Lattice-based Anomaly Rectification for Sport Video Annotation

Aftab Khan, David Windridge, Teofilo de Campos, Josef Kittler, William Christmas
CVSSP, University of Surrey, Guildford, United Kingdom
{*Aftab.Khan, D.Windridge, T.Decampos, J.Kittler, W.Christmas*}*@surrey.ac.uk*

***Abstract*—Anomaly detection has received much attention within the literature as a means of determining, in an unsupervised manner, whether a learning domain has changed in a fundamental way. This may require continuous adaptive learning to be abandoned and a new learning process initiated in the new domain. A related problem is that of anomaly *rectification*; the adaptation of the existing learning mechanism to the change of domain. As a concrete instantiation of this notion, the current paper investigates a novel lattice-based HMM induction strategy for arbitrary court-game environments. We test (in real and simulated domains) the ability of the method to adapt to a change of rule structures going from tennis singles to tennis doubles. Our long term aim is to build a generic system for transferring game-rule inferences.**

I. INTRODUCTION

There is a well-established requirement for treating anomalies in machine learning. Artificial cognitive systems, in particular, should be able to autonomously extend capabilities to accommodate anomalous input as a matter of course (humans are known to be able to establish novel categories from single instances [5]). Typically, the anomaly detection problem is one of distinguishing novel (but meaningful) input from misclassification error within existing models i.e. by defining a new learning domain. By extension, the *treatment* of anomalies so determined typically involves the attribution of suitable class designators to the novel input, along with an appropriate method for extending (i.e. generalizing) this categorization. The composite system should thus be capable of inferring novel representations — 'bootstrapping' — via the interaction between the bottom-up processes of anomaly detection and the top-down processes of novel object categorization. Such composite techniques have been applied, for example, to the problem of segmentation [2]. Often, bottom-up description will also explicitly consider *context*, rather than specific objects of classification interest as means of generating high-level domain description [4], [3].

In this paper we consider anomaly rectification in the context of sporting events, focusing on Markovian modeling of anomalous high-level (i.e. abstract, rule-like) state transitions such that the inference system must detect *how* the rules of game-play should change. As a test-bed for this idea, we start with a system trained on 'singles' tennis matches, and then change the input material for doubles tennis matches. On the assumption that a suitable detection system has already flagged the gameplay anomaly and collected suitable quantities of data in the newly defined domain, the problem then is to adapt the existing rule structure accordingly. We define our approach in terms of observed state transition probabilities defined in the two different rule domains, initially testing the method on simulated state transition data and later on testing on real data deriving from an existing system that employs court line detection, homography, player/serve detection, and ball detection via tracklet propagation for singles tennis annotation [1], [6].

In the following section we discuss the problem formulation, with bootstrapping mechanisms described in Section 3. An experimental validation on real and simulated data is given in Section 4; Section 5 concludes.

II. ANOMALY RECTIFICATION

The tennis annotation system described in [1], [6] does not identify individual players, so that scoring is primarily determined via ball movements with respect to designated play areas; these are the play area (PA), near play area (NPA), far play area (FPA), ball out area (BO, subdivided into BO1, BO2) and near/far serve areas (NSA, FSA). Each of these areas is associated with a 4-tuple box designation, b, given in terms of the ordered set of horizontal and vertical screen lines, $H = \{(h_1, h_2, \ldots h_{n_h})\}$ and $V = \{(v_1, v_2, \ldots v_{n_v})\}$. Thus $b \in \{(h_\alpha, v_\alpha, h_\beta, v_\beta)\}$, with $h_\alpha, h_\beta \in H$ and $v_\alpha, v_\beta \in V$. Applying the constraint $v_\alpha < v_\beta$ and $h_\alpha < h_\beta$, each box has a unique b designated in terms of its bottom left and top right corner coordinates; i.e. (h_α, v_α) and (h_β, v_β). The complete set of boxes $\{b\}$ forms a *lattice*, having *joins* (also known as least upper bounds or suprema) and *meets* (also known as greatest lower bounds or infima) analogous to intersection, union and complementation etc in set theory[1]. This allows for complex relationships between designated play areas, e.g. overlaps and subset relations (such as PA/FPA=NPA). This notion of a lattice clearly generalizes to any other rectilinear court structures such as those of badminton (and indeed our method as a whole is intended to generalize to any such domain, so that learning-transfer is possible between superficially different game types). From this perspective, the distinction between singles and doubles tennis is characterized by a change in definition of the play area (PA); $(PA \rightarrow b_o) \rightarrow (PA \rightarrow b_n)$ with $b_o, b_n \in \{(h_\alpha, v_\alpha, h_\beta, v_\beta)\}$. As a step towards a fully general sport-rule annotation induction system capable of transferring learning from one domain to another, our aim is to detect this transition and thereby identify both the old and new play area definitions i.e. b_o and b_n.

[1]A set equipped with a partial order relation for all elements is automatically a lattice if this relation is reflexive, antisymmetric and transitive.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.1063

IEEE computer society

This situation is made inherently complex by the fact that ball state transitions in terms of which the high-level game description is given (e.g. $SA \to PA \to BO$ for a typical serve) are not directly observed. Instead, we see only transitions in the occupancies of the various boxes within the lattice (e.g. $b_1 \to b_2 \to b_3$) to which they correspond, such that the high-level state space can be regarded as the *hidden* states of a Hidden Markov Model (HMM). Moreover (making this analogy exact), we find that the transition structure is inherently ambiguous within the observable state space because of the possibilities of inclusion and intersection within the lattice. We will thus in general have a large set of box transitions within the lattice (e.g.$\{(b_1 \to b_2 \to b_3), (b_5 \to b_8 \to b_3), \ldots\}$) consistent with any given sequence of key play areas. The task of determining which high-level play area has undergone re-definition in the transition from singles to doubles gameplay requires that we obtain a method for treating this ambiguity. We do this via a Minimum Description Length (MDL)-like approach in which we favor the smallest parametric change (e.g. the single transformation $(key\_area \to b_A) \to (key\_area \to b_B)$ required to bring-about the appropriate high-level re-description of the game mechanics (key areas being the main rule-designated areas of play). Note that key area transitions in an arbitrary court game can be between any boxes of *any* size, for instance a transition that goes from the serve area (SA) to the far play area (FPA) is generally a transition from an area of 1 'court unit' to an area of several court units in size (if a court unit is the smallest delineatable region defined by the court lines).

Consistent with a fully-unsupervised approach, we will initially assume no prior knowledge of the injective mapping $\mathcal{P} \to \{b\}$ where $\mathcal{P}$ is the set of play areas $\mathcal{P} = \{PA, NPA, FPA, NSA, FSA, BO1, BO2\}$ and $b \in \{(h_\alpha, v_\alpha, h_\beta, v_\beta)\}$ (i.e. we will not assume knowledge of even the initial single play areas). However, for the purposes of experimental application, we will later relax this assumption in order to recast the approach as one of learning transfer (which can be treated as a subset of the above problem).

III. Methodology

We assume that game play can be modeled via an HMM in which the hidden states are the rule-designated play areas $\mathcal{P}$ and the emission states are the least elements of the lattice $\mathcal{P}$ (i.e. the 'smallest' indivisible boxes of the court such that $b \in \{(h_\alpha, v_\alpha, h_{\alpha+1}, v_{\alpha+1})\}$. The game play is thus described by key points of the ball's trajectory (serves, hits and bounces) which are described by the system as having occurred at a particular time within one of these 'small' (i.e. indivisible) court units. An HMM-based game-play description of this kind is sufficient to enable the existing hardwired tennis annotation system to provide accurate score annotation of singles games.

Within such a Markovian framework we consequently assume that there exists a transition probability matrix $M_P(\mathcal{P}_{in}, \mathcal{P}_{out})$ describing the probability of transition between key play areas. In particular, this matrix is sufficient to capture the notion that a certain fraction of the serves will be returned, with the remainder resulting in either a point award (i.e. $FPA \to BO$) or an 'out ball' (i.e. $SA \to BO$). The returned balls will either go out, be awarded a point or enter into a further rally recursion, with some particular probability captured by the matrix M_P.

In addition to this matrix, game-play characterization also requires the injective mapping $f(\mathcal{P}) \to \{b\}$ that gives the actual *definitions* of the play areas (f is thus the mapping between the key-area labels and the corresponding boxes within the lattice). Consequently, the transition from singles to doubles tennis gameplay may be characterized by a transition from this mapping to some other specific mapping i.e. $f \to f'$ (i.e we assume that the basic gameplay structure remains the same in terms of the key-area transition probabilities, with only the mapping into the lattice undergoing change) . In our later simulation of the single to doubles transition, only a single element of this mapping (relating to the play area) will undergo change: i.e. $(PA \to b_o) \to (PA \to b_n)$ where b_o is the *old play area* and b_n is the *new play area*, with $b_o, b_n \in \{(h_\alpha, v_\alpha, h_\beta, v_\beta)\}$.

However, the only evidence for this transition in the definition of play area (PA) that we are presented with is in terms of the *observed* matrix of box transitions defined over the entire lattice $M(b_1 \in \{b\}, b_2 \in \{b\})$ (note that M is the histogram of lattice transitions for a given set of play sequences, rather than a true row-normalized transition matrix like M_P). A change to the transition matrix M_P can thus only be detected by compiling multiple observations in the differing domains (resulting in e.g. a singles transition matrix, M^s, and a doubles transition matrix, M^d). However, even with sufficient sampling of M^s, we do not directly know which play-area box-mapping has undergone transition, since in general a large fraction of other boxes in the lattice will experience correlated activity as a result of the transition. In order to determine precisely which key area redefinition has taken place our first goal is thus to determine the matrix transform, T, parameterized by the key play-area transform, $(b_o \to b_n)$, that brings about M^d i.e. we require a T such that $T(M^s, b_o, b_n) = M^d$.

Without further analysis, it is not clear *a priori* that this is a well-posed problem, in the sense that the transform may be non invertible if the resultant matrix $M^d = T(M^s, b_o, b_n)$ loses information about the individual lattice components b_o, b_n. In addition to this difficulty, we also have the potentially inadequate sampling of the probabilities in the underlying Markovian play-area transitions of M_P manifested in M^s and M^d. We therefore seek instead to minimize the *residual* of the parameterized transform $T(M, b_A, b_B)$ with respect to M^d, rather than directly inverting it:

$$(b_o^{est}, b_n^{est}) = \underset{(b_A, b_B)}{\operatorname{argmin}} \left[D(T(M^s, b_A, b_B), M^d) \right] \quad (1)$$

where D is an appropriate distance measure (see below); the superscript est denotes the estimated lattice value. The transform T, itself, is derived as follows:

The aggregate 'ball-event activity' associated with any given box b in the lattice can be separated into 'into', $M(.,b)$, and 'out-of', $M(b,.)$ transition components. We can also define a coarse aggregate activity measure $A(x)$ by summing over all of the observed transitions *into* and *out of* the box x for every single box within the lattice

$$A(x) = \Sigma_{m=1}^{|b|}(M(x,m) + M(x,b)) \tag{2}$$

The rationale for doing so is we can thereby obtain an approximate means for estimating the effect of redefining a key area (e.g. $(PA \to b_o) \to (PA \to b_n)$) by *translating* the activity associated with a box b_o to b_n: i.e. such that $A^{new}(b_n) = A(b_o)$. However, it is not simply the case that we can transfer activity in this way without also explicitly considering interactions *within* the lattice structure.

A measure of this lattice interaction can be defined in terms of the proportional overlap of one box with respect to another. The expectation of the coarse activity measure A in box b_1 due to activity in box b_2 for uniformly distributed ball events is thus:

$$E[A(b_1|b_2)] = \frac{|b_1 \cap b_2|}{|b_2|}.A(b_2),\ |b| = (h_\alpha - h_\beta)(v_\alpha - v_\beta)$$

This is also true for both the 'into' and 'out of' of activity components (thus, for example, given an isolated 'into' component $M(.,b)$, we expect a second, potentially overlapping, box b' to have an 'into' component $\frac{|b' \cap b|}{|b|}M(.,b)$). Consequently, to a first order of approximation, the play area redefinition $(PA \to b_o) \to (PA \to b_n)$ has the effect on the matrix M of subtracting a '*lattice interaction*' matrix, M_{sub}, that removes activity attributable to box b_o, while adding another lattice interaction matrix, M_{add}, that displaces this activity to box b_n. Hence:

$$M_{add}^{(o,n)}(x,y) = M(x,b_o)E[A(y|b_n)] + M(b_o,y)E[A(x|b_n)]$$
$$M_{sub}^{(o,n)}(x,y) = M(x,b_o)E[A(y|b_o)] + M(b_o,y)E[A(x|b_o)]$$

That is, we obtain M_{add} and M_{sub} by multiplying all 'into' and 'out of' transitions of the box in question by the expected overlap of activity. To the first order, the transform T can, thus, be approximated by:

$$T(M, b_o, b_n) = M(.,.) + M_{add}^{(o,n)}(.,.) - M_{sub}^{(o,n)}(.,.)$$

However, this does not take into account the fact that activity in b_o and b_n have a certain likelihood of influencing each other at the outset; i.e. we cannot say that *all* of the activity in M attributable to b_o should be transferred to b_n. Moreover, we cannot say that all activity in b_o is attributable *specifically* to b_o; it could equally apply to an intersecting box. We therefore introduce a free parameter representing the appropriate proportion of activity to transfer for inclusion within the optimization i.e. we specify:

$$T(M, b_A, b_B, \gamma) = M(.,.) + \gamma(M_{add}^{(A,B)}(.,.) - M_{sub}^{(A,B)}(.,.))$$

such that the optimization function becomes:

$$(b_o^{est}, b_n^{est}) = \underset{b_A,b_B}{\operatorname{argmin}}\left[\underset{\gamma}{\operatorname{argmin}}\ D(T(M^s, b_A, b_B, \gamma), M^d))\right] \tag{3}$$

The ready optimisability of the above equation lies in the fact that the matrices M^s and M^d are essentially sparse when the effects of lattice interaction are removed from consideration, with occupancy dictated by the size of the game-play transition matrix, $M_P(\mathcal{P},\mathcal{P})$ (i.e. $\mathcal{P} \times \mathcal{P}$), rather than the size of the lattice transition matrix, $|b| \times |b|$. We can thus regard M as a convolution of the individual components (g_x^i, g_y^i) of $M_P(f(\mathcal{P}), f(\mathcal{P}))$ with an activity 'point-spread function' $E[A(x|g_x^i)].\delta(y - g_y^i) + E[A(y|g_y^i)].\delta(x - g_x^i)$.

The full optimization function for the transform, using an activity-normalized RMS (root mean square) residual difference measure, is thus:

$$(b_o^{est}, b_n^{est}) = \underset{b_A,b_B}{\operatorname{argmin}}\left[\underset{\gamma}{\operatorname{argmin}}\ RMS((M^d - (M^s + \gamma(M_{add}^{(A,B)} - M_{sub}^{(A,B)}))) \circ M_{\text{norm}})\right] \tag{4}$$

where the normalization matrix is defined $M_{\text{norm}}(a,b) = \left(\frac{|h_\alpha^a - h_\beta^a|.|v_\alpha^a - v_\beta^a|.|h_\alpha^b - h_\beta^b|.|v_\alpha^b - v_\beta^b|}{(|h_1 - h_{n_h}|.|v_1 - v_{v_h}|)^2}\right)^{-1}$.
($\circ$ is the Hadamard product).

The above optimization is still based on finding a single, optimal substitution $b_o \to b_n$; however, denoting this optimization $O(p_o^{est}, p_n^{est})$, it can be seen that a fully general optimization function for arbitrary matrix transforms, O_{gen}, can be obtained by concatenating sequences of individual box redefinitions:

$$O_{gen}(M_1, M_2) = \Sigma_{b=1}^{m} O(b_{2m-1}, b_{2m}) \tag{5}$$

However, in this case it is necessary to balance the allocation of parametric freedom (essentially governed by $|m|$) with the cumulative RMS residuals. This requires an empirical cross-validation or *a priori* MDL-like criterion to accomplish. Such a generalization of the current approach could potentially transfer learning from tennis to badminton, since much of the serve/return game-play structure is consistent between the two, with only the court area definitions differing between them.

IV. Experimental results

We simulate a simplified tennis game by choosing $M_P(\mathcal{P},\mathcal{P})$ with the following transition probabilities (we omit NPA and FPA transitions to give a single-box lattice-transformation problem):

$p(NSA \to PA) = .9,\ p(NSA \to BO1) = .1,\ p(PA \to PA) = .2$
$p(PA \to BO1) = .7,\ p(PA \to BO2) = .1,\ p(others) = 0$

i.e. we capture the possibility that a serve may or may not be returned; a low rally probability is also included. We also have the following ordered 4-tuple play area definitions for simulated 'doubles' play (omitting center lines for simplicity):

$NSA \to (1,1,3,2),\ PA \to (2,2,5,5)$
$BO1 \to (1,5,6,6),\ BO2 \to (1,1,6,2)$

For 100 simulated serves this generates the lattice transition matrix depicted in Figure 1 (left).

Singles play is simulated (in this simplified scenario) by changing the PA key area description to $(PA = (3,2,4,5))$ and keeping all the remaining values. This represents the

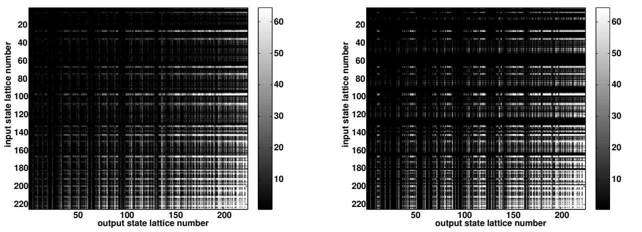

Figure 1. *Gray-scale histogram of Singles (M^s, left) and Doubles (M^d, right) transition counts over the lattice (ordered by box size and count-number, respectively).*

fact that the 'tram-lines' are no longer part of the legitimate play area, so that any ball bouncing in this area is not automatically out. The resulting lattice transition matrix depicted for observations of 100 simulated serves is depicted in Figure 1 (right).

Carrying out the optimization in Equation 4 by considering all possible transitions $(PA \to b_o^{est}) \to (PA \to b_n^{est})$ and iterating over γ, we obtain an estimate of this gameplay area redefinition. (Note that for the transfer learning problem, we need only consider the redefinitions of *known* play areas such that the search space is of size $|\mathcal{P}|$ rather than $|b|$, i.e. $b \in f(\mathcal{P})$).

A general performance metric for proposed play-area redefinitions of this type can be obtained by taking the total ordinal difference between proposed and actual transitions. Thus, for a 'ground-truth' box redefinition:

$PA(h_1^a, v_1^a, h_2^a, v_2^a) \to PA(h_3^a, v_3^a, h_4^a, v_4^a)$

and a proposed box redefinition supplied by the optimization method:

$PA(h_1^p, v_1^p, h_2^p, v_2^p) \to PA(h_3^p, v_3^p, h_4^p, v_4^p)$

We have:

$Error = \left(\frac{1}{max(Error)}\right) \Sigma_{x=1}^{4} |h_x^a - h_x^p| + |v_x^a - v_x^p|$

Figure 2 (left) thus gives the resulting average prediction error for a given number, x, of complete Markov chains obtained by Gibbs sampling of the indicated singles and doubles play area transition matrices (with error bars given by the standard error of mean determined from 20 samples). It may be observed that $x \approx 10$ complete gameplay sequences is sufficient to identify the play area redefinition involved in transiting from singles to doubles for the specified game parameters.

We also test on real data derived from the Toray Pan Pacific Open 2009 womens singles match between M. Rybarikova and A. Radwanska with a total of 58 Playshots with 58 Serves, giving 343 events in total (excluding hits) and 285 Bounces. Doubles play is simulated by multiplying the baseline (x-axis) by 1.33 (centralized at the court centre) so as to extend the legal play into the tram-lines. We fold the court along it's symmetric x and y axes around the court center to provide better statistical sampling (generating a lattice of 36 elements) and also introduce a weighting proportional to the physical size of of court box to ensure the validity of the 'within box' uniform distribution assumption as far as possible. For this, we use the following horizontal and vertical ordinate values:

horzontalLineSet = [40, 100, 127, 208, 289, 316, 376];
verticalLineSet = [-10, 50, 158, 284, 410, 518, 578];

In the above experiment the method returns the estimated transform (in the folded coordinate system):

$(PA \to (2,3,4,4)) \to (PA \to (2,2,4,4))$

That is, the system has correctly identified the original play area and made a correct identification of its redefinition (differing in no ordinate values); a residual graph is given in figure 2 (right). The method is hence sufficiently robust to accommodate any systematic deviations from uniformity in the play sequence distribution.

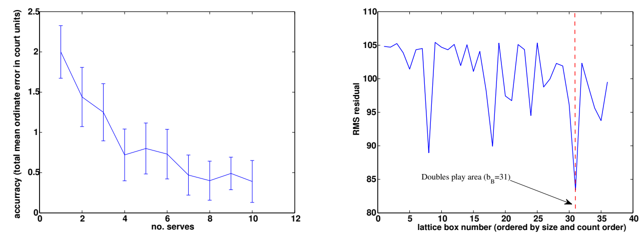

Figure 2. *Left: Mean prediction error of simulated game area transformation for a given number of complete singles/doubles serve sequences (x-axis), Right: RMS residual over b_B for b_A =play area.*

V. Discussion and Conclusions

We set out, within the context of sport video annotation, to address the problem of anomaly *rectification*; the adaptation of an existing learning mechanism to a change of domain. Consequently, we proposed a novel HMM induction strategy tuned for court-game environments that maps 'hidden' gameplay states into a court lattice using a deconvolution-like strategy. The system was able to correctly determine transitions in the definition of a play area on both real and simulated data. When extended to accommodate arbitrary numbers of play-area redefinitions, it is intended that the system will be coupled with an anomaly detector in order to build a generic system for sport annotation and learning transfer. **Acknowledgement:** The research leading to these results has received funding from the EC 7^{th} Framework Programme FP7/2007-2013 under grant agreement n^o 215078. We also gratefully acknowledge the support of EPSRC through grant EP/F069626/1.

References

[1] J. Kittler, W J Christmas, F Yan, I Kolonias, and D Windridge. A memory architecture and contextual reasoning for cognitive vision. In *Proc. SCIA*, pages 343–358, 2005.

[2] Bastian Leibe, Aleš Leonardis, and Bernt Schiele. Robust object detection with interleaved categorization and segmentation. *IJCV*, 77(1):259–289, May 2008.

[3] Y. Liu and K.N. Ngan. Embedded wavelet packet object-based image coding based on context classification and quadtree ordering. 21(2):143–155, February 2006.

[4] S. H. K. Parthasarathi, B. Motlicek., and H. Hermansky. Exploiting contextual information for speech/non-speech detection. In *LNCS - TSD*, volume 5246, pages 451–459, 2008.

[5] T. Tommasi and B. Caputo. The more you know,the less you learn: from knowledge transfer to one-shot learning of object categories. In *Proc. BMVC*, 2009.

[6] F. Yan, W. J. Christmas, and J. Kittler. Layered data association using graph-theoretic formulation with application to tennis ball tracking in monocular sequences. *Trans. PAMI*, 2008.