].

To choose an appropriate face recognition algorithm for our EAR system, we compare the PCA based approach, AdaBoost with LDA as weak classifier method, and the LBP based technique for face recognition [8, 9]. We observe that the PCA (*Eigenfaces*) approach is slightly more precise in classifying variation in pose, while AdaBoost has a clear edge when there are low-resolution images. The LBP only correctly classifies frontal face images. The experiments are carried out on Multi-PIE [24] database. In our system, the low-resolution is a primary concern because the images are captured during sports in large sports arena[2],[3]. Therefore, we selected the AdaBoost algorithm [31], [32] with Linear Discriminant Analysis (LDA) [37]. It is basically applied to extract feature from the face image. The classification is performed using the Classic Nearest Neighbor Classifier (CNNC). Additional advantage of the AdaBoost based approach is that it does not require large set of training images. The algorithm is tested with a set of face images having a resolution of 5×5 pixels and a pose variation from frontal up to $\pm35^0$. The employed AdaBoost based approach outperforms standard PCA based approach in terms of correct recognition rate [38] considering both pose and low-resolution. Using the AdaBoost algorithm for player face recognition, we formulate the task of learning from samples as:

Suppose we have training set, $Z = \{Z_i\}_{i=1}^C$ containing C classes with each class $Z_i = \{(\mathbf{z}_{ij}, y_{ij})\}_{j=1}^{Li}$ consisting of a number of samples $\mathbf{z}_{ij}$ and their corresponding class labels y_{ij}, a total of $N = \sum_{i=1}^C Li$ samples are available in the set. Let Z be the sample space: $\mathbf{z}_{ij} \in Z$, and $Y = \{1,., C\}$ be the label set: $y_{ij} (=i) \in Y$. Now taking as input such a set Z, the objective of learning is to estimate a function or classifier $h(z) : Z \rightarrow Y$, i.e. h will correctly classify unseen samples (z,y). AdaBoost algorithm works by repeatedly applying a given weak learner to a weighted version of the training set in a series of rounds $t = 1, \ldots T$, and then linearly combining these weak classifiers $\{h_t\}_{t=1}^T$ constructed in each round into a single, strong and accurate classifier h_f as shown in **Equation (3). Algorithm 2** shows the Pseuodocode of the employed face recognition algorithm [37].

$$h_f(\mathbf{z}) = \underset{y \in \mathbb{Y}}{\arg\,max} \; \Sigma_{t=1}^T \left(\log \frac{1}{\beta_t}\right) h_t(\mathbf{z}, y) \qquad (3)$$

where $h_t(z, y)$ are weak classifiers and β_t is error ratio. In our applied face recognition algorithm, performance of LDA based methodology is enhanced by including it in the boosting structure. Each step in boosting simplifies a new LDA subspace that concentrates on face samples that are incorrectly classified in the preceding LDA subspace. Final classifier is an ensemble of different specific LDA solutions. Such type of collective approach takes advantage of both Boosting and LDA (BLDA) and outperforms the traditional LDA based

---

[2]Melbourne Sports Arena, Australia. Field size 171 m $\times$ 146 m, [3]Sydney Cricket Ground, Australia. Field size 155 m $\times$ 136 m

schemes in face recognition tasks. Furthermore, this approach is mainly effective in situations where fewer training face images for each person are available compared to the small sample size [37] problem. Once the player image is known to be in the database, in the next step personal information of the player is displayed.

Alogrithm 2. Pseuodocode of player face recognition algorithm

---

Input: A set of training images $Z = \{(\mathbf{z}_{ij}, y_{ij})_{j=1}^{\mu}\}_{i=1}^{C}$ with labels $y_{ij}, = i \in \mathbb{Y}$, where $\mathbb{Y} = \{1,..., C\}$; a LDA-style learner; and the iteration number, T

Let $B = \{(\mathbf{z}_{ij}, y): \mathbf{z}_{ij} \in Z, \mathbf{z}_{ij} \in \mathbb{Z} = \mathbb{R}^{J} \cdot y \in \mathbb{Y}, y \neq y_{ij}\}$

Initialize $Y_1(\mathbf{z}_{ij}, y) = \frac{1}{|B|} = \frac{1}{N(C-1)}$, the mislabel distribution over B

(For simplicity, we denote the LDA-based feature extractor as a function $\mathcal{L}(.)$, which has $(\Psi_t, \{\bar{z}_{i,t}\}_{i=1}^{C}) = \mathcal{L}(\mathcal{R}_t, \hat{D}_t, A_t)$

Do for $t = 1, \dots, T$:

1. Update the pseudo sample distribution: $\hat{D}_t(\Upsilon_t)$
2. **If** $t = 1$ **then**
   randomly choose r samples per class to form a learning set $\mathcal{R}_1 \subset Z$
   else choose r hardest samples per class based on $\hat{D}_t$ to form $\mathcal{R}_t \subset Z$
3. Train a LDA-style feature extractor with $\mathcal{L}(\mathcal{R}_t, \hat{D}_t, A_t)$ to obtain $(\Psi_t, \{\bar{z}_{i,t}\}_{i=1}^{C})$
4. Build a gClassifier $h_t = d(\Psi_t, \{\bar{z}_{i,t}\}_{i=1}^{C})$ training set Z, and get back corresponding hypotheses,
   $h_t: \mathbb{R}^{J} \times \mathbb{Y} \rightarrow [0, 1]$
5. Calculate the pseudo-loss produced by h_t:
   $$\hat{\epsilon}_t = \frac{1}{2} \sum_{(\mathbf{z}_{ij}, y) \in B} \Upsilon_t(\mathbf{z}_{ij}, y) \left(1 - h_t(\mathbf{z}_{ij}, y_{ij}) + h_t(\mathbf{z}_{ij}, y)\right)$$
6. Set $\beta_t = \hat{\epsilon}_t / (1 - \hat{\epsilon}_t)$. If $\beta_t = 0$, then set $T = t - 1$ and abort loop.
7. Update the mislabel distribution Υ_t:
   $$\Upsilon_{t+1}(\mathbf{z}_{ij}, y) = \Upsilon_t(\mathbf{z}_{ij}, y) \cdot \beta_t^{\left(1 + h_t(\mathbf{z}_{ij}, y_{ij}) - h_t(\mathbf{z}_{ij}, y)\right)/2}$$
8. Normalize Υ_{t+1} so that it is a distribution,
   $$\Upsilon_{t+1}(\mathbf{z}_{ij}, y) \leftarrow \frac{\Upsilon_{t+1}(\mathbf{z}_{ij}, y)}{\sum_{(\mathbf{z}_{ij}, y) \in B} \Upsilon_{t+1}(\mathbf{z}_{ij}, y)}$$

Output the final composite gClassifier,
$$h_f(\mathbf{z}) = \operatorname*{arg\,max}_{y \in \mathbb{Y}} \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t}\right) h_t(\mathbf{z}, y)$$

---

5. Player Statistics Retrieval

The final step of our developed EAR system is to display the statistics of the player after the player's face is matched with a face in the database of players face images. We keep players' related information in a database together with the identities of the players. For a detected face, the face recognition algorithm computes a similarity metric called *score* of the input face to all face images in the database. The score is a maximum matching value that is produced when a gallery and probe pair results in a maximum similarity. The score essentially shows that how much probe image is similar to the gallery image. The identity of the player corresponding to the *maximum score* is passed on to a module that retrieves the players' statistics. Finally, a rectangle is drawn nearby the player facial region and the statistics are displayed as shown in **Fig. 3.**

(a)   (b)   (c)   (d)

Fig. 3. An example output of our developed Enhanced Augmented Reality (EAR) system.
(a) Original image **(b)** Enhanced player image **(c)** Player detection **(d)** Player face detection, face recognition, and display of statistics (image captured using smart phone during ATP US Open 2015, USA)

6. Simulation Results

As described in Section 1.4, the EAR system comprises five independent modules: **(1)** player image enhancement, **(2)** player detection, **(3)** player face detection, **(4)** player face recognition, and **(5)** display of player statistics. Besides, evaluating the accuracy of the whole system as a *black box*, it is also important to evaluate and analyze the accuracy of each module. An example of the output produced by each module of the system is shown in **Fig. 3.** We collect a database with diverse kind of images in terms of lighting conditions and players' pose. In subsequent sub-sections, we present results of the image enhancement, detection (player and face), and face recognition. Moreover, impact of the image enhancement on detection and recognition modules is also discussed. Furthermore, accuracy of the EAR system and Classification Error Rate (CER) is also presented before and after player image enhancement. In addition, complete execution time of each of the module is also shown.

6.1 Image Enhancement Module

Fig. 4 shows the superiority of the applied MSR algorithm when tested on 256×256 pixels outdoor images. Clearly, the companion techniques have either washed out appearance or blurred output. The proposed technique effectively enhances parameters, such as grass/sky color, board below the person feet, and roof inside the hall. **Table 2** shows the comparative analysis of average values of entropy, PSNR, and execution time.

Definition 5: Entropy: Entropy is defined as the average amount of information contained in each image.

$$H(x) = \sum_{i=1}^{k} p(i) \log_2 \frac{1}{p(i)} - \sum_{i=1}^{k} p(i) \log_2 p(i) \tag{4}$$

where $H(x)$ is the discrete entropy. Discrete entropy is considered in current calculations, because the pixel values are always discrete in any digital image and $p(i)$ is the probability distribution taking all histogram counts. Entropy values of other techniques are too high that results in washed out appearance. The proposed method has great contrast in output image.

Definition 6: 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:

$$PSNR_i(x,y) = 10 log_{10} \left(\frac{255 \times 255}{\frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(x,y) - I\prime(x,y)]2} \right) \text{dB} \tag{5}$$

(a). Original Images   (b). LB_MSR   (c). YIQ_MSR   (d). NTF   (e). Proposed Method

Fig. 4. Comparison of Image enhancement techniques

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 Equation(5) represents the Mean Squared Error (MSE) of the original and enhanced images.

The results shown in **Table 2** show the preeminence of the applied MSR technique with respect to aforementioned parameters as well as for object detection and face recognition. Fig. 5 shows some enhanced player images using the proposed image enhancement algorithm. In the left column of **Fig. 5,** original images shown are degraded due to very low or strong sunlight. These images have unclear details with distorted colors. These effects occur because digital cameras lack in color constancy, color interpretation, and broad dynamic range [30]. In the right column the enhanced images are shown that have better visual appeal and improved perceptual quality. In the enhanced images, the players and other objects are clearer compared to the original images. The enhancement algorithm is very useful in such applications. Because, it is expected that images can be degraded by strong sunlight. More player image enhancement results have been shown in **Fig. 6, 8, and 10. In Table 3,** we present a quantitative summary of the accuracy of individual 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. After the player image is enhanced, in the next step player and face are detected. **Fig. 6** shows some example test images with detected players, which shows that the enhancement algorithm is robust to different kind of image degradations. **Table 3** also shows a quantitative analysis and impact of the image enhancement module. It can be concluded that the effect of the enhancement on the individual components as well as on the whole system is significant.

6.2 Players' Detection Module

The player detection module yields much improved results after the images are enhanced and successfully detects players' in various poses, such as standing, running, throwing, waiving hands, and diving as shown in right column of **Fig. 6.** For difficult cases, such as audience and advertisement boards behind the players, the detection module is up to the mark.

Table 2. Comparison of different image enhancement techniques

Figure/ Technique	LB_MSR [40]	YIQ_MSR [41]	NTF [42]	Proposed Method
Entropy	7.4401	7.4909	7.8990	**7.4233**
PSNR (dB)	24.6704	23.7928	24.8692	**36.8596**
Execution Time (seconds)	1.6216	2.1780	4.5289	**0.0291**
Memory Usage (MB)	19.8900	21.3449	21.0109	**18.2437**
Average Player and Face Detection accuracy(%)	91.8100	90.0099	94.9000	**95.1800**
Average Face Recognition Accuracy (%)	88.9999	87.0192	89.8961	**90.7500**

(a) Original images   (b) Enhanced images

Fig. 5. Player image enhancement module
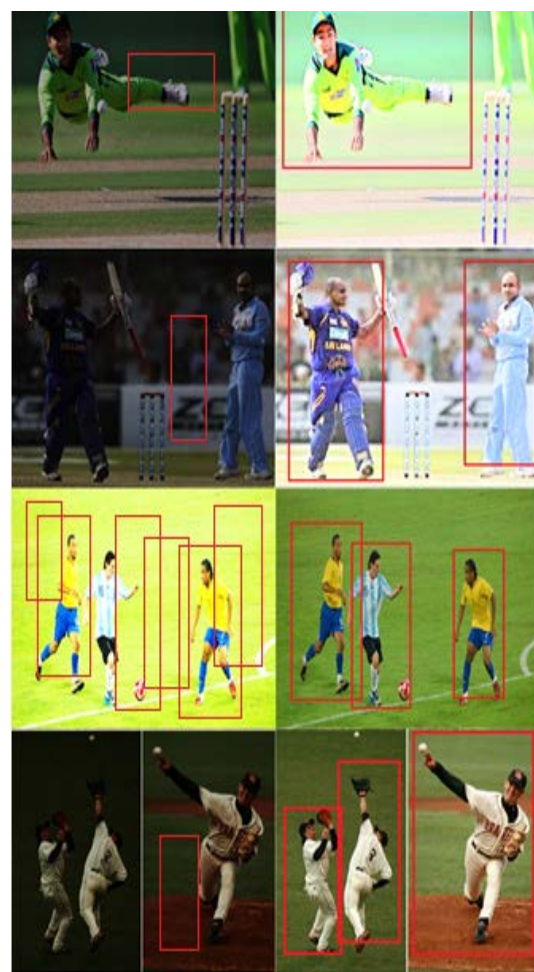
(a) Original images   (b) Enhanced images

Fig. 6. Illustration of the output of the player detection module

(images captured using digital camera)

This is the case in the second figure in the right column in **Fig. 6.** However, the case where only a small part of the player body is visible cannot be handled in most cases. This is visible in the example test image in the first row of **Fig. 6.**

6.3 Players' Face Detection Module

Player face detection module processes detected players as its input to find the face of the player. This module accurately detects a face size of up to even 5×5 pixels. **Fig. 7** shows sample face database, while **Fig. 8** indicates player face detection results. We observe that the player face detection module also performs better after player image enhancement. It detects faces in all cases except when the players' faces are very small sized or when there is a large variation in pose. This is the case in the lower right image in **Fig. 8.** There exist more robust face detection and recognition algorithms that are designed for detecting and recognizing faces with large variation in pose, which can be investigated to improve system's accuracy in future. However, in our current system, we handle pose variation from frontal (0^0) to $\pm35^0$. The design of our system easily allows for such improvement because of its modular nature where

Fig. 7. (Left). Face database of 140 different players
Fig. 8. (Right). Top row: Original images. Bottom row: Enhanced images. Fig. 8 also explains performance of player face detection module before and after image enhancement (images are captured using digital camera/smartphone during various sports)

different implementation of one module does not affect the overall system.

6.4 Players' Face Recognition Module

Face Recognition is defined as given an input face image of an unknown subject and a database containing face images of known subjects, the task is to determine the identity of the subject in the input image [25]. For the EAR system, face detection/recognition experiments were performed on 2096 images that contained 750 images from Cricket, 650 images from Soccer, 300 images from Hockey, 240 images from Tennis, and 156 images from Baseball. Out of total 2880 facial images, 1200 images were of 40×40 pixels or above, 1300 images were of 10×10 pixels or above, and 380 facial images were 5×5 pixels. It is important to state here that the images, which contained face images of 50×50 pixels or above have image resolution of 340×420 pixels. While, for facial images of size 10×10 pixels or below, image resolution was found to be 232×302 pixels. (In **Fig. 7,** a sample of 140 player face database is shown from a total of 2880 facial images). Moreover, 2710 images contained player's faces that had orientation of ±35^0 or below. We perform empirical study of the EAR system before and after image enhancement. To perform the face recognition experiments, we partition the database randomly into training and test subsets. After the player face recognition algorithm is trained, the resulting face recognizer is used on test images to perceive Classification Error Rate (CER). Taking a subset of 22 faces from the database, 8 face images are used from each player for training. It is important to state that from a series of experiments, we observe that from 200 different player images, we count 515 detected faces out of 550. Therefore, 515 face images are compared with the 22 faces in the database. From **Table 3,** it is evident that the image enhancement module has a significant impact on the face recognition module as well. **Fig. 9** shows comparison and accuracy of the player face recognition module when the numbers of training images are varied. We observe as the number of training images are increased, the face recognizer is able to extract more facial features from diverse images, which results in rapid increase of recognition ability along with decrease of CER. This kind of experiment is useful because in a certain situation the number of training images per player available for training might be limited and the system should be able to perform well even if

Table 3. Accuracy of the developed Ear System

No of Players per image	% Accuracy of the EAR System Before and After Player Image Enhancement (NP: Not Processed)					
	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
Average	**55.15**	**95.18**	**53.43**	**90.75**	**56.12**	**94.22**

there are less training images available. For the cases, where face recognition fails, no information about the player is displayed. In this case only player/face detection is achieved as shown in 4th row in **Fig. 10**. Moreover, in **Fig. 10**, players with displayed statistics depict success of all the modules of the EAR system. It is obvious that in few cases only the player detection and face detection modules succeed. While in some cases only player detection succeeds and both face detection and recognition flop. The images in which player face is invisible/unidentifiable are left unprocessed and in such circumstances no statistics of the player are shown. For exceptionally low-resolution and players overlapping images, face recognition algorithm does not perform well. In these cases, the detection module still shows significant accuracy. Moreover, it is also verified from **Fig. 9** that as we add training images, accuracy rises rapidly for enhanced player images while stays low for un-enhanced images.

6.5 Additional Exploration of Face Recognition Module

We accomplish extensive experiments to confirm that the player face recognition algorithm is reliable. One feature of machine learning based algorithms is the compulsory number of training samples of individual subject to obtain the classifier. In another series of experiments, we take a subgroup from the whole player database.
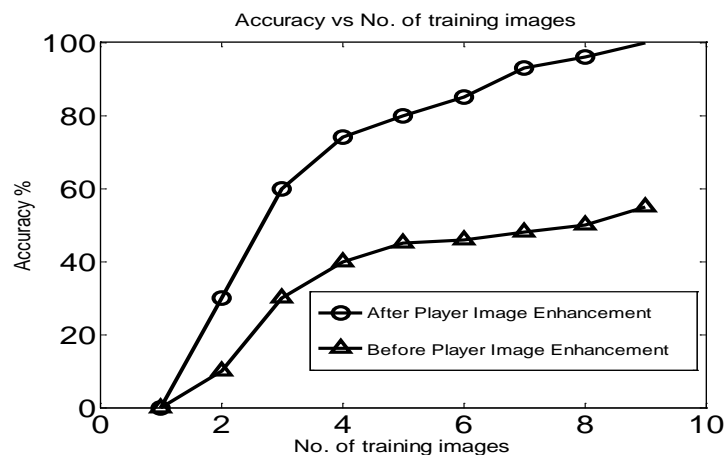
Fig. 9. Comparison of the accuracy of the face recognition module by varying the number of training

(a)   (b)   (c)   (d)

Fig. 10. Output of the developed EAR system on sample images (a) Original input player images
(Player and face not detected)   (b) Player image enhancement using the MSR (c) Player detection
(d) Player face detection, face recognition, and display of player statistics
(images captured using digital camera/smartphone during various sports)

Table 4. Average minimum Classification Error Rate

Combination	Mean Error Rate %	
	Before Enhancement	After Enhancement
(2tr) & (8tst) e-ʍ (T*)	26.00 (21)	15.90 (15)
(3tr) & (7tst) e-ʍ (T*)	22.22 (29)	12.08 (17)
(4tr) & (6tst) e-ʍ (T*)	14.44 (11)	4.14 (1)
(5tr) & (5tst) e-ʍ (T*)	11.20 (30)	3.00 (10)
(6tr) & (4tst) e-ʍ (T*)	6.87 (8)	3.20 (6)
(7tr) & (3tst) e-ʍ (T*)	6.80 (8)	2.05 (5)
(8tr) & (2tst) e-ʍ (T*)	5.32 (51)	1.29 (9)
(9tr) & (1tst) e-ʍ (T*)	4.37 (46)	0.00 (1)

The subset contains 30 different players where each player has 10 different images. Different combinations of the gallery and probe samples are used to obtain Classification Error Rate (CER). **Table 4** shows results of the sensitivity of the CER 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 CER. In **Table 4**, the e-ʍ shows obtained CER for specific training and testing face pair, while (T*) shows the iteration number on which minimum error rate is obtained. The iteration number was varied between 30 and 100 to perceive the CER.

In general, as we increase the number of training images, the CER decreases. For enhanced images, the CER decreases drastically and is close to zero when there are nine training and one testing sample. **Fig. 11** summarizes our results based on the 2096 test images. It can be observed that image enhancement potentially boosts the face recognition accuracy. We observe that a key challenge to achieve higher accuracy from the player face recognition module is due to the variation in the pose of the input face image. The player face recognition module works practically well for even 5×5 pixels of the face region even with only 2 training samples. This is possible because of the very small number of different players in the database and the total number of players in most sports does not exceed 22. In a typical experiment in face recognition, the size of the database is very large, usually in thousands. This, however, is not the case in our application and therefore we are able to achieve such a high accuracy even with small size of the face image. For player face size of 231×251, 200×200, 160×180, 100×140, 50×50, and 40×40 pixels, the face recognition accuracy is found to be same. Therefore, we only show the face recognition accuracy of 40×40 pixels in **Fig. 11**. The image enhancement algorithm is also more useful for improving the accuracy of the face recognition algorithm when faces are of small size, such as 30×30, 20×20, 10×10, and 5×5 pixels. In general, the image enhancement has a significant impact on the player face recognition accuracy that results in the high performance of the EAR system.
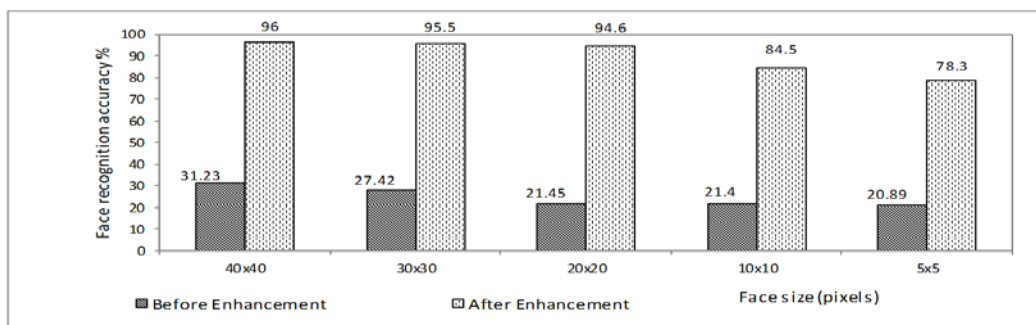

Fig. 11. Performance of player face recognition module by varying player face image resolution

Table. 5. Results of Face Size: 231×251, 180×200, 120×140, 100×100, 80×100, 60×80, 50×50, 40×40, 20×20, 10×10, and 5×5 Pixels

No of Players per image	Player Detection Accuracy (%)	Player Faces Pose and Detection Accuracy (%)				Player Faces Pose and Recognition Accuracy (%)							Complete System Accuracy from Frontal (0^0) to ±35^0 (%)
		±35^0	±25^0	±15^0	0^0 (frontal)	+35^0	+25^0	+15^0	0^0 (frontal)	-15^0	-25^0	-35^0	
1	100	100	100	100	100	90	91	100	100	100	90	85	96.33
2	100	100	100	100	100	90	91	100	100	100	89.76	84.75	96.29
3	100	100	100	100	100	88	89.88	100	100	100	88.88	86	96.10
4	100	100	100	100	100	88	89.95	100	100	100	86	86.84	95.90
5	100	100	100	100	100	83	84.39	100	100	100	85	84.75	94.76
6	100	88.88	100	100	100	83	84.90	100	100	100	85.50	81.20	93.62
7	100	81.59	100	100	100	81.57	82.61	100	100	100	84	80	92.48
9	88.88	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
14	81.23	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
Average	**96.67**	**93.82**	**100**	**100**	**100**	**86.22**	**87.68**	**100**	**100**	**100**	**87.02**	**84.01**	**95.07**

6.6 Face Angle Analysis for Face Detection and Recognition

To detect and recognize players' faces under different pose is always a challenging task. **Table 5** shows the detailed face detection and recognition results for various angles. From **Table 5,** key observations are in order:

- Player's face recognition module has 100% recognition accuracy from 0^0 (frontal) to ±15^0 rotated face up to seven player faces that appear in the input image.
- For the same angle but opposite face orientation, such as +25^0 and -25^0 face recognition accuracy is not the same. It might be due to: **(i)** different angle while face image is captured, which results in non-uniform illuminations on face and **(ii)** differences in two sides of a face caused by moles, scars, or occlusion. Various aspects cause the look of a face to change, such as non-uniform illuminations, expressions, and occlusions. It is important to state here that in our database all face images (gallery) are frontal (0^0). While, captured images (probe) contain different orientations up to ±35^0.
- Player face detection module has 100% detection accuracy from 0^0 (frontal) to ±35^0 face orientation across the entire range of samples from 231×251 down to 5×5 pixels. The player face detection module can handle pose variation, which is the main reason for the improved accuracy of the overall system.
- Face image resolution, for example, 231×251, 180×200, 140×180, 120×140, 100×100, 80×100, 60×80, 50×50, 40×40, and 5×5 pixels does not affect the recognition accuracy of the AdaBoost-LDA based face recognition algorithm.
- Face detection and recognition module is unable to detect and recognize any face angle greater than 35^0.

6.7 Computational Complexity

We evaluate the EAR system on Super micro Super-Server (SYS-7047GR-TRF) with 128 GB of RAM containing 96 cores. **Table 6** shows execution time of each module. As can be seen from **Table 6,** to completely process even six players in an image, the developed EAR system consumes about **1.0872 seconds**. The processing time can be further reduced by parallel programming, since all detected and recognized players can be processed independently from each other.

Table 6. Summary of execution time of the developed EAR system

No of Players	Execution Time (Seconds) (NP: Not Processed)				
	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	0.0385	NP	NP	NP
14	0.0275	0.0813	NP	NP	NP
17	0.0279	0.1365	NP	NP	NP

6.8 Memory Requirements

Table 7 lists the memory requirements for the EAR system. As listed in Table 2, the MSR consumes less memory because of the processing of only luminance channel. The complex object detection and recognition algorithms require memory of 4.48 GB and 1.96 GB, respectively. In future, detection and recognition algorithms can be optimized to be run on smart phone.

6.9 Discussion

Performance and Limitation: In the EAR system, detecting, recognizing, and displaying the player(s) relevant statistics adds a level of information over the captured input images. We evaluate the developed EAR system on still images for three reasons: **(a)** spectators strive hard to keep the smartphone balanced without accidentally trembling the device while concurrently interacting with the augmented contents, **(b)** we believe that still images contain higher resolution than image frames that are obtained from videos, and **(c)** still images provide a more accurate rendering of players' information. **Fig. 10** depicts that the EAR system works well in real life scenarios. Display of the statistics besides the player region in the input image validates the success of all modules. However, for the cases where complete players' body is not visible, player and face detection modules fail. This is the example in bottommost right image in **Fig. 8**. Similarly, for the cases where player face is not visible, face detection module fails as shown in 4th row in **Fig. 6,** bottom right image in **Fig. 8**, and 4th row in **Fig. 10**. In this case only player detection succeeds and no subsequent processing is performed. Moreover, we observed for occluded faces, such as cap over the head (first player in 4th row in **Fig. 6**), sun cream, or sunglasses around the player's face, the player's face detection and recognition modules have very low identification rate. The EAR system performs exceptionally well for frontal (0^0) up to $\pm35^0$ player facial pose images, such as shown in 1st, 2nd, 3rd, and 5th rows in **Fig. 10**. In situations, where player face pose exceeds $\pm35^0$, as is the situation for players in the right corner in 4th row in **Fig. 10**, face recognition module fails to recognize.

Since, the employed face recognition module is unable to recognize any face pose greater than $\pm35^0$. In future, we also aim to explore and utilize Deep Learning Models (DLM) to handle the larger face pose variation [43]-[44]. The DLMs are composed of multiple processing layers, which earn representations of face data with multiple levels of abstraction.

Table 7. Memory Consumption

Algorithm	Memory Consumption
Player Detection (Algorithm + training data)	4.48 GB
Face Detection (Algorithm + training data)	4.48 GB
Face Recognition (Algorithm + training data)	1.96 GB

Future: Currently, the EAR system has been designed and evaluated using smart phone or digital camera images. Recently, developed techniques, for example, [1], [3]-[6], [13], [15]-[16] are nice efforts in the AR domain. We believe that in future, the EAR system will be potentially very useful to develop state-of-the-art AR application.

6. Conclusions and Future Work

We presented the feasibility of computer vision based approaches to develop an enhanced augmented reality system for sports applications. We developed a system, which incorporates a robust enhancement approach to improve the accuracy of the detection and recognition. It is observed that under bad lighting, conditions performance of a robust object detector and face recognizer drops up to 55.15% and 53.43%, respectively. The main contribution of the work is to develop a complete system for sports entertainment applications that processes and enhances naturally degraded images due to strong sunlight. This enhancement is tested and verified in the context of the presented system. After the enhancement was carried out, object (player and face) detector and face recognizer accuracy is increased to 95.18% and 90.75%, respectively. We effectively applied machine learning and computer vision algorithms to detect/recognize players/faces and display the relevant statistics of each player that appeared in the input image. Under bad lighting conditions, the final accuracy of developed EAR system has been increased to 94.22% from 56.12%.

One application of the developed system is when viewers' are sitting in a stadium where they do not have access to TV. In such situations, viewers can retrieve players' statistics on their smartphone. We observed that one of the major constraints to develop such applications is the robust face recognition across pose. Applying a robust face recognition algorithm to handle occlusion, pose, or low-resolution can significantly improve the accuracy of the system. Future research can be carried out to solve problems, such as face pose and occlusion. Moreover, we aim to modify the EAR system to process videos and deal motion blur. Furthermore, we intend to shift the EAR system to mobile cloud computing environments.

References

[1] Bielli S, Harris C, "A mobile augmented reality system to enhance live sporting events," in *Proc. of 6th Augmented Human International Conference ACM*, USA, pp. 141-144, 2015. Article (CrossRef Link).

[2] Yang M and Liao W, "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, pp. 107-117, 2014. Article (CrossRef Link).

[3] Billinghurst M and Dunser A, "Augmented reality in the classroom," *IEEE Comput.*, vol. 45, no. 7, pp. 56–63, 2012. Article (CrossRef Link).

[4] Kuzmanovic et-al, "Hybrid broadcast broadband TV implementation in java based applications on digital TV devices," *IEEE Trans. Consumer Electroics*, vol. 58, no. 3, pp. 1056–1062, 2012. Article (CrossRef Link).

[5] Kim et-al, "Design and implementation for interactive augmented broadcasting system," *IEEE Transactions on broadcasting*, vol. 60, no. 2, pp. 217-226, 2014. Article (CrossRef Link).

[7] Cavallaro R, Hybinette, M, White, and T. Balch, "Augmenting live broadcast sports with 3-D tracking information," *IEEE Multimedia*, vol. 18, no. 4, pp. 38–47, 2011. Article (CrossRef Link).

[8] Mahmood Z, Ali T, and Khan S U, "Effects of pose and image resolution on automatic face

recognition," *IET Biometrics*, Vol. 5, No. 2, pp. 111-119, 2016.
Article (CrossRef Link).

[9] Mahmood Z, Ali T, Khattak S, and Khan S U, "A comparative study of baseline algorithms of face recognition," in *Proc. of 12th International Conference on Frontiers of Information Technology (FIT)*, Pakistan, pp. 263-268, 2014. Article (CrossRef Link).

[10] Han J and Farin D, "A real-time augmented reality system for sports broadcast video enhancement," in *Proc. of MM'07*, Augsburg, Germany, pp. 1-4, 2007.
Article (CrossRef Link).

[11] Liang et-al. "Video2cartoon: generating 3-D cartoon from broadcast soccer video," in *Proc. of ACM Multimedia*, pp.217-218, 2005.

[12] Yu X, Yan X, Chi T, and Cheong L, "Inserting 3-D projected virtual Content into broadcast tennis Video," in *Proc. of ACM Multimedia*, pp.619-622, 2006. Article (CrossRef Link).

[13] Matsui et-al, "Soccer image sequence computed by a virtual camera," in *Proc. of CVPR*, pp. 860-865, 1998.

[14] Hervas R, Bravo J, and Fontecha J, "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, pp. 368-374, 2014. Article (CrossRef Link).

[15] Cavallaro et-al. "Augmenting live broadcast sports with 3D tracking information," *IEEE MultiMedia*, pp. 38-47, 2011.

[16] Jungong H and Farin D, "A real-time augmented-reality system for sports broadcast video enhancement," in *Proc. of 15th International Conference on Multimedia*, ACM, 2007.

[17] Xiangzeng et-al. "Tennis ball tracking using a two-layered data association approach," *IEEE Transactions on Multimedia*, vol. 17, no. 2, pp. 145-156, 2015. Article (CrossRef Link).

[18] Hanzra B and Rossi R, "Automatic cameraman for dynamic video acquisition of football match," in *Proc. of 2nd International Conference on Image Information Processing (ICIIP)*, pp. 142-147, 2013. Article (CrossRef Link).

[19] Liu et-al. "Automatic player detection, labeling and tracking in broadcast soccer video," *Pattern Recognition Letters*, pp. 103-113, 2009.

[20] Venkateshan, Kishore, A, Shekar, and Saha S, "Baseball hand tracking from monocular video," *Advances in Computing, Communications and Informatics (ICACCI)*, pp. 953-961, 2013.

[21] Hu et-al. "Robust camera calibration and player tracking in broadcast basketball video," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 266-279, 2011. Article (CrossRef Link).

[22] Mackowiak, et-al, "A complex system for football player detection in broadcasted video," in *Proc. of International Conference on Signals and Electronic Systems (ICSES)*, pp. 119-122, 2010.

[23] 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 Processin*, pp. 1652-1667, 2011.

[24] Gross et-al, "Multi-pie", Image Vis. Comput, vol. 28, no. 5, pp. 807-813, 2010.
Article (CrossRef Link).

[25] Mahmood et-al, "Automatic player detection and identification for sports entertainment applications," *Pattern Analysis and Applications*, vol. 18, no. 4, pp. 971-982, 2015.
Article (CrossRef Link).

[26] Mahmood Z, Ali T, and Khattak S, "Automatic player detection and recognition in images using AdaBoost," in *Proc. of 9th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 64-69, 2012. Article (CrossRef Link).

[27] Setty S, Srinath N, and Hanumantharaju M, "Development of multiscale retinex algorithm for medical image enhancement based on multi-rate sampling*," in Proc. of International Conference on Signal Processing, Image Processing and Pattern Recognition (ICSIPR)*, pp. 1-6, 2013.
Article (CrossRef Link).

[28] Poursaberi A, Svetlana N, and Gavrilova M, "Modified multiscale vesselness filter for facial feature detection," in *Proc. of 4th International Conference on Emerging Security Technologies*, pp. 21-24, 2013. Article (CrossRef Link).

[29] Changyan X, Staring M, and Wang Y, "Multiscale Bi-Gaussian filter for adjacent curvilinear structures detection with application to vasculature images," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 174-188, 2013. Article (CrossRef Link).

[30] Mahmood et al., "A color image enhancement technique using multiscale retinex," in *Proc. of 11th International Conference on Frontiers of Information Technology (FIT)*, pp. 119-124, 2013. Article (CrossRef Link).

[31] Viola P and Jones M, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, pp. 137–154, 2004. Article (CrossRef Link).

[32] Kuo, Chorng S, Lin C, and Peng C, "Using AdaBoost method for face detection and pedestrian-flow evaluation of digital signage,"in *Proc. of International Symposium on Computer, Consumer and Control (IS3C)*, pp. 90-93, 2014.

[33] Shylaja S, Balasubramanya K, and Natarajan S, "Dimensionality reduction techniques for face recognition," *Department of Information Science and Engineering, PEC Institute of Technology*.

[34] Turk M and Pentland A, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71-86, 1991. Article (CrossRef Link).

[35] Ahonen, Timo, Hadid A, and Pietikainen M, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006. Article (CrossRef Link).

[36] Belhumeur P N, Hespanha J P, and Kriegman D, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997. Article (CrossRef Link).

[37] Lu J, Plataniotis K, Venetsanopoulos A, and Li S, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 166-178, 2006. Article (CrossRef Link).

[38] Mahmood et-al, "A parallel framework for object detection and recognition for secure vehicle parking," in *Proc. of 17th International Conference on High Performance Computing and Communications (HPCC)*, USA, pp. 892-895, 2015.

[39] Gao Q B, and Wang Z Z, "Center-based nearest neighbor classifier," *The Journal of the Pattern Recognition Society*, vol. 40, pp. 346-349, 2007. Article (CrossRef Link).

[40] Sun B, Tao W, and Chen W, "Luminance based MSR for color image enhancement," *International Congress on Image and Signal Processing*, pp. 358-362, 2008. Article (CrossRef Link).

[41] Jiao Z and Xu B, "An image enhancement approach using retinex and YIQ," in *Proc. of International Conference on Information Technology and Computer Science (ITCS)*, pp. 476-479, 2009. Article (CrossRef Link).

[42] Ghimire D and Lee J, "Nonlinear transfer function-based local approach for color image enhancement," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 2, pp. 858-865, 2011. Article (CrossRef Link).

[43] Cun L, Yann, Bengio Y, and Hinton G, "Deep learning," *Nature*, 521.7553. pp. 436-444, 2015.

[44] Parkhi, Omkar M, Vedaldi A, and Zisserman A, "Deep Face Recognition," BMVC, vol. 1. no. 3, pp. 1-12, 2015. Article (CrossRef Link).

[45] Abdi H and Williams L J, "Principal component analysis," *Computational statistics*, vol. 2, no. 4, pp. 433-459, 2010. Article (CrossRef Link).

[46] Crane R, "Simplified approach to image processing: classical and modern techniques in C," *Prentice Hall PTR*, pp. 13-23, 1996.

Zahid Mahmood received a Ph.D. degree from North Dakota State University, USA in 2015. Currently, he is Assistant Professor in Electrical Engineering Department at COMSATS Institute of Information Technology, Abbottabad, Pakistan. Zahid Mahmood's research expertise encompasses topics, such as Digital Image Processing, pattern recognition, and Signal Processing. In these areas, he has published over 15 papers.

Tauseef Ali received Ph.D degree from University of Twente, Netherlands in 2014. Tauseef Ali's research interest includes digital image processing and digital signal processing. He is the recipient of Higher Education Commission, Government of Pakistan scholarship award for MS in Electrical Engineering.

Nargis Bibi received a Ph.D. degree in Computer Science, the School of Computer Science, University of Manchester, UK, in 2014. Currently, she is Assistant Professor at the Department of Computer Science, Fatima Jinnah Women University (FJWU), Rawalpindi, Pakistan. Her interests are digital signal processing, OFDM, coding theory, and information theory.

Muhammad Imran Shahzad is pursuing a Ph.D. degree in Electrical Engineering from from COMSATS Institute of Information Technology, Abbottabad, Pakistan. Imran Shahzad's research expertise encompasses topics, such as pattern recognition and Signal Processing.

Shoaib Azmat received a PhD degrees in Electrical and Computer Engineering from Georgia Institute of Technology, USA in 2014. Currently, he is Assistant Professor in Electrical Engineering Department at COMSATS Institute of Information Technology, Abbottabad, Pakistan. Shoaib Azmat's research expertise encompasses topics, such as Digital Image Processing, pattern recognition, and Signal Processing.