Discovering reoccurring motifs to predict opponent behavior

Auke Wiggers
Informatics Institute
University of Amsterdam

Arnoud Visser
Informatics Institute
University of Amsterdam

Abstract— **In contrast to human soccer players, autonomous robot soccer players often move according to a limited set of predefined behavioral rules. This knowledge can be used advantageously: If the opponent's behavioral rules are learned, it will be possible to detect these during a match and react accordingly. A method for autonomous activity mining in videos, called Probabilistic Latent Sequential Motifs, is used to discover optical flow patterns in videos of a robot soccer player during a penalty shootout. The discovered patterns are used by a humanoid goalkeeper to predict and anticipate opponent behavior. Effectiveness of the method is tested by comparing the performance of this goalkeeper with predictive behavior to that of an existing goalkeeper that only reacts when the ball approaches at sufficient speed. The performance is measured based on the ratio of number of goals to number of goals prevented. Results show that the goalkeeper with predictive behavior could prevent a fair amount of goals, but that it loses in performance to the existing goalkeeper. Methods that may improve performance are discussed.**

I. INTRODUCTION

RoboCup is an international research and education initiative, attempting to foster Artificial Intelligence and robotics research by providing a standard problem where a wide range of technologies can be integrated and examined. One of these problems is teaching a robot to play soccer. At the RoboCup, typically a rule-based approach is applied [1]. During matches, a reaction from the goalkeeping robot is triggered by the perception of a ball which approaches at sufficient speed. Other features in the scene, such as presence of opponent players, could be relevant but are not taken into account to keep the decision rules simple.

A more flexible solution enables the goalkeeper to find the preconditions for these logical rules autonomously and in an unsupervised manner. A method for unsupervised activity mining in videos called Probabilistic Latent Sequential Motifs (PLSM) is introduced by Varadarajan et al [2]. It uses of topic models to find temporal activity patterns in video data. Each of these patterns represents an activity in a video, and relates to the video as combinations of syllabi would relate to a word. A model that describes recurring patterns in a dataset can be discovered offline. The detection of these patterns during a match will enable a goalkeeper to anticipate the opponent's behavior. The effectiveness of the method is tested by comparing performance in a penalty shootout between a regular rule-based goalkeeper and the goalkeeper that predicts to the opponent's behavior using patterns discovered by using PLSM.

II. RELATED WORK

Activity mining is a field of research that is often associated with surveillance scenarios [3], e.g., to detect violent behavior in a crowd or to analyze busy traffic scenes. In activity mining, the use of topic models [4] is an approach that has proven to be quite successful [5], [6]. This model is a statistical type that enables discovery of so-called 'topics', which are abstract occurrences in a document. Often, topic model-based approaches first convert a video to documents containing bags-of-words, where each word is a representation for quantized pixel motion at certain locations in the image [2], [5], [6], [7]. Note that the 'words', the visual features, are only a tag corresponding to the correct optical flow vector, and do not contain the vector itself. One document is thus the result of analysis of optical flow data in one video. An activity pattern in a video can be represented as a set of relative movement vectors along with their respective starting positions and starting times. These activity patterns are the topics that are to be learned. To remain faithful to the terminology used by Varadarajan et al, these topics will be referred to as 'motifs' in this paper.

When a regular bag-of-words approach is used, the temporal information is lost in the process [7]. To take this information into account as well, methods have been proposed where both the motifs and their starting times are jointly learned by complementing the visual features with their respective timestamps [2], [5]. In a method introduced by Emonet et al [5] both the total number of motifs and the motifs themselves are learned in an unsupervised manner using the Hierarchical Dirichlet Process (HDP) [8], which allows for an infinite amount of motifs on multiple levels. In the problem addressed in this paper, two levels can be learned: On the first, the total number of motifs that is shared by all documents; on the second, the individual motifs and their starting times in each document. The PLSM method can only be applied to the second level, the method requires a priori knowledge (e.g., the number of motifs) for the first level. Although the HDP-based approach may be more complete than PLSM, it would seem that the latter is sufficient to solve the problem addressed in this paper: The number of strategies for robot soccer may be theoretically infinite, in practice; it is always limited to a certain number of re-occurring actions. For example: Shooting at the goal can be described as a motion in an infinite number of directions and can occur on every position on the field. However, it only occurs when at

least one robot is moving and there are only as many possible shooting directions as there are quantization categories. If the number of possible activities is learned in an automatic manner, it is likely that this number will be too high to be useful. Instead, we will use a small number of motifs and let PLSM determine what they are.

PLSM has been used in pedestrian and traffic analysis [2], in scenarios in which the camera is stationary, to find recurring activities in video data. In this paper, PLSM will be applied in a new area. Instead of a camera placed above the scene, all video data will come from the goalkeeper's point of view.

III. APPROACH

Several steps are performed to solve the addressed problem. The first step is the creation of temporal documents, derived from video data. From these documents, a model that describes the motifs can be derived using PLSM [2]. Next, a reaction that enables the goalkeeper to anticipate the opponent's activity has to be created. The final step is the prediction of an opponent robot's activity pattern, based on detection of the first frames of a learned motif.

A. Probabilistic Latent Sequential Motifs

To find recurring activity patterns, a temporal document d, that represents the image data as a bag-of-words, has to be derived. This document is of size $V \times T$, where V is the vocabulary size and T is the number of timesteps the document covers. At each timestep, the document d is filled based on the presence of visual words w (See Fig. 1). PLSM is applied on this document to find motifs z, and their starting times t_s.

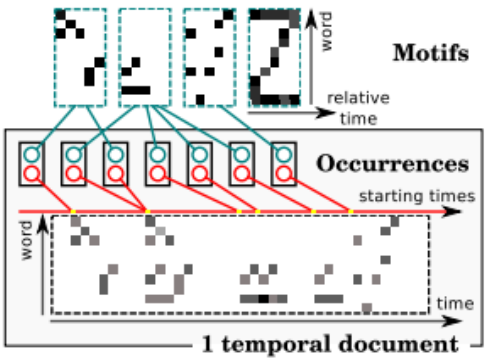


Fig. 1: Graphical representation of a temporal document. Image courtesy of Emonet et al [5].

The main assumption of the PLSM model is that given a motif z, the occurrence of words within the document is independent of the time of occurrence. Note that there is a deterministic relation between the time variables $t_a = t_s + t_r$, enabling the use of variable t_r to denote the relative time since the start of a motif: In this model, occurrence of a word depends only on the motif and the time it occurs in the topic, not on the absolute time of occurrence t_a. The joint distribution of the model is given by:

$$p(w, t_a, d, z, t_s) = p(d)\, p(z|d)\, p(t_s|z, d)\, p(w|z) \dots \quad (1)$$
$$p(t_a - t_s|w, z) \quad (2)$$

The Expectation-Maximization algorithm can be used to estimate the set of model parameters Θ, by maximizing the log-likelihood of the model for the observed data. It is an iterative algorithm that is initialized using random values as parameters, and is stopped when the increase in log-likelihood is too small. The log-likelihood is given by:

$$E[L] = \sum_{d=0}^{D} \sum_{w=0}^{N_w} \sum_{t_a=0}^{T_d} \sum_{z=0}^{N_w} \sum_{t_s=0}^{T_{ds}} n(w, t_a, d) \dots \quad (3)$$
$$p(z, t_s|w, t_a, d)\ \log p(w, t_a, d, z, t_s) \quad (4)$$

where the normalized correlation $n(w, t_a, d)$ is the output of the generative model as described in next section (equation (13)).

The posterior distribution of variables t_s and z is calculated in the E-step of the Expectation-Maximization algorithm, given w, the absolute time t_a and the document d:

$$p(z, t_s|w, t_a, d) = \frac{p(w, t_a, d, z, t_s)}{p(w, t_a, d)} \quad (5)$$

$$\text{with } p(w, t_a, d) = \sum_{z=1}^{N_w} \sum_{t_s=1}^{T_{ds}} p(w, t_a, d, z, t_s) \quad (6)$$

Next, in the M-step of the Expectation-Maximization algorithm, Θ is updated accordingly:

$$p(z|d) \propto \sum_{t_s=1}^{T_{ds}} \sum_{t_r=0}^{T_z-1} \sum_{w=1}^{N_w} n(w, t_s + t_r, d) \dots$$
$$p(z, t_s|w, t_s + t_r, d) \quad (7)$$
$$p(t_s|z, d) \propto \sum_{w=1}^{N_w} \sum_{t_r=0}^{T_z-1} n(w, t_s + t_r, d) \dots$$
$$p(z, t_s|w, t_s + t_r, d) \quad (8)$$
$$p_w(w|z) \propto \sum_{d=1}^{D} \sum_{t_s=1}^{T_s} \sum_{t_r=0}^{T_z-1} n(w, t_s + t_r, d) \dots$$
$$p(z, t_s|w, t_s + t_r, d) \quad (9)$$
$$p_{t_r}(t_r|w, z) \propto \sum_{d=1}^{D} \sum_{t_s=0}^{T_{ds}} n(w, t_s + t_r, d) \dots$$
$$p(z, t_s|w, t_s + t_r, d) \quad (10)$$

The motifs and their starting times can be discovered in the optimized distributions $p(z|d)$ and $p(t_s|z, d)$.

B. From images to documents

The content of temporal documents that will serve as input for PLSM depends on how the visual words w that form the vocabulary are defined. These words will be based on optical flow in the image, so first, optical flow is computed in areas of interest, namely the areas containing the ball and opponent robot (See Fig. 2). If we set the robot so its camera is stationary, the majority of the noise due to camera motion will be eliminated. The resulting optical flow vectors are quantized into four general directions (up, down, left and right) or marked as static if the norm is not sufficiently large. Based on their respective starting locations in the image of 640 $\times$ 480, they are also quantized into 64 $\times$ 48 non-overlapping cells of 10 $\times$ 10 pixels. It is possible to use these categorized vectors and their timestamp as low-level features in the temporal document. However, the resulting documents are quite large, as the vocabulary would consist of $64 \times 48 \times 5 = 15360$ words.

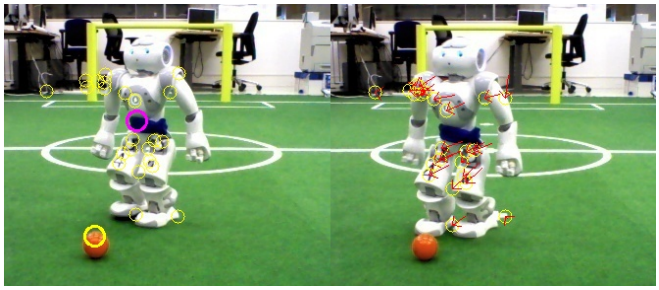


Fig. 2: Left: Selected features in frame f are indicated by thin yellow circles. Only features in the area of interest are considered, which is determined by location of the ball (thick yellow circle) and waistband (thick purple circle). Right: Optical flow vectors between frame f and frame $f+1$, indicated by red arrows.

Instead, Probabilistic Latent Semantic Analysis (PLSA) [9], a dimensionality reduction method that makes use of latent classes, will be used to reduce the size of this vocabulary. This method was used by Varadarajan et al as well, in order to reduce computation time [2]. Its generative model, describing how each variable in the distribution can be sampled, is given by Fig. 3. PLSA models the probability of co-occurrence of words and documents as a mixture $p(\omega, d)$, creating the assumption that the occurrence of a word ω is independent of the video document d it belongs to, given a latent class c:

$$p(\omega, d) = \sum_c p(c)p(d|c)p(\omega|c) \tag{11}$$

The documents d are defined as word count matrices of size $f \times V$ extracted from overlapping clips of f frames. The parameters of the model $P(\omega|z), P(c)$ and $P(d|c)$ are estimated using the maximum likelihood principle. Given a set of training documents $\mathcal{D}$, the log-likelihood of the parameters Θ is given by:

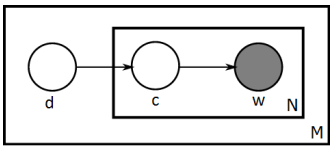


Fig. 3: The generative model of PLSA.

$$\mathcal{L}(\Theta|\mathcal{D}) = \sum_{d \in \mathcal{D}} \sum_{\omega} n(d, \omega)\, log(p(\omega|d)) \tag{12}$$

Similar to the PLSM model, the parameters are optimized using the EM algorithm [10], which allows to learn distributions $p(\omega|c)$ for every latent class. The found distributions are then used to define words w for PLSM, meaning that $w = c$ and $N_w = N_c$. By specifying the desired number of latent classes N_c beforehand, we are able to reduce the size of the vocabulary to any given number. To keep the runtime low, $N_c = 25$ was chosen. The presence of each word w at an absolute time t_a in a document d is then defined as the normalized correlation:

$$n(d, t_a, w) = \frac{1}{\sum_{\omega \in W_c} n(d, \omega)} \sum_{\omega} n(d, \omega) p(\omega|c) \tag{13}$$

In this equation, W_c indicates the set of all words in the distribution $p(\omega|c)$ with a non-zero probability. This ensures us that the found presence of each word w is independent of activities elsewhere in the scene.

A resulting temporal document, that is to be used by PLSM, is of size $N_w \times T$, where N_w is the vocabulary size and T the range of absolute times (see Fig. 1). The documents are used as input for the PLSM method. As stated before, each word w in the vocabulary is a representation of a temporal pattern from the distribution $p(\omega|c)$, as found by PLSA.

C. Prediction task: anticipation

The final step is detection of activities, and reacting if one has been detected. The optical flow is calculated in areas containing the ball and the waistband of the opponent robot at one frame per second. The resulting vectors are quantized and converted to the bag-of-words representation. The found PLSA model is used to calculate presence for every latent class c, resulting in a document of exactly one timestep. The probability of d_{new} being a part of a motif is calculated as follows, with $0 \leq \alpha \leq 1$:

$$\begin{aligned} p(z, t_n|d_{\text{new}}) &= (1-\alpha) * p(z, t_{n-1}|d_{\text{new}}) \ldots \\ &\quad + \alpha * \sum_{w}^{N_w} n(w, d_{\text{new}}) p(z|w, d_{\text{new}}) \end{aligned} \tag{14}$$

In this equation, the probability that a new document of one timestep is part of a motif z depends on previous probabilities $p(z, t_{1\ldots n-1}|d_{\text{new}})$. If, for several timesteps, this probability is higher than a predefined threshold, we assume that the rest of the activity will be similar to the rest of the motif,

and the goalkeeper will react accordingly. Note that the goalkeeper's reactions are intended to anticipate the end of their respective corresponding motifs: How the reaction is executed is independent of which of the timesteps in the motif is perceived.

Robot soccer behavior is often limited to several different strategies. Therefore, it is highly likely that mirroring will enable us to reduce the number of motifs even further: The reaction to an activity perceived on the right side of the goalkeeper can be mirrored, this will result in a reaction that can be used if the same activity, but mirrored, is perceived on the left side of the goalkeeper. However, in practice, mirroring the patterns results in more motifs, which increases computation time (also see V-B). This renders it useless in real-time applications for the Nao, which only has a single ATOM processor. It will be possible to apply this method for humanoids with greater computing power.

IV. RESULTS

The used robot is the Aldebaran Nao 4.0 [11] which is a humanoid robot with an Intel ATOM 1,6ghz CPU. It has two cameras, capable of taking snapshots of 640×480 pixels. The dataset used for training consists of 2370 images, taken at one frame per second, from one of these cameras. Higher framerates are possible, but 1 frame per second is chosen to guarantee that there is enough difference between the frames. The set of images is split into 43 groups, each of which is between 20 and 128 frames long and describes a single recording of a penalty shootout by one opponent robot, as seen from the goalkeeper's point of view. An average duration of a penalty shootout recording is 55 seconds. PLSA was applied on the optical flow detected in these images as described in subsection III-B, with $N_c = 25$, resulting in 43 temporal documents. PLSM was applied on these documents to find reoccurring motifs. For comparison, activity patterns were discovered in the data for various numbers of motifs N_z and its maximum duration T_z, as activities in the image sequence may vary in length (see Fig. 4).

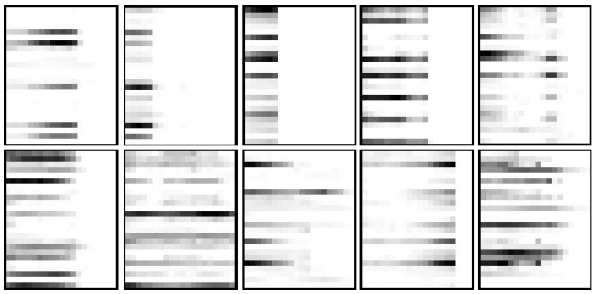
For instance, one can see that the last motif of Fig. 4.b is mainly due to movement detected at the beginning of the activity, while the motif directly left from it is mainly due to movement detected at the end of the activity.

A fitting reaction was created manually for each motif after analyzing the pattern of optical flow that it describes. The reaction consists of at least one action: Either walking, diving or a combination of the two. The walking direction and distance and the diving direction (left or right) were adjusted as well. As the reactions themselves are not part of the training, there is no guarantee that these are the optimal anticipating reactions for a specific motif.

The optical flow vectors that correspond to a motif were found by taking distribution $p(\omega|z)$ and, for each motif, calculating the presence of a word in the scene at a timestep t, given the motif, as:

$$n(\omega, t, z) = p(\omega|z)\, n(\omega, t) \tag{15}$$

(a) Reoccurring temporal motifs for number of motifs $N_z = 5$ and maximum number of timesteps $T_z = 10$.



(b) Reoccurring temporal motifs for number of motifs $N_z = 10$ and maximum number of timesteps $T_z = 20$.

Fig. 4: Graphical representations of two of the reoccurring sets of motifs, for different numbers of motifs N_z and maximum duration T_z. Motifs are of size $T_z \times N_c$, with the number of PLSA patterns $N_c = 25$.

During the anticipation task, $p(z)$ is calculated for real scenes once every second. If, for any motif, this probability exceeds a predefined threshold for five successive seconds, the goalkeeper assumes this motif is perceived and executes the corresponding reaction regardless of what happens after the start of the reaction. The effectiveness of these reactions was compared to performance of a 'regular' goalkeeper, which only reacts if the ball approaches at sufficient speed by diving, through a total of 15 penalty shootouts. So, the 'regular' goalkeeper makes its decision after the ball is shot, the goalkeeper trained with PLSM can anticipate its decision. These shootouts were conducted by a Nao with a basic behavior model: Find the ball, walk towards it, locate the goal and shoot. In theory this behavior will always result in the same movements, in practice there is enough variation in how the robot positions itself behind the ball and how it shoots, due to sensor and actuator noise. Fig. 5 depicts the starting positions of both robots and the ball in a penalty shootout.

We say that a goalkeeper interferes if the ball would have gone into the goal, were it not for the goalkeeper. A miss is when the ball does not reach the goal at all. A hit is, obviously, when the ball passes the keeper, in between the poles of the goal. If the keeper does not react but prevents the hit by standing still, this is counted as a hit. The rationale behind this decision is that standing still was not selected as a fitting reaction for any of the motifs. Interestingly, the goalkeepers that were the result of PLSM did not react in case of a miss, whereas the regular goalkeeper did, although this may be due to chance. There is no time limit for a single shootout: It ends when the player scores, the ball is kicked out of the field (behind the

Fig. 5: Typical start of a penalty shootout. The opposing player (left) starts at the center line, with a distance of 3 meters from the goal, walks towards the ball and kicks it. The goalkeeper starts at the goalline.

goal line), or the goalkeeper has made a save. Details on specifications the field and rules of the penalty shootout can be found in [12]. The results can be seen in Table I:

N_z / T_z	Hit	Miss	Goalkeeper interferes
5 / 10	8	3	4
5 / 20	9	4	2
10 / 10	9	3	3
10 / 20	11	4	0

TABLE I: The interception results for a trained goalkeeper.

From these results, we can conclude that the most effective goalkeeper that was reacting on activity patterns discovered using PLSM uses a small number of motifs. A likely cause for this is that PLSM is forced to map a large number of reoccuring patterns to a limited and predefined number of motifs: The smaller the number of motifs, the more general they will be. Results also indicate that effectiveness is increased if T_z is low. This seems counterintuitive, as the longer motif contains more timesteps that could match the new document. However, the activities that are represented by motifs of $T_z = 10$ are forced to represent short and concise patterns that match typical recurring actions, whereas the motifs of $T_z = 20$ may represent an entire penalty shootout. An indication for this is that, when shown a real-life scene, $p(z)$ is large for almost every motif, for $T_z = 20$, whereas there is more contrast for $T_z = 20$ (see Fig. 7).

The goalkeepers that were trained using PLSM were compared to the regular goalkeeper model. This model bases its actions on the speed of the ball. If this speed exceeds a certain threshold, the direction is calculated and the keeper dives left or right. The direction is based on the expected location where the ball will cross the goal line. Results are shown in Table II:

Clearly, the results of the regular goalkeeper show that

	Hit	Miss	Goalkeeper interferes
Regular goal-keeper	2	4	9

TABLE II: The interception results for a regular goalkeeper.

the parameters of this behavior are optimized to perform well during soccer games at the yearly RoboCup competition. The results of the goalkeeper trained with PLSM are still poor when compared to those of the regular goalkeeper. This can partly be due to the limited training set of 40 minutes and partly due to the noise and redundant information that each motif contains, as Fig. 6 indicates.

Fig. 6: Two consecutive timesteps of a noisy motif z for $N_z = 5$, $T_z = 10$. Circles indicate cells where optical flow is detected, with the thickness of the circle corresponding to presence $p(w|z)$. As optical flow is detected throughout almost the entire image, this motif is too general to be of use.

The noise may render a motif too general to be useful, as features that do not contribute to the action represented by the motif are included in it as well. Such a motif generally has a high $p(z)$, even if the activity it represents is not perceived. The opposite can also be true: Fig. 7 shows us that the fourth motif is, compared to the other motifs, a motif with very low probability of occurring, given the document d_{new} that describes the scene. There are two possible causes: Either the motif is the result of a specific activity that does not occur in the scene, or the motif is the result of noisy optical flow and will therefore always have low probability. In the second case, this motif will not be of any use, which would influence goalkeeper performance.

V. DISCUSSION

In this paper, PLSM is applied in an environment in which an activity can occur at every position on the field, with different orientations. This in contrast to, for example, traffic analysis scenarios, in which cars are always on the road and can move in a limited number of directions. It seems that for this problem, the location cue can be discarded: Optical flow vectors on which the activity patterns are based can be translated and scaled to find the pattern that describes the activity when it is perceived at a different location. If this technique was applied, an action can be represented by optical flow alone. Whether this will improve results, and

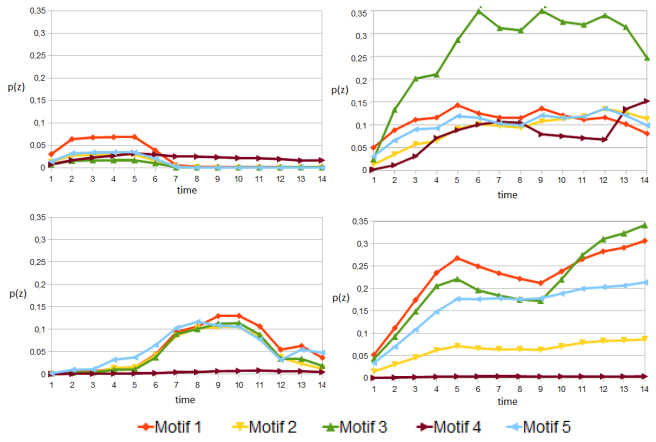


Fig. 7: Probability $p(z)$ for $N_z = 5$, for two identical static scenes (left) and two similar non-static scenes (right), for 14 timesteps. Top: Results for $T_z = 10$. Bottom: Results for $T_z = 20$.

several other factors that are likely to influence results, are discussed in this section.

A. Location quantization and scaling

As mentioned in III-B, the possibility exists that the same activity is perceived at different locations. As a result, a set of perceived patterns, in reality consisting of only one activity but seen at different locations, will not necessarily be mapped to the same motif. One reason for this phenomenon is difference in distance: An opponent robot that is close to the goalkeeper appears larger in the image than when it is far away, and will thus occupy a larger part of the grid. Possible solutions for this problem involve use of a cell grid that uses a logarithmic scale, or dividing the image into larger cells, both of which reduce the number of possible features. However, each activity pattern represents a possible activity of an opponent robot, and these activities affect the game differently if they occur on unusual positions on the field. Therefore, it is impossible to group patterns that are a result of optical flow detected on an unusual position, with the already learned motifs, even if they describe the same action by the robot player.

For example: When a player shoots the ball to the left of the goalkeeper and is close, the goalkeeper may want to move left as well to defend the goal, while the same action, when perceived at a greater distance, may not pose a threat at all and not require a reaction. A second reason is that the number of features may be reduced, but it is hard to tell whether we can afford to lose these features: This method may cause PLSM to find motifs that are too general to be used.

B. Mirroring

If we assume that a perceived pattern is the result of an action in arbitrary reaction, it should be possible to reduce the number of motifs by mapping patterns that are mirrored versions of each other to a single motif. The disadvantage of this approach is that it increases computation time for the prediction, as each motif that is mirrored requires calculation of $p(z)$. Even if the mirrored versions are mapped to the same motif, it will be necessary to check which of the two versions is perceived before a fitting reaction can be given. Additionally, there may be exceptions to the rule, or the mirrored motif may be too similar to a different motif, causing the goalkeeper to react wrongly.

C. Motion

In this paper, the difficulty of the problem is greatly reduced by removing camera motion. The calculation of optical flow is quite different for a non-stationary camera, even if the absolute camera motion is known. However, the Nao does not stand perfectly still due to the servos not being able to keep a pose exactly the same. As a result, there is always some noise present in calculation of optical flow in a static scene. This slight movement also makes it hard to use background subtraction [13], a method which may improve results due to the removal of objects that are not of interest. If the Nao is able to stand perfectly still during motif detection, it will still be hard to use background subtraction: Camera calibration is needed after the robot has moved (e.g., walking towards an opponent, diving to stop the ball) and this will cost time. Additionally, the Nao's camera moves inside the head after an impact or sudden movement, making any method that assumes that the camera is stationary very susceptible to noise. It is imperative that the goalkeeper reacts as fast as possible, and therefore, background subtraction was not applied.

VI. FUTURE RESEARCH

The RoboCup consists of several leagues, each league having its own rules and fields of research. The league in which the Nao is used is the Standard Platform League, a soccer competition for humanoid robots. Although optimization of soccer behavior is one field of research in the league, detection of opponent's behavioral patterns has not been attempted before. The main reason for this is that participating teams are constantly adjusting behavior; it would be futile to analyze the opponent to create opponent-specific strategies manually.

Automatic activity mining has been used in the Small Size League of the RoboCup, where an overhead camera is available. For example by Ball et al [14], who use Bayesian approaches to find patterns in opponent behavior. Similar to the goal of the research in this paper, the patterns are used to predict the opponent's behavior, and effective strategies are created to exploit the opponent's weak points. An important difference between this league and the Standard Platform League is that data is retrieved from two stationary cameras that are placed above the playing surface, whereas in the humanoid leagues, the camera feed of the robots is the only available source of visual information. It is likely that results for PLSM will improve when, instead of point-of-view image data from the goalkeeper, an external camera that gives a top view is used. However, as the Standard Platform League

rules do not allow use of sensor data other than that of the robots [12], any method that uses external sensors cannot be applied in an official match.

The use of similar automated activity mining methods in official competitions may prove to be advantageous: Detection and prediction of an opponent's movement will make complex actions such as defending the goal by blocking, or passing the ball without it being intercepted, a lot simpler. It is likely that eventually, the detection of an opponent's behavioral patterns will become a necessary part of any robot soccer team, as adapting to the opponent's strategy is a basic element of soccer.

This study demonstrates that Probabilistic Latent Sequential Motifs can be applied to predict opponent's behavior in soccer. Previously, its value was already demonstrated in the analysis of traffic and pedestrian streams. This successful application indicates that this method could also be attractive for other applications.

VII. CONCLUSION

This paper describes the application of automatic unsupervised activity mining in videos for a humanoid soccer robot, and its effectiveness when used for the prediction and anticipation of opponent actions. A topic model-based method called Probabilistic Latent Sequential Motifs [2] is used to find recurring patterns of optical flow, referred to as motifs, in a dataset of short image sequences. This is the first time that Probabilistic Latent Sequential Motifs are used in this setting. The number of motifs and the maximum length of such a motif is specified beforehand. The discovered activity patterns are then used by the goalkeeper to predict an opponent's action and react accordingly. The effectiveness of the used method was tested by comparing performance (i.e., the ratio of the number of prevented goals to the number of scored goals) of the resulting goalkeeper to that of a simpler goalkeeper model, that only reacts when the ball is approaching, through a penalty shootout. Results indicate that setting the desired number of motifs relatively low contributes to performance, but that a goalkeeper with a predefined set of behavioral rules performs better overall. Nevertheless, the paper shows that automatic activity mining is a promising field of research in robot soccer.

ACKNOWLEDGMENT

We would like to thank Jagannadan Varadarajan, Remi Emonet and Jean-Marc Odobez for the distribution of their Matlab source code for Probabilistic Latent Sequential Motifs.

REFERENCES

[1] T. J. de Haas, T. Laue, and T. Rofer, "A scripting-based approach to robot behavior engineering using hierarchical generators," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4736–4741.

[2] J. Varadarajan, R. Emonet, and J. Odobez, "Probabilistic latent sequential motifs: Discovering temporal activity patterns in video scenes," in *Proceedings of the British Machine Vision Conference*, September 2010.

[3] J. Varadarajan and J. Odobez, "Topic models for scene analysis and abnormality detection," in *ICCV 12th International Workshop on Visual Surveillance*, 2009, pp. 1338–1345.

[4] C. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, "Latent semantic indexing: A probabilistic analysis," in *Proceedings of ACM PODS*, 1998.

[5] R. Emonet, J. Varadarajan, and J. Odobez, "Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2011, pp. 3233–3240.

[6] D. Kuettel, M. D. Breitenstein, L. van Gool, and V. Ferrari, "What's going on? Discovering spatio-temporal dependencies in dynamic scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010.

[7] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 31, no. 3, 2000.

[8] Y. W. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, 2006.

[9] T. Hofmann, "Probabilistic latent semantic indexing," in *In Proceedings of the 22th International Conference on Research and Development in Information Retrieval (SIGIR)*, 1999.

[10] ——, "Unsupervised learning by probability latent semantic analysis," *Machine Learning*, vol. 42, pp. 177–196, 2001.

[11] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, "Mechatronic design of nao humanoid," in *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, May, pp. 769–774.

[12] R. T. Committee, "Robocup standard platform league (nao) rule book," May 2012.

[13] A. McIvor, "Background subtraction techniques," in *Image and Vision Computing New Zealand*, 2000.

[14] D. Ball and G. Wyeth, "Classifying an opponents behaviour in robot soccer," in *Proceedings of the 2003 Australasian Conference on Robotics and Automation*, 2003.