

Automatic player detection and identification for sports entertainment applications

Zahid Mahmood · Tauseef Ali · Shahid Khattak ·
Laiq Hasan · Samee U. Khan

Received: 25 November 2013 / Accepted: 7 September 2014 / Published online: 28 September 2014
© Springer-Verlag London 2014

Abstract In this paper, we develop an augmented reality sports broadcasting application for automatic detection, recognition of players during play, followed by display of personal information of players. The proposed application can be divided into four major steps. In first step, each player in the image is detected. In the second step, a face detection algorithm detects face of each player. In third step, we use a face recognition algorithm to match the faces of players with a database of players' faces which also stores personal information of each player. In step four,

personal information of each player is retrieved based on the face matching result. This application can be used to show the viewers' information about players such as name of the player, sports record, age, highest score, and country of belonging. We develop this system for baseball game, however, it can be deployed in any sports where the audience can take a live video or images using smart phones. For the task of player and subsequent face detection, we use AdaBoost algorithm with haar-like features for both feature selection and classification while player face recognition system uses AdaBoost algorithm with linear discriminant analysis for feature selection and nearest neighbor classifier for classification. Detailed experiments are performed using 412 diverse images taken using a digital camera during baseball match. These images contain players in different sizes, facial expressions, lighting conditions and pose. The player and face detection accuracy is high in all situations, however, the face recognition module requires detected players' faces to be frontal or near frontal. In general, restricting the head rotation to $\pm 30^\circ$ gives a high accuracy of overall system

Z. Mahmood (✉)
Electrical Engineering, North Dakota State University, Fargo,
USA
e-mail: zahid.mahmood@ndsu.edu

T. Ali
Faculty of Electrical Engineering, Mathematics and Computer
Science, University of Twente, Enschede, Netherlands
e-mail: t.ali@utwente.nl

S. Khattak
Electrical Engineering Department, COMSATS IIT, Abbottabad,
Pakistan
e-mail: skhattak@ciit.net.pk

L. Hasan
Computer Engineering Laboratory, Faculty of Electrical
Engineering Mathematics and Computer Science, Delft
University of Technology, Delft, The Netherlands
e-mail: L.Hasan@tudelft.nl

L. Hasan
Department of Computer Systems Engineering, University of
Engineering and Technology, Peshawar, Pakistan

S. U. Khan
Department of Electrical and Computer Engineering, North
Dakota State University, Fargo, USA
e-mail: samee.khan@ndsu.edu

Keywords AdaBoost · Face detection and recognition ·
LDA · NNC

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. In recent times, computer-generated visualizations are increasingly used in sports broadcasting

to enhance the viewers' experience such as displaying simple data and current scores. Augmented reality (AR) is a technique for overlaying virtual objects onto the real world. AR of real sports games facilitates viewers to engage and immerse into the action. In 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.

Until now, there is limited research on the development of AR applications. It has been applied to some kinds of entertainment applications using vision-based tracking technique [1]. In [2], technique is presented to generate a visual enhancement for TV-broadcasted court net sports. It had a major drawback as system generates redundant virtual scenes which abruptly affect system performance. In [3], 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], 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 [5], authors tried to insert and manipulate 3D virtual contents into broadcasted tennis video. This technique also fails in situations like invisibility of one net. In [6], 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 3 poses eliminates improvement in system performance. All these previous efforts have yet to reach the acceptable level of reality. In a nutshell, 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 [5–7]. The developed system is free from problems in aforementioned techniques such as synchronization, generation of virtual redundant scenes, and player pose.
- The developed technique does not have problems like 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 does not put any restriction on the input

images. Player face recognition requires frontal or near frontal faces; however, it is quite robust to recognize small resolution faces and faces as small as 5×5 pixel can also be successfully recognized.

In this paper, we present a complete framework for displaying player statistics in still images which can easily be extended to real-time videos. Figure 1 shows the block structure of our proposed system. In step 1, we detect players using AdaBoost [8] and Haar-like features. In step 2, we use the same algorithm to detect face [9], [10], in each player image. Step 3 matches the face with a database of player faces. Step 4 retrieves statistics of players such as name of the player, height, age, sports record, and country of belonging. Face recognition is performed using AdaBoost with LDA [11] as weak learner and NNC.

The remainder of the paper is organized as follows: In Sect. 2, we discuss our proposed player detection scheme which is based on AdaBoost algorithm. Section 3 presents face detection process based on algorithm described in step II. Section 4 explains the employed player face recognition approach. Section 5 explains retrieval of player statistics from player face database. Detailed experimental results are presented in Sect. 6. Finally, conclusions and future research directions are given in Sect. 7.

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 [8] approach. 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 combine them into a strong classifier. The strong classifier $g(x)$ is defined as follows:

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

where x is the input image, k is the number of weak classifiers, θ is the decision threshold, and $g(x)$ is the final strong classifier. Table 1 lists important notations used in our paper.

Table 2 illustrates the complete pseudo-code and full details of AdaBoost algorithm for obtaining strong classifier of Eq. 1. More details of this employed approach can

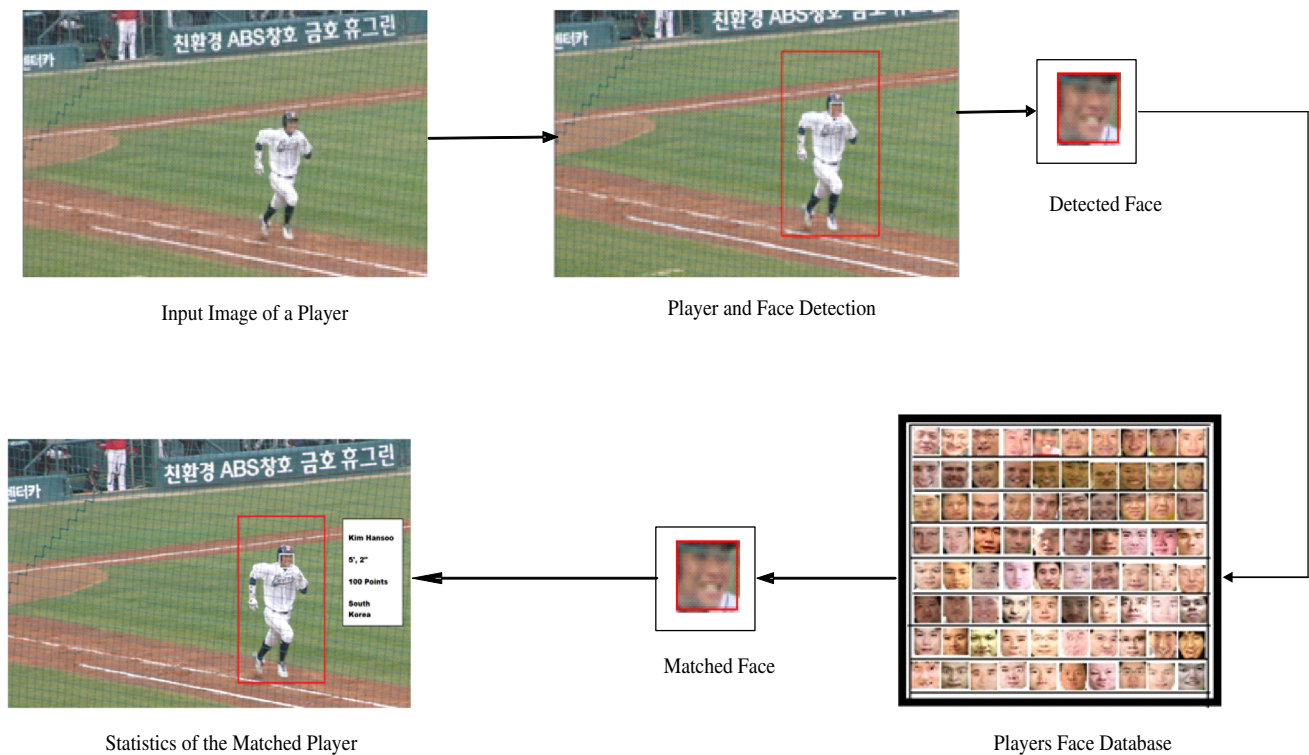


Fig. 1 Framework of our proposed application (Image taken at Jamshil Stadium, Seoul, South Korea)

Table 1 Important notations used in this paper and their description

Notation	Description
$g(x)$	Classifier obtained to detect players in input image
w_t	Probability distribution
$h_f(z)$	Player face classifier
$\hat{\epsilon}(t)$	Pseudo-loss
\mathcal{T}_{t+1}	Mislabel distribution
$\mathcal{L}(R_p, D_r^A, A_t)$	LDA-based feature extractor

be found in [9]. From Table 2, conclusion can be drawn that a correct classification of a sample makes its corresponding weight drop for the next iteration, otherwise it is unchanged. The smaller the error rate of a weak classifier, the smaller is its corresponding β and the greater is the α . Therefore, the main idea of AdaBoost classifier is to learn a series of weak classifiers and threshold θ that is adjusted to meet the detection rate goal. The Haar-like rectangle features are computed using “Integral Image” representation. The cascade method quickly filters out non-player image areas.

A set of 630 positive samples and 1100 negative samples are used to train the classifier. The training data are selected from our own image database of baseball players. To speed up the computation time, all player images are first converted into gray scale and size normalized to

Table 2 Adaboost Algorithm

- Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,j} = \frac{1}{2m \cdot 2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t=1, \dots, T$:
 - Normalize the weights,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$
 So that w_t is a probability distribution.
 - For each feature, j , train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t , is

$$e_j = \sum w_i |h_j(x_i) - y_i|$$
 - Choose the classifier, h_t , with the lowest error e_t .
 - Update the weights: $w_{t+1,i} = w_{t,i} B_i^{1-e_i}$ where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and

$$\beta_t = \frac{e^t}{1 - e^t}$$
- The final strong classifier is:

$$g(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad \text{Where } \alpha_t = \log 1/\beta_t$$

20×40 pixels. Some examples of positive and negative data are shown in Fig. 2. Negative 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



Fig. 2 Example of positive (*left*) and negative (*right*) training samples

of detection window is set to half of the height. This improves accuracy and searching time. An image captured in the sports ground contains players as well as other objects such as audience and advertising boards. So the sole purpose of player detection was to confine the AR application processing area by eliminating the redundant information as mentioned above. Detecting player also gives viewers the exact position of their favorite players, as well as increases the esthetic sense.

3 Player face detection

Face detection employs boosting methodology which is one of state-of-the art approaches for object detection [10]. 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. Figure 3 shows the two, three and four rectangle features.

The rectangle features are computed using Integral Image, which is sum of the pixels above and to the left of

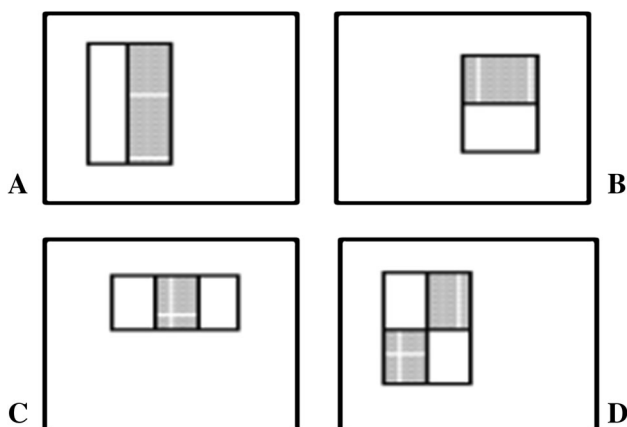


Fig. 3 Example of rectangle features shown relative to the enclosing detection window

the point (x, y) . The Integral Image $I(x, y)$ is defined in Eq. 2.

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

where $I(x, y)$ is Integral Image at points (x, y) and $i(x, y)$ is original image. For the task of player face detection, 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 those of 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 Player face recognition

Currently, efforts are ongoing for efficient storage and retrieval of face images. Various techniques exist to handle the curse of dimensionality out. Its two wide categories are: linear methods and nonlinear methods. PCA, LDA, LPP are some popular linear methods, while nonlinear methods include ISOMAP and Eigenmaps learning [12]. PCA and LDA are the two most widely used subspace learning techniques for face recognition. Jun Yu et al presented an adaptive hyper graph scheme with application in image classification [13]. Proposed algorithm generates hyperedges by linking images with nearest neighbors by varying the size of the neighborhood. Interestingly, the proposed method simultaneously learns the labels of unlabeled images and the weights of hyperedges. Therefore, it automatically modulates the effects of different hyperedges. The scheme produced better results than its baselines schemes. Full details can be seen in [13].

A research group in [14], presented a complex object correspondence construction technique in 2D animation for effective face recognition. A semisupervised scheme of patch alignment framework for complex object correspondence construction was introduced. It constructs local patches for each point on an object and aligns these patches in a new feature space, in which correspondences between objects can be detected by the subsequent clustering. Experiments conducted on MPEG-7 dataset revealed its

performance comparable with that of a prominent algorithm published in T-PAMI 2009. Detailed description can be found in [14].

An interesting semisupervised multiview distance metric learning for cartoon synthesis is presented in [15], where authors introduced three visual features, color histogram, shape context, and skeleton, to characterize the color, shape, and action, respectively, of a cartoon character. A novel system was developed which successfully decomposes the modules of multiview cartoon character

classification, multiview graph-based cartoon synthesis, and multiview retrieval-based cartoon synthesis. The developed technique was found extremely efficient in cartoon applications.

Combining multiple view features for cartoon character retrieval and clip synthesis was presented in [16]. Researchers concluded experimental evaluations based on cartoon character retrieval and clip synthesis. Proposed scheme demonstrated the effectiveness of the developed technique for cartoon applications. A semisupervised

Table 3 Adaboost algorithm for face recognition

Input: A set of training images $Z = \{(z_{ij}, y_{ij})\}_{j=1}^{L_i}$ with labels $y_{ij} (=i) \in Y$, where $Y = \{1, \dots, C\}$; a LDA-style learner; and the iteration number, T .
 Let $B = \{(z_{ij}, y): z_{ij} \in Z, z_{ij} \in \mathbb{R}^J, y \in Y, y \neq y_{ij}\}$.

Initialize

$\gamma^1(z_{ij}, y) = \frac{1}{|B|} \equiv \frac{1}{N(C-1)}$ the mislabel distribution over B .
 (For simplicity, we denote the LDA-based feature extractor as a function $\mathfrak{F}(\cdot)$, which has $(\psi_t, \{\bar{z}_{i,t}\}_{i=1}^C) = \mathfrak{F}(R_t, D_t, A_t)$).

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

- Update the pseudo sample distribution: $D_t^\wedge(\gamma_t)$, and the PCDD: A_t with eq.2.
- If** $t = 1$ then
 randomly choose r samples per class to form a learning set $R_t \subset Z$.
else choose r hardest samples per class based on D_t^\wedge to form $R_t \subset Z$.
- Train a LDA-style feature extractor with:
 $\mathfrak{F}(R_t, D_t^\wedge, A_t)$ to obtain $(\psi_t, \{\bar{z}_{i,t}\}_{i=1}^C)$.
- Build a gClassifier $h_t = d(\psi_t, \{\bar{z}_{i,t}\}_{i=1}^C)$ with eq.8, apply it into the entire training set Z , and get back corresponding hypotheses, $h_t: \mathbb{R}^J \times Y \rightarrow [0, 1]$.
- Calculate the pseudo-loss produced by h_t as:
 $\varepsilon^t = \sum_{(z_{ij}, y) \in B} \gamma(z_{ij}, y)(1 - h_t(z_{ij}, y) + h_t(z_{ij}, y))$
- Set $\beta_t = \varepsilon^t / (1 - \varepsilon^t)$. If $\beta_t = 0$, then set $T = t - 1$ and abort loop.
- Update the mislabel distribution: $\gamma_{t+1}(z_{ij}, y) = \gamma_t(z_{ij}, y) \cdot \beta_t^{(1 + h_t(z_{ij}, y)) - h_t(z_{ij}, y)/2}$.
- Normalize γ_{t+1} so that it is a distribution,
 $\gamma_{t+1}(z_{ij}, y) \leftarrow \frac{\gamma_t(z_{ij}, y)}{\sum_{(z_{ij}, y) \in B} \beta_t \gamma_t(z_{ij}, y)}$
- Output** the final composite gClassifier,
 $hf(z) = \operatorname{argmax}_{y \in Y} \sum_{t=1}^{t=T} \log\left(\frac{1}{\beta^t}\right) h_t(z, y)$

multiview subspace learning (semi-MSL) algorithm was also introduced to encode different features in a unified space. Details can be found in [16, 17]. In our work, we utilized AdaBoost [8–10] algorithm with LDA [11] as a weak learner for feature selection in LDA subspace while classification is performed using a classic Nearest Neighbor Classifier (NNC). In pattern classification, the task of learning from examples can be stated as:

Suppose we have training set, $Z = \{Z_i\}_{i=1}^C$ containing C classes with each class $Z_i = \{(z_{ij}, y_{ij})\}_{j=1}^{L_i}$ consisting of a number of samples z_{ij} and their corresponding class labels y_{ij} , a total of $N = \sum_{i=1}^C L_i$ samples are available in the set. Let Z be the sample space: $z_{ij} \in Z$, and $Y = \{1, \dots, C\}$ 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 $t = 1, \dots, T$, 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 . Eq. (3) shows the final strong classifier:

$$h_f(z) = \arg \max_{y \in Y} \sum_{t=1}^{T} \log \left(\frac{1}{\beta_t} \right) h_t(z, y) \quad (3)$$

The 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. This kind of ensemble-based approach takes advantage of both LDA and boosting and outperforms only LDA-based systems in complex face recognition tasks. This is particularly important where less number of training samples for each subject are available compared to number of dimensions of samples, i.e., Small-Sample-Size (SSS) problem [18] and when nonlinear variations are present in facial images. Our employed face recognition system is more robust when

performing recognition of low-resolution face images. This result is also verified by the authors in [11]. The pseudo-code of employed face recognition system is given in Table 3.

5 Player statistics retrieval

The final goal of this work is to display relevant statistics for player(s) once an image is captured or obtained from a video. This depends on the accuracy of detection and recognition results. Recently, various object recognition approaches have been presented for automatic retrieval of data [19–21]. In [22], a general classification algorithm for image-based object recognition is presented. This new framework provided nice insights into two crucial issues in face recognition: feature extraction and robustness to occlusion. The algorithm was reported to have significant output with varying illumination and expression. There were no experiments performed on low-resolution images. Researchers in [23] presented a novel multiview hypergraph-based method that effectively integrates click data with varied visual features. It considers a pairwise discriminative constraints from click data that maximally distinguish images with high clicks counts. Therefore, a semantic manifold is constructed. Experiments performed on real datasets revealed the effectiveness of the proposed scheme than state-of-the-art image re-ranking methods. We store players' statistics in our system together with the corresponding players' facial images. For a given input face image, the face recognition system calculates distance metric of the input face to all face images in database. The pair resulting in the maximum value of this distance metric (called matching score) is considered as the true match. The true matched player face image is obtained by simple comparison of all the matching scores obtained from the system while the test image is being compared to all possible players' faces in the database. This type of AR application can be used for enhancing viewers' experience where a viewer is using smart phone or a camera to capture photos or video of a live sport.



Fig. 4 An example outputs of our developed system

6 Experimental results

Our proposed system is composed of 3 modules: Player detection, face detection and face recognition. It is necessary to analyze the results of each module separately besides considering the performance of the whole system

Table 4 Summary of our system accuracy

No of Players/ image	Player and face detection accuracy (%)	Face recognition accuracy (%)	System accuracy (%)
1	100	100	100
2	100	100	100
3	100	100	100
4	100	66.66	88.88
5	100	50	83.33
6	100	Not processed	Not processed
7	100	Not processed	Not processed
14	78.57	Not processed	Not processed
17	70	Not processed	Not processed

as a ‘black box’. Figure 4 shows an example of the system output for each of the module.

We use a dataset of 412 diverse images with different number of players in various lighting conditions and small pose variation of face from frontal. To better evaluate the system feasibility for real-world applications, we use two types of images: images taken with a smart phone and images taken with a good quality digital camera. There are a total of 80 different subjects (players) in the dataset shown in Fig. 7. In Table 4, we give a quantitative summary of the performance of each module. The number of players per image essentially demonstrates the size of detected player. Fewer numbers of players per image imply that image is taken from a small distance and player(s) in image has high resolution which helps improve the performance of subsequent steps. In the subsequent sections we discuss and analyze each module developed in our work.

6.1 Player and face detection

Figure 5 shows some example images with detected players by employed player detection methodology and

Fig. 5 Sample images illustrating performance of player detection module (Images taken using smart phone at Jamshil stadium, Seoul, South Korea)





Fig. 6 Player face detection results (Images taken at Jamshil stadium, Seoul, South Korea)

feasibility of our approach for real-world scenarios. Face detection and face recognition is performed 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 4×4 could not be detected by player detection module.

Player face detection module takes detected players as its input and detects the faces of the players up to face size of 5×5 . Figure 6 shows some examples of player face detection results. It can be observed that face detection module performs well in most cases except when detected players are very small sized. This is the case in lower right image in Fig. 6. Face detection module does not process extremely blurred and invisible player face images. In such cases, it is generally the case when human eye is also unable to locate such faces.

6.2 Player face recognition

We partition the database randomly into two subsets: the training set Z and test set Q . Face Recognition system is first trained with Z training images and the resulting face recognizer is then applied to Q test images to observe classification error rate (CER). Figure 7 shows database of 80 different player faces. Once the input face is matched to a face in the database, personal information of the matched player such as 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.

Player and face detection is implemented in C while face recognition module is implemented in MATLAB.



Fig. 7 Face database of 80 different players

Implementing the whole system in low-level language will make it usable in real-time application such as processing images from a video. Table 5 shows execution time of each module.

Figure 8 shows the final output of the proposed system on some sample images from our database of 412 images. Most of the images contain several players at various poses and lighting conditions. In Fig. 8, players with statistics displayed show success of all the modules of the system. In some cases, only player detection and face detection succeed but recognition is fails. In some cases only player detection succeeds and both subsequent face detection and recognition fails. Examples of such results can be seen in lower left image in Fig. 8. The images in which player face

Table 5 Execution Time

No of players	Player detection time (ms)	Player face detection time (ms)	Player face recognition time (s)	System execution time (s)
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	1,211.86	301.61	2.35	3.86
6	1,211.86	301.61	2.35	3.86
7	1,211.86	305.77	2.37	3.81
14	1,221.62	0	0	0
17	1,251.78	0	0	0

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 system does not work. In such cases, player and his face detection still exhibits considerable accuracy. To further assess the performance of our face recognizer, we also used BioID database [18] in our experiments for testing and training purposes besides using player database that we developed in our lab.

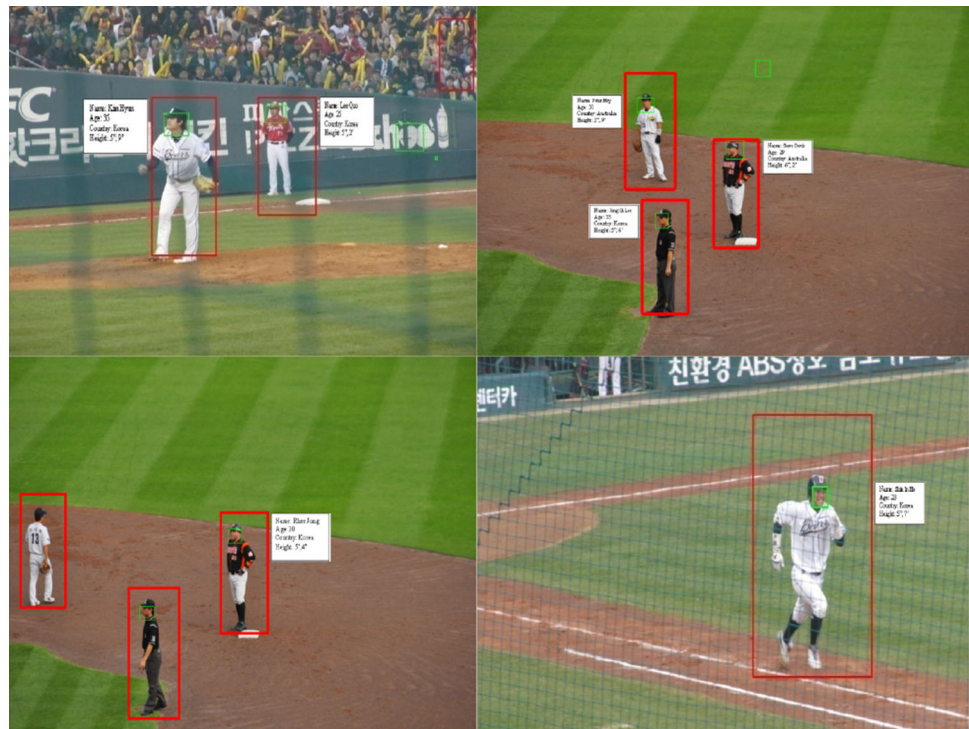
Figure 9 shows accuracy of face recognition module when the numbers of training images are varied. This kind of experiment is difficult to perform with players' database

because of the difficulty to obtain many images of a single player during play particularly when we want to have more players in our database such as we consider 80 players in our experiments. It is clear as we increase the number of training images, accuracy increases and reaches almost 100 %.

6.3 Further analysis of face recognition module

We perform several tests to ensure employed face recognition system is robust. One aspect of learning-based algorithm is the required number of training samples per subject to train the classifier. We take a subset from the whole player database by taking only 10 different baseball players with each player having 10 different images. Different numbers of training and testing samples are used to measure the CER. In Table 6 we show results of the sensitivity of the CER measured for various SSS learning tasks arising from different database partitions $\rho(L)$ to various weakness extents of gClassifiers in each task $\rho(r)$. Each combination of training and test samples was executed five times and minimum CER was obtained in each run along with iteration number as shown in Table 5. The \bar{e}_{20} shows that iteration number was set to 20 and (T^*) shows the iteration number on which we found lowest CER. We can see as we increase the number of training samples correspondingly CER is improved.

Fig. 8 Output of proposed system on sample images from database (Images taken at Jamshil Stadium, Seoul, South Korea)



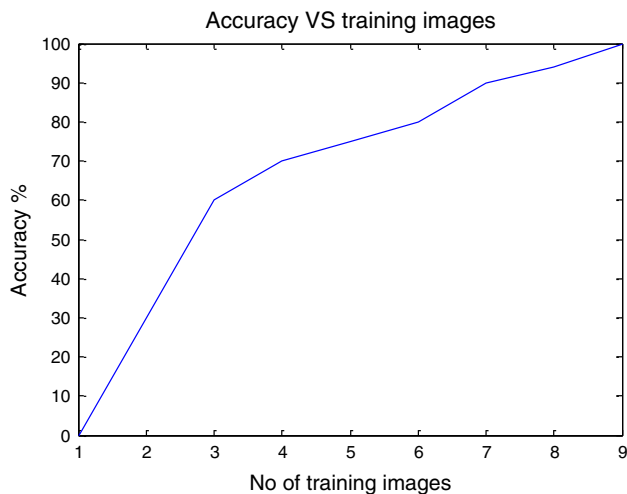


Fig. 9 Accuracy with number of training images used per subject

Table 6 Minimum CER With iteration number

Combination	% CER in each run					Mean error
	(1)	(2)	(3)	(4)	(5)	
(2tr) and (8tst) $e_{20}^-(T^*)$	20.00 (1)	20.00 (1)	12.50 (1)	10.00 (1)	17.50 (2)	16.00
(3tr) and (7tst) $e_{20}^-(T^*)$	20.00 (1)	8.57 (2)	20.00 (1)	5.71 (1)	7.14 (1)	12.28
(4tr) and (6tst) $e_{20}^-(T^*)$	3.33 (1)	3.33 (4)	8.33 (2)	8.33 (1)	10.00 (1)	4.64
(5tr) and (5tst) $e_{20}^-(T^*)$	4.00 (1)	4.00 (1)	0.00 (10)	6.00 (1)	2.00 (29)	3.20
(6tr) and (4tst) $e_{20}^-(T^*)$	2.50 (1)	0.00 (2)	2.50 (2)	5.00 (2)	5.00 (1)	3.00
(7tr) and (3tst) $e_{20}^-(T^*)$	6.66 (1)	3.33 (19)	0.00 (1)	0.00 (1)	0.00 (1)	2.00
(8tr) and (2tst) $e_{20}^-(T^*)$	1.50 (1)	0.00 (1)	5.00 (2)	0.00 (20)	0.00 (1)	1.30
(9tr) and (1tst) $e_{20}^-(T^*)$	0.00 (1)	0.00 (1)	0.00 (1)	0.00 (1)	0.00 (1)	0.00

In Table 7 we show generalization loss with number of training (tr) examples per subject to total samples as well as estimate of best r (r^* = The combination of training (tr) and testing (tst) samples on which we had lowest CER) and worst r (r^- = The combination of training and testing samples on which we had highest CER).

Detailed experiments are performed to test performance of face recognition module when the input face image is of small size. Table 7 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 and not the size of the face or illumination

Table 7 Generalization loss and estimation

Database partition ρ (L)	Total subjects/total images in database (10/100)
(2 tr) and (8 tst)	0.166
(3 tr) and (7 tst)	0.128
(4 tr) and (6 tst)	0.046
(5 tr) and (5 tst)	0.032
(6 tr) and (4 tst)	0.030
(7 tr) and (3 tst)	0.020
(8 tr) and (2 tst)	0.013
(9 tr) and (1tst)	0.00
Estimation	r^* = (9 Tr and 1 Testing) r^- = (2 Tr and 8 Testing)

Table 8 Recognition performance with different size of face images

# Players	Error (%) face size of 20×20		Error (%) face size of 10×10		Error (%) face size of 5×5	
	(1 tr and 5 tst)	(2 tr and 5 tst)	(1 tr and 5 tst)	(2 tr and 5 tst)	(1 tr and 5 tst)	(2 tr and 5 tst)
30	15.5	5.5	16.4	11.75	18.5	11.22

condition. As we increase the number of training samples corresponding CER is reduced. The face recognition module works well for as small as size of 5×5 with even 2 training samples. Player face databases range from size 231×251 to 10×10 . Therefore, the system performs well in most of recognition tasks.

7 Conclusions

In this paper, we propose 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 taken 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, nationality, etc. This work shows the feasibility of developing a real-time augmented reality application for enhanced users' experience which is based only on computer vision and image processing techniques [24]. 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 frame in real time.

References

- Looser J, Grasset R, Billinghurst M (2007) A 3D flexible and tangible magic lens in augmented reality. In: Proceedings of IEEE and ACM international symposium on mixed and augmented reality (ISMAR 2007), pp 51–54. ISBN: 978-1-4244-1749-0
- Han J, Farin D (2007) A real-time augmented-reality system for sports broadcast video enhancement. In: *MM'07*, September 23–28, 2007, Augsburg, Bavaria, Germany
- Inamoto N, Saito H (2004) Free viewpoint video synthesis and presentation of sporting events for mixed reality entertainment. In: Proceedings of ACM ACE, vol 74, pp 42–50
- Liang D, Liu Y, Huang Q, Zhu Q, Jiang S, Zhang Z, Gao W (2005) Video2cartoon: generating 3D cartoon from broadcast soccer video. In: Proceedings of ACM multimedia, pp 217–218
- Yu X, Yan X, Chi T, Cheong L (2006) Inserting 3D projected virtual content into broadcast tennis video. In: Proceedings of ACM multimedia, pp 619–622
- Matsui K, Iwase M, Agata M, Tanaka T, Ohnishi N (1998) Soccer image sequence computed by a virtual camera. In: Proceedings of CVPR, pp 860–865
- Mahmood Z, Ali T, Khattak S (2012) Automatic player detection and recognition in images using AdaBoost. In: 9th international bhurban conference on applied sciences and technology (IB-CAST). IEEE, pp 64–69
- Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57:137–154
- Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. *Comp Vision Pattern Recogn Conf* 2001(1):511–518
- Lang L, Gu W (2009) Study of face detection algorithm for real-time face detection system. In: 2nd international symposium on electronic commerce and security, pp 129–132
- Lu J, Plataniotis K, Venetsanopoulos A, Li S (2006) Ensemble-based discriminant learning with boosting for face recognition. *IEEE Trans Neural Networks* 17(1):166–178
- Shylaja SS, Balasubramanya KN, Natarajan S. Dimensionality reduction techniques for face recognition. Department of Information Science and Engineering, PEC Institute of Technology, India. <http://www.intechopen.com>
- Yu J, Tao D, Wang M (2012) Adaptive hypergraph learning and its application in image classification. *IEEE Trans Image Process* 21(7):3262–3272
- Yu J, Liu D, Tao D (2011) Complex object correspondence construction in two-dimensional animation. *IEEE Trans Image Process* 20(11):3257–3269
- Yu J, Wang M, Tao D (2012) Semisupervised multiview distance metric learning for cartoon synthesis. *IEEE Trans Image Process* 21(11):4636–4648
- Yu J, Liu D, Tao D (2012) On combining multiple features for cartoon character retrieval and clip synthesis. *IEEE Trans Syst Man Cybern* 42(5):1413–1427
- Yu J et al (2013) Pairwise constraints based multiview features fusion for scene classification. *Pattern Recogn* 46(2): 483–496
- Raudys SJ, Jain AK (1991) Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans Pattern Anal Mach Intell* 13(3):252–264
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139
- Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
- Kirchberg KJ, Jesorsky O, Frischholz RW (2002) Genetic optimization for hausdorff-distance based face localization. In: Intl. workshop on biometric authentication, Denmark, pp 103–111
- Wright J, Yang AY, Ganesh A, Sastry SS (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31(2):210–227
- Yu J, Rui Y, Chen B (2014) Exploiting click constraints and multi-view features for image re-ranking. *IEEE Trans Multimedia* 16(1):159–168
- Mahmood Z, Ali T, Khattak S, Aslam M, Mehmood H (2013) A color image enhancement technique using multiscale retinex. In: 11th international conference on frontiers of information technology (FIT). IEEE, pp 119–124



Zahid Mahmood Zahid Mahmood received his B.S. degree in Electrical and Computer Engineering from COMSATS Institute of Information Technology, Abbottabad, Pakistan, in 2007 and MS in Electrical Engineering from Hanyang University, South Korea in 2011. Currently, he is a Ph.D student at North Dakota State University, USA. His research expertise encompass topics such as digital image processing and signal processing. He is the recipient of Higher Education Commission, Government of Pakistan scholarship award for MS and Ph.D. in Electrical Engineering.



Tauseef Ali Tauseef Ali received his B.S. degree in Computer System Engineering from NWFP UET Peshawar, Pakistan in 2006, MS from Myungji University South Korea in 2008, and currently pursuing Ph.D from University of Twente, Netherlands. His research interests include 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.



Shahid Khattak Shahid Khattak received his B.Sc degree in Electrical Engineering from NWFP UET Peshawar, Pakistan in 1993, MS in Prude University USA in 1996 and Ph.D from Germany in 2008. Currently, he is serving as Professor and Chairman of Department of Electrical Engineering CIIT Abbottabad, Pakistan. His research interests include cognitive radios and wireless communication.



Laiq Hasan Laiq Hasan received his B.S. degree in Electrical Engineering from NWFP, UET Peshawar, Pakistan in 2000, M.Sc. degree in 2007, and Ph.D. degree in Computer Engineering from Delft University of Technology, Delft, Netherlands in 2011. He is a visiting researcher in the Faculty of Electrical Engineering Mathematics and Computer Science at Delft University of Technology, Delft, and an Assistant Professor in the

Department of Computer Systems Engineering at NWFP University of Engineering and Technology, Peshawar, Pakistan.



Samee U. Khan Samee U. Khan received his B.S. degree from GIK Institute of Engineering Sciences and Technology, Pakistan, in 1999 and a Ph.D. degree from the University of Texas in 2007. Currently, he is Associate Professor at the North Dakota State University. His research expertise encompasses topics such as sustainable computing, social networking, and reliability. In these areas, he has published over 230 papers.

He is the recipient of the Sudhir Mehta Memorial International Faculty Award, Chinese Academy of Sciences Young International Scientist Fellowship, CEA Researcher of the Year Award, ACM/IEEE GreenCom Best Paper Award, and Nortel Outstanding Doctoral Dissertation.