Background Modelling on Tensor Field for Foreground Segmentation

Rui Caseiro
ruicaseiro@isr.uc.pt
Pedro Martins
pedromartins@isr.uc.pt
Jorge Batista
batista@isr.uc.pt

Institute for Systems and Robotics,
Faculty of Science and Technology,
University of Coimbra.
3030 Coimbra - Portugal

Abstract

The paper proposes a new method to perform foreground detection by means of background modeling using the tensor concept. Sometimes, statistical modelling directly on image values is not enough to achieve a good discrimination. Thus the image may be converted into a more information rich form, such as a tensor field, to yield latent discriminating features. Taking into account the theoretically well-founded differential geometrical properties of the Riemannian manifold where tensors lie, we propose a new approach for foreground detection on tensor field based on data modeling by means of Gaussians mixtures directly on tensor domain. We introduced a online Kmeans approximation of the Expectation Maximization algorithm to estimate the parameters based on an Affine-Invariant metric. This metric has excellent theoretical properties but essentially due to the space curvature the computational burden is high. We propose a novel Kmeans algorithm based on a new family of metrics, called Log-Euclidean, in order to speed up the process, while conserving the same theoretical properties. Contrary to the affine case, we obtain a space with a null curvature. Hence, classical statistical tools usually reserved to vectors are efficiently generalized to tensors in the Log-Euclidean framework. Theoretical aspects are presented and the Affine-Invariant and Log-Euclidean frameworks are compared experimentally. From a practical point of view, results are similar to those of the Affine-Invariant framework but are obtained much faster. Theoretic analysis and experimental results demonstrate the promise and effectiveness of the proposed framework.

1 Introduction

Foreground segmentation is the process that subdivides an image into regions of interest and background. This task usually relies on the extraction of suitable features that are highly discriminative. Most of the foreground detection techniques are based on intensity or color features. However there are situations, where these features may not be distinct enough (e.g. dynamic scenes). Texture is one of the most important features, therefore its consideration can greatly improve image analysis. The structure tensor [4] [9] has been introduced for such texture analysis providing a measure of the presence of edges and their orientation. Over the years, a considerable number of background models for foreground detection have been proposed. These models can be broadly classified into *pixel-wise* and *block-wise*.

The pixel-wise models relie on the separation of statistical model for each pixel and the pixel model is learned entirely from each pixel history. The background model usually can be parametrically derived using a mixture of Gaussians or through Bayesian approaches [36] [34] [33] [29]. Once the per-pixel model was derived, the likelihood of each incident pixel color is calculated and labeled as belonging to the background or not. In situations where the density that describe the pixel data is more complex and cannot be modeled parametrically, the probability distribution of background model can be approached by non-parametric estimation methods [11] [35] [22]. In case of block-wise models, the background model of a pixel depends not only on that pixel but also on the nearby pixels. These models consider spatial information an essential element to understand the scene structure [25] [24] [31]. One major disadvantage of these methods is that the boundary of the foreground objects cannot be delineated exactly. In recent years researchers have been concentrating more on incorporating spatial aspect into background modeling to take advantage of the correlation that exists between neighbouring pixels. Thus the background model of a pixel also depends on its neighbors [32] [28][3]. Some researchers have also used texture based methods to incorporate spatial aspect into background models. Spatial variation information, such as gradient feature, helps improve the realiability of structure change detection [15] [16] [37] [14] [38].

The tensor space does not form a vector space, thus standard linear statistical techniques do not apply. We propose to account for the Riemannian geometry of the tensor manifold when computing the probability distributions used in segmentation, preserving the natural properties of the tensors. Although, the classical Euclidean operations are well adapted to general square matrices, they are practically/theoretically unsatisfactory for tensors, which are very specific matrices (symmetric positive-definite). These problems have led to the use of Riemannian metrics as an alternative (for more information see [1]). To fully circumvent these difficulties an Affine-Invariant metric [26] [27] has been proposed as a rigorous and general framework for tensors. This metric has excellent theoretical properties and provide powerful processing tools, but essentially due to the curvature induced on the tensor space the computational burden is high. To overcome this limitation, a new family of metrics called Log-Euclidean was presented in [1], while preserving excellent theoretical properties. This new approach is based on a novel vector space structure for tensors. In Log-Euclidean framework, Riemannian computations become classical Euclidean computations in the domain of matrix logarithms. This leads to simple efficient extensions of the classical tools of vector statistics to tensors. From a practical point of view yields similar results, but with much simpler and faster computations, with an experimental computation time ratio of at least 2 and sometime more in favor of the Log-Euclidean framework. In order to exploit the information present in all the components of the structure tensor, a background modeling method for tensor data is presented based on the definition of mixture of Gaussians over tensor field. Theoretical aspects are presented and the frameworks are compared experimentally.

2 Structure Tensor

The combination of color/texture features can improve the segmentation. In order to extract suitable information from an image, we focus on the combination of these features, encoded by means of a structure tensor [9]. With regard to the extraction of the texture information, we use a gradient-based method. For vector-valued images (in our work RGB) the structure tensor is defined as $[\mathbf{T} = K_\rho * (vv^T)]$ with $v = [I_x, I_y, I_r, I_g, I_b]$, where I is a vector-valued image, K_ρ is a Gaussian kernel with standard deviation ρ, (I_x, I_y) are the partial derivates of the gray image, and (I_r, I_g, I_b) are the color components. The tensor field obtained directly from the images is noisy and needs to be regularized before being further analyzed. A naive but simple and often efficient regularization method is smoothing with a Gaussian kernel.

3 Riemannian Manifolds

Henceforth, $\mathbf{S}_d^+$ denotes a symmetric positive definite matrix, $\mathbf{S}_d$ is a symmetric matrix and $\mathbf{N}_d$ is a Gaussian distribution of dimension d with zero mean. It well-known that $\mathbf{S}_d^+$ do not conform to Euclidean geometry, because the $\mathbf{S}_d^+$ space is not a vector space, e.g., the space is not closed under multiplication with negative scalers. Instead, $\mathbf{S}_d^+$ lies on a Riemannian manifold (*differentiable manifold* equipped with a Riemannian metric) [6] [17] [26] [27] [20]. A manifold is a topological space which is locally similar to an Euclidean space. Let M be a topological n-manifold. A *coordinate chart* on M is a pair (U,φ), where U is an open set of M and $\varphi\colon U \to \tilde{U}$ is a homeomorphism from U to an open set $\tilde{U} = \varphi(U) \subset \Re^n$. Given a chart (U,φ) the set U is called a *coordinate domain*. The map φ is denominated as *(local) coordinate map*, and the component functions of φ are called *local coordinates* on U, i.e., for any point $\mathbf{p} \in M$, $\varphi(\mathbf{p}) = \mathbf{x} = (\mathbf{x}^1,\ldots,\mathbf{x}^n)^T$ is the local coordinate representation of $\mathbf{p}$.

A Riemannian manifold (M,G) is a differentiable manifold M endowed with Riemannian metric G. A Riemannian metric is a collection of inner products $< .,. >_{\mathbf{p}}$, defined for every point $\mathbf{p}$ of M, on the tangent space $(T_{\mathbf{p}}M)$ of M at $\mathbf{p}$. The tangent space $T_{\mathbf{p}}M$ is simply the vector space, attached to $\mathbf{p}$, which contains the tangent vectors to all curves on M passing through $\mathbf{p}$, i.e., the set of all tangent vectors at $\mathbf{p}$. More precisely, let $\gamma(t)\colon I = [a,b] \subset \Re \to M$ denote a curve on the manifold M passing through $\gamma(a) = \mathbf{p} \in M$. The tangent vector at $\mathbf{p}$ is representated by $\dot{\gamma}(a) = d\gamma(a)/dt$. The derivatives of all possible curves compose the $T_{\mathbf{p}}M$.

Let $G_{\mathbf{p}}$ be *the local representation of the Riemannian metric* at $\mathbf{p}$. The inner product of two tangent vectors $u,v \in T_{\mathbf{p}}M$ is then expressed as $< u,v >_{\mathbf{p}} = u^T G_{\mathbf{p}} v$, inducing a norm for the tangent vectors in the tangent space such that $||u||_{\mathbf{p}}^2 = < u,u >_{\mathbf{p}}$. Distances on manifolds are defined in terms of minimum length curves between points. The *geodesic* between two endpoints $\gamma(a)$ and $\gamma(b)$ on a Riemannian manifold is locally defined as the minimum length curve $\gamma(t)\colon I = [a,b] \subset \Re \to M$ over all possible smooth curves on the manifold connecting these endpoints. This minimum lenght is called *geodesic/intrinsic distance*. The tangent vector $\dot{\gamma}(t)$ defines the instantaneous velocity of the curve and its norm $|\dot{\gamma}(t)| = \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}^{1/2}$ is the instantaneous speed. The *geodesic distance* can be calculated integrating $|\dot{\gamma}(t)|$ along γ. Taking $I = [0,1]$ for simplicity, and let $\gamma(0) = \mathbf{p}$, given a tangent vector $\dot{\gamma}(0) \in T_{\mathbf{p}}M$ there exists a unique geodesic $\gamma(t)$ starting at $\mathbf{p}$ with initial velocity $\dot{\gamma}(0)$. Therefore the geodesic $\gamma(t)$ is uniquely defined by its starting point $\mathbf{p}$ and its initial velocity $\dot{\gamma}(0)$. The endpoint $\gamma(1)$ can be computed by applying the *exponential map* at $\mathbf{p}$, such that $\gamma(1) = \exp_{\mathbf{p}}(\dot{\gamma}(0))$.

Two maps are defined for mapping points between the manifold M and a tangent plane $T_{\mathbf{p}}M$. The exponential map $\exp_{\mathbf{p}}\colon T_{\mathbf{p}}M \to M$, defined on the whole $T_{\mathbf{p}}M$, is a mapping between the $T_{\mathbf{p}}M$ and the corresponding manifold M. It maps the tangent vector $\dot{\gamma}(0)$ at point $\mathbf{p} = \gamma(0)$ to the point of the manifold $\mathbf{q} = \gamma(1)$ that is reached by the geodesic at time step one. The inverse of the exponential map is given by the logarithm map and denoted by $\log_{\mathbf{p}}\colon M \to T_{\mathbf{p}}M$. It maps any point $\mathbf{q} \in M$ to the unique tangent vector $\dot{\gamma}(0)$ at $\mathbf{p} = \gamma(0)$ that is the initial velocity of the unique geodesic $\gamma(t)$ from $\mathbf{p} = \gamma(0)$ to $\mathbf{q} = \gamma(1)$. In other words, for two points $\mathbf{p}$ and $\mathbf{q}$ on the manifold M the tangent vector to the geodesic curve from $\mathbf{p}$ to $\mathbf{q}$ is defined as $\dot{\gamma}(0) = \log_{\mathbf{p}}(\gamma(1))$. It follows that, the *geodesic distance* $\mathbf{D}(\mathbf{p},\mathbf{q})$ is given by

$$\mathbf{D}(\gamma(0),\gamma(1)) = \langle \dot{\gamma}(0), \dot{\gamma}(0) \rangle_{\gamma(0)}^{1/2} \qquad \dot{\gamma}(0) = -\nabla_{\gamma(0)} \mathbf{D}^2(\gamma(0),\gamma(1)) \tag{1}$$

The velocity $\dot{\gamma}(0)$ is computed from the squared distance gradient with respect to $\gamma(0)$ [20].

4 Tensor Statistics

We define an *image-tensor* $\mathbf{T}$ as $\mathbf{T}\colon \Omega \subset \Re^3 \mapsto \mathbf{S}_d^+$, where Ω is the original color image (3^{rd} dimension represent the color channels), $\mathbf{T}(x,y)$ is a *pixel-tensor* in image position (x,y) and $\mathbf{S}_d^+$ denotes the space of $(d \times d)$ symmetric positive definite matrices ($d = 5$ in our case). Since $(d \times d)$ symmetric matrices have only $n = (d/2)(d+1)$ independent components, applying a local coordinate chart $\varphi\colon \mathbf{S}_d^+ \mapsto \Re^n$ it is possible to associate to each tensor $\mathbf{T}(x,y) \in \mathbf{S}_d^+$ its n independent components, such that $\mathbf{S}_d^+$ is isomorphic to $\Re^n$ ($n = 15$ in our case). A tensor can be understood as the parameters (covariance matrix) of a d-dimensional normal distribution. Therefore our goal is to define statistics between multivariate normal distributions and apply it to tensor data. In order to achieve this goal, we need first to define, the mean and covariance matrix over a set of tensors, and the respective probability density function. See more details about this section in, [20] [6] [7] [21] [18] [19]. As defined by Fréchet in [13] the empirical mean tensor $\bar{\mathbf{T}}$, over a set of N random tensors $\{\mathbf{T}_i\}$, is defined as the minimizer $\mathbf{T} = \bar{\mathbf{T}}$ of the expectation $E[\mathbf{D}^2(\mathbf{T},\mathbf{T}_i)]$ and the empirical covariance matrix Λ, with respect to the mean tensor $\bar{\mathbf{T}}$, it is estimated as

$$E[\mathbf{D}^2(\mathbf{T},\mathbf{T}_i)] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{D}^2(\mathbf{T},\mathbf{T}_i) \qquad \Lambda = \frac{1}{N}\sum_{i=1}^{N}\varphi(\beta_i)\varphi(\beta_i)^{\mathbf{T}} \tag{2}$$

$$p(\mathbf{T}_i|\bar{\mathbf{T}},\Lambda) = \frac{1}{\sqrt{(2\pi)^n|\Lambda|}}\exp\left(-\frac{\varphi(\beta_i)^{\mathbf{T}}\Lambda^{-1}\varphi(\beta_i)}{2}\right) \tag{3}$$

where $\beta_i = -\nabla_{\bar{\mathbf{T}}}\mathbf{D}^2(\bar{\mathbf{T}},\mathbf{T}_i)$ and $p(\mathbf{T}_i|\bar{\mathbf{T}},\Lambda)$ define the Gaussian law on the tensor manifold. This characterization of $\mathbf{S}_d^+$ through its statistical parameters allow us to derive statistics on tensors based on different metrics. Next, we will apply these concepts to the three metrics studied in this paper. Namely, we will study the conventional Euclidean metric ($\mathbf{D}_e$), then we describe the geometry of $\mathbf{S}_d^+$ equipped with a Affine-Invariant Riemannian metric derived from the Fisher information matrix [30], from which can be induced a geodesic distance ($\mathbf{D}_a$) and finally we exploit the properties of a new family of metrics, called Log-Euclidean ($\mathbf{D}_l$).

4.1 Euclidean Metric

Using a Euclidean metric the dissimilarity $\mathbf{D}_e(\mathbf{X},\mathbf{Y})$ between tensors $\mathbf{X},\mathbf{Y} \in \mathbf{S}_d^+$ is given by the Frobenius norm of the difference

$$\left[\mathbf{D}_e(\mathbf{X},\mathbf{Y}) = |\mathbf{X}-\mathbf{Y}|_F = \sqrt{\mathrm{tr}((\mathbf{X}-\mathbf{Y})(\mathbf{X}-\mathbf{Y})^{\mathbf{T}})}\right] \qquad \left[\nabla_{\mathbf{X}}\mathbf{D}_e^2(\mathbf{X},\mathbf{Y}) = \mathbf{X}-\mathbf{Y}\right] \tag{4}$$

The gradient of the squared distance $\nabla_{\mathbf{X}}\mathbf{D}_e^2(\mathbf{X},\mathbf{Y})$ can be proved to correspond to the *difference* tangent vector. The empirical mean tensor $\bar{\mathbf{T}}_e$ over a set of N tensors $\{\mathbf{T}_i\}$, is estimated as $\bar{\mathbf{T}}_e = (1/N)\sum_{i=1}^{N}\mathbf{T}_i$. The respectively covariance matrix Λ_e can be estimated plugging $\beta_i = -\nabla_{\bar{\mathbf{T}}_e}\mathbf{D}_e^2(\bar{\mathbf{T}}_e,\mathbf{T}_i) = (\mathbf{T}_i - \bar{\mathbf{T}}_e)$ into equation 2.

4.2 Affine-Invariant Metric

Using the fact that the manifold $\mathbf{N}_d$ can be identified with the manifold of $\mathbf{S}_d^+$ matrices, a Riemannian metric on $\mathbf{S}_d^+$ can be introduced in terms of the Fisher information matrix [30]. Thus it is possible to induce several properties of $\mathbf{S}_d^+$ and derive a Gaussian law on that manifold. An invariant Riemannian metric [7] for the space of multivariate normal distributions with zero mean $\forall \mathbf{X} \in \mathbf{S}_d^+$ is given by

$$\langle u,v\rangle_{\mathbf{X}} = \frac{1}{2}\mathrm{tr}(\mathbf{X}^{-1}u\mathbf{X}^{-1}v) \tag{5}$$

where $\langle u,v\rangle_{\mathbf{X}}$ is the inner product for any tangent vectors $u,v \in \mathbf{S}_d$, in tangent space $\mathbf{T}_{\mathbf{X}}M$, relative to point $\mathbf{X}$. Let γ: $[0,1] \subset \Re \to M$ be a curve in $\mathbf{S}_d^+$, with endpoints $\gamma(0) = \mathbf{X}$ and $\gamma(1) = \mathbf{Y}$, $\forall \mathbf{X},\mathbf{Y} \in \mathbf{S}_d^+$. The geodesic defined by the initial point $\gamma(0) = \mathbf{X}$ and the tangent vector $\dot{\gamma}(0)$ can be expressed [23] as

$$\gamma(t) = \exp_{\mathbf{X}}[t\dot{\gamma}(0)] = \mathbf{X}^{\frac{1}{2}} \exp\left[(t)\mathbf{X}^{-\frac{1}{2}}\dot{\gamma}(0)\mathbf{X}^{-\frac{1}{2}}\right]\mathbf{X}^{\frac{1}{2}} \tag{6}$$

which in case of $t = 1$ correspond to the exponential map $\exp_{\mathbf{X}} : \mathbf{T}_{\mathbf{X}}M \to M$ with $\gamma(1) = \exp_{\mathbf{X}}(\dot{\gamma}(0))$. The respective logarithm map $\log_{\mathbf{X}} : M \to \mathbf{T}_{\mathbf{X}}M$ is defined as

$$\dot{\gamma}(0) = \log_{\mathbf{X}}(\mathbf{Y}) = -\mathbf{X}\log(\mathbf{Y}^{-1}\mathbf{X}) \tag{7}$$

Notice that these operators are point dependent where the dependence is made explicit with the subscript. The geodesic distance $\mathbf{D}_a(\mathbf{X},\mathbf{Y})$ between two points $\mathbf{X},\mathbf{Y} \in \mathbf{S}_d^+$, induced by the Affine-Invariant Riemannian metric, derived from Fisher information matrix was proved (*Theorem: S.T.Jensen, 1976* , see in [2]) to be given as

$$\left[\mathbf{D}_a(\mathbf{X},\mathbf{Y}) = \sqrt{\frac{1}{2}\mathrm{tr}(\log^2(\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}}))}\right] \qquad \left[\nabla_{\mathbf{X}}\mathbf{D}_a^2(\mathbf{X},\mathbf{Y}) = \mathbf{X}\log(\mathbf{Y}^{-1}\mathbf{X})\right] \tag{8}$$

The gradient of the squared distance $\nabla_{\mathbf{X}}\mathbf{D}_a^2(\mathbf{X},\mathbf{Y})$, is equal to the negative of initial velocity $\dot{\gamma}(0)$ that define the geodesic [23]. This metric exhibits all the properties necessary to be a true metric such that, positivity, symmetry, triangle inequality and is also affine invariant and invariant under inversion. Using this metric as soon as $N > 2$, a closed-form expression for the empirical mean $\bar{\mathbf{T}}_a$ of a set of N tensors $\{\mathbf{T}_i\} \in \mathbf{S}_d^+$ cannot be obtained. The mean is only *implicitly* defined based in the fact that the Riemannian barycenter exists and is unique for the manifold $\mathbf{S}_d^+$. In the literature [26], this problem is solved iteratively, for instance using a Gauss-Newton method (gradient descent algorithm) given by

$$\left[\bar{\mathbf{T}}_a^{t+1} = \exp_{\bar{\mathbf{T}}_a^t}(\mathbf{V})\right] \qquad \left[\mathbf{V} = -\frac{1}{N}\sum_{i=1}^{N}\nabla_{\bar{\mathbf{T}}_a^t}\mathbf{D}_a^2(\bar{\mathbf{T}}_a^t,\mathbf{T}_i) = -\frac{1}{N}\bar{\mathbf{T}}_a^t\sum_{i=1}^{N}\log(\mathbf{T}_i^{-1}\bar{\mathbf{T}}_a^t)\right] \tag{9}$$

where $\mathbf{V}$ is a tangent vector ($t = 1$), given by the gradient of the variance. The covariance Λ_a can be estimated plugging $\beta_i = -\nabla_{\bar{\mathbf{T}}_a}\mathbf{D}_a^2(\bar{\mathbf{T}}_a,\mathbf{T}_i) = -\bar{\mathbf{T}}_a\log(\mathbf{T}_i^{-1}\bar{\mathbf{T}}_a)$ into equation 2.

4.3 Log-Euclidean Metric

We now present the framework for the tensor space endowed with the Log-Euclidean metric [1]. Based on specific properties of the matrix exponential/logarithm on tensors, it is possible to define a *vector space structure* on tensors. The important point here is that the logarithm of a tensor $\mathbf{X}$ is unique, well defined and is a symmetric matrix $V = \log(\mathbf{X})$. Conversely, the exponential of any symmetric matrix V yields a tensor $\mathbf{X} = \exp(V)$, i.e. each symmetric matrix is associated to a tensor by the exponential. This means that under the matrix exponentiation operation, there is a one-to-one correspondence between symmetric matrices and tensors. Since there is a one-to-one mapping between the tensor space and the vector space of symmetric matrices, one can transfer to tensors the standard algebraic operations (addition $+$ and scalar multiplication .) with the matrix exponential. This defines on tensors the *logarithmic multiplication* $\odot$ and the *logarithmic scalar multiplication* $\circledast$, given by:

$$\mathbf{X}\odot\mathbf{Y} = \exp[\log(\mathbf{X}) + \log(\mathbf{Y})] \qquad \lambda \circledast \mathbf{X} = \exp[\lambda.\log(\mathbf{X})] = \mathbf{X}^{\lambda} \tag{10}$$

The operator $\odot$ is *commutative* and coincides with matrix multiplication whenever the two tensors $\mathbf{X}$,$\mathbf{Y}$ commute in the matrix sense. With $\odot$ and $\circledast$ the tensor space has by construction a *vector space structure*, which is not the usual structure directly derived from addition and scalar multiplication on matrices.

Among Riemannian metrics in Lie groups, the most suitable in practice, when they exist, are *bi-invariant* metrics, i.e., metrics that are invariant by multiplication and inversion. These metrics are used in differential geometry to generalize to Lie groups a notion of mean that is consistent with multiplication and inversion. For our tensor Lie group, *bi-invariant* metrics exist and are particularly simple [1]. Their existence simply results from the commutativity of logarithmic multiplication between tensors, and since they correspond to Euclidean metrics in the domain of logarithms are called *Log-Euclidean* metrics. The tensor vector space with a Log-Euclidean metric is in fact *isomorphic* (algebraic structure of vector space is conserved) and *isometric* (distances are conserved) with the corresponding Euclidean space of symmetric matrices. Hence, the Riemannian framework for statistics is extremely simplified. Results obtained on logarithms are mapped back to the tensor domain with the exponential. The Log-Euclidean metric for the space of $\mathbf{S}_d^+$, defined on the tangent space $\mathbf{S}_d$ is given $\forall \mathbf{X} \in \mathbf{S}_d^+$ as follows

$$\langle u,v\rangle_{\mathbf{X}} = \langle \partial_{\mathbf{X}}\mathbf{log}.u, \partial_{\mathbf{X}}\mathbf{log}.v\rangle_{\mathbf{Id}} \tag{11}$$

where $\langle u,v\rangle_{\mathbf{X}}$ is the inner product for any tangent vectors $u,v \in \mathbf{S}_d$, in tangent space $\mathbf{T_X}M$, relative to point $\mathbf{X}$. The operator $\partial_{\mathbf{X}}\mathbf{log.}$ correspond to the *differential* of the matrix logarithm. Let γ: $[0,1] \subset \Re \rightarrow M$ be a curve in $\mathbf{S}_d^+$, with $\gamma(0) = \mathbf{X}$ and $\gamma(1) = \mathbf{Y}$, $\forall \mathbf{X},\mathbf{Y} \in \mathbf{S}_d^+$. The geodesic defined by the $\gamma(0) = \mathbf{X}$ and the tangent vector $\dot{\gamma}(0)$ can be expressed as

$$\gamma(t) = \exp_{\mathbf{X}}[t\dot{\gamma}(0)] = \exp[\log(\mathbf{X}) + \partial_{\mathbf{X}}\mathbf{log}.[t\dot{\gamma}(0)]] \tag{12}$$

which in case of $t = 1$ correspond to the exponential map $\exp_{\mathbf{X}}\colon \mathbf{T_X}M \rightarrow M$ with $\gamma(1) = \exp_{\mathbf{X}}(\dot{\gamma}(0))$. The respective logarithm map $\log_{\mathbf{X}}\colon M \rightarrow \mathbf{T_X}M$ is defined as

$$\dot{\gamma}(0) = \log_{\mathbf{X}}(\mathbf{Y}) = \partial_{\log(\mathbf{X})}\mathbf{exp}.[\log(\mathbf{Y}) - \log(\mathbf{X})] \tag{13}$$

where the operator $\partial_{\mathbf{X}}\mathbf{exp.}$ correspond to the *differential* of the matrix exponential. Since the Log-Euclidean metrics correspond to Euclidean metrics in the domain of logarithms, the shortest path going from the tensor $\mathbf{X}$ to the tensor $\mathbf{Y}$ is a straight line in that domain. Hence, the interpolation between two tensors is simplified, and is expressed as

$$\gamma(t) = \exp[(1-t)\log(\mathbf{X}) + t\log(\mathbf{Y})] \tag{14}$$

The geodesic distance $\mathbf{D}_l(\mathbf{X},\mathbf{Y})$ between these points $\mathbf{X},\mathbf{Y} \in \mathbf{S}_d^+$, induced by this metric is also extremely simplified as follows

$$\left[D_l(\mathbf{X},\mathbf{Y}) = \sqrt{\mathrm{tr}[(\log(\mathbf{Y}) - \log(\mathbf{X}))^2]}\right] \qquad \left[\nabla_{\mathbf{X}} D_l^2(\mathbf{X},\mathbf{Y}) = -\dot{\gamma}(0)\right] \tag{15}$$

We consider that the gradient $\nabla_{\mathbf{X}}\mathbf{D}_l^2(\mathbf{X},\mathbf{Y})$, is equal to the negative of initial velocity $\dot{\gamma}(0)$ that define the geodesic. As one can see, the Log-Euclidean distance is much simpler than the equivalent Affine-Invariant distance where matrix multiplications, square roots, and inverses are used. The greater simplicity of Log-Euclidean metrics can also be seen from the mean in the tensor space. In this case the Fréchet mean of a set of N tensors $\{\mathbf{T}_i\} \in \mathbf{S}_d^+$ is a direct generalization of the geometric mean of positive numbers and is given *explicitly* by

$$\bar{\mathbf{T}}_l = \exp\left[\frac{1}{N}\sum_{i=1}^{N}\log(\mathbf{T}_i)\right] \qquad \Lambda = \frac{1}{N}\sum_{i=1}^{N}\varphi(\beta_i)\varphi(\beta_i)^{\mathbf{T}} \tag{16}$$

This closed form equation makes the computation of Log-Euclidean means straightforward. Pratically, one simply uses the usual tools of Euclidean statistics on the logarithms and maps the results back to the tensor vector space with the exponential. This is theorically fully justifed because the tensor Lie group endowed with a bi-invariant metric (i.e. here a Log-Euclidean metric) is *isomorphic*, *diffeomorphic* and *isometric* to the additive group of symmetric matrices [1]. When the Fréchet expectation is uniquely defined, one can also compute centered moments of superior order like the covariance. The covariance matrix

Λ_l can be estimated plugging $\beta_i = -\nabla_{\bar{\mathbf{T}}_l} \mathbf{D}_l^2(\bar{\mathbf{T}}_l, \mathbf{T}_i)$ into equation 2. In terms of elementary operations like distance, geodesics and means, the Log-Euclidean case provides much simpler formulae than in Affine-Invariant case. However, we see that the Riemannian exponential/logarithm mappings are complicated in the Log-Euclidean case by the use of the differentials of the matrix exponential/logarithm. For general matrices, one has to compute the series

$$\partial_{\mathbf{X}} \mathbf{exp.}(u) = \sum_{k=1}^{+\infty} \frac{1}{k!} \left[\sum_{i=0}^{k-1} u^i \mathbf{X} u^{(k-i-1)} \right] \tag{17}$$

This cost would probably be prohibitive if we had to rely on numerical approximation methods. However, in the case of symmetric matrices, the differential is simplified. Using spectral properties of symmetric matrices, one can compute an explicit and very simple/efficiently closed-form expression for the differential of both matrix logarithm and exponential (see more details in [12]). Let $u = RDR^T$ where D is a diagonal matrix, and consider $Z = R\mathbf{X}R^T$. As D is diagonal, one can access the (l,m) coefficient of the resulting matrix as :

$$\left[\partial_{\mathbf{X}} \mathbf{exp.}(u) = R^T \partial_{\mathbf{Z}} \mathbf{exp.}(D) R \right] \qquad \left[[\partial_{\mathbf{Z}} \mathbf{exp.}(D)]_{(l,m)} = \frac{\exp(d_l) - \exp(d_m)}{d_l - d_m} [Z]_{(l,m)} \right] \tag{18}$$

5 Background Modeling

In order to model the background we use a mixture of K Gaussians on tensor domain as proposed in [8] for DT-MRI segmentation. Based on the definition of a Gaussian law on tensor space, we can define a mixture of Gaussians (GMM) as follows

$$p(\mathbf{T}_i | \Theta) = \sum_{k=1}^{K} \omega_k \frac{\exp\left(-(1/2) \varphi(\beta_{i,k})^T \Lambda_k^{-1} \varphi(\beta_{i,k}) \right)}{\sqrt{(2\pi)^n |\Lambda_k|}} \tag{19}$$

where each gaussian density $\mathcal{N}(\mathbf{T}_i | \bar{\mathbf{T}}_k, \Lambda_k)$ is a *component of the mixture*. Each component is characterized by, a *mixing coeficient* ω_k (prior), a mean tensor $\bar{\mathbf{T}}_k$ and a covariance matrix Λ_k. Θ denotes the vector containing all the parameters of the given mixture. The matrix $\beta_{i,k} = -\nabla_{\bar{\mathbf{T}}_k} \mathbf{D}^2(\bar{\mathbf{T}}_k, \mathbf{T}_i)$ depends on the chosen metric. A general technique for finding maximum likelihood estimators in latent variable models is the Expectation-Maximization (EM) algorithm [10]. An exact EM algorithm implementation as proposed in [8] for the Affine-Invariant case can be a costly procedure. In order to reduce the processing time we propose a algorithm based on an online K-means approximation of EM, adapted from the version presented in [34]. The foreground detection is performed in the same way as in [34].

5.1 Kmeans - Euclidean Metric

In the Euclidean case, the algorithm proposed is similar to the Stauffer's algorithm [34] except for the fact that the pixel is modeled using tensors instead of vectors (color). The new mixture parameters combine the prior information with the observed sample. The model parameters are updated using an exponential decay scheme with learning rates (α and ρ). The mixture weights are updated using $\omega_k^t = (1-\alpha)\omega_k^{t-1} + (\alpha)(M_k^t)$, where M_k^t is 1 for the model which matched and 0 for the remaining models. The parameters of the distribution which matches the new observation $(\mathbf{T}_i)$ are updated as follows

$$\left[\bar{\mathbf{T}}_k^t = (1-\rho)\bar{\mathbf{T}}_k^{t-1} + \rho \mathbf{T}_i \right] \qquad \left[\Lambda_k^t = (1-\rho)\Lambda_k^{t-1} + \rho \varphi(\beta_{i,k}^t)^T \varphi(\beta_{i,k}^t) \right] \tag{20}$$

$$\left[\beta_{i,k}^t = -\nabla_{\bar{\mathbf{T}}_k^t} \mathbf{D}_e^2(\bar{\mathbf{T}}_k^t, \mathbf{T}_i) = \mathbf{T}_i - \bar{\mathbf{T}}_k^t \right] \qquad \left[\rho = \alpha \mathcal{N}(\mathbf{T}_i | \bar{\mathbf{T}}_k^{t-1}, \Lambda_k^{t-1}) \right] \tag{21}$$

5.2 Kmeans - Affine-Invariant Metric

The tensor mean ($\bar{\mathbf{T}}_k^t$) update equation presented (20) can only be directly applied in the Euclidean case. As mentioned previously we need to take into account the Riemannian geometry of the tensor manifold to apply the geodesic metrics (Affine-Invariant, Log-Euclidean). We propose an approximation method to update the tensor mean, based on the concept of interpolation between two tensors. The tensors interpolation can be seen as a walk along the geodesic joining the two tensors. In the Affine-Invariant case, a closed-form expression is given by the exponential map (6). In order to simplify we change the notation as follows

$$\left[\mathbf{Z} = \bar{\mathbf{T}}_k^t\right] \qquad \left[\mathbf{X} = \bar{\mathbf{T}}_k^{t-1} = \gamma(0)\right] \qquad \left[\mathbf{Y} = \mathbf{T}_i = \gamma(1)\right] \tag{22}$$

Let $\gamma(t)\colon [0,1] \subset \Re \rightarrow M$ be the geodesic with $\gamma(0) = \mathbf{X}$ and $\gamma(1) = \mathbf{Y}$. $\mathbf{Z}$ is the interpolation between $\mathbf{X}$ and $\mathbf{Y}$ at $t = \rho$

$$\left[\dot{\gamma}(0) = -\mathbf{X}\log(\mathbf{Y}^{-1}\mathbf{X})\right] \qquad \left[\mathbf{Y} = \exp_{\mathbf{X}}(\dot{\gamma}(0))\right] \qquad \left[\mathbf{Z} = \exp_{\mathbf{X}}(\rho\dot{\gamma}(0))\right] \tag{23}$$

Plugging $\dot{\gamma}(0)$ into (6), the point $\mathbf{Z}$ on the manifold that is reached by the geodesic $\gamma(t)$ at time $t = \rho$ is estimated as

$$\mathbf{Z} = \gamma(\rho) = \mathbf{X}^{\frac{1}{2}} \exp\left[(\rho)\mathbf{X}^{-\frac{1}{2}}\left[-\mathbf{X}\log(\mathbf{Y}^{-1}\mathbf{X})\right]\mathbf{X}^{-\frac{1}{2}}\right]\mathbf{X}^{\frac{1}{2}} \tag{24}$$

$$\beta_{i,k}^t = -\nabla_{\bar{\mathbf{T}}_k^t}\mathbf{D}_a^2(\bar{\mathbf{T}}_k^t, \mathbf{T}_i) = -\nabla_{\mathbf{Z}}\mathbf{D}_a^2(\mathbf{Z}, \mathbf{Y}) = -\mathbf{Z}\log(\mathbf{Y}^{-1}\mathbf{Z}) \tag{25}$$

5.3 Kmeans - Log-Euclidean Metric

In the Log-Euclidean case, a closed-form expression for interpolation between two tensors is given by the equation (14). Let $\gamma(t)\colon [0,1] \subset \Re \rightarrow M$ be the geodesic with $\gamma(0) = \mathbf{X}$ and $\gamma(1) = \mathbf{Y}$. The interpolation point $\mathbf{Z}$ between $\mathbf{X}$ and $\mathbf{Y}$ on the manifold that is reached by the geodesic $\gamma(t)$ at time $t = \rho$ is estimated as

$$\mathbf{Z} = \gamma(\rho) = \exp[(1-\rho)\log(\mathbf{X}) + \rho\log(\mathbf{Y})] \tag{26}$$

$$\beta_{i,k}^t = -\nabla_{\bar{\mathbf{T}}_k^t}\mathbf{D}_l^2(\bar{\mathbf{T}}_k^t, \mathbf{T}_i) = -\nabla_{\mathbf{Z}}\mathbf{D}_l^2(\mathbf{Z}, \mathbf{Y}) = \partial_{\log(\mathbf{Z})}\mathbf{exp.}[\log(\mathbf{Y}) - \log(\mathbf{Z})] \tag{27}$$

6 Results

In order to analyze the effectiveness of the proposed methods, we conduct several experiments on two sequences presented in previous literature. The first scene (Sequence1) is the HighWayI sequence from ATON project (http://cvrr.ucsd.edu/ aton/shadow/). The second scene (Sequence2) is the moving camera sequence from [32]. The groundtruth foreground was obtained by manual segmentation. In this section several results of applying the proposed methods to these sequences are presented. Two widely-used vector space methods, namely Mixture of Gaussians (GMM) [34] and Kernel Density Estimation (KDE) [11] are employed to compare with the proposed tensor framework (using the three metrics presented). We analyze the performance of both vector methods (GMM,KDE) using two types of features sets, namely a set with RGB data (I_r, I_g, I_b) and a set incremented with image gradients (I_r, I_g, I_b, I_x, I_y). It is stressed that no morphological operators were used.

Traditional background modeling methods assume that the scenes are of static structures with limited perturbation. Their performance will notably deteriorate in the presence of dynamic backgrounds. In dynamic scenes, although some pixels significantly changes over time, they should be considered as background. As shown in Fig. 1, the traditional vector methods (GMM,KDE) can not accurately detect moving objects in dynamic scenes. They label large numbers of moving background pixels as foreground and also output a huge amount of false negatives on the inner areas of the moving object. However, the proposed framework can accurately distinguish moving background pixels and true moving objects. Our method handles small dynamic background motions, since the proposed procedure integrates spatial texture (pixel based and region based information by tensor matrices), considering the correlation between pixels. It uses features which effectively models the spatial correlations of neighbors pixels, which is very important to accurately label those moving background pixels. The vector GMM methods at the beginning of the sequences which do not include foreground objects performs poorly and detected as foreground a lot of background pixels. The reason for this, is because these methods exploit only simple features, and so need to take longer time to train the background models than the proposed methods in order to accurately detect the foreground pixels. On the other hand, the proposed framework handles dynamic motions immediately and achieves accurate detection at the beginning of the sequences. The spatial correlations provide the substantial evidence for labeling the center pixel and they are exploited to sustain high levels of detection accuracy.

	Sequence 1		**Sequence 2**	
Methods	TPR	FPR	TPR	FPR
GMM [**Ir, Ig, Ib**]	55.20	16.07	58.00	17.12
GMM [**Ir, Ig, Ib, Ix, Iy**]	61.30	13.50	63.20	14.65
KDE [**Ir, Ig, Ib**]	62.80	12.83	65.60	13.32
KDE [**Ir, Ig, Ib, Ix, Iy**]	69.10	10.95	71.50	11.26
GMM [**Tensor**] Euclidean	72.50	8.72	76.40	8.95
GMM [**Tensor**] Affine-Invariant	93.00	2.08	90.50	2.24
GMM [**Tensor**] Log-Euclidean	91.40	2.20	89.60	2.45

Table 1: True positive ratio (**TPR**) , False positive ratio (**FPR**)

7 Conclusions

We proposed a novel method to perform background modeling using the tensor concept. The tensor was used to convert the image into a more information rich form, encoding color and texture data. We review the geometrical properties of the tensor space and focus on the characterization of the mean, covariance and generalized normal law on that manifold. In order to exploit the information present in all the tensor components and taking into account the natural Riemannian structure of the tensor manifold, GMM on tensor fields have been introduced to approximate the probability distribution of tensor data. This probabilistic modeling directly on tensor domain was employed to formulate foreground segmentation on tensor field. As this work shows, new points of view on the tensor space can lead to significantly faster and simpler computations. We proposed a new K-means approximation of the EM algorithm to estimate the mixture parameters based on a new Riemannian metric, called Log-Euclidean, in order to speed up the process. Based on a novel vector space structure for tensors, the Log-Euclidean framework transforms Riemannian computations on tensors into Euclidean computations on vectors in the logarithms domain. This leads to simple efficient extensions of the classical tools of vector statistics to tensors. This new metric also has the same excellent theoretical properties as the Affine-Invariant metric. From a practical point of view yield very similar results and are obtained much faster, with an experimental computation time ratio of at least 2 and sometime more in favor of the Log-Euclidean.

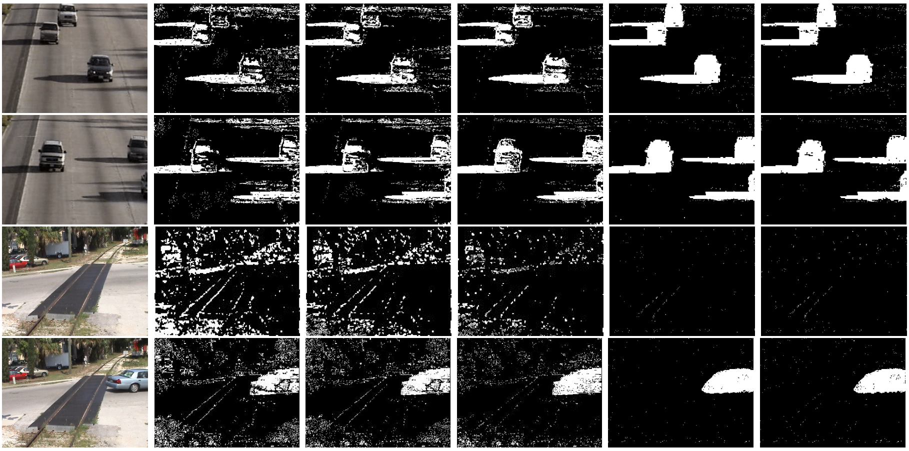

Figure 1: Column left to right : Original frames , GMM [**Ir,Ig,Ib,Ix,Iy**] , KDE [**Ir,Ig,Ib,Ix,Iy**] , **Tensor** Euclidean , **Tensor** Affine-Invariant , **Tensor** Log-Euclidean

Acknowledgements

This work was supported by BRISA, Auto-estradas de Portugal, S.A.

References

[1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2006.

[2] C. Atkinson and A. F. Mitchell. Rao distance measure. *Sankhya: The Indian Journal of Statistics*, 43(3):345–365, 1981.

[3] S. Babacan and T. Pappas. Spatiotemporal algorithm for joint video segmentation and foreground detection. *European Signal Processing Conference - (EUSIPCO)*, 2006.

[4] J. Bigun and J. Wiklund. Multidimensional orientation estimation applications texture analysis optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence - (TPAMI)*, 13(8):775–790, 1991.

[5] Manfredo P. Carmo. *Riemannian Geometry*. Birkhauser, 1992.

[6] Carlos A. Castaño-Moraga, C. Lenglet, and R. Deriche. A fast rigorous anisotropic smoothing method DT-MRI. *IEEE International Symposium on Biomedical Imaging - (ISBI)*, pages 93–96, 2006.

[7] Carlos A. Castaño-Moraga, C. Lenglet, and R. Deriche. A Riemannian approach to anisotropic filtering of tensor fields. *Signal Processing*, 87(2):263–276, 2007.

[8] R. de Luis Garcia and C. Alberola-Lopez. Mixtures of gaussians on tensor fields for DT-MRI segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention - (MICCAI)*, 2007.

[9] R. de Luis Garcia, M. Rousson, and R. Deriche. Tensor processing for texture and colour segmentation. *Scandinavian Conference on Image Analysis*, 2005.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1): 1–38, 1977.

[11] A. M. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction. *European Conference on Computer Vision - (ECCV)*, 2000.

[12] P. Fillard, V. Arsigny, X. Pennec, and N. Ayach. Joint Estimation and Smoothing of Clinical DT-MRI with a Log-Euclidean Metric. *Technical Report - INRIA*, (RR-5607), 2005.

[13] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310, 1948.

[14] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence - (TPAMI)*, 28(4):657–662, 2006.

[15] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld. Detection and location of people in video images using adaptive fusion of color and edge information. *IEEE Computer Vision and Pattern Recognition - (CVPR)*, 2000.

[16] O. Javed, K. Shafique, and M. Shah. A hierarchical approach to robust background subtraction using color and gradient information. *IEEE Workshop on Motion and Video Computing*, 2002.

[17] John M. Lee. *Introduction to Smooth Manifolds*. Springer, 2003.

[18] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics Multivariate Normal Distributions: Geometric Approach Application Diffusion Tensor MRI. *Technical Report - INRIA*, (RR-5242), 2004.

[19] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Toward Segmentation 3D Probability Density Fields Surface Evolution: Application Diffusion MRI. *Technical Report - INRIA*, (RR-5243), 2004.

[20] C. Lenglet, M. Rousson, and R. Deriche. A statistical framework for DTI segmentation. *IEEE International Symposium on Biomedical Imaging - (ISBI)*, pages 794–797, 2006.

[21] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on Manifold Multivariate Normal Distributions Theory Application Diffusion Tensor MRI Processing. *Journal Mathematical Imaging Vision - (JMIV)*, 25(3):423–444, 2006.

[22] A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. *IEEE Computer Vision and Pattern Recognition - (CVPR)*, 2004.

[23] Maher Moakher. A Differential Geometric Approach to Geometric Mean Symmetric Positive-Definite Matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3): 735–747, 2005.

[24] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh. Background modeling and subtraction of dynamic scenes. *IEEE International Conference on Computer Vision - (ICCV)*, 2003.

[25] N. Oliver, B. Rosario, and A. P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence - (TPAMI)*, 22(8):831–843, 2000.

[26] X. Pennec, P. Fillard, and N. Ayache. A Riemannian Framework for Tensor Computing. *International Journal Computer Vision - (IJCV)*, 66(1):41–66, 2006.

[27] Xavier Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal Mathematical Imaging Vision - (JMIV)*, 25(1):127–154, 2006.

[28] R. Pless. Spatio-temporal background models for outdoor surveillance. *EURASIP - Journal on Applied Signal Processing*, 2005:2281–2291, 2005.

[29] F. Porikli and O. Tuzel. Bayesian background modeling for foreground detection. *International Workshop on Video Surveillance & Sensor Networks*, 2005.

[30] C. R. Rao. Information accuracy attainable in estimation statistical parameters. *Bulletin Calcutta Mathematical*, 37:81–91, 1945.

[31] M. Seki, T. Wada, H. Fujiwara, and K. Sumi. Background subtraction based on cooccurrence of image variations. *IEEE Computer Vision and Pattern Recognition - (CVPR)*, 2003.

[32] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence - (TPAMI)*, 27(11): 1778–1792, 2005.

[33] A. Shimada, D. Arita, and R. Taniguchi. Dynamic control of adaptive mixture-of-gaussians background model. *IEEE International Conference on Video and Signal Based Surveillance - (AVSS)*, 2006.

[34] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. *IEEE Computer Vision and Pattern Recognition - (CVPR)*, 1999.

[35] T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi. A fast algorithm for adaptive background model construction using parzen density estimation. *IEEE International Conference on Video and Signal Based Surveillance - (AVSS)*, 2007.

[36] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence - (TPAMI)*, 19(7):780–785, 1997.

[37] J. Yao and J. M. Odobez. Multi-layer background subtraction based on color and texture. *IEEE Computer Vision and Pattern Recognition - (CVPR)*, 2007.

[38] S. Zhang, H. Yao, and S. Liu. Dynamic background modeling and subtraction using spatio-temporal local binary patterns. *IEEE International Conference on Image Processing - (ICIP)*, 2008.