Injury Mechanism Classification in Soccer Videos

O. V. Ramana Murthy[1]
[1]Vision & Sensing, HCT Research Centre,
ESTeM, University of Canberra
O.V.RamanaMurthy@ieee.org

Roland Goecke[1,2]
[2]IHCC, RSCS, CECS,
Australian National University
roland.goecke@ieee.org

Abstract

Soccer is a very popular sport but also has a high rate of injuries. In this paper, player falling events in soccer videos are classified into five major categories. These categories have been identified by soccer coaches as the major mechanisms behind player injuries. Automatic detection of these events will be useful to coaches to plan specific training modules and to impart individual training to the players that will enhance their physical strength and also their playing style. A Bag-of-Words framework is used and a baseline classification accuracy is established that will serve as a reference point for further work.

1. Introduction

With roughly 200,000 professional soccer players and around 240 million amateur soccer players, soccer is a game with worldwide appeal. However, it has been reported [23] that injuries found in soccer are more frequent than in field hockey, volleyball, handball, basketball, rugby, cricket, badminton, fencing, cycling, judo, boxing, sub-aqua and swimming. This can be a considerable problem for the player, the team, the club, and for society at large. Health consequences are seen not just in the short term but also as a risk of career break and early-onset health problems such as osteoarthritis. Generally, 50-80% of soccer injuries occur at the feet and legs, such as a sprained ankle or torn knee ligaments. The ACL injury (Anterior Cruciate Ligament) is the most frequently reported severe injury in any season. While head injuries account for 4-22%, concussions are usually rare, making up only 2-3% of all soccer injuries. Most severe head injuries are caused by collisions that could be with other players, goalposts, the ground, or the ball. While most injuries to the foot and ankle joint can be treated conservatively, complex injuries require anatomic reconstruction to allow for quick rehabilitation and return to play as early as possible.

Tysvaer [16] extensively reviewed soccer injuries in the head and neck, while Wong and Hong [23] reviewed on lower extremities such as hip, groin, upper leg, knee, lower leg, ankle, and foot. Valderrabano *et al.* [17] surveyed foot and ankle joint injuries in soccer players and divided the risk factors into two divisions – intrinsic (self-inflicted) factors and extrinsic (external forces). Intrinsic factors are influenced by individual, biological or psychological attributes of the soccer player. Extrinsic factors are related to the role that the environment plays. Examples of intrinsic factors are previous injuries, stress level, inadequate rehabilitation after injury,etc. These can be assessed only from the past medical history and/or post injury medical/physical tests. On the other hand, extrinsic risk factors are due to foul play, physical activities involving sudden forced impact on the body due to collision with another player or object. Injuries can be also caused by a collision when jumping for a header or when landing. Soccer shoes are equipped with cleats for better grip on the turf. Getting stuck on turf leads to unusually high load and torque in the knee and ankle joints, leading to serious injuries. With the recent breakthroughs in computer vision and pattern recognition techniques, video analysis can be helpful in analysing the external risk factors.

Different definitions of injury exist in the literature [23]. Some defined injury as any condition that caused a player to be removed from a game, miss a game, or to be disabled enough to seek medical treatment, while others defined injury as one received during training or competition, which prevented the injured player from participating in normal training or competition for more than 48 hours, not including the day of the injury. Some studies counted injuries occurring in competition only, while others counted both competition and training injuries. In this paper, we use a broad perspective, taking player falling events in competition as the base for analysis. We hypothesise that a player falling event is a precursor to direct injury or down performance of the player. (Note, not all events of a player falling lead to injury.) Automatic identification of such events from the matches will be helpful for coaches for timely substitution of the player (in-match) and training programme (post-match). The contributions of this paper in this regard are:

- *Identifying events based on potential external risk fac-*

tors
Player falling events in soccer match videos are identified into five broad categories, as shown in Table 1. These five classes are based on the most common injury mechanisms as identified by Wong and Hong [23].

- *Baseline for classification*
The five categories in the falling events are very similar to each other. Current state-of-the-art feature vectors in-use for general human action recognition are used to establish the baseline performance.

Automatic detection of player falling events can be used to develop training programs for the players accordingly. A good example is the "**FIFA 11+**" [1] [14] warm-up training developed by leading sports medicine professionals to prevent injury. The training program **FIFA 11+** is based on the hypothesis that the human body has some natural defence mechanisms against injuries. If a player is trained properly, that individual can become more resistant to injuries. For example, training certain muscles helps to stabilise joints, while training one's balance makes one less susceptible to a loss of balance and subsequent falls. There are also other techniques, such as how one jumps or lands, that protect one from getting injured in these critical situations. In a scientific study with almost 2,000 female youth players, teams using the **FIFA 11+** training program at least twice a week had $30-50\%$ fewer injured players than teams who warmed up as usual.

In the remainder of the paper, Section 2 contains a review of studies on injury analysis in soccer players and low-level features for action recognition. Section 3 describes the overall framework and details of experiment settings, such as the interest points, local feature descriptors, codebook generation, classifier and the dataset used for validation. Section 4 presents and discusses the results obtained on the dataset. Finally, conclusions are drawn in Section 5.

2. Related Work

In this section, we briefly review significant studies in soccer video analysis and low-level features for action recognition in real-world videos.

2.1. Injury Analysis in Soccer Players

Gouttebarge *et al.* [6] investigated the effect of an intrinsic factor – previous injury – on knee and ankle osteoarthritis (OA) in former professional football players and reported that it does have negative consequences on Sports and daily/work activities too. del Pozo *et al.* [4] investigated and reported that professional status can also be determining factor as regards injuries in Spanish football players and advocated developing injury prevention strategies to reduce the overall risk to clubs and players.

Correa *et al.* [3] investigated the occurrence of incidents involving the craniofacial region during Brazilian Professional Soccer League matches. The observance of game rules and "fair play" spirit by both athletes and referees seems to be the better strategy to reduce the number of incidents. The use of protective appliances could be taken into account for players in specific at-risk position such as goalkeepers and other players that are involved in the defense zone.

There are several studies in vision based analysis of soccer injuries. Due to their unobtrusive nature, vision-based solutions have an edge over other solutions such as wearable sensors (e.g. GPS or RFID sensors). Players or balls need not be instrumented prior and during matches. Andersen *et al.* [1] analysed videotapes of 313 matches from Norwegian and Icelandic elite football during the 1999 to 2000 seasons w.r.t ankle injuries. Although only 57% of the actual injures were identified in the videotapes, they concluded that systematic video analysis provides detailed information on the mechanisms for ankle injuries in football – for lateral ligament sprains and for the condition dubbed "footballers ankle". Bjørneboe *et al.* [2] investigated possible injury incidents from the 2000 season to the 2010 season in Norwegian male professional football using video analysis. After studying 414 matches, they found 1287 incidents. They reported an increased rate of non-contact and opponent-to-player contact incidents in both heading and tackling duels in the 2010 season compared with 10 years earlier, even if there was no increase in the frequency of player-to-player contact situations. Waldn *et al.* [18] assessed videos of all professional football injury surveillance videos between 2001 and 2011. Five analysts independently reviewed all videos to estimate the time of initial foot contact with the ground and the time of ACL tear. All videos were then analysed according to a structured format describing the injury circumstances and lower limb joint biomechanics. They concluded that 85% of the ACL injuries in male professional football players resulted from non-contact or indirect contact mechanisms. The most common playing situation leading to injury was pressing, followed by kicking and heading. In all these studies, only few (in the tens) situations of injury incidents were found after manually analysing the videomatches repositories.

Distinct events such as a goal scored or a player injury are normally annotated by a human, which is tedious and error prone. Automatic detection of such events can alleviate this burden and help coaches to focus on higher-level tasks such as strategy analysis. Wei *et al.* [22] proposed a two-layer hierarchical approach to detect events such as *in-play* (when the match being played), *stoppage* (when the ball is out, fouls, player injury, substitution, etc.), *out-for-corners*,

[1] http://f-marc.com/11plus/home/

out-for-goal-kicks, *Foul Freekicks* and *Out-for-throw-in*. They ignore player substitutions and player injury. Ramana Murthy and Goecke [12] employed a temporal modelling technique to classify player falling from player running/playing normal situations. Along these lines, we identity five major mechanisms that can potentially lead to injury incidents.

2.2. Low-level Features for Action Recognition

Human action recognition by low-level features, such as spatio-temporal interest points (STIP) [9], improved dense trajectories [20] from real-world videos, has made rapid progress especially in the last 3-5 years. Remarkable performance of 66.79% [11], 87.9% [10] and 92.3% [10] have been reported on benchmark datasets **HMDB51** [7], **UCF101** [15] and **UCF50** [13], respectively. These datasets contain 50-101 different classes and videos are recorded when people are performing these actions in their routine/daily life. Further, the range of actions is very wide, beginning form facial actions, such as *smiling* or *talking*, to body movements with others or objects, such as *sword fight*, *ride horse* or *marching*.

Peng *et al.* [10] provide a comprehensive study of all steps in a Bag-of-Visual Words pipeline and different fusion methods to produce a state-of-the-art action recognition system. Specifically, they explored two kinds of local features – STIP [9] and improved dense trajectories [20], ten kinds of encoding methods, eight kinds of pooling and normalization strategies, and three kinds of fusion methods. In our work, we pick up their best features and encoding techniques (Fisher Vector). Their observations are consistent with other studies [21], the top performers in action recognition with large number of classes.

3. Overall Framework

The overall layout of the framework used is shown in Figure 1. It is based on a Bag-of-Words (BoW) approach. Initially, interest points – Spatio Temporal Interest Points (STIP) or trajectories of moving objects – are detected separately. Local descriptors are computed around these detected interest points. Fisher Vector encoding is used to construct a feature vector per video, which is nearly of length $\sim 100K$. This feature vector is used to learn a classifier (for each action class).

3.1. Spatio Temporal Interest Points

In their seminal work, Laptev *et al.* [8] proposed the usage of Harris 3D corners as an extension of traditional (2D) Harris corner points for spatio-temporal analysis and action recognition. These interest points are local maxima of a function of space-time gradients. They compute a spatio-temporal second-moment matrix at each video point in different spatio-temporal scales. This matrix essentially captures space-time gradients. The interest points are obtained as local maxima of a function of this second-moment matrix. We use the original implementation available online.[2] We compute the local descriptors histograms of gradient orientations (HOG) and histograms of optical flow (HOF). While the former captures the local motion and appearance, the latter captures the temporal changes.

3.2. Trajectories of Moving Objects

Wang *et al.* [19] proposed dense trajectories of moving objects to model human actions. Interest points are sampled at uniform intervals in space and time, and tracked based on displacement information from a dense optical flow field. Improved dense trajectories (iDT) [20] are an improved version of the dense trajectories obtained by estimating the camera motion. Wang and Schmid [20] use a human body detector to separate motion stemming from humans movements from camera motion. The estimate is also used to cancel out possible camera motion from the optical flow. For trajectories of moving objects, we compute these iDT. In our experiments, we only use the online version[3] of camera motion compensated iDT, without any human body detector. The local descriptors computed on these trajectories are HOG, HOF, motion boundary histograms (MBH) and trajectory shape. MBH are descriptors based on motion boundaries and are computed by separate derivatives for the horizontal and vertical components of the optical flow. The trajectory shape descriptor encodes local motion patterns (in terms of displacement vectors of points (x, y) of subsequent frames).

3.3. Feature Encoding and Classification

We build Fisher Vectors [20] for each descriptor separately. In this technique, a Gaussian Mixture Model (GMM) is fitted to a randomly selected (250,000) descriptors from the training set. Let the parameters obtained from the GMM fitting be defined as $\theta = (\pi_j, m_j, \sum_j; j = 1, 2, ..., k)$ where π_j, m_j and $\sum_j$ are the prior probability, mean and covariance of each distribution. The mean (u_{jk}) and deviation vectors (v_{jk}) for each mode k are computed and concatenated to yield Fisher Vector (FV). The FV is then normalised by the 'power-law normalisation' defined as $v_j = |v_j|^\alpha \times sign(v_j)$ with $\alpha = 0.5$. Finally, the vector is L_2-normalised as $v = \frac{v}{||v||}$ to yield the FV vector. For classification, we concatenate all Fisher Vectors (of different descriptors) and use linear SVM LIBLINEAR [5]. We apply the one-versus-all approach in all cases and select the class with the highest score.

[2] http://www.di.ens.fr/~laptev/download.html/#stip

[3] http://lear.inrialpes.fr/people/wang/improved_trajectories

Figure 1. Overall Framework: Starting with an input video, first, interest points are generated and local descriptors computed. Codewords are obtained via GMM and Fisher vectors are generated from them to serve as feature vector. Finally, a SVM classifier is used to predict the action label.

3.4. Dataset

To investigate our hypothesis, we use video clips from the publicly available **Soccer Events** dataset [12]. The **Soccer Events** dataset consists of 480 High Definition (HD) (1920 × 1080 pixels) videos clips captured by static cameras mounted in the four corners of the field. It has only two classes – player running and falling. We select only few clips from the player falling category and identify them into five classes as described in the Table 1. We name this subset as **Injury Mechanism** dataset and will release the annotations for use by other researchers. These five classes are based on the most common injury mechanisms as identified by Wong and Hong [23] and are as follows:

1. **Tackling** is when players try to get possession of the ball from the opponents. As players cannot respond quickly enough to avoid rapid and unpredictable movements of the opponent players, the lower extremities are often injured during tackling. (See Figure 2 (a) & (b) and supplementary video 1).

2. **Running** is when a player is not in contact with any other player, nor the ball. (See Figure 2 (c) & (d) and supplementary video 2).

3. **Twisting and Turning** is when a player has suddenly lost balance due to twisting the ankle. (See Figure 2 (e) & (f) and supplementary video 3).

   Often, **Running** and **Twisting and Turning** injuries are due to inferior playing surfaces and inappropriate footwear. Uneven playing surfaces may result in more loading on the ligaments and muscles. When external loading is greater than what the ligaments and muscles can tolerate, injury usually follows. Incorrect footwear that cannot provide sufficient frictional force will eventually lead to slipping. On the other hand, too much frictional force will produce large torque when **Twisting and Turning**, which may also lead to injury.

4. **Shooting** is when a player hits the ball and loses balance. In this situation, there is no contact with any other player. (See Figure 2 (g) & (h) and supplementary video 4).

5. **Jumping and Landing** injuries occur during a header and goalkeeping. Causes of such injuries are incorrect landing technique and collisions between players after take off and before landing. (See Figure 2 (i) & (j) and supplementary video 5).

Table 1. Injury mechanism events

Class	No. of samples
Tackling	92
Running	12
Twisting and Turning	10
Shooting	12
Jumping and Landing	31
Total	157

This data has been collected from four matches (90min each, 4 cameras). Thus, a total of 24h of video has been indexed and only 157 clips (total of 15.56min) were found to containing potential injury incidents. That is only 1.08%. Thus, our hypothesis of automation would greatly be of assistance to coaches.

4. Results and Discussion

In this section, the results obtained on the **Injury Mechanism** dataset are presented and analysed. The average accuracies over a five-fold cross-validation are presented in Table 2. The performance of low-level features – STIP and iDT – is tested for different numbers of Gaussians ($k = 128, 256, 512$) for each descriptor type in all our experiments.

Table 2. Performance of low-level features

Description	No. of Gaussians	Accuracy
STIP	128	56.7%
	256	**57.4%**
	512	56.1%
iDT	128	52.8%
	256	**59.3%**
	512	58.2%

It is observed that iDT performed better than STIP by 2.1% (absolute). This might be due to the motion changes detected by iDT (that track for motion changes for every 15 frames span). Further, the size of codebook (number of Gaussians) that yielded highest performance is found to be 256, which has been observed to be best experimental setting [10]. The results obtained by STIP and iDT define the baseline performance against which the results obtained in future work can be compared.

5. Conclusions and Future Work

Five categories are identified as the major mechanisms behind most potential injury incidents in soccer. A Bag-of-Words framework is used and a baseline classification accuracy is established. The highest recognition rate observed was 59.3%. Although the event of 'falling' is the same for each class, the mechanism behind that fall is different. Whether computer vision techniques can detect those mechanisms is an interesting challenge and this dataset will be useful for that research. In future, we would like to investigate effective feature vectors for improving classification accuracy and also localise the event in time and space for more detailed analysis.

References

[1] T. E. Andersen, T. W. Floerenes, A. Arnason, and R. Bahr. Video analysis of the mechanisms for ankle injuries in football. *The American Journal of Sports Medicine*, 32(1):69S – 79S, 2004. 2

[2] J. Bjrneboe, R. Bahr, and T. E. Andersen. Video analysis of situations with a high-risk for injury in norwegian male professional football; a comparison between 2000 and 2010. *British Journal of Sports Medicine*, 48:774 – 778, January 2013. 2

[3] M. B. Correa, C. B. Knabach, K. Collares, P. C. Hallal, and F. F. Demarco. Video analysis of craniofacial soccer incidents: A prospective study. *Journal of Science and Medicine in Sport*, 15(1):14 – 18, 2012. 2

[4] L. B. del Pozo, C. A. Prez, G. R. Benzanilla, A. M. Fernndez, T. F. Villa, and V. M. Snchez. Influence of the soccer players professional status on the frequency and severity of injuries: A comparative pilot study. *Apunts. Medicina de l'Esport*, 49(181):20 – 24, 2014. 2

[5] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008. 3

[6] V. Gouttebarge, H. Inklaar, and M. H. Frings-Dresen. Risk and consequences of osteoarthritis after a professional football career: a systematic review of the recent literature. *Journal of Sports Med Phys Fitness*, 54(4):494 – 504, 2014. 2

[7] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A large video database for human motion recognition. In *International Conference on Computer Vision (ICCV)*, 2011. 3

[8] I. Laptev and T. Lindeberg. Space-Time Interest Points. In *International Conference on Computer Vision (ICCV)*, pages 432–439, Oct 2003. 3

[9] M. Marszalek, I. Laptev, and C. Schmid. Actions in context. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. 3

[10] X. Peng, L. Wang, X. Wang, and Y. Qiao. Bag of Visual Words and Fusion Methods for Action Recognition: Comprehensive Study and Good Practice. *ArXiv e-prints*, May 2014. 3, 5

[11] X. Peng, C. Zou, Y. Qiao, and Q. Peng. Action recognition with stacked fisher vectors. In *European Conference on Computer Vision(ECCV)*, pages 581–595. 2014. 3

[12] O. V. Ramana Murthy and R. Goecke. Player falling event detection in soccer videos. In *ACM Conference on Knowledge Data and Discovery (KDD) Workshops, Large Scale Sports Analytics*, Aug. 2015. 3, 4

[13] K. K. Reddy and M. Shah. Recognizing 50 Human Action Categories of Web Videos. *Machine Vision and Applications*, 24:971–981, Sept 2012. 3

[14] T. Soligard, G. Myklebust, K. Steffen, I. Holme, H. Silvers, M. Bizzini, A. Junge, J. Dvorak, R. Bahr, and T. E. Andersen. Comprehensive warm-up programme to prevent injuries in young female footballers: cluster randomised controlled trial. *British Medical Journal*, 337, 2008. 2

[15] K. Soomro, A. R. Zamir, and M. Shah. UCF101: A Dataset of 101 Human Action Classes from Videos in the Wild. In *CRCV-TR-12-01*. November 2012. 3

[16] A. Tysvaer. Head and neck injuries in soccer. *Sports Medicine*, 14(3):200–213, 1992. 1

[17] V. Valderrabano, A. Barg, J. Paul, G. Pagenstert, and M. Wiewiorski. Foot and ankle injuries in professional soccer players. *Sport-Orthopdie - Sport-Traumatologie - Sports Orthopaedics and Traumatology*, 30(2):98 – 105, 2014. 1

[18] M. Waldn, T. Krosshaug, J. Bjrneboe, T. E. Andersen, O. Faul, and M. Hgglund. Three distinct mechanisms predominate in noncontact anterior cruciate ligament injuries in male professional football players: a systematic video analysis of 39 cases. *British Journal on Sports Medicine*, April 2015. 2

[19] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu. Action recognition by dense trajectories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 3

[20] H. Wang and C. Schmid. Action Recognition with Improved Trajectories. In *Intenational Conference on Computer Vision (ICCV)*, 2013. 3

[21] H. Wang and C. Schmid. LEAR-INRIA submission for the THUMOS workshop. In *THUMOS:ICCV Challenge on Action Recognition with a Large Number of Classes, Winner*, 2013. 3

[22] X. Wei, L. Sha, P. Lucey, S. Morgan, and S. Sridharan. Large-scale analysis of formations in soccer. In *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8, Nov 2013. 2

[23] P. Wong and Y. Hong. Soccer injury in the lower extremities. *British Journal in Sports Medicine*, 39(8):473–82, Aug 2005. 1, 2, 4

Figure 2. Cropped portions from sample frames of different injury mechanisms. *Left hand side* : Just before falling. *Right hand side*: just after falling. (a) & (b) Tackling (See Supplementary Video 1); (c) & (d) Running (See Supplementary Video 2); (e) & (f) Twisting and Turning (See Supplementary Video 3); (g) & (h) Shooting (See Supplementary Video 4); (i) & (j) Jumping and Landing (See Supplementary Video 5)