Mr. Referee, it was goal!

Juan Bernabé Moreno
University of Granada, Department of Computer Science and Artificial Intelligence

Abstract— **This work recreates the article of Ancona et all where they provide a very innovative use of a SVM classifier to solve a very old problem in the history of football: the ghost goals.**

***Index Terms*— SVM, pattern recognition, classification**

I. INTRODUCTION

When Helmut Haller put Germany 1-0 up, it was the first time England had been behind in the tournament. But a header by Hurst soon cancelled that out, and the scoreline stayed at 1-1 until 12 minutes from time. "Geoff had a shot blocked by Hottges, and it fell to me about eight yards out," said Peters.
"I thought, 'whatever you do, just hit the target, don't hit it over'
"I connected with it well, the goalkeeper Tilkowski went one way, the defender Schnellinger went the other and my shot went straight down the middle and into the net.
"The feeling was amazing, it was as if I had been struck by lightning, but I never had any selfish thoughts about how it might be my goal that won the World Cup.

"I was thinking about winning it for the team, for the forgotten heroes like Jimmy Armfield and Ron Springett, who were in the squad but not playing."
But as every football fan knows, Peters' goal did not win the World Cup

The problem is basically detecting when a goal takes place in a football match without modifying the football or setting up additional devices in the goalposts.
It is not unlikely that in certain actions neither the referee nor his linemen are able to determine if the football trespassed the goal line. This is known as ghost goals… and there have been many in the football history… One of the most prominent is the one

Optical devices like standard TV cameras can be used to tackle the problem because they don't interfere the match or modify the game, and can be placed externally, yet providing sharp snapshots from the goal if equipped with the required optical accessories (like zoom and filters removing the noise). Another huge advantage of the cameras usage is that images are real-time recorded and can be provided within seconds for analysis and decision support to for example, the famous forth referee.

There has been an attempt to solving the problem of goal detection using an uncalibrated binocular vision system[4]. The method relies on exploiting two images of the field taken simultaneously from two different viewpoints, where both goal area and goal are visible.

From the two video sequences of an incident captured from different view-points, they compute a novel (overhead) view using pairs of corresponding images -the computation of the vertical vanishing point in both images and of the homography between the two- images. Using projective constructs we determine the point at which the vertical line through the ball pierces the ground plane in each frame, and the distance of this point to the goal line determines if the goal occurred.

Limitations are the detection of the ball in both images to enable the triangulation and the fact that the method works only in 2D, as the third coordinate of the ball can't be calculated

Ancona et al propose a very innovative approach to detect ghost goals, as we will see in the subsequent sections

II. COMPLEXITY REDUCTION

Empirically, it is proven that the best place a linesman can be to assist the referee in his decision is close to the corner flag (see Figure 1). Only determining if the ball is/has been on the left (right) side of the goalpost will decide to give the goal to the scoring team or rather leave the match continue.

Having said that, the problem is being reduced to ball detection in images.

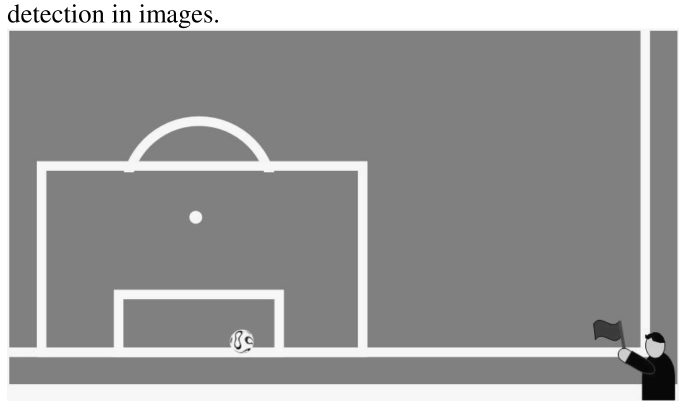

Figure 1

The problem of detecting the ball occurrence in images taken from a particular place is just an instance of a generic problem of detecting 3D objects by using the image projected by the object on the sensing plane of the standard camera.

This problem has been traditionally tried to be solved using standard computer vision algorithms, like region growing, edge detection, snakes, texture analysis, etc all of them aiming at determining if a given 3D object is present in an image. The fact is that this problem is very unlike to be solved by the mentioned techniques due to the complexity derived from the countless pattern variations in the images.

Ancona's suggestion exploits a new technique based on an example learning approach: determining if an object is present in one image is done taking into account a lot of views of the object the algorithm has been trained with. The most suitable schemas to be applied are based on supervised learning. Being a bit more specific, the current problem can be considered a classification problem, as we have to distinguished goal and not-goal situations. Actually what we are looking for is a separating surface (optimal only under certain conditions), which is able to separate object views from image patterns that are not instances of the object.

III. Support Vector machines to solve the problem

It is known that the general of classification, can be interpreted as the problem of approximating a multivariate function from sparse data [2], where the data are in the form of (input, output) pairs, obtained by random sampling the unknown function in the presence of noise.

This problem is clearly ill-posed, since it has an infinite number of solutions and, in order to choose one particular solution, we need to have some a priori knowledge of the function that has to be reconstructed

Vapnik [3] introduced a new learning schema based in the statistical learning theory, called SVM for approaching classification and regression problems. Vapnik's theory relies on the idea of regularization (for a finite set of training examples, the search of the best model has to be constrained by an approximately small hypothesis space that is the set of functions the machine implements. Taking a too large space will lead to have functions that are fitting exactly the data, but with a very poor generalization capabilities.

Roughly speaking, for a given learning task, with a given finite amount of training data, the best generalization performance will be achieved if the right balance is struck between the accuracy attained on that particular training set, and the "capacity" of the machine, that is, the ability of the machine to learn any training set without error. A machine with too much capacity is like a botanist with a photographic memory who, when presented with a new tree, concludes that it is not a tree because it has a different number of leaves from anything she has seen before; a machine with too little capacity is like the botanist's lazy brother, who declares that if it's green, it's a tree. Neither can generalize well.
The exploration and formalization of these concepts has resulted in one of the shining peaks of the theory of statistical learning [3]

In this section we review the basic concepts of SVM for two classes' classification problems for the general case of not linearly separable classes with linear and not linear surfaces.

We are given a training set

$S = \{(x_i, y_i)\}_{i=1}^{l}$ of size l where $x_i \in R^n$ and $y_i \in \{-1,1\}$, for i=1,2,…,l. That is, all samples belong to either of two classes.

The hypothesis of having 2 non-linearly separable classes, the optimal separating hyperplane

$w^* \cdot x + b^* = 0$ found by SVM is solution of the following quadratic problem with linear constraint:

$$\min_{w,b,\xi} \frac{1}{2} w \cdot w + C \sum_{i=1}^{l} \xi_i$$

Fulfilling that

$$y_i (w \cdot x_i + b) + \xi_i \geq 1 \quad i = 1,2,\ldots,l \qquad \xi \geq 0$$

Where:

- C is a positive number
- ξ_i is a non-negative slack variable

(one for each point in S that measures the amount of misclassification of the point x_i with respect to the optimal separating hyperplace (will be 0 if the point is correctly classified)
C is called regularization parameter:

- if C is too large, the optimal hyperplane tends to minimize the non correctly points of S.
- if C is too small, the optimal hyperplane tends to maximize the distance of the closest point of S

Thus, for intermediate values of C, the solution of the quadratic problem is a tradeoff between maximum margin and minimum number of misclassified points

We can also solve the problem with linear constraints by using the Lagrange multiploers. At this aim, we introduce *l* non negative slack variables λ_i relative to the constraints $\xi_i \geq 0$. If we call $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ and $\mu = (\mu_1, \mu_2, \ldots, \mu_l)$ the 2l Lagrange multipliers relative to the constraints of the quadratic problem we formulated above, solving it is equivalent to determining the saddle point of the Lagrangian function:

$$L = \frac{1}{2} w \cdot w + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \lambda_i [y_i (w \cdot x_i + b) + \xi_i - 1] - \sum_{i=1}^{N} \mu_i \xi_i$$

Where $L=L(w,b,\xi,\lambda,\mu)$. So the optimal w* is:

$$w^* = \sum_{i=1}^{N} \lambda^*_i x_i$$

where λ^* is the solution to the dual problem of the quadratic problem.

$$\max_{\lambda} \; -\frac{1}{2} \lambda \cdot D \lambda + \sum_{i=1}^{l} \lambda_i$$

Subject to

$$\sum_{\substack{i=1 \\ 0 \le \lambda_i \le C}}^{l} \lambda_i y_i = 0 \qquad i = 1, 2, ..., l$$

Where D is the matrix of size lxl, with $D_{ij}=y_i y_j x_i \cdot x_j$
Per i,j=1,2,…,l. Moreover, the optimal b* can be computed using the Kuhn-Tucker conditions:

$$(C - \lambda_i^*)\xi_i^* = 0 \quad i = 1, 2, ..., l$$

$$\lambda_i^*\left[y_i(w \cdot x_i + b) + \xi_i^* - 1\right] = 0 \quad i = 1, 2, ... l$$

where ξ_i^* are the values of ξ_i at the saddle point. In fact, from the Kuhn Tucker condition we have that:

$$b^* = y_i - w^* \cdot x_i \quad \forall i \ni' 0 < \lambda_i^* < 0$$

The points x_i with $\lambda_i^* > 0$ are called support vectors. The classification of a new data x involves the evaluation of the decision function:

$$f(x) = sign\left(\sum_{i=1}^{l} \lambda_i^* y_i (x_i \cdot x) + b^*\right)$$

Where the solution is expressed evaluating the dot product between the data and some elements (support vectors) of the training set S

A. Extension to non linear separating surfaces

This is done by mapping the input vectors x in a higher dimensional space, called feature space and looking for the optimal separating hyperplane in this new space. Let Φ(x) be the image of the point x in the feature space, with:

$$\Phi(x) = \left(a_1\Phi_1(x), a_2\Phi_2(x), ..., a_n\Phi_n(x)\right)$$

where $\{a_n\}\infty_{n=1}$ are real numbers and $\{\Phi_n\}\infty_{n=1}$ are real functions. In the feature space induced by the mapping Φ, the optimal separating hyperplane found by SVM has the form:

$$f(x) = sign\left(\sum_{i=1}^{l} \lambda_i^* y_i \Phi(x_i) \cdot \Phi(x) + b^*\right)$$

Where the inner product of vectors in the feature space is:

$$\Phi(x) \cdot \Phi(y) = \sum_{n=1}^{\infty} a_n^2 \Phi_n(x) \cdot \Phi_n(y)$$

Let K a function of two variables x and y of the input space which estimates the inner product of their corresponding images, Φ(x) and Φ(y), in the feature space, that is:

$$K(x, y) = \Phi(x) \cdot \Phi(y)$$

Then, the optimal separating hyperplane in the feature space can be written as a non linear separating surface in the input space:

$$f(x) = sign\left(\sum_{i=1}^{l} \lambda_i^* y_i K(x_i, x) + b^*\right)$$

represented as a linear combination of kernel functions centered on the support vectors only. The Mercer's theorem establishes general conditions for a kernel function K to estimate inner products in Hilbert spaces. In fact, suppose K a continuous symmetric function, kernel of the positive definite integral operator:

$$(T_K f)(x) = \int K(x, y) f(y) dy$$

Then K admits an expansion of the form:

$$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \Phi_n(x) \Phi_n(y)$$

where Φ_n are the mutually orthogonal eigen-functions and λ_n the corresponding eigen values of the integral operator Tk, that is they are solution of the following integral equation:

$$\int K(x, y) f(y) dy = \lambda \Phi(x)$$

It is important to point out that the mutually orthogonal functions Φ_n (features) span a Hilbert space in which the optimal classifier lives. In other words, specifying the kernel function K used in SVM is equivalent to specify the set of all possible classifier that the machine implements, or the complexity of the function space in which the final classifier lives.

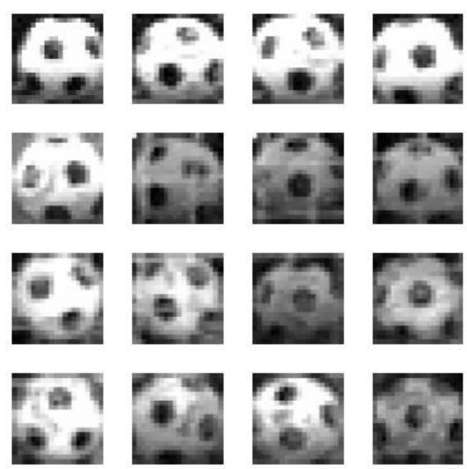

Figure 2

IV. Experimental results

The experiment has been designed in three phases: data collection, SVM training and performance estimation of the resulting classifier.

To take the images, Ancona et all used a standard TV camera with a zoom lense with f=75 mm focal length. The shooting time was set to 1/10000 sec to reduce the motion blurring.

The camera was placed externally to the football ground, the height of its optical center was 1.5 m roughly and its optical axis was manually aligned with the goal line. The distance of the camera with respect to the center of the two goal posters was 48 m, as shown in Figure 3

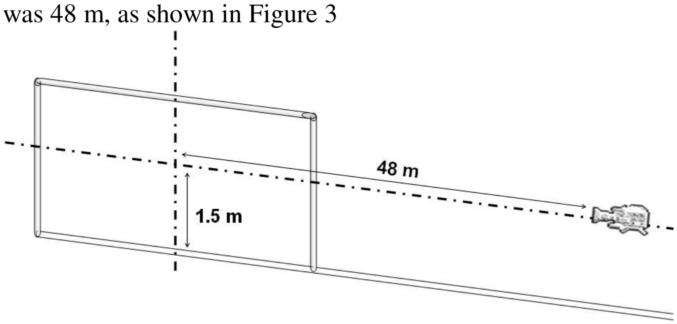

Figure 3

Having chosen the camera position as describes brings a huge advantage: the football size projected on the camera is almost constant when the football moves inside the area being monitored by the camera. Actually, for 3D points belonging to this area, the image formation process can be described, from a geometric point of view, by using orthographic projection, instead of perspective projection. It is a matter of fact, that the size of objects perceived by the camera is invariant under the orthographic projection. For us this means that the size of the football stays the same and whatever algorithm put in place to detect the occurrence of the ball in the goalmouth area doesn't have to manage scale variations of the football.

What Ancona and his team did was taking 2004 images of the textured standard football (in different attitudes and illumination conditions, in different positions inside the area being monitored). Samples have been chosen with a size of 20x20pixels. For all positive examples, the football was totally visible (detecting partially occluded footballs has not been taken into account in the current implementation).

About negative examples have been collected following these steps:

- Images were acquired, where the football wasn't present (people, advertising, players containing all potential false positive patterns)
- Each 20x20pixels sub-image of the previous ones is a negative example and can be considered in the training process.
- The exponential growth of the negative examples motivated the usage of some technique that low the number of negative examples and at the same time select image patterns relevant for the problem (detecting balls in the present context). For this purpose Ancona collected 1230 negative examples, sampling in a sampling on a regular grid a negative examples' image. A training set composed of 3234 examples was used for training an SVM for classification with a second degree polynomial kernel $K(x, y) = (1 + x \cdot y)^2$ and a regularization parameter of C = 200. The classifier was tested in one image not containing instances of the football, and 3647 false positive image patterns. The selected negative examples were added to the training set and the training process was repeated, by using a total of 6881 examples and the same kernel function and regularization parameter used before. This procedure of search of negative examples relevant for the problem at hand was iterated several times, each time using different negative example images. In particular, the 5 successive images produced 277 negative image patterns and the last 37 images produced 2817 negative image patterns. The final classifier was so obtained training an SVM on 9975 positive and negative examples by using the same kernel function and regularization parameter. Notice that the refinement process involved 43 false positive images for a total of over 4 millions of negative examples.
- All the examples were appropriately preprocessed before training. First, pixels close to the boundary of each example window were removed in order to eliminate parts belonging to the background. Then histogram equalization was applied to reduce variations in image brightness and contrast. The resulting pixels were used as input to the classifier.
- For measuring the generalization capabilities of the learning machine, that is the ability of the machine to correctly classifying image patterns never seen before, we tested the classifier on 900 images acquired under different illumination conditions. Each test image was exhaustively scanned and all the sub-images with size 20_20pxl were classified as instance of the football or not. The Figure 4 shows a typical image used for testing. For better understanding the performances of the classifier, we analyzed all the test images checking for the visibility of the football, before of the classification process. We counted the images in which the football was visible, occluded and partially occluded, with occlusion less than or greater than 50%. The ROC curves show the performances of the classifier on images with fully visible footballs (upper curve) and on images with occluded footballs (lower curve). In the first case we had a detection rate of 98:3% with a false positive rate of 0:2%; in the second case, where occluded footballs were considered too, we had a detection rate of 76:2% with a false positive rate of 2:6% (see Figure 5)

Figure 4

V. Conclusions

I have presented a very innovative use of the SVM to solve an old historic football problem. Where traditional computer vision algorithms failed, SVM provides a solid solution after applying some complexity reduction assumptions on the original problem.

Ancona et all aimed at creating a kind of electronic linesman who helps the referee to establish the occurrence of a goal. In fact, they success of their approach relies very much on the outrageous procedure they established to carefully select the negative training samples.

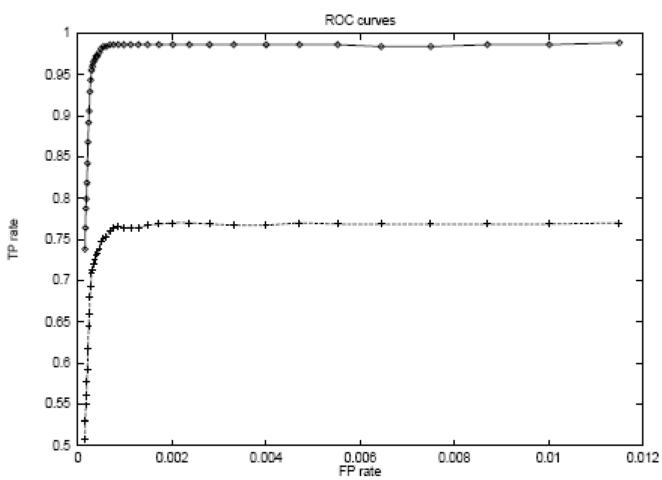

Figure 5

References

[1] Julian Shea, Pride of the three lions. BBC Sport, available at http://news.bbc.co.uk/sport2/hi/football/world_cup_2006/teams/england/5003426.stm

[2] T. Poggio and F. Girosi. A theory of networks for approximation and learning. Technical Report A.I. Memo No. 1140, Arti_cial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[3] V. Vapnik. The Nature of Statistical Learning Theory. Springer Verlag, 1995

[4] Ian Reid and Andrew Zisserman. Goal-directed video metrology. In 4th European Conference on Computer Vision '96, Cambridge, April 1996

[5] N. Ancona, G. Cicirelli, A. Branca and A. Distante, *Ball recognition in images for detecting goals in football.* Proceedings of ANNIE 2001, St. Louis, Missouri, USA, November 4-7, 2001