


Fast Support Vector Classifier for Automated Content-based Search in Video Surveillance

Cătălin Mitrea[1,2], Ionuț Mironică[1], Bogdan Ionescu[1,3], Radu Dogaru[1]

[1] LAPI & Natural Computing Labs, University "Politehnica" of Bucharest, 061071, Romania

[2] UTI GRUP, Bucharest, 020492, Romania

[3] LISTIC, Université Savoie Mont Blanc, 74940 Annecy-le-Vieux, France

Email: *catalin.mitrea@uti.ro, {imironica,bionescu}@imag.pub.ro, radu d@ieee.org*

12-th International Symposium on Signals, Circuits and Systems, Iasi, Romania. 9-10 July.

ISSCS 2015

Outline

- Introduction
- Related work
- The proposed system
- Experimental results
- Conclusions and future work

Introduction

- High volume of video acquisition (>4mil CCTV cameras only in UK, ~500k only in London);
- Limited human resources.

Solutions

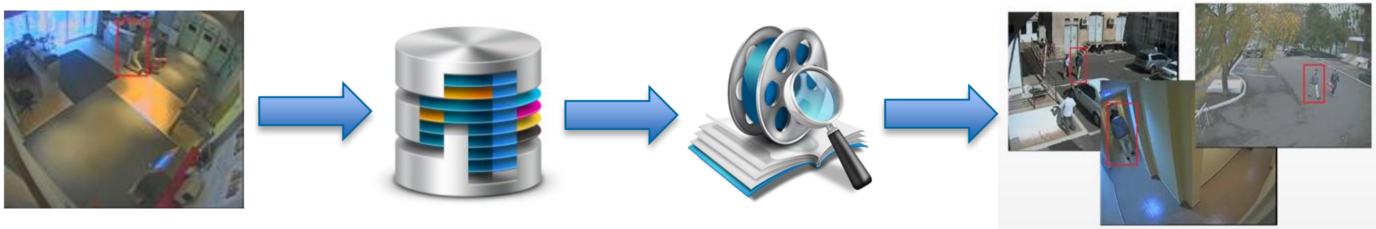
- Intelligent video surveillance techniques represent an important research domain:
  - Real-time identification and tracking of object of interest;
  - Behavior and incident detection;
  - Crowd analysis;
  - **Content-based multiple-instance searching and indexing of objects (humans).**

Research area

Introduction (cont.)

Problem statement

- Starting from a small sample (few frames) of the object to-be-found (human) => find (search) all relevant instances into a vast multisource video database.



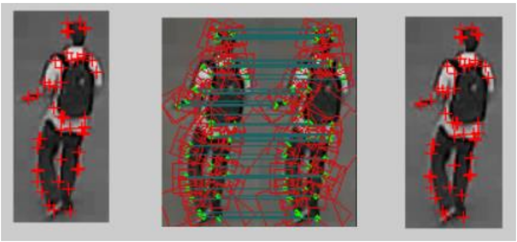
Objectives:

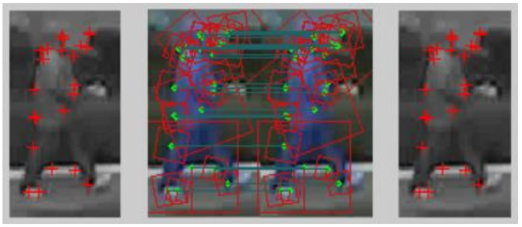
- Develop a system for providing content-based search capabilities within multi-source video surveillance footage.
- Investigate and propose specialized “decisioning” systems (Fast Support Vector Classifier – FSVC).

Related work

Main methods and directions

- Large video databases processing techniques [Snoek,IEEE 2010]; BigData technologies;
- Content descriptors extraction (color, texture, shape, temporal and motion, audio [Ionescu, LNCS 2011]); Feature points (SIFT, SURF [Stottinger, IEEE 2010]);
- Dictionaries (bag-of-words and fisher kernel representations [Mironica, ACM 2013]).
- Classifiers' parameter optimization during the training process [Chapell, JMLR 2008, Wright, CLW 2012]





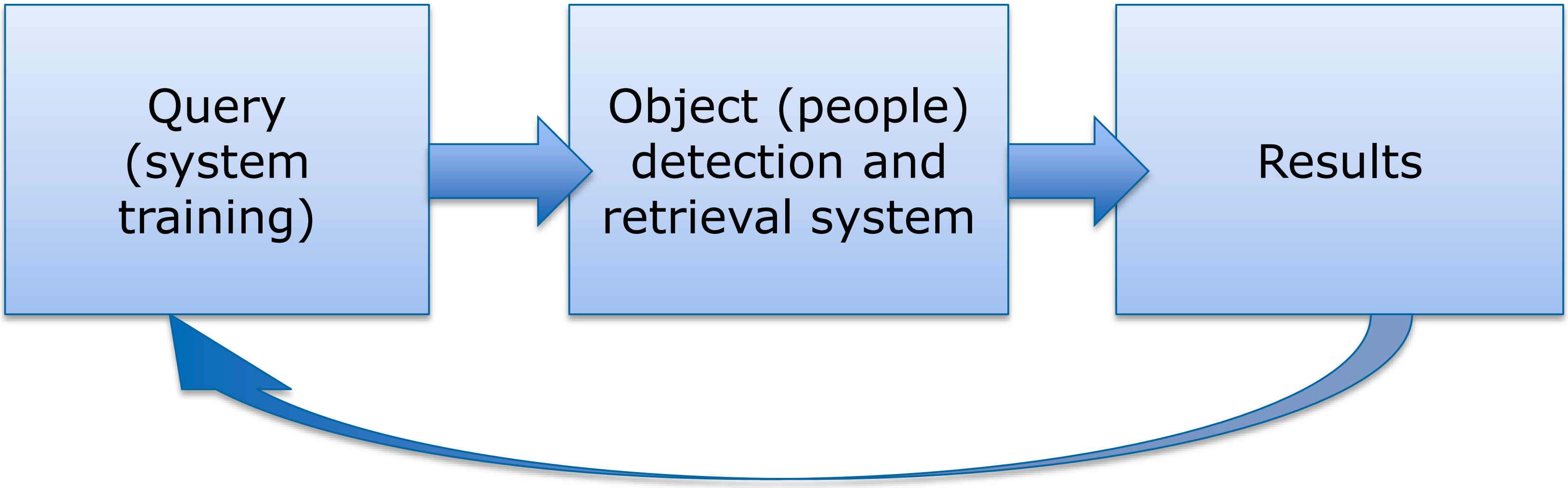
Drawbacks

- Computation complexity
- Difficult to implement for "real field" systems
- Not all methods are suitable for video surveillance datasets

perspectives - e.g., multiple source cameras, different weather conditions, different setups - e.g., indoor vs. outdoor, appearances, etc.

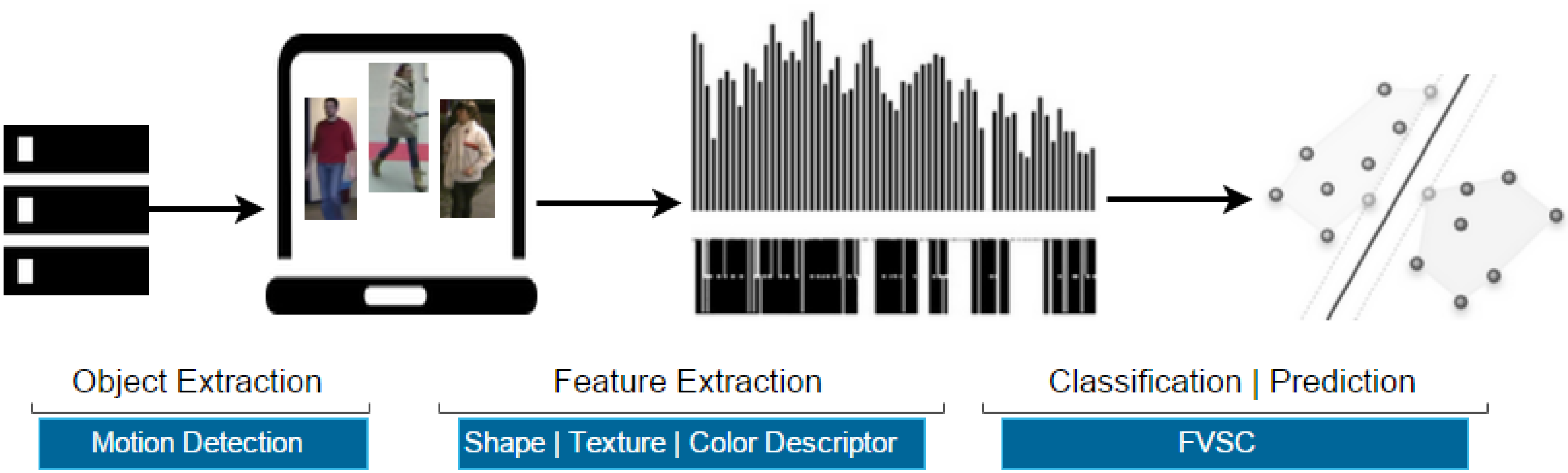
Proposed system

- Block diagram



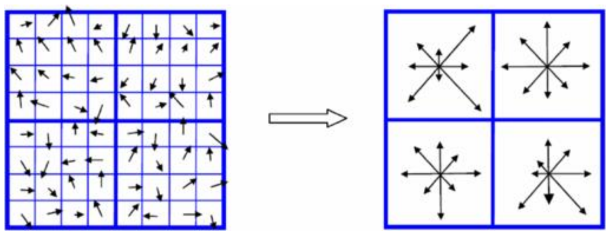
Proposed system (2)

- Object (people) detection and retrieval system



Content descriptors

- **HoG** features (Histogram of Oriented Graphs – shape-based descriptor, 81 values) - [Dalal, CVPR 2005]



- **CN** features (Color Naming histogram – color descriptor, 11 dimensions) [Van De Weijer, CVPR 1994]

11 colors distribution:

"black", "blue", "brown", "gray", "green", "orange", "pink", "purple", "red", "white" and "yellow".

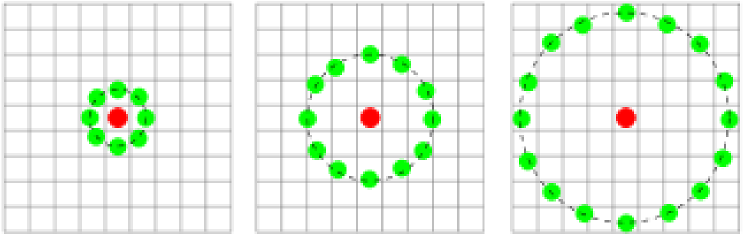
Content descriptors

- **CM** (Color moments –color descriptor, 225 dimensions) [Stricker, SPIE 1995];

Color similarity:

Three central moments of an image’s color distribution: mean, standard deviation and skewness.

- **LBP** (Local Binary Pattern – texture descriptor, 256 dimensions) [Ojala, ICPR 1994]



Reasoning system

- Powered by classifiers

- Established:

Support Vector

Machines (SVM) and

- Proposed:

Fast Support

Vector Classifier (FSVC).

E.g. of 12 training samples of true class (Scouter dataset)

E.g. 7 training samples of false class

Fast Support Vector Classifier (FSVC)

- FSVC was first introduced in [Dogaru, ICMNN 1996] as RBF-M - Modified Radial Basis Function Network.
- Based on simple arithmetic operators and employs simple Least Mean Squares (LMS) training in an expanded feature space generated by Radial Basis Function (RBF) kernels centered on support vectors selected via a simple algorithm.
- Needs only one single epoch to select the support vectors among feature vectors in the training samples. Then, simple Adaline training is performed in the expanded space formed of RBF kernels centered on the previously discovered support vectors.

FSVC – Algorithm overview

- *m ←1; k ←1; j_k ←1; // select the center of the first RBF unit as the first sample of the dataset;*
- *ov ← 1; // set overlapping coefficient between two RBF units;*
- *for j=2 to N // all training samples*
  - *act* ← $\sum_{k=1}^{m} K(x^j, x^{j_k})$; *// compute activity level*
  - *if (act < ov)*
  - *k ← k +1; j_k ← j; m ← m+1; // create a new RBF unit*
- *end if*
- *end for*

-> output calculated as $y = w_0 + \sum_{k=1}^{m} w_k K(x, x^{j_k})$
where $w_0, \dots, w_m$ represents the weights of an outputted *Adaline* trained with LMS and x^{j_k} represents the center vector selected as the sample j_k from the *N* training samples. A sample of either test or training set is defined as a pair (x^k, d^k), where $x^k$is an input vector (scaled between [0; 1]) with size *n* and $d^k$is the training label that belongs to the set {-1, 1}, $d^k = 1$ indicating that x^k belongs to the search class.

FSVC - Computational complexity

- The training is much simpler, i.e., Adaline training, while only one epoch suffice to identify the support vectors;

- Unlike in the SVM, where kernels must satisfy the Mercer's condition there is no such restriction for the FSVC. Consequently, simple triangular kernels may replace Gaussian ones and Manhattan distances may replace the Euclidian one with no significant performance loss;

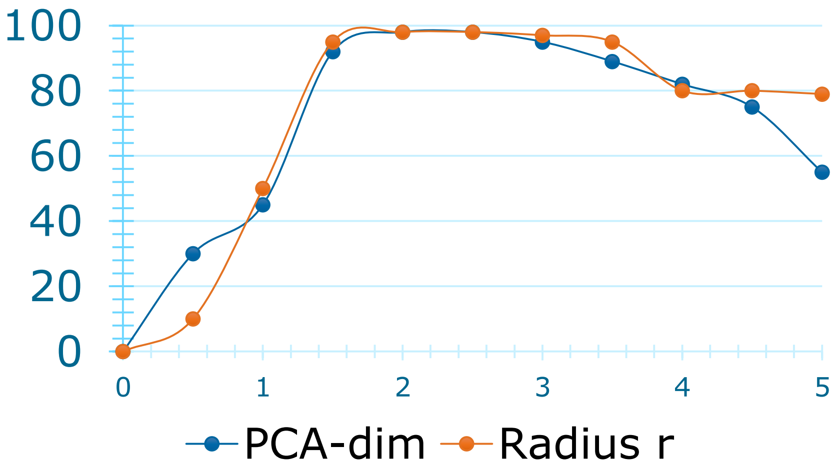
- Unlike SVM where for multiclass problems different sets of support vectors are generated for each class, in the FSVC there is only one set of support vectors (and the same number of RBF-units) for all classes, since each class is only assigned a different output Adaline (with the same nonlinear kernel for all classes) => This results in a significantly lower number of kernel units than in the case of SVM. The effect is a much more compact classifier structure;

- All the above makes FSVC a very attractive architecture for real-time task implementations in automated video surveillance => This solution has the advantage of a significantly lower cost of implementation due to simple arithmetic modules which map conveniently in common digital but also analogue implementation technologies.

FSVC - Parameter tuning

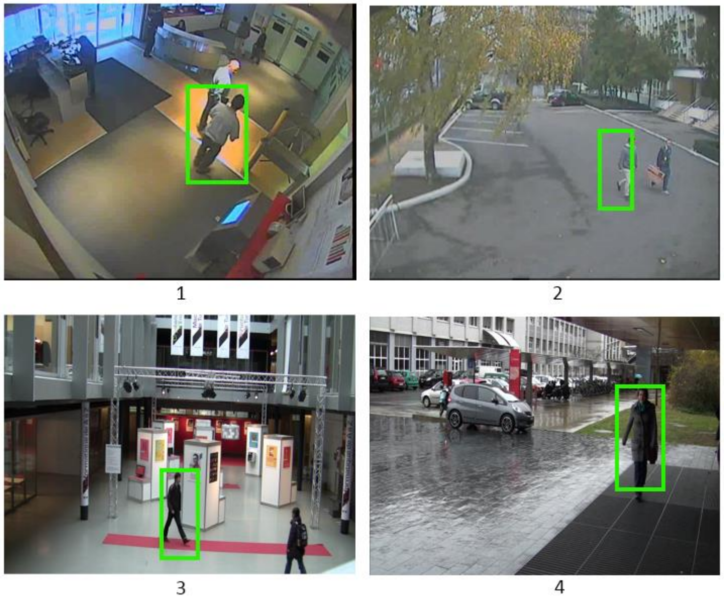
- In order to achieve the best generalization performance for proposed FSVC, different training parameters need to be properly adjusted.

- Variation of FSVC precision (%) during training in relation to the Radius *r* (spreading radius of the activation function) and PCA dimension (x10).



Datasets and Evaluation

- The approach was evaluated on two standard datasets (accounting for 16 people searching scenario on ca. 53000 labeled frames). Performance in terms of F2-Score attained promising results while dealing with our current task.



- Precision = TP/(TP + FP)
- Recall = TP/(TP + FN)
- F2Score = 5*Precision*Recall/(4*Precision + Recall)

- TP - True Positives
- FP - False Positives
- TN - True Negatives
- FN - False Negatives

The dataset was publicly released an can be downloaded at: **http://uti.eu.com/pncd-scouter/rezultate.html**

Experimental results

	F2-Score (%)			
Clasifier	SVM		FSVC	
Database	SCOUTER	PEVID-HD	SCOUTER	PEVID-HD
HoG	41.03	40.19	**42.48**	**47.37**
CM	38.87	45.00	**41.77**	**46.52**
LBP	42.26	47.26	**44.70**	**52.74**
CN	37.64	34.53	**38.35**	**41.10**
Fused	**44.08**	**48.56**	40.20	34.10

- best F2-score is obtained by FSVC – LBP pair (52.74% on PEVID-HD dataset) and (44.7% on SCOUTER dataset).
- lower performance is obtained by HoG-FSVC pair while lowest scores are obtained by color descriptors (CN – SVM pair - 34.53% on PEVID-HD dataset).
- as we address the issue of real-time processing, we consider the LBP - FSVC pair implementation is more suitable for our current task.

Examples of system classification responses

Cam0014 – set 2
frame 144

Cam0001 – set 1
frame 79

Cam0002 – set 2
frame 1351

Cam0015 – set 1
frame 119

Cam0001 – set 2
frame 313

Cam0002– set 1
frame 845

Cam0010– set 2
frame 174

Cam0007– set 2
frame 248

Cam0016 – set 1
frame 1224

Cam0010 – set 3
frame 1489

Cam0008– set 2
frame 197

Cam0010 – set 3
frame 1570

Correct classification. Results and Ground Truth agrees

True Positives examples (TP)

Cam0015 – set 1
frame 132

Cam0011 – set 1
frame 2046

Cam0014 – set 3
frame 1699

Cam0016 – set 2
frame 247

Misclassification reasons: signal noise, illumination conditions (insufficient, over exposed), partial object view (out of frame, junction with another object) or dimension too low.

False Negatives examples (FN)

Conclusions and future work

- The classification-based approach seems a suitable perspective to solve multi-instances object retrieval (search).
- The FSVC classifier is considered as a low complexity alternative to SVM for the use of multiple instance human retrieval task.
- Evaluation results revealed similar or better performance when compared with established classifiers as Support Vector Machines.

Conclusions and future work

- Drawbacks
  - Although the FSVC obtains better results that SVM on both selected databases, it shows a sensitivity of the performance to the range of the analyzed input data.
  - The performance of the system is closely related to the number of frames and the diversity of training sample (different perspective, object size, the quality of the images)
- Future work
  - Future work will address and investigate techniques to enhance further FVSC performance by employing specialized methods as ensemble learning and co-training which are adapted to the situation when very few training samples are available.

Acknowledgement

- This work is supported by the SCOUTER project (PN-II-IN-DPST-28DPST/30.08.2013);

Thank you!