Manifold-based Fingerprinting for Target Identification

Kang-Yu Ni, Terrell N. Mundhenk, Kyungnam Kim, Yuri Owechko
HRL Laboratories
3011 Malibu Canyon Road, Malibu, CA 90265
{kni,tnmundhenk,kkim,yowechko}@hrl.com

Abstract

In this paper, we propose a fingerprint analysis algorithm based on using product manifolds to create robust signatures for individual targets in motion imagery. The purpose of target fingerprinting is to re-identify a target after it disappears and then reappears due to occlusions or out of camera view and to track targets persistently under camera handoff situations. The proposed method is statistics-based and has the benefit of being compact and invariant to viewpoint, rotation, and scaling. Moreover, it is a general framework and does not assume a particular type of objects to be identified. For improved robustness, we also propose a method to detect outliers of a statistical manifold formed from the training data of individual targets. Our experiments show that the proposed framework is more accurate in target re-identification than single-instance signatures and patch-based methods.

1. Introduction

Large amounts of video data have become more available to defense and security analysts as unmanned aerial vehicles (UAVs) with sensor payloads are used in many ISR (intelligence, surveillance, and reconnaissance) operations and large numbers of networked cameras are used for urban surveillance. Analysts are looking for significant events/targets that may be of importance for their surveillance missions. However, most of this video footage will be eventless and therefore of no interest to the analysts responsible for checking it. If a computer system could scan the videos for potential events/targets of interest, it would greatly lessen the analysts' workload, allowing them to focus only on the events of possible importance.

Recent video surveillance paradigms for multi-object detection and tracking [15, 23] in video sequences are developed as an integrated framework to not only spot unknown targets and track them, but also handle target reacquisition and target handoff to other cameras [4, 13, 19, 24]. A common process for these paradigms can be divided into four steps: detection of objects of interest, extraction of features or signatures, dimensionality reduction of the feature space, and target template or signature matching. A variety of different tracking or fingerprinting techniques have been proposed to address one or more of these key capabilities.

The tracking problem can be very challenging under unconstrained environment [7], because the target may change its appearance quickly, due to lighting conditions and frame rate, there might be other objects with similar appearance, and the background could be cluttered. In addition, occlusions and other nuisance factors of targets can further add the complexity to the problem [1]. Similar to single-object tracking, many existing multiple-object tracking methods are feature-based [15, 23] or image-patch based [17]. The latter applies sparse representation to predefined templates and random projection to speed up the solutions. Using motion and spectral clustering of trajectories instead of feature extraction is an alternative approach for tracking [10].

To handle target exiting and reappearing in videos or camera handoff situations, there are several methods working towards a unified framework of detection, tracking and identifying targets. For example, [4] demonstrates tunnel surveillance applications and shows the ability to follow vehicles through tunnels with non-overlapping cameras. Other methods that handle multi-camera networks are [13] for person re-identification, [19] for vehicles, and [24] for wide-area surveillance settings. Compact fingerprints are extracted from targets in these methods, such as Haar-features [4], covariance [24], implicit shape models, and SIFT features [13]. To handle significant scale variations and partial occlusions, [16] detects and tracks an object from a training video sequence by matching salient components of local descriptors and salient locations on a new image.

In this paper, we present a video exploitation framework to spot unknown targets, track them, segment the target blobs to obtain useful features, and then create signature fingerprints for the targets so that they can be reacquired. We focus on developing a manifold-based framework of robust and compact signature fingerprinting for the target re-identification task. Fingerprinting is the acquisition of a distinct description of an object of interest

978-1-4673-1612-5/12/$31.00 ©2012 IEEE

so that the object can be identified after occlusion or camera handoff. The fingerprinting problem is to capture the essence/invariance of the target so that it can be identified after leaving the video and then re-entering, with different illumination conditions and viewpoints.

Fingerprinting methods based on image patches [6, 9, 11] in general suffer from changes in appearance, such as rotation, translation, and scaling, and other unconstrained environment variants unless the amount of training data is sufficiently large. Moreover, since the patches of targets in general contain background, the performance of this approach is hampered when the background is cluttered or changes rapidly. Feature-based methods [4] in general deal with changes in appearance. However, they are sometimes not compact or are tailored to track specific types of objects, hence not universal. Compact and universal fingerprints are preferred because they allow computations online and transmissions to other cameras without significant delay, in addition to being able to deal with a variety type of objects. Our approach to fingerprinting is statistics-based, which has the benefit of being compact, robust to appearance change, and independent of the type of objects being tracked.

Following this introductory section, Section 2 describes a framework of moving target spotting, frame-to-frame tracking, and blob segmentation for feature extraction. Section 3 is the heart of the paper, where the proposed manifold-based methods for robust fingerprinting are presented with a few examples of compact signatures. Evaluations of target identification using different fingerprinting methods including random projection (RP), single instance (S), manifold-based (M), and product manifold (PM) are shown in Section 4, followed by conclusions and future work in Section 5.

2. Unsupervised target acquisition and tracking

Since the manifold-based fingerprinting method relies on target segmentation as inputs, we describe here the tracking system that is used to obtain the input data in our experiments for evaluation of target re-identification. More details can be found in [21]. The tracking system is unsupervised and allows for a moving camera platform. Initially, it finds a target of interest using motion. Unknown objects are spotted by subtracting global motion from local motion. The objects are then precisely segmented with the EDISON [12] mean shift [5] segmentation tool. Whole segments from EDISON which have common motion and which are contiguous to each other are grouped into a single master object. Once an initial acquisition has been created, segments can also be merged based on their match to a fingerprint. The tracked objects from adjacent frames are checked between adjacent frames for consistency, by calculating the Kullback-Leibler divergence between the GLM signatures described in Section 3.

In Figure 1, the left image shows an example of the segmentation of the spotted objects using prior object fingerprints the tracker has acquired. The right image is its original video frame from the DARPA VIVID video set [25] which includes a variety of moving vehicles under different viewpoints. Some of them are very similarly colored in their appearance and some are disappearing from the scene and re-entering the video. One can see that the segmentation is not exactly accurate. Therefore, the evaluation of fingerprinting in our experiments shown in Section 4 is performed on this real system, and as such we do not assume that the input data is completely ideal.

3. Manifold-based fingerprinting

Given target segmentation described in the previous section, we present our method of extracting fingerprints for individual targets. First, we set up the notations and describe single-instance fingerprinting, which is denoted by (**S**) throughout the paper. Let $I: \Omega \to \mathbb{R}^n$ be a given image, where $\Omega \subset \mathbb{R}^2$ is the image domain, and let $y = I(x)$ denote the image value at pixel x. Given the segmented region $\Lambda \subseteq \Omega$ of a target, we denote the signature fingerprint of the target by $f(y; I(\Lambda))$ for some pre-defined statistic f. In other words, each segmented target detected in the video is initialized by its model signature, $f(y; I(\Lambda))$. To identify a query target, its fingerprint is first extracted and then matched to the target whose signature it is closest to, with an appropriately chosen signature distance $d(f, g)$ between a pair of signatures f and g. The benefits of statistics-based fingerprints are that they significantly reduce the dimensionality of the segmented image data, $I(\Lambda)$, and at the same time are able to well characterize the object of interest for the re-identification task.

3.1. Manifold-based fingerprinting framework

Since single-instance signatures in general only work well for target fingerprinting within a short period of time, i.e., a short shelf life, we propose a manifold-based signature analysis method to extend the shelf life, as well as to improve the signature robustness to changes in appearance and lighting conditions. The idea is to form a signature manifold (**M**) from the collection of signatures of an object $\{f(y; I_i(\Lambda_i)) | i = 1, 2, \dots, T\}$ as it is being tracked. A signature space, a family of all statistics of a particular type, can be considered as a statistical manifold or information geometry. For example, it is known that the family of normal distributions forms a manifold with constant curvature -1. Therefore, the collection of signatures of all targets forms a statistical manifold. For individual targets, their respective signatures should lie in

disjoint lower-dimensional submanifolds for the re-identification problem to be solvable. Therefore, the statistic for fingerprinting needs to be properly chosen, so that all signature manifolds of individual targets are well separated.

The collection of signatures of a tracked target from a training set is viewed as sample points drawn from the statistical manifold $\mathcal{M}$ of the target, which can be parametrized by time. As explained in [6], for the purpose of re-identification, it is not necessary to explicitly represent the lower-dimensional submanifold, since the submanifold can be implicitly represented by the sample points. Therefore, assuming the number of samples is sufficient, one only needs to calculate the distance from the query signature g to each target manifold $\mathcal{M}$: $\mathrm{dist}(g, \mathcal{M}) = \min_{f \in \mathcal{M}} d(g, f)$ in order to determine which object the query target is identified with.

In the next section, we give a few examples of compact signature space, which can be parametric or nonparametric and spatial-aware or spatial-agnostic. Since the submanifolds of individual targets from a particular type of signature space may not always be disjoint due to the compactness of the signature, we further extend a single statistical manifold to a product manifold through joint probability, in order to better separate submanifolds of individual targets. Suppose we have different types of statistics $f_j, j = 1, 2, \dots, J$ for a single segmented target and $\mathcal{M}_j$ denotes the corresponding statistical manifold. To capture different characteristics described by different types of signatures, we combine them by joint probability $(f_1, f_2, \cdots, f_J) \in \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_J$ as a Cartesian product of the manifolds. Let d_j denote the distance between two probabilities in $\mathcal{M}_j$. We define the distance between two probabilities in $\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_J$ by $d = \sqrt{\sum_{j=1}^{J} d_j^2}$. This is the product manifold fingerprinting method (**PM**). A related approach is the joint manifold framework for data fusion [6] based on image manifolds, i.e. families of images generated by articulation of one or several objects in a scene [8]. Our framework is different because it is based on statistical manifolds, rather than image manifolds. Besides the advantage of being able to achieve much lower dimensionality, individual targets' image patches do not have to be rectangular and have a fixed size.

A practical issue of the manifold framework is that it needs to be robust to noise and outliers. This is especially important in a real tracking scenario, since the statistics used for re-identification depend on the frame-to-frame tracking data, in which the segmentation may not be accurate. To determine whether a statistic extracted from the tracking data is reliable, we first calculate the distance matrix D, whose ij^{th} entry is defined as the distance from signature at time i to signature at time j of the tracked target: $D_{ij} = d\left(f\left(y; I_i(\Lambda_i)\right), f\left(y; I_j(\Lambda_j)\right)\right)$. By this construction, the distance matrix D is symmetric and zero-diagonal. Since the signatures are ordered according to time and the target generally does not change drastically within nearby video frames, its statistics should have small distances within nearby video frames as well. To detect unreliable signatures, we take the local sum of the distance matrix D centered at D_{ii} for each time $i = 1, 2, \dots, T$, $s(i) = \sum_{j=i-t}^{j=i+t} \mathrm{D}_{ij}$, for some small number t. The local sum that is larger than the mean by a multiple of the variance, i.e. $s(i) > \mathrm{E}[s] + c \cdot \mathrm{Var}(s)$, indicates an outlier and its signature $f\left(y; I_i(\Lambda_i)\right)$ is eliminated from the sample collection. In addition to detecting unreliable samples from a single submanifold perspective with the method described here, additional outliers are detected by sparse subspace clustering [9] of training data to ensure that submanifolds of individual targets are disjoint.

3.2. Examples of signature space

In this section, we give a few examples of statistics for compact fingerprinting, as follows:

1) Generalized linear model (GLM) – parametric/ spatially agnostic
2) Spatial generalized linear model (SGLM) – parametric/ spatially aware
3) Histogram – non-parametric/spatially agnostic
4) Spatiogram – non-parametric/spatially aware

The GLM [22] signature is obtained by fitting the segmented data **y** with a multivariate normal distribution:

$$f(y;\ \mu, \Sigma) = \frac{1}{(2\pi)^{N/2}\det(\Sigma)^{1/2}} e^{-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)}.$$

Specifically, the data of each pixel, $\{a, b, a_e, b_e\}$, includes the croma components of L*a*b color [20] and spatial entropy of the a and b components, a_e and b_e, taken from a 9x9 patch. The SGLM signature in addition uses relative location $\{u, v\}$ as features, which is the pixel location after normalizing into [0, C], with a constant C. The SGLM fits $\{a, b, a_e, b_e, u, v\}$ with a multivariate normal distribution. The histogram and spatiogram signatures use $\{a, b, a_e, b_e\}$ and $\{a, b, a_e, b_e, u, v\}$ as data, respectively.

Parametric methods, S1 and S2, are more compact but may not be ideal since the underlying statistics may not be a single Gaussian. For instance, texture-rich images may not be well approximated with a normal distribution. On the other hand, non-parametric methods, S3 and S4, can take arbitrary statistics but are less compact. Spatial-agnostic methods, S1 and S3, are invariant to rotation, which can be significant after long time. On the other hand, spatial-aware methods, S2 and S4, have the advantage of separating features [2] so that where the feature is on an object becomes a factor. For instance, this

might help if the front of a car is red, but the back is black.

The difference between two GLM signatures (or two SGLM signatures) can be measured using the symmetric Kullback-Leibler (KL) divergence, since the standard KL divergence is not symmetric [14]. The symmetric KL divergence is defined as:

$$\frac{\mathrm{KL}[f(y;\mu_1,\Sigma_1)\|f(y;\mu_2,\Sigma_2)]+\mathrm{KL}[f(y;\mu_2,\Sigma_2)\|f(y;\mu_1,\Sigma_1)]}{2},$$

which has been used in [26]. An advantage of using the KL divergence is that there is a closed form for the distance between two Gaussian distributions:

$$\mathrm{KL}[f(\boldsymbol{y};\mu_1,\Sigma_1)\parallel f(\boldsymbol{y};\mu_2,\Sigma_2)] = -\frac{1}{2}\Big[\ln\Big(\frac{\det(\Sigma_2)}{\det(\Sigma_1)}\Big) + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) - N\Big],$$

where N is the dimension of y. Histogram similarity can be computed by the Bhattacharyya distance [3]:

$$d(h_1,h_2) = \sum_y \frac{\sqrt{h_1(y)h_2(y)}}{\sqrt{(\sum_z h_1(z))(\sum_z h_2(z))}}.$$

The spatiogram is a variant that weights the sum of the histogram with a Gaussian Mahalanobis distance $\psi(y)$ for the average spatial location for a bin [18]:

$$d(h_1,h_2) = \sum_y \psi(y)\frac{\sqrt{h_1(y)h_2(y)}}{\sqrt{(\sum_z h_1(z))(\sum_z h_2(z))}}.$$

4. Experimental results

We evaluate the performance of our proposed signature fingerprinting methods along with others. In order to see how well they create unique and robust signatures for individual targets, we simulate a large number of target losses in 4 real aerial videos of 1000 frames long, each from the DARPA VIVID video set. These videos contain multiple moving vehicles from various viewpoints. Some of the vehicles have similar appearance and are occasionally occluded, thus making the task challenging. The fingerprinting performance is evaluated in terms of the shelf life metric, which is the reacquisition rate for a fingerprint stored away between target appearances.

To test the proposed (M) and (PM) methods, the input segmented vehicle data is initially obtained by the tracking method in Section 2. Then, the first 500 frames of the input data are used as training data, i.e. points on the statistical manifold of a target, and the second 500 frames are used for re-identification evaluation. Therefore, we have simulated a large number of target disappearance/loss up to 500 frames. For each unknown target, the distances from its signature g to each target manifold $\mathcal{M}_{\mathrm{obj}}$, for $\mathrm{obj} = 1, 2, \ldots$, are calculated, denoted by $dist(g, \mathcal{M}_{\mathrm{obj}})$. The target is then identified by the criterion: $\arg\min_{\mathrm{obj}} dist(g, \mathcal{M}_{\mathrm{obj}})$. Figure 2 shows the identification rate in percentage averaged over all 4 videos and over each frame group, which are frames 501-667, 668-834, and 835-1000. Equivalently, the groups are organized by disappearance interval (in frames), or number of frame losses: 1-167, 168-334, and 335-500. As expected, the identification rate decreases as the interval of disappearance increases. The proposed (M) method in Figure 2-(c) implies that performance of spatial-agnostic signatures, S1 and S3, degrades much slower than the spatial-aware ones, S2 and S4, because they are invariant to rotation. The evaluation of the proposed (PM) method is performed similarly, where the signatures are combined pairwise: GLM + Histogram, GLM + Spatiogram, SGLM + Histogram, SGLM + Spatiogram, and all four signatures are combined. In these experiments, we chose $c = 2.5$ in $s(i) > \mathrm{E}[s] + c \cdot \mathrm{Var}(s)$ for the outlier removal method described in Section 3.

For fair comparison, the identification rate of the (S) method in Figure 2-(b) uses single signatures obtained at frame 500, so that the disappearance interval is equivalent. We observe that (M) performs better than (S), and (PM) in average performs better than (M).

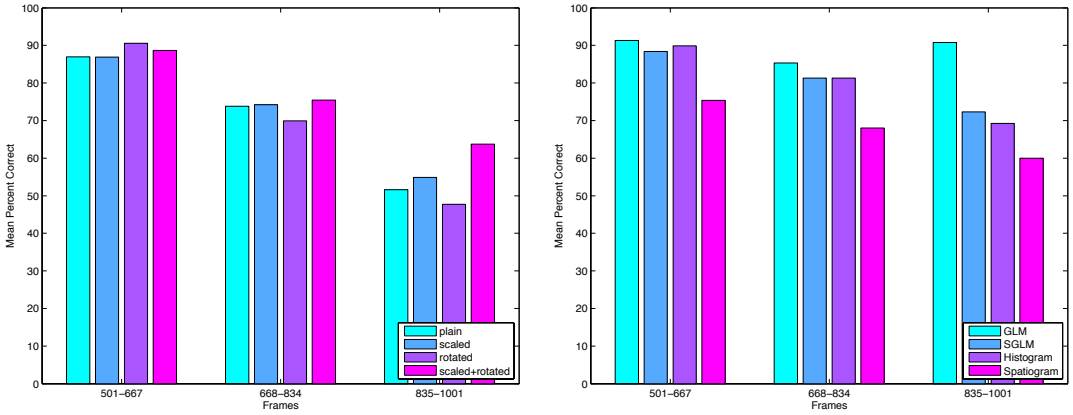
Figure 2-(a) shows results of patch-based methods.. We used the manifold-based random projection (RP) method, which was recently demonstrated for object identification in [6], because (RP) is proven to reduce the dimensionality without losing information for image manifolds. Specifically, each image patch is first vectorized, and then multiplied by a pre-generated random matrix with i.i.d Gaussian entries. The random matrix has much fewer rows than columns so that the dimension of the data is reduced. Since this method requires image patches to be rectangular and of the same size, we extract patches of size 45x45 centered at targets. For fair comparison, we further minimize the number of pixels in background by scaling and cropping to fit the target as close to the patch edge as possible. Another method to minimize the appearance of background is through rotation and cropping. In addition, we combine scaling and rotation to further optimize the conditions of targets patches. The dimension of each patch is 45x45x3, since 3 color channels used. The dimension after random projection is 150x3, which is still much larger than the dimensions of the statistical manifolds in (M) and (PM), which are 14 for GLM and 27 for SGLM.

Table 1 shows the identification rates of the best methods for each of the (RP), (S), (M), and (PM) methods. All statistics-based methods outperform (RP), showing the robustness to nuisance factors arising in real-videos over image-patch based methods.

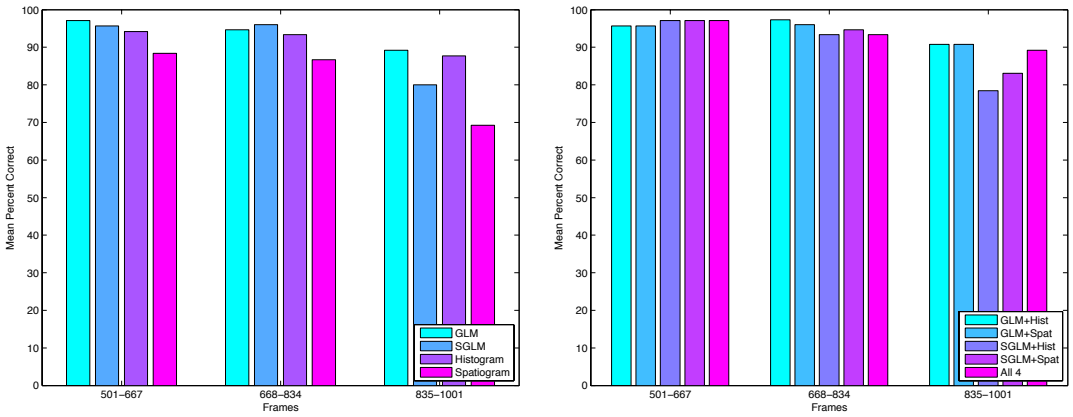
Disappear Interval	(RP) scaled + rotated	(S) GLM	(M) GLM	(PM) GLM+hist
1-167	88.68%	91.30%	97.10%	95.65%
168-334	75.42%	85.33%	94.67%	97.33%
335-500	63.74%	90.77%	89.23%	90.77%

Table 1. Comparison of best identification rates of each of the (RP), (S), (M), and (PM) methods.

Figure 1. Left: an example of segmentation obtained by applying the tracking method [21]. Right: the corresponding video frame from the DARPA VIVID video set [25].



(a) (RP) (b) (S)



(c) (M) (d) (PM)

Figure 2. Identification rate averaged over all 4 videos and each disappearance interval group with various fingerprinting methods. (a) Image patch-based random projection (RP) method. (b) Single-instance signature (S). (c) Manifold-based (M) method. (d) Product manifold (PM) method.

5. Conclusions and Future work

In this work, we propose a statistical manifold-based fingerprinting framework. The submanifolds of individual targets can be based on either single compact signatures or fully incorporated multiple signatures. In the experiments, the reacquisition rates are over 90% using the product manifolds with the single Gaussian model (GLM) even after 500 frames of target loss. The accuracy is very high although re-identification is dependent on a real tracking system, in which the segmentation of targets may not be completely precise. In the future, we plan to extend our product manifold framework to multimodal manifold framework, for data fusion from multiple sensors through the joint manifold concept [6]. In particular, we believe that multimodal manifold framework for registered EO (color) and IR data can further increase the robustness for target re-identification and extend the high accuracy in re-identification to much higher frame loss time.

References

[1] B. Amberg and T. Vetter, "GraphTrack: fast and globally optimal tracking in videos", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[2] S. T. Birchfield and S. Rangarajan, "Spatiograms Versus Histograms for Region-Based Tracking," Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2005.

[3] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society,* vol. 35, pp. 99-109, 1943.

[4] R. R. Cabrera, T. Tuytelaars, and L. V. Gool, "Efficient multi-camera detection, tracking, and identification using a shared set of Haar-features", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[5] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Towards Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 1-18, 2002.

[6] M. A. Davenport, C. Hegde, M. F. Duarte, and R. G. Baraniuk, "Joint manifolds for data fusion", *IEEE Transctions on Image Processing*, vol. 19, no. 10, pp. 2580-2594, 2010.

[7] T. B. Dinh, N. Vo, and G. Medioni, "Context tracker: exploring supporters and distractters in unconstrained environments", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[8] D. L. Donoho and C. Grimes, "Image manifolds which are ismetric to Euclidean space", Journal of Mathematical Imaging and Vision, vol. 23, 2005.

[9] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[10] K. Fragkiadaki and J. Shi, "Detection free tracking: exploiting motion and topology for segmenting and tracking under entanglement", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[11] Y. Guo, S. Hsu, Y. Shan, H. Sawhney, and R. Kumar, "Vehicle Fingerprinting for Reacquisition & Tracking in Videos", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2005.

[12] http://coewww.rutgers.edu/riul/research/code/EDISON/

[13] K. Jüngling, C. Bodensteiner, and M. Arens, "Person re-identification in multi-camera networks", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[14] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics,* vol. 22, pp. 79-86, 1951.

[15] C.-H. Kuo and R. Nevatia, "How does person identify recognition help mulri-person tracking? ", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[16] T. Lee and S. Soatto, "Learning and matching multiscale template descriptors for real-time detection, localization and tracking", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[17] H. Li, C. Shen, and Q. Shi, "Real-time visual tracking using compressive sensing", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[18] P. C. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India,* vol. 2, pp. 49-55, 1936.

[19] B. C. Matei, H. S. Sawhney, and S. Samarasekera, "Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[20] K. McLaren, "The development of the CIE 1976 (L*a*b*) uniform colour-space and colour-difference formula," *Journal of the Society of Dyers and Colourists,* vol. 92, pp. 338-341, 1976.

[21] T. N. Mundhenk, R. Sundareswara, D. R. Gerwe, and Y. Chen, "High Precision Object Segmentation and Tracking for use in Super Resolution Video Reconstruction", Proc. SPIE 7878, 78780G, 2011.

[22] J. A. Nelder and R. W. M. Wedderburn, "Generalized Linear Models," *Journal of the Royal Statistical Society A,* vol. 135, pp. 370-384, 1972.

[23] H. Pirsiavash, D. Ramana, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[24] K. Sankaranarayanan, J. W. Davis, "Object association across PTZ cameras using logistic MIL", Proc. of *IEEE International Conference on Computer Vision and Pattern Recognition,* 2011.

[25] http://www.darpa.mil/i2o/Programs/vivid/vivid.asp.

[26] Z Yao, Z Lai, W Liu, "A symmetric KL divergence based spatiogram similarity measure", *Proc. of IEEE International Conference on Image Processing*, 2011.