Waving detection using the FuzzyBoost algorithm and flow-based features*

Plinio Moreno and José Santos-Victor
{plinio, jasv}@isr.ist.utl.pt

Instituto Superior Técnico & Instituto de Sistemas e Robótica
1049-001 Lisboa - Portugal

Abstract. We present an application of the FuzzyBoost learning algorithm, where the weak learners select spatio-temporal groups of features for waving detection. The features encode the spatial distribution of the optic flow of a tracked person, considering the polar sampling of the flow for each instant. The FuzzyBoost algorithm selects groups of features that discriminate better than any single feature, bringing robustness and generalization over the TemporalBoost algorithm.

1 Introduction

Previous works have considered people as helpers of surveillance systems [1] by signaling emergency, dangerous or suspicious situations with a universal alerting gesture: waving. The waving detector of [2] acts as an emergency signal which was tested on indoors, outdoors and several camera position with respect to the people. The waving detector relies on the TemporalBoost algorithm [3], which was recently generalized to the FuzzyBoost learning algorithm [4]. This paper presents the application of the FuzzyBoost learning algorithm on the waving detection, improving the results of the waving detector of [2].

The analysis of spatio-temporal patterns for human activity recogntion is a challenging area of research due to its computational complexity [5]. Exahustive search is not feasible, so several works have addressed ways to consider just subsets of all the possible patterns such as: space-time interest point detection [6], segmentation of the spatio-temporal volume using mean shift [7] and more recently the construction of random trees where each leave contains the information of a spatio temporal pattern [8], amongst others. The main objective of these works is to achieve good recognition rates, disregarding the real-time performance of the system. In this article we present an approach that analyses the spatio temporal patterns using the FuzzyBoost, which allows to detect the waving patterns in real-time (20fps).

The main difference between our previous work [2] and this paper is the search for spatio-temporal patterns (i.e. searching for several feature dimensions along frames). In our initial approach, the TemporalBoost algorithm assumes just temporal patterns (i.e. a single feature dimension along frames) for each dimension of the data samples.

* This work was supported by FCT (ISR/IST plurianual funding through the PIDDAC Program), partially funded by High Definition Analytics (HDA), QREN - I&D em Co-Promoção 13750 and and by the project CMU-PT/SIA/0023/2009 under the Carnegie Mellon-Portugal Program.

Although the TemporalBoost algorithm improves the classification performance, it is ignoring the information contained in the spatio-temporal patterns. In this work we present the advantages of the FuzzyBoost algorithm, which finds both the temporal and spatio-temporal patterns that improve the classification performance.

The remaining components of this work are the same as [2], namely: (i) the frame-based waving pattern extraction, which utilizes the Focus Of Attention (FOA) features; (ii) the optic flow computation [9] and (iii) the segmentation and labeling of moving targets in the image, which utilizes the LOTS method [10] for segmentation and hungarian assigment [11] for labeling. We show the generalization properties of the FuzzyBoost, which improves the performance of TemporalBoost.

In section 2, we describe the waving model, followed by the results in section 3 and conclusions in section 4.

2 Waving model

The waving model includes the optic flow based features and the FuzzyBoost learning. The optic flow features are collected in a spatio-temporal cuboid. The FuzzyBoost algorithm selects feature subsets (i.e. weak learners) from the cuboid at each learning round.

2.1 Spatio-temporal Focus Of Attention (FOA) features

FOA features encode the motion patterns of parts of the body with respect to its center [12]. This representation is based on the mean value of the optical flow for a set of cells, which correspond to the detected targets. Assuming that the centroid of the bounding box corresponds to the center of the person's body, the bounding box of the target is divided into polar sampled cells. Then, for each cell the mean value of the optical flow is projected onto the middle radial and normal directions of the polar cell. Particular gestures involve motion of body parts within a limited range of angles.

Various body movements will activate different cells in different ways, constructing patterns that represent motions of the human limbs, such as rising/putting down arms, bending, sitting, etc. The response on each cell of the FOA at each time instant is as follows:

$$\text{FOA}^t = \left[\text{FOA}^t_{1_R}\ \text{FOA}^t_{1_T}\ \ldots\ \text{FOA}^t_{i_R}\ \text{FOA}^t_{i_T}\ \ldots\ \text{FOA}^t_{n_R}\ \text{FOA}^t_{n_T}\right] \in \mathbb{R}^{2\cdot n},$$

where $\text{FOA}^t_{i_R} \in \mathbb{R}^2$ denotes the FOA computed at cell i and frame t in the radial direction (R) of the cell. $\text{FOA}^t_{i_T}$ denotes the tangential direction (T) of the i-th cell. The spatial patterns included by the FOA^t are augmented with temporal patterns, stacking all the FOA^t vectors in the previous $\tau - 1$ frames,

$$x_i = \left[\text{FOA}^t\ \ldots\ \text{FOA}^{t+\tau-1}\right] \in \mathbb{R}^{2\cdot n\cdot \tau}. \quad (1)$$

The spatio-temporal cuboid of the waving patterns is contained in the feature vector x_i of Eq. (1). In the following section we explain how to find useful patterns for classification, by selecting sets of components from x_i.

1. Input: $(x_1, y_1), \ldots, (x_N, y_N)$ where $x_i \in X$, $y_i \in Y = \{-1, +1\}$, set $H(x_i) := 0$, initialize the observation weights $w_i = 1/N$, $i = 1, 2, \ldots, N$
2. Repeat for $m = 1, \ldots, M$
   (a) Find the optimal weak classifier h_m over (x_i, y_i, w_i).
   (b) Update weights for examples $i = 1, 2, \ldots, N$, $w_i := w_i e^{-y_i h^*_m(x_i)}$
3. Output: Compute the strong classifier as $H(x_i) = \sum_m^M h^*_m(x_i)$ and classify the sample x_i according to $\operatorname{sgn} H(x_i)$

Fig. 1. GentleBoost algorithm

2.2 The FuzzyBoost algorithm

Boosting algorithms computes a linear combination of (weak) models into a strong classifier, $H(x_i)$. The final model is learned by minimizing, at each round, the weighted squared error,

$$J = \sum_{i=1}^{N} w_i (y_i - h_m(x_i))^2, \tag{2}$$

where $w_i = e^{-y_i h_m(x_i)}$ are the weights and N the number of training samples. At each round, the optimal weak classifier is added to the strong classifier and the sample weights adapted, increasing the weight of the misclassified samples and decreasing for the correctly classified ones [13]. Figure 1 shows the steps of the GentleBoost algorithm.

Decision stumps are the usual choice for GentleBoost: $h_m(x_i) = a\delta\left[x_i^d > \theta\right] + b\delta\left[x_i^d \leq \theta\right]$, where d is the dimension index and δ is the indicator function (i.e. δ[*condition*] is one if *condition* is *true* and zero otherwise). Decision stumps choose either branch a or b according to the threshold θ and feature value x_i^d. At each round, the the set of parameters $\{a, b, d, \theta\}$ that minimizes J w.r.t. h_m must be found. In the case of GentleBoost, there is a closed form for the optimal a and b, while the pair $\{d, \theta\}$ is found through exhaustive search [13].

2.3 Fuzzy weak learners optimization

Moreno et. al. [4] propose to augment the search space, looking for a set of dimensions instead of just one dimension, as follows:

$$h^*_m(x_i) = a\,\frac{F^T\delta\left[x_i > \theta\right]}{||F||} + b\,\frac{F^T\delta\left[x_i \leq \theta\right]}{||F||}. \tag{3}$$

where $x_i \in \mathbb{R}^D$ and the vector $F \in \mathbb{Z}_2^D$, denotes a D dimensional vector with binary components, and the non-zero components of F define a feature set. The vector F selects a group of dimensions that cope with the indicator function constraints of Eq. (3). Note that the feature sets of classic decision stump are

$$\mathcal{F} = \{F_1, \dots F_d \dots, F_D\}, \text{where } F_d = \left[0 \dots 1_d \dots 0\right]^T. \quad (4)$$

Therefore, the vector F generalizes GentleBoost by considering additional feature dimensions. We remark that selector F of Eq. (3) is replacing the indicator function (i.e. a true or false decision) by an average of decisions. The new functions are:

$$\Delta_+(x_i, \theta, F) = \frac{F^T \delta\left[x_i > \theta\right]}{||F||}, \quad \Delta_-(x_i, \theta, F) = \frac{F^T \delta\left[x_i \leq \theta\right]}{||F||}. \quad (5)$$

The functions Δ_+ and $\Delta_- = 1 - \Delta_+$ of Eq. (5) sample the interval $[0\ 1]$ according to the number of features selected (i.e. non-zero entries of F), which are above and below the threshold θ. The new weak learners, the fuzzy decision stumps, are expressed as

$$h_m^*(x_i) = a\Delta_+ + b\Delta_-. \quad (6)$$

The substitution of the fuzzy stumps of Eq. (3) into the error minimzation of Eq. (2), yields the optimal decision parameters a and b,

$$a = \frac{\bar{\nu}_+\bar{\omega}_- - \bar{\nu}_-\bar{\omega}_\pm}{\bar{\omega}_+\bar{\omega}_- - \left(\bar{\omega}_\pm\right)^2}, \quad b = \frac{\bar{\nu}_-\bar{\omega}_+ - \bar{\nu}_+\bar{\omega}_\pm}{\bar{\omega}_+\bar{\omega}_- - \left(\bar{\omega}_\pm\right)^2}, \quad (7)$$

with $\bar{\nu}_+ = \sum_i^N w_i y_i \Delta_+^T$, $\bar{\nu}_- = \sum_i^N w_i y_i \Delta_-^T$, $\bar{\omega}_+ = \sum_i^N w_i \Delta_+^T$, $\bar{\omega}_- = \sum_i^N w_i \Delta_-^T$, $\bar{\omega}_\pm = \sum_i^N w_i \Delta_-^T \Delta_+^T$.

There is no closed form to compute the optimal θ and F, thus exhaustive search is usually performed. On one hand, finding the optimal θ is a tractable problem. On the other hand, the search for the best F is NP-hard. In previous work, we assumed the temporal similarity of each feature dimension in order to build the feature sets $\mathcal{F}$ [3].

Algorithm 1: Generation of feature sets $\mathcal{F}$ using the TemporalBoost algorithm [3, 2]. Line 3 sets as value one at components $F_i(d)$ where feature cell and frame conditions are fulfilled.

input : Spatio-temporal feature, such as FOA in Eq. (1) with nC cells
output: $\mathcal{F} = \{F_{11}, \dots F_{jt} \dots, F_{nC\tau}\}$

```
1 for each time window w_t   t = 1 ... τ do
2 |   for each cell c_j   j = 1 ... nC do
3 |   |   F_jt(d) = δ[d_c = c_j ∧ d^t ∈ w_t = {1, ..., t}];
4 |   end
5 end
```

Alg. 1 shows the feature set selection of TemporalBoost, a heuristic that builds temporal threads in the spatio-temporal feature volume and was used previously on the same problem [2]. In this work we address the search for sets in the full spatio-temporal volume, guiding the search and reducing the number of possible candidates through dimensionality reduction algorithms.

Dimensionality reduction algorithms, as explained below, provide a projection matrix that we explore in order to find feature set candidates. Figure 2 shows the FuzzyBoost algorithm, which relies on the sets of features

$$\mathcal{F} = \{F_{11}, \ldots F_{ij} \ldots, F_{n_{\text{rows}} n_s}\}, \tag{8}$$

provided by a feature search on a linear projection matrix L with n_{rows} rows and a predefined number of intervals n_s. In the following section we present the algorithm that searches for $\mathcal{F}$ using a linear dimensionality reduction technique.

1. Given:
   $(x_1, y_1), \ldots, (x_N, y_N)$ and $\mathcal{F} = \{F_{11}, \ldots F_{ij} \ldots, F_{n_{\text{rows}} n_s}\}$. Data $x_i \in X$, $y_i \in Y = \{-1, +1\}$ and feature sets $\mathcal{F}$ provided by a feature search on a linear projection matrix L with n_{rows} rows and a predefined number of intervals n_s.
   Set $H(x_i) := 0$, initialize the observation weights $w_i = 1/N$, $i = 1, 2, \ldots, N$
2. Repeat for $m = 1, \ldots, M$
   (a) Find the optimal weak classifier h_m over (x_i, y_i, w_i) using the feature sets $\mathcal{F}$.
   (b) Update weights for examples $i = 1, 2, \ldots, N$, $w_i := w_i e^{-y_i h^*_m(x_i)}$
3. Compute the strong classifier as $H(x_i) = \sum_m^M h^*_m(x_i)$ and classify the sample x_i according to $\operatorname{sgn} H(x_i)$

Fig. 2. FuzzyBoost algorithm

2.4 The search space for the feature set

The follow-up work of the TemporalBoost attempted to search for groups of features in a local neighborhood [14]. The main drawback of the local search is that GentleBoost can reach its performance by considering a very large number of iterations. We addressed recently a global search for groups of features [4], which relies on linear dimensionality reduction techniques in order to find good set candidates for the FuzzyBoost. The linear mapping

$$x^* = Lx \tag{9}$$

contains relevant information about the correlations between dimensions of the original feature space. Moreno et al. analyze independently each row of the matrix L (row projection vector), clustering vector components with similar values. Their approach is based on clusters of weights: if the weight of a dimension in the (row) projection vector is similar to other dimension(s) in that vector, this implies some correlation level between those dimensions. Amongst the three dimensionality reduction algorithms, the Multiple Metric Learning for large Margin Nearest Neighbor (MMLMNN) Classification [15] provided better results than Linear Discriminant Analysis (LDA) and Principal Component Analylsis (PCA). The MMLMNN method aims to learn a linear transformation of the input space, such that each training input should share the same labels as

its k nearest neighbors, and the training inputs with different labels should be widely separated.

Given the linear mapping L computed by MMLMNN, each row of the matrix is considered separately in order to extract feature set candidates. The sets are built by selecting the components of the row vector having very similar values and discarding components having very low values (see Alg. 2). The quantitative measures of closeness and low values are: the size of the similarity interval (Δ_s in Alg. 2) and the lower threshold (s_0 in Alg. 2). The values of the projection matrix are scaled as follows: $\mathcal{L}_{ij} = \frac{|L_{ij}|}{\max(L)}$, which ensures that $0 < \mathcal{L}_{ij} \leq 1$.

Algorithm 2: Generation of feature sets F from a scaled linear mapping $\mathcal{L}$

input : s_0 lower threshold, n_s number of intervals, $\mathcal{L}$ normalized projection matrix
output: $F_{ij} \quad i = 1 \ldots n_{\text{rows}} \quad j = 1 \ldots n_s$

```
1 for each projection (row) vector L_i do
2     compute Δ_s = (max(L_i) − s_0)/n_s;
3     for j = 1 ... n_s do
4         compute s_j = s_0 + (j − 1)Δ_s;
5         F_ij = δ[s_j ≤ L_i < s_j + jΔ_s];
6     end
7 end
```

The threshold $s_0 \in [0, 1[$ removes components of $\mathcal{L}_i$ having very low weights, which are the less meaningful dimensions. The amount of intervals $n_s \in \mathbb{N}$ defines the size of the similarity interval Δ_s (line 2 of Alg. 2), where the dimension weights inside the interval are grouped into a feature set (line 5 of Alg. 2).

Our model for waving detection has a performance comparable to the state-of-the-art with the advantage of a very low computational load at detection time. We have implementation running in real time (20fps) on full sized images (640x480).

3 Experiments and results

We compare the performance of TemporalBoost and FuzzyBoost in two datasets: The KTH actions dataset [6] and waving *vs* not waving dataset. Figure 3 shows samples of the waving *vs* not waving dataset, which was introduced first in [2]. The FOA feature sampling is $\Delta\theta = \pi/4$. The support window of the TemporalBoost and the FuzzyBoost algorithms is 20 frames. The event window size is 4s (20 frames), considering a waving event if at least 60% of the single-frame classifications are positive.

Figure 4 shows the improvements of the FuzzyBoost over TemporalBoost. In the case of the KTH dataset the FuzzyBoost improvement is around 1%. Although the recognition results in the KTH dataset are below some recent approaches (like [8]), our system has the advantage of real-time performance. In the case of the waving *vs.* not waving dataset the FuzzyBoost improvement is around 2%. These improvements follow the trend of [4], showing that the spatio-temporal search of FuzzyBoost generalizes

Fig. 3. Examples of data samples from the waving *vs* not waving dataset. Positive and negative samples of the training set (First row). Samples of waving events correctly detected (Second row). Samples of the negative class detected correctly (Third row).

Related work	Accuracy
Moreno et al. [2]	91.7%
Our method	**92.6%**
Schuldt et al. [6]	73.6%
Ke et al. [16]	91.7%
Niebles et al. [17]	93%
Yao et al. [8]	97%

	single-frame	Event
TemporalBoost [2]	85.95%	94.43%
FuzzyBoost	91.21%	96.47%

Fig. 4. The left side table shows the performance of several approaches in the KTH dataset [6]. The right side table compares the performance of the TemporalBoost and the FuzzyBoost for the dataset illustrated in Figure 3.

better than the temporal strips of TemporalBoost. Recent tests in a multi-camera setup with both traditional cameras and HD cameras show the real-time performance of the FuzzyBoost waving detector. The computational overheads of the FuzzyBoost during classification are very small, so the FuzzyBoost waving detector has practically the same real-time performance of the TemporalBoost on the same samples. [1]

4 Conclusions

We have applied the FuzzyBoost algorithm on the detection of waving gestures. The FuzzyBoost algorithm searches for spatio-temporal patterns in the spatio-temporal cuboid of optic-flow based features. Our approach improves the results of TemporalBoost, showing the generalization properties of FuzzyBoost. Furthermore, FuzzyBoost classifies the waving patterns as fast as the TemporalBoost, mantaining the real-time execution of the waving detector.

[1] http://youtu.be/9eakYtZADu8

References

1. Sanfeliu, A., Andrade-Cetto, J.: Ubiquitous networking robotics in urban settings. In: Workshop on Network Robot Systems. Toward Intelligent Robotic Systems Integrated with Environments. Proceedings of IROS2006. (2006)
2. Moreno, P., Bernardino, A., Santos-Victor, J.: Waving detection using the local temporal consistency of flow-based features for real-time applications. In: Proc. of ICIAR - 6th International Conference on Image Analysis and Recognition. (2009)
3. Ribeiro, P.C., Moreno, P., Santos-Victor, J.: Boosting with temporal consistent learners: An application to human activity recognition. In: Proc. of 3rd International Symposium on Visual Computing. (2007) 464–475
4. Moreno, P., Ribeiro, P.C., Santos-Victor, J.: Feature set search space for fuzzyboost learning. In: Proceedings of the IbPRIA 2011. (2011)
5. Aggarwal, J., Ryoo, M.: Human activity analysis: A review. ACM Comput. Surv. **43**(3) (2011) 16:1–16:43
6. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local svm approach. In: Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Volume 3. (2004) 32–36 Vol.3
7. Ke, Y., Sukthankar, R., Hebert, M.: Spatio-temporal shape and flow correlation for action recognition. In: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on. (2007) 1–8
8. Yao, A., Gall, J., Van Gool, L.: A hough transform-based voting framework for action recognition. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. (2010) 2061 –2068
9. Ogale, A.S., Aloimonos, Y.: A roadmap to the integration of early visual modules. International Journal of Computer Vision **72**(1) (2007) 9–25
10. Boult, T.E., Micheals, R.J., Gao, X., Eckmann, M.: Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. Proceedings Of The IEEE **89**(10) (2001) 1382–1402
11. Ahuja, R., Magnanti, T., Orlin, J.: Network Flows. Prentice Hall (1993)
12. Pla, F., Ribeiro, P.C., Santos-Victor, J., Bernardino, A.: Extracting motion features for visual human activity representation. In: Proceedings of the IbPRIA'05. (2005)
13. Torralba, A., Murphy, K., Freeman, W.: Sharing visual features for multiclass and multiview object detection. IEEE Transactions On Pattern Analysis and Machine Intelligence **29**(5) (2007) 854–869
14. Ribeiro, P., Moreno, P., Santos-Victor, J.: Introducing fuzzy decision stumps in boosting through the notion of neighbourhood. Computer Vision, IET **6**(3) (2012) 214–223
15. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10** (2009) 207–244
16. Ke, Y., Sukthankar, R., Hebert, M.: Efficient visual event detection using volumetric features. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Volume 1. (2005) 166–173 Vol. 1
17. Niebles, J.C., Wang, H., Fei-Fei, L.: Unsupervised learning of human action categories using spatial-temporal words. International Journal of Computer Vision **79**(3) (2008) 299–318