The International Society for Optical Engineering

SPIE Newsroom

10.1117/2.1.

Making Silicon A Little Bit Less Blind: Seeing and Tracking Humans

Fatih Porikli

An automatic human detection and tracking method that uses a novel object descriptor, called as covariance matrix, is described.

Object detection and tracking, as our eyes do so innately, are entitled to be the most essential components of computer vision from consumer electronics to smart weapons. In video surveillance, these methods facilitate understanding of motion patterns to uncover suspicious events. Navigation systems need them to keep vehicles in lane and avoid collisions. Traffic management systems employ them to control the flow, so we spend less time on the road. Video broadcasting makes use of them to better compress the data, so we wait less on line. In medical field, we use them in analysis of tumors and cellular entities to obtain accurate diagnosis. Still, robust detection and tracking of a deforming, non-rigid, and fast moving object, e.g. human body, presents a challenge.

Many different object descriptors, from aggregated statistics to appearance models, have been used by computers to translate the images of real world to the world of numbers. Histograms are among the most popular representations. However they disregard the spatial arrangement of features. Moreover, they do not scale to higher dimensions. Appearance models, on the other hand, are highly sensitive to noise and shape distortions. To overcome these shortcomings, we have developed a novel object descriptor, bag of covariance matrices, to represent an image window. We use this representation to automatically detect and track any target object in video images[1].

Basically, covariance is a measure of how much two variables vary together. By constructing the covariance of different features of an image window such as coordinate, color, gradient, edge, texture, motion, etc. as illustrated in Fig. 1, we capture the information embodied in both histograms and appearance models. By using a *bag* of such covariance matrices, we improve robustness against pose and shape changes. The bag of covariance matrix descriptor provides a natural way of fusing multiple features. Unlike histograms, it has a very low dimensionality. It is

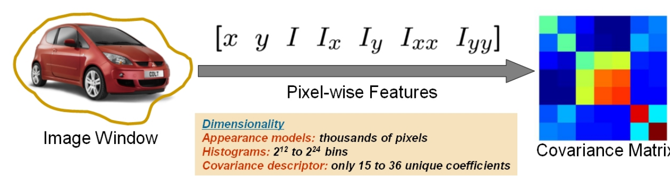

Figure 1. *Any region can be represented by a covariance matrix. Size of the covariance matrix is proportional to the number of features used.*

scale and illumination independent. Noise corrupting individual samples is largely filtered out during the covariance computation. We compute covariance matrix very fast using integral images[2]. This technique significantly accelerates the computation time by taking advantage of the spatial arrangement of the points.

To detect a specific object, e.g. human, in a given image, we

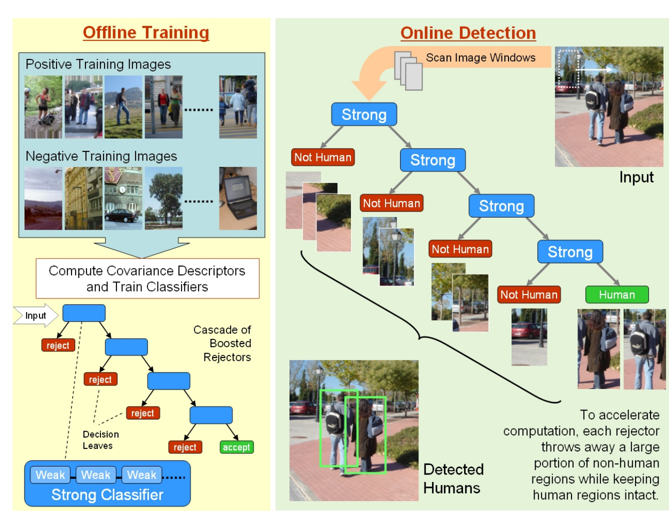

Figure 2. *Classifier is trained with positive (depicting humans) and negative (non-human) examples. Each weak classifier makes its estimation based on a single matrix from the bag of covariance matrices.*

Continued on next page

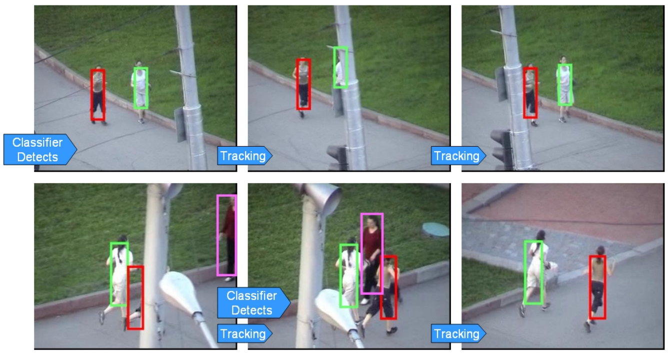

Figure 3. *Detection initiates objects and tracking resolves the identity correspondence problems.*

train a boosted classifier offline using covariance descriptors of the positive and negative training samples as shown in Fig. 2. We then apply the classifier online at each candidate image window to determine whether the window depicts the target object or not. To track a given object, we compare the covariance descriptor of object and candidate windows in the consecutive video frames using an eigenvector based distance metric[3]. We select the window that has the minimum distance and assign it as the estimated location. Covariance tracker does not make any assumption on the motion. In other words, it keeps track of objects even if the motion is erratic and fast as sample results are given in Fig. 3.

We employ a cascade of rejectors and a boosting framework to increase the speed of detection process. Each rejector is a strong classifier, and consists of a set of weighted linear weak classifiers. The number of weak classifiers at each rejector is determined by the target true and false positive rates. Each weak classifier corresponds to a region in the training window and it splits the high-dimensional input space with a hyperplane. Boosting works by sequentially fitting weak classifiers to reweighted versions of the training data. We fit an additive logistic regression model by stage-wise optimization of the Bernoulli log-likelihood.

Objects undergo appearance changes in time, thus, we construct and update a temporal kernel of covariance descriptors corresponding to the previously estimated object regions. From this set, we compute an intrinsic mean matrix that blends all the descriptors in the kernel. Since the space of covariance descriptors is not Euclidean space, we transform their manifold onto its tangent space where the relation between the vectors on the tangent space and the geodesics on the manifold are given by an exponential map. This enables us to define the dissimilarity between covariance descriptors by the sum of the squared logarithms of their generalized eigenvalues[4].

We are currently integrating this promising technology into surveillance products for multi-camera setups. We also plan to further improve the computational complexity by hardware implementation. Since our method works on any object type, we expect it to be applied to other target detection and tracking tasks.

Author Information

Fatih Porikli
Mitsubishi Electric Research Labs - MERL
Cambridge, MA

Dr. Fatih Porikli is leading research on the areas of computer vision, pattern recognition, data-mining, robust optimization for multi-camera surveillance and intelligent transportation systems at MERL. He is serving as associate editor for Journal of Real-Time Image Processing. He has authored over 50 papers and invented over 30 patents. He is a senior IEEE, ACM, SPIE member.

References

1. F. Porikli, O. Tuzel, and P. Meer, *Covariance tracking using model update based means on Riemannian manifolds*, **Proc. IEEE Conf. on Computer Vision and Pattern Recognition,** New York, NY.
2. F. Porikli, *Integral histogram: a fast way to extract histograms in Cartesian spaces*, **Proc. IEEE Conf. on Computer Vision and Pattern Recognition,** San Diego, CA.
3. O. Tuzel, F. Porikli, and P. Meer, *Region covariance: a fast descriptor for detection and classification*, **Proc. 9th European Conf. on Computer Vision,** Graz, Austria.
4. W. Förstner and B. Moonen, *A Metric for Covariance Matrices* tech. rep., Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999.