

A Random Field Model for Improved Feature Extraction and Tracking

Xiaotong Yuan, Stan Z. Li

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Science
Beijing, China, 100080

Abstract

This paper presents a novel method for illumination-invariant and contrast preserving feature extraction, aimed at improving performance of tracking under complex light condition. Features to be extracted are represented as a weight field. An energy function of the field is defined as an approximate variance in robust statistics. A simple non-linear iterative rule is derived to compute the optimal field. The optimal field is shown to be invariant to global illumination switching, and preserving target/background contrast. We incorporate the feature extraction method into a mean-shift tracker and this achieves reliable results on real-world sequences in complex scenes and varying illumination.

1 Introduction

Random field modeling (e.g. Markov Random Fields, MRF) has been used for solving many image analysis problems, including restoration, segmentation, edge-preserving filtering, reconstruction in inverse problems, target detection and tracking, etc.[9, 1, 12]. Roughly speaking, there are three advantages of such techniques: Firstly, it provides a solution for extracting some intrinsic information (e.g. pixel class labels represented as a discrete MRF) from high dimensional observation data. Secondly, prior constraints about the shape and average size of homogeneous regions in an image can be incorporated into the model in a systematic way. Finally, even when exact optimal estimate cannot be precisely computed for the field, it is still possible to approximate the estimate that work well in many cases.

Online optimal tracking feature extraction is an important aspect in active visual tracking. Decades of research have yielded an arsenal of powerful algorithms. While a full review on this topic is beyond the scope of this paper, we focus on two kinds of tracking features that have received much attention in the last few decades of research, especially on template-matching trackers:

- Illumination-invariant photometric features (IIPF)
- Discriminative tracking features (DTF)

One necessary condition for successful template-matching based trackers is the temporal stability of targets' feature such as dominant color and probability distribution. However, the commonly used template-matching trackers are based primarily on tracking photometric variables (such as intensity, color, or texture), which are illumination dependent and may change rapidly and drift away from targets in situation of illumination switch. The extraction of IIPF, which aims to cope with the feature drift problem, is therefore a key issue in building such trackers. Linear subspace illumination models were used in [6] to model illumination changes. Incremental learning of eigenspace representation of original image is adopted in [10] to reflect the illumination change of target. An illumination-invariant optical flow field, which is robust to shadow, is constructed for visual correspondence and implemented through graph cuts algorithm in [5]. IIPF also received much attention in face detection and recognition [2]. However, all these methods need an off-line training stage (e.g. [6, 10]), or they may be computationally too complicated for real-time performance (e.g. [2, 5]). These limit their application in online visual tracking.

At the same time, in many applications, the tracking performance depends highly on how the appearance of a target differs from the its nearby distractors (e.g. background, group targets). The DTF extraction [11, 3], which aims to enhance the discriminability of target and its distractors, provides an optimal way for object/distractor classification. They lead to improved tracking performance in situation of low object/distractor contrast.

In this work, we develop a novel feature extraction approach for visual target tracking using random field modeling techniques. We define a cost function that includes a *robust variance term*, a *preference term*, and a *regularization term*. In terms of random field theory, the regularization term corresponds to a smoothness prior, the variance term and the preference term together correspond to a model of

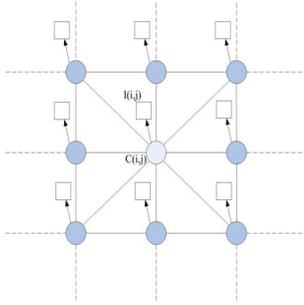


Figure 1. A hidden random field representation (with eight-neighborhood system) for pixels' weights.

outlier indicators. A simple nonlinear iterative rule, which maps the input image into a weight field, is derived using this technique. The optimal weight field has the advantages of both IIPF and DTF.

While the energy functional used in [7] is of a similar form to ours, their main concern is performing background modeling and the motion segmentation in a coupled way. In our method, we take into consideration both illumination change and contrast enhancement and model them in the energy function.

The resulting field model can be incorporated with a wide variety of template-matching trackers based on appearance models (e.g. probability distribution). Here, we consider applying it into Mean-Shift tracker [4], which is well suited for tracking deformable and connected objects.

2 Random Field Modeling

Random field modelling provides a mathematical foundation to make a global inference using local information. In this paper, we construct a hidden weight field to represent tracking feature. Let I denotes the current image frame observed from the pixel lattice Ω and indexed by coordinates (i, j) . The raw feature $I(i, j)$ is represented in a d -dimensional vector space. The desired feature field is modelled as an undirected graph $G(C) = \{\Omega, E\}$, where each site $(i, j) \in \Omega$ represents the point-wise hidden weight $C(i, j)$ to be inferred. In this model, $C(i, j)$ is real-valued in the interval $[0, 1]$. Each hidden node is connected to its neighborhood nodes, thus forming a field, as shown in fig.1.

2.1 Definition of Objective Function

Our object is to recover a nonlinear projection function $g : I \mapsto [0, 1]$ that maps high dimensional image feature vector into a real weight value. The inference is based on the viewpoint of robust statistics by estimating the pixels'

robust mean value μ , by regarding C as a 2D outlier indicator function. In this context, we also call $G(C)$ *outlier indicator field*. For better readability, we consider hereafter the 1-dimensional feature space, and all the analysis can be directly extended for high dimensional cases. We search for the robust mean μ and outlier indicator function C as the solution of the following minimization problem

$$\begin{aligned} \inf_{\mu, C} E(\mu, C) &= \iint_{\Omega} C^2 \frac{(I - \mu)^2}{\sigma_I^2} dx dy \\ &+ \eta \iint_{\Omega} (1 - C)^2 dx dy \\ &+ \alpha \iint_{\Omega} \varphi(|\nabla C|) dx dy \end{aligned} \quad (1)$$

where $\varphi(\cdot)$ is some convex and strictly decreasing function, σ_I^2 is the variance of current image I , η and α are positive parameters that control the tradeoff among these terms.

There are three terms in the objective functional and getting the minimum of the functional means that we want each term to be small, having in mind the phenomena of the compensations. The first term means that μ should equals to the mean value of pixels whose $C(i, j)$ are close to 1. The second term means that we want the field site $C(i, j)$ to be close to one, that is, we give a preference to the inlier. However, if the data $I(i, j)$ is too faraway from the supposed mean value μ , then the difference $\frac{(I(i, j) - \mu)^2}{\sigma_I^2}$ will be high, and to compensate this value, the minimization process will force $C(i, j)$ to be zero. Therefore, the function $C(i, j)$ can be interpreted as the indicator of outlier pixels. The third term is a regulation term, by which the smoothness can be promised for outlier indicator field $G(C)$. The proposed energy function (1) can also be interpreted by a generalized Bayesian rule for field model[12], that is, the three terms separately corresponds to the *local condition likelihood*, *the local prior* and *the neighborhood prior*. An analysis of mathematical properties for problem (1) is beyond the scope of this paper and we therefore refer readers to [7] for details

2.2 Optimization Algorithm

The energy function in (1) can be minimized via half quadratic method [1, 7]. By introducing the dual variable d and starting from an initial estimate (μ^0, C^0, d^0) , the algorithm consists in minimizing the Euler-Lagrange equations, which can be written as

$$\mu^{n+1} = \frac{\iint_{\Omega} (C^n)^2 I dx dy}{\iint_{\Omega} (C^n)^2 dx dy} \quad (2)$$

$$C^{n+1} (I - \mu^{n+1})^2 + \eta \sigma_I^2 (C^{n+1} - 1) - \alpha \sigma_I^2 \text{div}(d^n \nabla C^{n+1}) = 0 \quad (3)$$

$$d^{n+1} = \frac{\varphi'(|\nabla C^{n+1}|)}{2|\nabla C^{n+1}|} \quad (4)$$

with discretized Neumann condition at the boundaries. Notice that (4) gives explicitly d^{n+1} while for a fixed d^n , C^{n+1} is a solution of a linear equation. After discretizing in space we have that $(C^{n+1}(i, j))_{(i, j) \in \Omega}$ is the solution of a linear system which can be solved iteratively by the Gauss-Seidel method. Besides variational method, stochastic algorithms, such as belief propagation[8], can also be adopted to obtain approximation to optimal solution of (1).

However, the numerical procedure mentioned above is rather time consuming, mainly due to the third smoothing term in (1). The smooth constrain is unnecessarily argued in each iteration step (3) until convergence (see sect.3). In the following work, we therefore set $\alpha = 0$ to speed up the algorithm and only make some post-processing to smooth the resulting optimal field. The following is the two-step iteration algorithm

$$\mu^{n+1} = \frac{\iint_{\Omega} (C^n)^2 I \, dx \, dy}{\iint_{\Omega} (C^n)^2 \, dx \, dy} \quad (5)$$

$$C^{n+1} = \frac{\eta \sigma_I^2}{\eta \sigma_I^2 + (I - \mu^{n+1})^2} \quad (6)$$

The iteration starts from $C^0(i, j) = 1, (i, j) \in \Omega$.

In this iteration algorithm, the only parameter need to be set is η . In the following section, we will provide a numerical study for the iteration described above and an empirical setting rule for η .

3 Numerical Study

When doing the iteration according to rules (5) and (6), two important questions arise: (i) is there a unique minimum point C (together with μ) for (1), and (ii) does the iteration rules converge?

We give in the following two propositions to answer these two questions and introduce the concept of *HRI²WF* Image. Due to the page limit, we do not give the proof here, and refer readers to [1] and [7] for details.

Proposition 1: A sufficient condition under which the minimization problem (1) has a unique solution is

$$\eta \geq \frac{3[\sup(I) - \inf(I)]^2}{\sigma_I^2}$$

Proposition 2: Denote $\mathcal{E}^n = \mathcal{E}(\mu^{n+1}, C^n)$, then the sequences $\{\mathcal{E}^n, n \geq 0\}, \{\mu^n, n \geq 0\}$ and $\{C^n, n \geq 0\}$ are convergent.

Remark: From proposition 1 and 2 we can tell that if $\eta \geq \frac{3[\sup(I) - \inf(I)]^2}{\sigma_I^2}$, then from any initial estimation, the two-step iteration algorithm presented in 2.2 will converge

to a unique minimizer for problem (1). From this point of view, the parameter η should be set large enough to promise the uniqueness of solution. On the other hand, from the interpretation of three terms of (1) in 2.1 we know that η plays the role of compensating the value $(I(i, j) - \mu)^2 / \sigma_I^2$ when it is relatively large. If we set $\eta \geq 3[\sup(I) - \inf(I)]^2 / \sigma_I^2$, then from (7) it can be easily seen that C^n will always greater than 0.75, and from (6) we can see that μ^n will approach just the arithmetic mean value of all the samples. In such a case, the term 2 is of little use for outlier detection, hence lose the robustness. Based on this skeptical analysis, we can tell that the value of η should not be set too large. Therefore, there exists contradiction for the selection of η . However, since the proposition 1 is just a sufficient condition, we could properly set the value of η lower than proposition 1 required and still make the sequence converges to a unique solution from any start point. From our implementation experience, as a compromising result, we recommend to set $\eta = K[\sup(I) - \inf(I)]^2 / \sigma_I^2$ and $K \in (0.1, 1)$.

3.1 Properties of Weight Field $G(C)$

From above discussion, we know that the two-step iteration algorithm (5) and (6) will generate a convergent sequence $\{C_n, n \geq 0\}$. Denote C^* be the convergent point of sequence $\{C^n, n \geq 0\}$ and μ^* the convergence point of sequence $\{\mu^n, n \geq 0\}$. In the following sub-sections, we discuss two useful properties of C^* for target tracking.

3.2 Illumination Invariant Property

C^* is almost invariant to rescaling and translation of illumination. Let the range of the photometric variable of interest (grayscale intensity, color or texture) be given by Λ . Let us denote the frame at time t and $t + 1$ by I_t and I_{t+1} , respectively. The sequences $\{\mu_t^n, n \geq 0\}, \{\mu_{t+1}^n, n \geq 0\}, \{C_t^n, n \geq 0\}, \{C_{t+1}^n, n \geq 0\}$ are similarly defined. Let the flow vector for the pixel (i, j) in frame I_t be given by (δ_i, δ_j) .

3.2.1 A Linear Model for Illumination Change

We adopt the reflection model

$$I(i, j) = gL(i, j)R(i, j) + b \quad (7)$$

where $I(i, j)$ is the image intensity at pixel (i, j) , $L(i, j)$ is the luminance, $R(i, j)$ is the reflectance, g and b are the camera gain factor and bias term, respectively. The two corresponding pixels in the adjacent frames can be represented as $I_t(i, j) = g_t L_t(i, j) R_t(i, j) + b_t$, $I_{t+1}(i + \delta_i, j + \delta_j) = g_{t+1} L_{t+1}(i + \delta_i, j + \delta_j) R_{t+1}(i + \delta_i, j + \delta_j) + b_{t+1}$. Suppose that reflectance keeps unchanged between two adja-

cent frames, that is $R_t(i, j) = R_{t+1}(i + \delta_i, j + \delta_j)$. After simple algebra

$$I_{t+1}(i + \delta_i, j + \delta_j) = \rho(i, j)I_t(i, j) + \varepsilon(i, j) \quad (8)$$

where $\rho(i, j) = \frac{g_{t+1}}{g_t} \frac{L_{t+1}(i+\delta_i, j+\delta_j)}{L_t(i, j)}$ and $\varepsilon(i, j) = b_{t+1} - \rho(i, j)b_t$. Hence, the illumination switch from time t to $t+1$ is modelled as a mapping $h : \Lambda \mapsto \Lambda$ with $h(I_t(i, j)) = I_{t+1}(i+\delta_i, j+\delta_j) = \rho(i, j)I_t(i, j) + \varepsilon(i, j)$, in which $\rho(i, j)$ is the rescaling term and $\varepsilon(i, j)$ is the translation term.

3.2.2 Study for Global Light Switch

To discuss the situation of global illumination switch, we assume that $\{\rho(i, j), (i, j) \in \Omega\}$ and $\{\varepsilon(i, j), (i, j) \in \Omega\}$ are i.i.d., with $\rho(i, j) \sim N(\lambda_1, \sigma_1^2)$ and $\varepsilon(i, j) \sim N(\lambda_2, \sigma_2^2)$, $N(\lambda_k, \sigma_k^2)$, $k = 1, 2$ denotes the Gaussian distribution with means value λ_k and variance σ_k^2 . Suppose that σ_k^2 is small enough, which means that all the pixels on the image plane Ω suffer from very similar extent of photometric feature value change. This can be seen as the simplest case of global illumination switch. Here, we assume that the motion scale in the scene of interest is very limited, therefore we get the following approximation:

$$\begin{aligned} \mu_{t+1}^{n+1} &= \frac{\iint_{\Omega} (C_{t+1}^n)^2 I_{t+1} dx dy}{\iint_{\Omega} (C_{t+1}^n)^2 dx dy} \\ &\approx \frac{\iint_{\Omega} (C_{t+1}^n)^2 (\rho I_t + \varepsilon) dx dy}{\iint_{\Omega} (C_{t+1}^n)^2 dx dy} \\ &\sim N(\lambda_1 \mu_t^{n+1} + \lambda_2, \sigma_3^2) \end{aligned} \quad (9)$$

$$\text{here } \sigma_3^2 = \frac{\iint_{\Omega} (C_{t+1}^n)^4 (\sigma_1^2 I_t^2 + \sigma_2^2) dx dy}{\iint_{\Omega} (C_{t+1}^n)^4 dx dy}.$$

Suppose that σ_3^2 is small enough, then

$$\begin{aligned} \frac{C_{t+1}^{n+1}(i + \delta_i, j + \delta_j)}{\eta \sigma_{I_{t+1}}^2} &= \\ \frac{\eta \sigma_{I_{t+1}}^2}{\eta \sigma_{I_{t+1}}^2 + (I_{t+1}(i + \delta_i, j + \delta_j) - \mu_{t+1}^{n+1})^2} &\approx \\ \frac{\eta \lambda_1^2 \sigma_{I_t}^2}{\eta \lambda_1^2 \sigma_{I_t}^2 + (\rho I_t(i, j) + \varepsilon - (\lambda_1 \mu^{n+1} + \lambda_2))^2} &\approx \\ \frac{\eta \lambda_1^2 \sigma_{h(I)}^2}{\eta \lambda_1^2 \sigma_{I_t}^2 + \lambda_1^2 (I_t(i, j) - \mu^{n+1})^2} &= C_t^{n+1}(i, j) \end{aligned} \quad (10)$$

Let $n \rightarrow \infty$, we will get $C_{t+1}^*(i + \delta_i, j + \delta_j) = C_t^*(i, j)$. In this case, the optical-flow consistency is preserved against the illumination switch.

3.3 Contrast preserving property

Another important property of C^* for tracking lies in that it is contrast preserving. $C^*(i, j)$ is temp to be close to one if the difference between $I(i, j)$ and robust mean μ^* is small, and zero if $I(i, j)$, as an outlier, is faraway from μ^* . That is, the contrast in photometric space among pixels on the Ω will be preserved in the extracted feature space. Such a contrast preserving property owns to the deviation of the model in the context of the robust statistics.

In fact, the weight field $G(C^1)$, which can be viewed to be a non-robust version of our model, is also illumination-invariant by similar analysis as in 3.2. However, we will show that it will lose the contrast preserving property. Here we give a simple visual example for explanation. As is shown in figure 2, (a) and (b) are original synthetic images before and after illumination switch. Here we set $\rho(i, j) \sim N(1, 1)$ and $\varepsilon(i, j) \sim N(100, 9)$. Figure 2 (c) and (d) are the C^1 images. It is obviously seen that the effect of light change has been eliminated in these two images. However, the contrast in original photometric space among adjacent vertical strides is also weakened, as can be expected for the reason that the original images have two dominant intensities, and the pixels numbers of both color groups are almost equal. C^1 is almost a constant function for the square term $(I(i, j) - \mu^1)^2$ is almost the same in (6). On the other hand, C^* (which is approached by C^{10} , $\eta = 0.6$) preserve the contrast in original photometric space, as is shown in figure 2 (e) and (f). Such property owns to the fact that the whole deviation is under a robust statistic framework.



Figure 2. Contrast preserving property of C^* . (a) and (b): original images before and after illumination switch; (c) and (d): the C^1 images; (e) and (f): the C^{10} images;

4 Applications in Tracking

In order to validate the proposed tracking feature extraction approach, we have performed some experiments on target tracking in situation of illumination switch and low object/background discriminability.

4.1 Illumination Invariant Tracking

The proposed feature image extraction method is incorporated with mean-shift algorithm to perform the illumination invariant tracking. For each coming frame, C^* image

is extracted from the input image and then the Mean-Shift process performs gradient ascent to find the nearest local mode.

The example is for an indoor human face tracking (320×240 , RGB, 246 frames). Figure 3 contains some resulting frames by a robust version of mean shift tracker [3] (row 1) and the proposed illumination invariant tracker (row 2). The light turns on and off during the tracking process and the illumination changes dramatically, e.g. from frame #194 ~ #197. The mean-shift tracker based on linear R-G-B combination feature [3] fails in this case while our method works well in such a scene under complex lighting condition. We set $\eta = 0.7$ in this experiment.

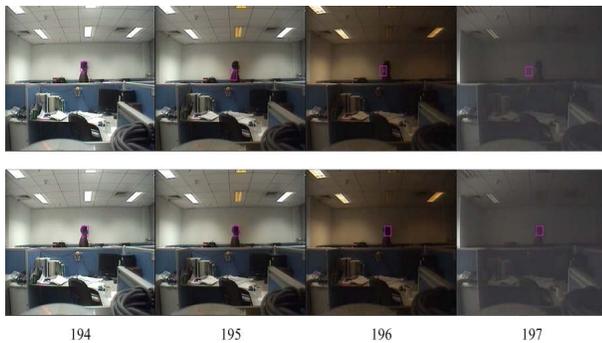


Figure 3. Indoor face tracking with illumination switch. Row 1: tracking results based on RGB photometric features. Row 2: tracking results with C^* extraction preprocessing.

4.2 Enhancement Tracking

In this section, we will further show that the contrast preserving property of C^* image can be applied to perform robust tracking in situation of low target/background contrast. Our experiments are performed on two real-world challenge sequences. Both sequences are captured at night with very low target/background discriminability in raw images. The first experiment is about the tracking of a running person before a building (320×240 , R-G-B, 126 frames). The results are shown in fig.4, where in each frame, the target is represented by a red rectangle and the local processing region by a green one. The C^* image of local processing region is shown on the top side area in each frame. It is clear to see that the contrast is significantly enhanced in extracted feature space compared to that in the RGB space. By incorporating C^* image with the mean-shift algorithm, the resulting tracker successfully locks on the moving person during the whole sequence.

The second experiment on enhancement tracking is performed on a even more challenging night sequence ($320 \times$

240 , R-G-B, 160 frames) with high noise and low contrast. Typical mean-shift tracker with R-G-B feature is sure to fail, as shown in the first row of fig.5. For our method, to deal with the noise, we consider the regulation term when extracting C^* . The corresponding tracking results are shown in the second and third row of fig.5. As illustrated in frame #50, the red rectangle and the green rectangle represent target and local processing region. The left image placed on the top is the extracted C^* without regulation while the right one do. It is clear to see that the C^* with regulation is contrast enhancing and more smooth. The corresponding tracker successfully tracks the target, a moving person, during the whole sequence.

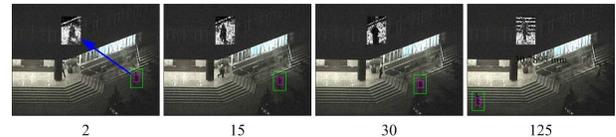


Figure 4. Track a running person at night.

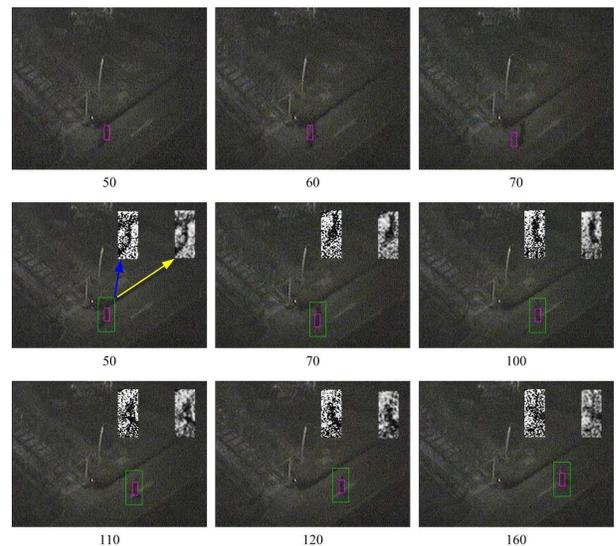


Figure 5. Low contrast and high noise tracking. First row: failure frames of traditional mean-shift tracker with R-G-B features. Second row and third row: tracking results of the proposed method.

5 Conclusion

We have presented a novel illumination-invariant and contrast preserving feature extraction mechanism for robust tracking in this paper. A weight field is generated from the input image data in photometric space to alleviate the effect of both global and local illumination switch. The whole

mechanism is well theoretically justified on a hidden field extraction framework through robust statistics. We embed the mechanism into the mean-shift tracker and the results are satisfying.

References

- [1] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *Image Processing, IEEE Transactions on*, 6:298–311, February 1997.
- [2] T. Chen, W. Yin, X. Zhou, D. Comaniciu, and T. Huang. Illumination normalization for face recognition and uneven background correction using total variation based image models. In *Computer Vision and Pattern Recognition*, volume 2, page 532. IEEE, 2005.
- [3] R. Collins, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27:1631–1643, October 2005.
- [4] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Computer Vision and Pattern Recognition*, volume II, pages 142–149. IEEE, 2000.
- [5] D. Freedman and M. Turek. Illumination-invariant tracking via graph cuts. In *Computer Vision and Pattern Recognition*, volume II, pages 10–17. IEEE, 2005.
- [6] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20:1025–1039, October 1998.
- [7] P. Kornprobst, R. Deriche, and G. Aubert. Image sequence analysis via partial differential equations. *Journal of Mathematical Imaging and Vision*, 11:5–26, January 1999.
- [8] B. Kschischang F.R., Frey and H. Loeliger. Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, 28, 2001.
- [9] S. Z. Li. In *Markov Random Field Modeling in Image Analysis*. Springer, 1995.
- [10] J. Limand, D. A. Ross, R.-S. Lin, and M.-H. Yang. Incremental learning for visual tracking. In *NIPS*, 2004.
- [11] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition*, volume I, pages 593–600. IEEE, 1994.
- [12] Y. Wu and T. Yu. A field model for human detection and tracking. *Pattern Analysis and Machine Intelligence, IEEE Tran.*, 47, 2006.