Iterative active querying for surveillance data retrieval in crime detection and forensics

D. Coppi, S. Calderara and R.Cucchiara

D.I.I. - University of Modena and Reggio Emilia, Italy

Keywords: Multimedia forensics, transductive learning, relevance feedback.

Abstract

Large sets of visual data are now available both, in real time and off line, at time of investigation in multimedia forensics, however passive querying systems often encounter difficulties in retrieving significant results. In this paper we propose an iterative active querying system for video surveillance and forensic applications based on the continuous interaction between the user and the system. The positive and negative user feedbacks are exploited as the input of a graph based transductive procedure for iteratively refining the initial query results. Experiments are shown using people trajectories and people appearance as distance metrics.

1 Introduction

Multimedia forensics is a new emerging discipline regarding the analysis and exploitation of digital data as a support for investigation and for the extraction of probative elements. In the context of forensics, visual data about people and people activities are becoming day by day more appealing, due to the availability of large video-surveillance footage. Despite this high availability of data, the resolution in terms of color, space and time is often low, moreover occluded viewpoint and bad luminance are additional challenging elements. For these reasons the useful elements about the people in the scene are only considered as soft biometric data, in contrast to hard biometric data such as fingerprints or retinal data. Working with the uncertainty of these identifiers, the experience of investigators becomes essential: the continuous knowledge transfer between users and machines and the role of human operator in evaluating queries results is central and user's deduction and feedback are invaluable elements that concur to gain a continuous improvement and refinement of automatic results. Most of the applications devoted to digital surveillance forensics allow to perform queries on some specific people related data, but so far the user involvement in the search process is limited to executing subsequent queries over the obtained results. This class of approaches can be referred to as **iterative passive querying**, conversely new generation of multimedia application can be named **iterative active querying**. In these new systems, user feedbacks are collected and adopted to improve the single query step [1, 2]. More precisely, results can be manually ranked and new positive and negative elements can be introduced by the operator. After the manual refinement the system must be able to improve the query results to better encounter the desires of the users.

Focusing on soft biometrics data, and, particularly on people appearance and people trajectories, large archives of these features can be available at the time of investigation and similarity measures have been deeply investigated by the research community. In this scenario, where the final user can choose among a plethora of robust similarity measures, the importance of user feedback on query results is growing dramatically. The activity of involving user in search process should not impact on the chosen similarity measure but would desirably act directly on query results improving the ordering of the results themselves as long as the user knowledge about the target is transferred to the system. Thus we propose the adoption of approaches of user relevance feedback, as defined for multimedia and CBIR systems, also in multimedia forensics. Relevance feedback algorithms are countless [3, 4], since the problem can be faced from several point of view.

We aim to present a link between machine learning, semi supervised classification and user relevance feedback to provide a general and mathematically well-founded methodology to include user feedbacks in the forensic query process. The goal is to improve the search results exploiting the notion of the user on the correctness of the automatically retrieved results. To this purpose we exploit *transductive learning*, a well-known approach in machine learning community [5, 6, 7], that recently gained attention in multimedia community too [8], with the final aim of incorporating such an inference element in the forensic query process.

The paper proposes a system for interactive active querying on forensics data, in particular on people appearance and trajectories. The reason for this choice is that people trajectories and appearances can carry powerful information about people moving in the scene, for example investigators may want to find people that follows specific paths and eventually searching for all the visual occurrences of a suspect in a video stream. The general framework of the system is a graph based transductive learning algorithm where the labeled input models are constituted by the user feedbacks. This approach has advantages in terms of flexibility, since it can be adopted on whichever feature similarity measure is available, and in terms of performance, because thanks to the adoption of some optimization strategies it can be applied to thousand of images to improve conventional

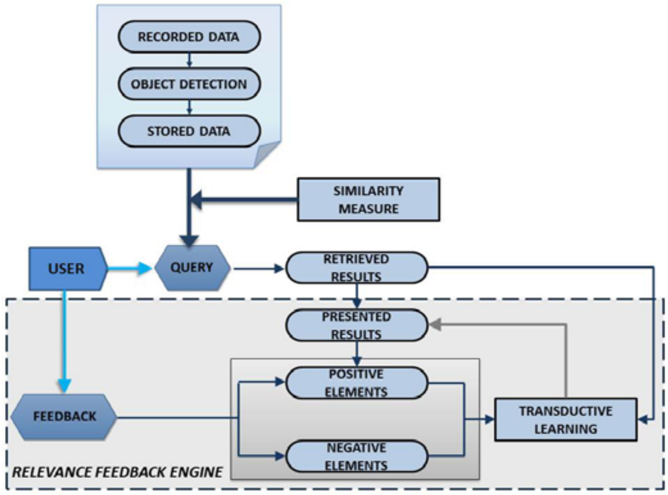

Figure 1. Overview of the Iterative Active Query System where user can interact with presented results in order to feed the transductive learner with positive and negative examples.

nearest neighbour queries.

2 Overwiev of the Iterative Active Querying System

This section gives a general overview of the proposed system, explaining the main building blocks and their properties. The overall schema of the application is depicted in Fig. 1. Our proposal is studied to work off-line on a set of previously recorded data offering a valid support to investigation and crime detection. Modern video surveillance systems can process acquired videos, automatically extracting several elements of interest, such as people silhouettes, their visual color appearance, geometric data etc. Among them we choose to use people appearance and trajectories since they can offer a valid support to analyse moving people on the scene. The goal of the system is thus to handle the interactive search for people appearances similar to a queried one or a people with similar trajectory. In both cases, after a proper **similarity measure** is defined, the first step of the system provides the user's choice of a query image q, and the query by example task retrieves a set of N results. The retrieved results are the first N Nearest Neighbours elements ranked with increasing distance from the query in the feature space according with the defined distance metric as in most of the content-based image retrieving systems. Between the n presented results, with $n < N$, the user can give a feedback on the quality of the response, marking elements as *relevant* or *not relevant*, i.e. indicating positive and negative elements in the retrieved set. The given feedback is exploited by the **relevance feedback engine** and positive and negative elements are used as input labeled models for the transductive learner. The transduction is performed over the labeled models and the unlabeled elements of the set of N previously retrieved results and the predicted label values for unlabeled elements are computed.

The iteration of multiple steps of the user feedback and the transductive process leads to a subsequent refinement of the results, offering an effective help in quickly processing large amount of data. The maximum number of iterations before reaching the final result is fixed to 5, since has been proved that after a certain number of iterations no satisfactory gains are achieved.

3 Relevance Feedback Engine

Our proposal, based on a joint use of transductive learning and relevance feedback, relies on the assumption that queries simply based on similarity between objects often show poor quality and do not offer an actual tool for improving search in video and achieving useful results. The basic idea of the system is to take advantage from the user feedback on the results retrieved from the first query and iteratively, and incrementally, use the user-classified set as the input labeled models of the transducer. At every iteration irrelevant samples are potentially moved away from the query center, while far away relevant samples, not considered at the beginning, are attracted toward the query center. This leads to more accurate and precise results allowing an effective support to investigations.

A first formulation of the transductive inference was given by Vapnik in [9], while the transductive method we propose to use is more inspired to a successive one proposed by Joachims [6]. Suppose to have a set of labeled instances $X_l = \{(x_i, y_i)\}_{i=1}^{l}$ with $y_i = \pm 1$ and a set of unlabeled data $X_u = \{x_i\}_{i=l+1}^{l+u}$. The complete dataset comprises both the model X_l and the candidates samples X_u and their associated label function y_i that takes non-zero values for the elements belonging to the model and zero value otherwise.

$$D(X, Y) = \{X_l \cup X_u, Y : y_i = \pm 1 \text{ iff } x_i \in X_l\} \quad (1)$$

Precisely the labeled data X_l are subdivided in two sets, the positive labeled samples X_{l+} with $y_i = +1$ corresponding to the target object we want to classify (i.e. the positive feedbacks), and the negative labeled samples X_{l-} with $y_i = -1$ corresponding to the labeled instances that differs from the target (i.e. the false positive results retrieved by the initial query and marked as negative by the user feedback) .

Following the postulates in [6] the transductive learning, based on positive and negative labeled elements, can be thought as the problem of estimating the label function values for the test set X_u while achieving a low leave-one-out (loo) on classification. Using a trivial K-Nearest-Neighbour rule the loo error of the classifier can be bounded by:

$$Err_{loo}^{KNN}(X, Y) \leq \sum_{i=1}^{N} (1 - \delta_i) \quad (2)$$

where δ_i is the KNN margin $\delta_i = y_i \frac{\sum\limits_{j \in KNN(x_i)} y_j w_{ij}}{\sum\limits_{m \in KNN(x_i)} w_{im}}$ with w_{ij} the similarity between x_i and x_j. The minimization of Eq. 2 can be obtained by maximizing the margin δ_i and imposing constrained values on the model labels, obtaining the following problem:

$$\begin{aligned} &\max_y (y^T A y) \text{ s.t.} \\ &y_i = \pm 1 \text{ if } x_i \in X_l \\ &y_j \in \{0, 1\} \end{aligned} \tag{3}$$

Where A is the affinity matrix computed over the samples X, and can be considered as the adjacency matrix of an undirected graph $G = (V, E)$ with the nodes V representing the samples X, and the edges E their similarities. G is in the form of a similarity-weighted k nearest-neighbour graph over X symmetrized by $A = A' + A'^T$, where:

$$A'_{i,j} = \begin{cases} \dfrac{w_{i,j}}{\sum\limits_{k \in KNN(x_i)} w_{i,k}} \text{ if } x_j \in KNN(x_i) \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

with $w_{i,j}$ an exponential symmetric function proportional to the distance $\rho(x_i, x_j)$ between samples x_i and x_j, $w_{ij} = \exp\left(-\dfrac{\rho(x_i, x_j)}{\sigma^2}\right)$ and σ a regularization parameter.

When A is considered as the affinity matrix of a graph G, as in our case, the problem in Eq. 3 is equivalent to finding the cut of the graph G that separates the two subgraphs G^+ constituted by the set of examples (i.e. vertices) with $y_i = +1$ and G^- constituted by the set of examples with $y_i = -1$. Although this problem can be solved using both, the s-t mincut algorithm or the transductive SVM, they easily lead to degenerate cuts when the number of labeled and unlabeled samples is not well balanced. Otherwise, including the cut size in the objective function to be maximized, the problem of Eq. 3 can be equivalently interpreted as a ratio-cut problem that can be efficiently solved exploiting the spectral properties of the graph Laplacian.

Let A be the affinity matrix explained above, and D be the diagonal degree matrix $D_{ii} = \sum_j A_{ji}$, the Laplacian graph can be computed as $L = D - A$ and L is symmetric and positive semi-definite, thus indirected. Graph Laplacians have recently been successfully adopted in image segmentation [10], spectral clustering and dimensionality reduction [11], since they represent a powerful manifold learning tool.

If we include the constraints in 3 in the ratio-cut problem, the supervised optimization problem becomes:

$$\begin{aligned} &\min_{\overrightarrow{z}} \overrightarrow{z}^T L \overrightarrow{z} + c(\overrightarrow{z} - \gamma)^T C(\overrightarrow{z} - \overrightarrow{\gamma}) \text{s.t.} \\ &\overrightarrow{z}^T \overrightarrow{z} = n \text{ and } \overrightarrow{z}^T 1 = 0 \end{aligned} \tag{5}$$

For each labeled example, the corresponding element of $\overrightarrow{\gamma}$ is respectively equal to $\gamma_+ = \sqrt{\frac{l_-}{l_+}}$ and $\gamma_- = \sqrt{\frac{l_+}{l_-}}$ for positive and negative examples, and it is zero for test examples, l_+ (l_-) is the number of the positive (negative) labeled training example as in [6]. Always referring to Eq.5 c is a parameter that trades off training error versus cut-value, and C is a diagonal cost matrix that allows different misclassification costs for each example. Taking the eigendecomposition $L = U \Sigma U^T$ of the Laplacian, one can introduce a new parameter vector $\overrightarrow{w}$ and substitute $\overrightarrow{z} = U\overrightarrow{w}$. Since the eigenvector of the smallest eigenvalue of a Laplacian is always $\overrightarrow{1}$, the constraint in Eq.5 becomes equivalent to setting $w1 = 0$. Letting Ev be the matrix with all eigenvectors U and EV the matrix with all eigenvalues Σ except the smallest one, the optimization problem can equivalently be written as

$$\begin{aligned} &\min_{\overrightarrow{w}} \overrightarrow{w}^T D \overrightarrow{w} + c(Ev\overrightarrow{w} - \gamma)^T C(Ev\overrightarrow{w} - \overrightarrow{\gamma}) \text{ s.t.} \\ &\overrightarrow{w}^T \overrightarrow{w} = n \end{aligned} \tag{6}$$

Defining $G = (EV + cEv^T CEv)$ and $\overrightarrow{b} = cEv^T C\overrightarrow{\gamma}$ the objective function can also be written as $\overrightarrow{w}^T G \overrightarrow{w} - 2\overrightarrow{b}^T \overrightarrow{w} + c\overrightarrow{\gamma}^T C \overrightarrow{\gamma}$, where the last term can be dropped since it is constant. Problem 6 is then minimized for $\overrightarrow{w}^* = (G - \lambda^* I)^{-1} \overrightarrow{b}$ where λ^* is the smallest eigenvalue of

$$\begin{bmatrix} G & -I \\ -\frac{1}{n}\overrightarrow{b}\overrightarrow{b}^T & G \end{bmatrix} \tag{7}$$

I is the identity matrix. The optimal value of Eq.5 is computed as

$$\overrightarrow{z}^* = Ev\overrightarrow{w}^* \tag{8}$$

producing a predicted value for each example in the test set. These values can be used to rank the test example or can be thresholded to have hard class assignment, where an obvious choice for the threshold can be the midpoint $\Theta = \frac{1}{2}(\gamma_+ + \gamma_-)$.

4 Similarity Measures for People Soft Biometrics Analysis

As introduced in Sec.2 we propose three different similarity measures for respectively people trajectories, based either on their shape or points coordinates, and for people snapshots. These measures can be used to perform queries by example and to build the affinity matrix of Eq. 4 depending on the feature we want to query. More precisely while snapshots can be queried based on their visual similarity that embodies both colour and textural information, trajectories can be compared either on their position in the scene (spatial analysis) or their shape (shape analysis).

In the following we analyse separately the three different similarity measures.

4.1 Trajectories Models for People Paths Analysis

The people trajectory projected on the ground plane is a very compact representation based on a sequence of 2D data coordinates $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, often associated with the motion status, e.g. the punctual velocity or acceleration. When data are acquired in a real system they should be properly modelled to account for tracking errors, noise in the support point extraction and inaccuracies; in particular, the modelling choice must take into account that when comparing two points belonging to different trajectories, small spatial shifts may occur and trajectories never exactly overlap point-to-point. Consequently, in our forensic application we adopted the spatial model proposed in [12], that combines a *statistical* representation of the data with a *point-to-point* approach to balance the computation cost and the accuracy. Briefly, given the k^{th} rectified trajectory projected on the ground plane $T_k = \{\mathbf{t_{1,k}} \dots \mathbf{t_{n_k,k}}\}$, where $\mathbf{t_{i,k}} = (x_{i,k}, y_{i,k})$ with n_k the number of points of trajectory T_k, a bi-variate Gaussian is centred on each data point $\mathbf{t_{i,k}}$ (i.e., having the mean equal to the point coordinates $\mu_{\mathbf{i},\mathbf{k}} = (x_{i,k}, y_{i,k})$) and with fixed covariance matrix C. Using a sequence of Gaussians, one for every point, allows to build an envelope around the trajectory itself, obtaining a slight invariance against spatial shifts. After assigning a Gaussian to every trajectory point, the trajectory can be modelled as a sequence of symbols corresponding to Gaussian distributions $\overline{T}_j = \{S_{1,j}, S_{2,j}, ..., S_{n_j,j}\}$, where every symbol $S_{i,j}$ is associated to its bi-variate Gaussian.

A completely different perspective of analysis consists in discarding the paths position in the scene focusing instead on their shape. This approach is oriented on discerning and synthesizing common and frequent motion patterns that are important indicators of people habits and interactions. Recently it has been proposed to perform the shape analysis in the directions domain, considering the trajectories as a sequence of angles and adopting a circular-defined statistic for modelling periodic angular data. Angular analysis seems a promising way to approach the shape comparison problem since sequences of angles are by definition location independent. We implemented the *statistical* model for shape analysis in [13], that exploits circular statistics to robustly model data points. In analogy with spatial model we aim at obtaining a sequence of symbols that statistically represent the sequence of angles that constitute he trajectory. Consequently, for handling angular data, circular distributions have been proposed in literature and among these the Von Mises distribution demonstrates to be the most suitable since it is circularly defined and correctly capture the periodic nature of angular data, [13]. Von Mises distribution is thus an ideal pdf to describe a trajectory T_j by means of its angles. However, in the general case a trajectory is not composed only of a single main direction, thus it should be represented by a multi-modal pdf, and thus the model consists of a mixture of Von Mises (MovM) distributions:

$$p(\theta) = \sum_{k=1}^{K} \pi_k \mathcal{V}(\theta | \theta_{0,k}, m_k) \tag{9}$$

When the mixture distributions components have to be learnt the EM algorithm is a natural solution to compute the parameters. In the case of MovM a full derivation of the EM has been proposed in [13]. Once the K components of the mixture are computed, every direction $\theta_{i,j}$ is encoded with a symbol $S_{i,j}$ with a MAP approach, where every symbol corresponds to the most probable MovM components.

Alignment Based Similarity Measure Once the data have been arranged in symbol sequences, the similarity measure is computed between two sequences using an alignment technique. The motivation resides in the fact that a plain comparison of the sequences would be imprecise due to their differences in length. In the case of statistical models, it has been suggested in [12] that the well-known Needleman-Wunsh (NW) alignment algorithm, is effective in comparing sequences of pdf while in [14] the efficiency of the Dynamic Time Warping(DTW) alignment algorithm in presence of very large datasets has been demonstrated. Independently from the chosen method, either NW or DTW, basically the alignment is performed by using dynamic programming with a computational complexity of $O(n * m)$, where m and n are sequences lengths, and exploiting specific recurrent relations, depending on the alignment algorithm, after the definition of a symbol-to-symbol distance measure; more specifically the Bhattacharyya distance, for the statistical model [12].

4.2 People Snapshots Model

In conjunction with path information, video surveillance system can automatically extract an additional important cue about people identity, the visual appearance. Investigators, during the analysis of video data, may want to search for people visually similar to a selected reference individual. For adding visual appearance capabilities to the proposed forensic application we implemented a snapshot similarity measure based on covariance matrix features descriptor, [15]. The covariance matrix is a square symmetric matrix $d \times d$, with d the number of selected features independently from the size of the image window, carrying the advantage of being a low dimensional data representation. Given the covariance matrix C its diagonal entries represent the variance of each feature and the non-diagonal entries represent the correlations; generally a single matrix extracted from a region is enough to match the region in different views and poses since the noise corrupting individual samples is largely filtered out with the average filter during covariance computation. Moreover covariance matrices can have scale and rotation invariance property and are independent to the mean changes such as identical shifting of color values. Let I be a three-dimensional color image and F be the $W \times H \times d$ dimensional feature image extracted from I, $F(x, y) = \Phi(I, x, y)$, where the function Φ can be any mapping such as intensity, color, gradients, filter responses, etc. Let $\{z_i\}_{i=1\dots N}$ be the d-dimensional feature points inside F, with $N = W \times H$. The image I is represented with the $d \times d$ covariance matrix of the feature points:

$$C_R = \frac{1}{N-1} \sum_{i=1}^{n} (z_i - \mu)(z_i - \mu)^T \tag{10}$$

where μ is the vector of the means of the corresponding features for the points within the region R.

In our case z_i is the feature vector composed for each pixel by its spatial, color and edge information. We use x and y pixel location in the image grid, RGB color values and Gx and Gy first order derivatives of the intensities calculated through Sobel operator w.r.t. x and y. Therefore each pixel of the image is mapped to a seven-dimensional feature vector $z_i = [\, x \;\, y \;\, R \;\, G \;\, B \;\, Gx \;\, Gy \,]^T$.

Appearance Based Similarity Measure Based on this features vector the covariance of a region is a 7×7 matrix.
To obtain the most similar region to the given object, we need to compute the distances between the covariance matrices corresponding to the target object and the candidate regions. However, the covariance matrices do not lie on the Euclidean space. Therefore an arithmetic subtraction of two matrices would not measure the distance of the corresponding regions. The distance metric between the covariance matrices is proposed in [16] as the sum of the squared logarithms of the generalized eigenvalues.

$$\rho(C_i, C_j) = \sqrt{\sum_{i=1}^{d} ln^2 \lambda_k(C_i, C_j)} \qquad (11)$$

where $\lambda_k(C_i, C_j)_{k=1\ldots d}$ are the generalized eigenvalues of C_i and C_j computed as $\lambda_k C_1 x_k - C_j x_k = 0$, $k = 0 \ldots d$, where x_k are the generalized eigenvectors. The distance measure ρ satisfies the metric axioms, positivity, symmetry, triangle inequality, for positive definite symmetric matrices.

5 Experiments

To evaluate the impact of relevance feedback on the querying system precision and recall we collected two different types of soft biometric data, namely people trajectories and people snapshots, from publicly available datasets. Trajectories have been acquired from Edinburgh Informatics Forum Pedestrian Database [1]. This dataset contains several days of people trajectories taken from a bird-eye view camera. Additionally, snapshots have been collected from the publicly available CAVIAR [2] and THIS [3] dataset extracting them from videos using a conventional HOG people detector, [17]. The dataset, for quantitative accuracy evaluation, consists of 4000 trajectories and 3000 people snapshots manually ground truthed.
In all the tests a query element was selected by the user and the first 30 results returned by the system. The baseline query method is the Nearest Neighbour classifier where the results are ranked according to the similarity w.r.t. the query element. For every query a maximum of 5 iteration of the relevance feedback engine of Sec. 3 are performed by the user, that can select either positive feedbacks or both positive and negative examples. Finally the improved ranking on results is given by the label function of Eq.8. The tests have been performed evaluating the relevance feedback importance in three different query types:

- Trajectory points query where we adopt a similarity measure for comparing people trajectories based on both their shape and coordinates in the image plane;
- Trajectory shapes query where trajectories are compared using a measure acting directly on their shape by comparing the sequence of directions(angles) that compose the trajectories;
- Snapshot appearance similarity query where given the image of a subject similar images are then returned on the basis of both color and textural elements.

Fig. 2 underlines the average results on 100 queries where the user could freely select positive and negative examples. The boost on performances obtained through relevance feedback (bars portions over the dashed lines) is evident and demonstrates the capability of the system to obtain satisfying results even when simple and fast similarity measures are employed to compare the elements. It is remarkable to note that the final average precision and recall are closer, in most of the considered cases, to 90% even when exploiting a reduced number of iterations of the transductive classifier.

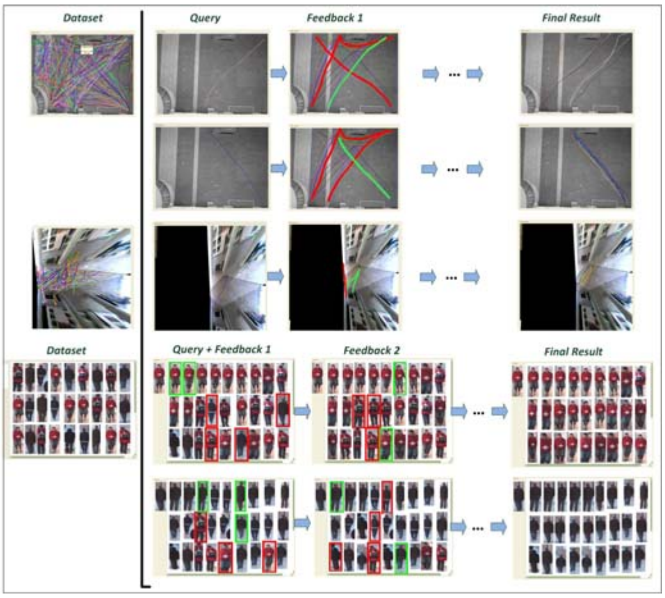

Figure 3. Examples of queries, first three queries are performed on people trajectories, while the last two are performed on people appearance. Green elements are positive selected feedbacks while red are the negative ones.

The developed active query engine has been integrated in our automatic video surveillance framework that is capable to extract people trajectories and snapshots from static video surveillance cameras. Examples of queries results using subsequent active iteration by transduction, on both publicly available data and data from a real experimentation carried out from cameras installed in our faculty, are depicted in Fig. 3.

6 Conclusions

In conclusion we presented an **iterative active querying** system for video surveillance forensic analysis. The proposed application consists of a transductive learning setting to easily incorporate user feedbacks in the query process as an aid for forensic investigation. The proposed solution is general and applies to whichever distance measure defined for comparing elements by their similarity. In particular we tested the system for three different similarity measures, namely people trajec-

[1] http://homepages.inf.ed.ac.uk/rbf/FORUMTRACKING
[2] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1
[3] http://www.openvisor.org

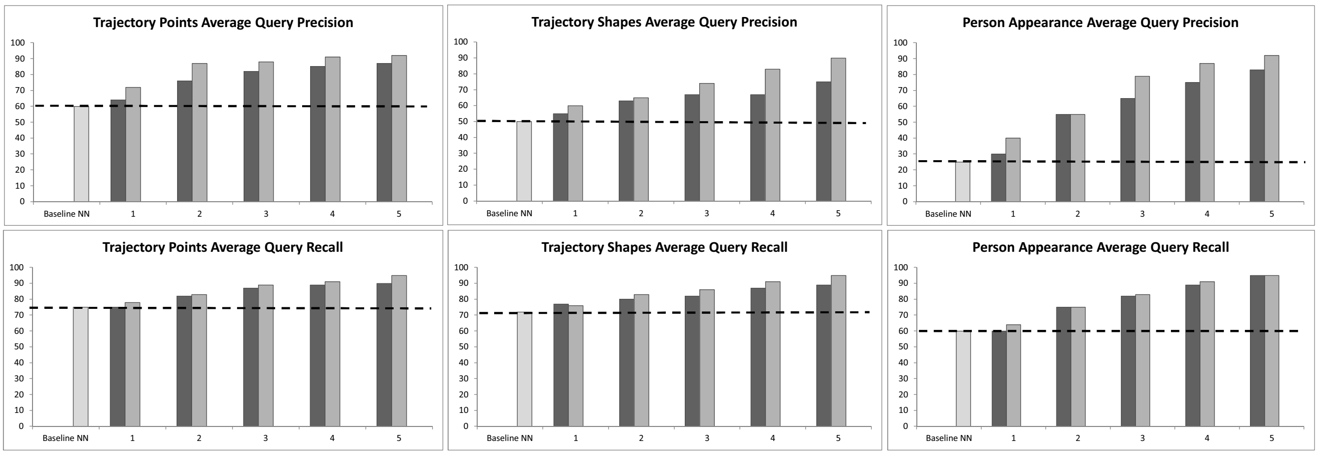

Figure 2. Queries average precision and recall on three different kinds of features. Dark grey bars refer to transduction with only positive feedbacks while light grey bars refer to transduction with both positive and negative feedbacks. First bars and dashed lines represent the values of precision and recall of the trivial KNN baseline method.

tories classified both, by shape and by points coordinates, and people appearance. Results are encouraging and demonstrate the effectiveness of both transduction and relevance feedback that can play a key role in forensic investigation process allowing an active approach that directly involves investigators knowledge and intuition in order to improve the results of automatic query systems.

References

[1] F. Hopfgartner, D. Vallet, M. Halvey, and J. Jose, "Search trails using user feedback to improve video search," in *Intl. conference on Multimedia (MM)*, 2008, pp. 339–348.

[2] M. J. Metternich and M. Worring, "Semi-interactive tracing of persons in real-life surveillance data," in *Intl. Workshop on Multimedia in Forensics, (MiFor)*, 2010.

[3] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, pp. 536–544, 2003.

[4] M. Crucianu, M. Ferecatu, and N. Boujemaa, "Relevance feedback for image retrieval: a short survey," in *State of the Art in Audiovisual Content-Based Retrieval, Information Universal Access and Interaction including Datamodels and Languages*, 2004.

[5] T. Joachims, "Transductive inference for text classification using support vector machines," in *Intl. Conf. on Machine Learning (ICML)*, 1999, pp. 200–209.

[6] ——, "Transductive learning via spectral graph partitioning," in *In Intl. Conf. on Machine Learning (ICML)*, 2003, pp. 290–297.

[7] X. Kong, M. Ng, and Z. Zhou, "Transductive multi-label learning via label set propagation," *Knowledge and Data Engineering, IEEE Transactions on*, no. 99, p. 1, 2011.

[8] H. Sahbi, J.-Y. Audibert, and R. Keriven, "Graph-cut transducers for relevance feedback in content based image retrieval," in *Intl. Conf. on Computer Vision (ICCV)*, 2007, pp. 1–8.

[9] V. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, 1998.

[10] O. Duchenne, J. Audibert, R. Keriven, J. Ponce, and F. Segonne, "Segmentation by transduction," *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

[11] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, 2007.

[12] S. Calderara, A. Prati, and R. Cucchiara, "Video surveillance and multimedia forensics: an application to trajectory analysis," in *Intl. Workshop on Multimedia in Forensics, (MiFor)*, 2009, pp. 13–18.

[13] ——, "Mixtures of von mises distributions for people trajectory shape analysis," *Circuits and Systems for Video Technology, IEEE Transactions on*, pp. 457 –471, 2011.

[14] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. J. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," *In Proc. of VLDB*, vol. 1, no. 2, 2008.

[15] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[16] W. Forstner, B. B. Moonen, and C. Gauss, "A metric for covariance matrices," 1999.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, 2005.