VARIATIONAL BAYESIAN INFERENCE FOR FORWARD-BACKWARD VISUAL TRACKING IN STEREO SEQUENCES

Giannis Chantas, Nikos Nikolaidis and Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki, GR 54124, Greece,
E-mail:{nikolaid,pitas}@aiia.csd.auth.gr

ABSTRACT

In this paper we propose a Bayesian framework for accurate object tracking in stereoscopic sequences. Object detection and forward tracking are first combined according to predefined rules to get a first set of tracked regions candidates. Backward tracking is then applied to provide another set of possible object localizations. Moreover, this strategy is applied herein in stereoscopic video. We introduce a Bayesian inference algorithm which is used to merge the information of both forward and backward tracking in order to refine the tracked region localization results. Experiments, performed on face tracking, show that the proposed method provides higher tracking accuracy than a forward tracker.

Index Terms— Stereo Tracking, Forward-Backward Tracking, Variational Inference

1. INTRODUCTION

Object tracking is an important problem in video semantic analysis, surveillance, etc. Many methods have been proposed to solve it [1]. The objective of a tracking algorithm is to locate in each frame a region of interest (ROI), usually depicting an object, a face or a human body. In the following, we use the term object to describe any such entity to be tracked. In many works, this problem is formulated in a stochastic Bayesian framework, [2], where the variables used to model the tracked ROIs are assumed to be indirect noisy observations of some variables that model the region trajectory. These variables are modeled as random variables that obey a stochastic model.

In this work, a Bayesian post-processing methodology is introduced that can refine the results of a forward-backward (FB) tracking strategy algorithm [3] when applied on a stereo video. The proposed methodology exploits the abundant information obtained by this FB tracking strategy which is richer than the information exploited by the standard forward tracking on single view videos. Indeed, the proposed methodology exploits and combines the results of both forward and backward tracking over time (combined with periodic object detection) on the left and right channel of a stereoscopic sequence of frames to increase the tracking accuracy.

We formulate the problem in a variational Bayesian inference framework [7], where it is assumed that the ROI coordinates derived by the above described FB tracking strategy are noisy observations of the coordinates of a single ideal ROI, for each frame in each channel. Additionally, an automatic relevance determination (ARD) model [7] used for the exclusion of observations coming from tracking failures is proposed. Moreover, the coordinates of the ideal ROIs are modeled as random variables, that are considered hidden (not directly observed), and a total variation (TV) prior [5] is imposed on them, in order to penalize abrupt changes in the estimated motion of the tracked object.

Given the observation and prior models, which are described in section 2 and 3 respectively, a variational Bayesian inference algorithm is derived in section 4 which estimates the model parameters and infers posteriors (though approximate, in order to bypass intractabilities) for the hidden variables (ideal ROI coordinates). In section 4, experiments are provided that demonstrate the efficiency of the proposed method.

2. OBSERVATION MODEL

In this section, we describe the proposed observation model for the forward-backward tracking strategy. The goal of the proposed methodology is to estimate the ideal (unknown) positions of the object ROIs in each video frame. Such ROI in the i-th frame is assumed herein to be a rectangular, defined uniquely by the upper-left and lower-right vertex coordinates, $[x_1(i), x_2(i)]^T$, and $[x_3(i), x_4(i)]^T$, respectively. These coordinates are assumed to be real numbers for the optimization convenience (see Section 4). We denote by $\mathbf{x}(i) = [x_1(i), x_2(i), x_3(i), x_4(i)]^T$ and $\mathbf{x} = [\mathbf{x}(1)^T, \mathbf{x}(2)^T, \ldots, \mathbf{x}(N)^T]^T$ the vector that contains the coordinates of all ideal ROIs to be estimated. N is the total

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n^o 287674 (3DTVS). The publication reflects only the authors' views. The EU is not liable for any use that may be made of the information contained herein.

number of frames of each channel in the stereo (L+R) video.

To introduce the observation model, we first describe our data/observations generation procedure. Object detection is performed periodically every n frames by using a suitably chosen object detector, e.g. the face detector in [9]. To track the detected ROI in intermediate frames between the frames where detection takes place, a tracking algorithm is applied, both in a forward and backward manner over time. The coordinates of the ROIs extracted by applying this procedure independently on the left and right channel are denoted by:

$$\mathbf{z}_k^D(i) = [z_{k,1}^D(i), z_{k,2}^D(i), z_{k,3}^D(i), z_{k,4}^D(i)], k = 1, \ldots, K_i^D,$$

for $D = L, R$ (L denotes left and R denotes right channel) and for $i = 1, \ldots, N$. K_i^L and K_i^R are the number of ROIs in each frame in a channel that vary with time. i denotes the frame number. We denote also by $\mathbf{z} = \{\mathbf{z}_k^L, \mathbf{z}_k^R\}$ the set of all extracted coordinates $\mathbf{z}^D = [\mathbf{z}_k^D(1)^T, \ldots, \mathbf{z}_k^D(N)^T]^T$ for $D = L, R$. We further assume that the extracted by tracking ROIs are noisy measurements of the ideall ROIs. Precisely, we assume that $p(\mathbf{z}|\mathbf{x}) = p(\mathbf{z}^L|\mathbf{h})p(\mathbf{z}^R|\mathbf{h})$, where

$$p(\mathbf{z}^L|\mathbf{h}) = \prod_{i,k} \exp\{-\frac{\lambda_{\mathbf{b}} d_k^L(i) b_k^L(i)}{2} \|\mathbf{z}_k^L(i) - \mathbf{x}(i)\|_2^2\}, \quad (1)$$

$$p(\mathbf{z}^R|\mathbf{h}) \propto \prod_{i,k} \exp\{-\frac{\lambda_{\mathbf{b}} d_k^R(i) b_k^R(i)}{2} \|\mathbf{z}_k^R(i) + \mathbf{m}(i) - \mathbf{x}(i)\|_2^2\}, \quad (2)$$

where $\mathbf{h} = \{\mathbf{x}, \mathbf{b}^R, \mathbf{b}^L, \mathbf{d}\}$ and $\mathbf{z} = [\mathbf{z}_k^L, \mathbf{z}_k^R]$. By $\mathbf{b}$ we denote the set of all the inverse variances $b_{k(i)}^{L,R}$ [7]:

$$\mathbf{b}^R = \{b_k^R(i) : \forall i, \forall k\}, \ \mathbf{b}^L = \{b_k^L(i) : \forall i, \forall k\},$$

$$\mathbf{b} = \{\mathbf{b}^L, \mathbf{b}^R\}, \ \mathbf{d} = \{d_k^R(i), d_k^L(i) : \forall i, \forall k\}.$$

To cope with the multiple distinct extracted ROIs in a frame, we introduce the binary variables $\mathbf{d}$ containing all $\mathbf{d}_k^D(i)$, where $\mathbf{d}_k^D(i) = 0$ or 1 (where $D = L, R$) and for constant i, $\mathbf{d}_k^D$ contains a single 1. These variables indicate which extracted ROI is really the noisy outcome of the ideal ROI. Although the variables $\mathbf{d}$ are binary, the algorithm uses their expected values, which are real numbers in the interval $[0, 1]$.

Furthermore, $\mathbf{m}(i) = [m_1(i), m_2(i), m_3(i), m_4(i)]^T$ is the vector containing the variables used to match the i-th ROI extracted from the right channel with its corresponding one in the right channel, by assuming that the reference coordinates system $\mathbf{x}$ is that of the left channel. Note also that they also vary with the frame number i. Herein, we further assume that $m_1(i) = m_3(i)$ and $m_2(i) = m_4(i)$.

A Gamma prior distribution [7] is imposed on each $b_k^D(i)$:

$$p(b_k^D(i)) \propto b_k^D(i)^{-0.5} \exp(-0.5 b_k^D(i)), \ D = L, R.$$

This is in essence the ARD model described in [7], used to model the varying nature of the inverse variances of the distribution in (1) and (2) to enable the model to ameliorate the influence of the ROIs coming from highly inaccurate tracking results (e.g. failures by occlusion). Indeed, a very small value of $b_k^D(i)$ excludes the $\mathbf{z}_k^D(i)$ ROI from the model. This is evident by inspecting their updates in (9) and (10): big differences of observed ROIs with the estimated lead to zero values of the estimated inverse variances, resulting in not to take inaccurate ROIs into account.

3. PRIOR MODEL

In this work, we impose a total-variation (TV) prior [5] on $\mathbf{x}$, aiming at obtaining a smoothed estimate of the ideal ROIs coordinates, since the observed coordinates $\mathbf{z}$ contain noise, due to tracking inaccuracies. This prior has been used successfully in image restoration, and the key of its success is the ability to provide smooth image estimates while preserving the image edges. Thus, when used in the present problem, the TV prior smooths the object trajectory (reducing noise), while in parallel preserves the possible abrupt changes of the tracked object position. We have:

$$p(\mathbf{x}) \propto \lambda_{\mathbf{x}}^{4N} \prod_{i=1}^{N-1} \exp\left(-\lambda_{\mathbf{x}} \|\mathbf{x}(i) - \mathbf{x}(i-1)\|_2\right). \quad (3)$$

Following the same reasoning, we use a TV prior for $\mathbf{m} = [\mathbf{m}(1)^T, \ldots, \mathbf{m}(N)^T]^T$. This means, as above, that:

$$p(\mathbf{m}) \propto \lambda_{\mathbf{m}}^{2N} \prod_{i=1}^{N-1} \exp\left(-\lambda_{\mathbf{m}} (\|\mathbf{m}(i) - \mathbf{m}(i-1)\|_2)\right). \quad (4)$$

4. VARIATIONAL BAYESIAN INFERENCE

The Bayesian paradigm dictates that we should estimate the variables of model $\mathbf{x}$, by taking their expectation with respect to their posterior using Bayes rule. However in our case, as in most models of interest, this is intractable. Thus, the variational Bayesian methodology is usually employed in order to obtain an approximate posterior for the model's hidden variables, [5], which provides tractable computations.

The goal of the variational Bayesian methodology is to compute a posterior, q, approximate to $p(\mathbf{x}, \mathbf{m}|\mathbf{z})$, by minimizing the Kullback-Leibler divergence [7] w.r.t to q:

$$KL(p||q) = \int q(\mathbf{x}, \mathbf{m}, \mathbf{b}) \log \frac{q(\mathbf{x}, \mathbf{m}, \mathbf{b})}{p(\mathbf{x}, \mathbf{m}, \mathbf{b}|\mathbf{z})} d\mathbf{x} d\mathbf{m} d\mathbf{b}.$$

We assume that $q(\mathbf{x}, \mathbf{m}, \mathbf{b}) = q(\mathbf{x})(\mathbf{m})q(\mathbf{b})$, i.e. $\mathbf{x}$, $\mathbf{b}$ and $\mathbf{m}$ are independent in the posterior. This approximation is what helps make the computations tractable.

However, due to the square roots in the priors for $\mathbf{x}$ and $\mathbf{m}$, the intractability still persists, and thus, we resort to minimizing an upper bound of the divergence [6], obtained by applying the inequality $\sqrt{w} \leq (w+u)/2\sqrt{u},\ \forall w,u > 0$ to all the squared terms in (3) and (4), and obtain lower bounds $\hat{p}(\mathbf{x}),\ \hat{p}(\mathbf{m})$ for $p(\mathbf{x}),\ p(\mathbf{m})$, described as follows:

$$\hat{p}(\mathbf{a}) \propto \prod_{i=1}^{N-1} \exp\left(-\frac{\lambda_{\mathbf{a}}}{2\sqrt{u_{\mathbf{a}}(i)}}\|\mathbf{a}(i)-\mathbf{a}(i-1)\|_2^2\right) \quad (5)$$

where $\mathbf{a}$ can be either $\mathbf{x}$ or $\mathbf{m}$. $u_{\mathbf{x}}(i)$ and $u_{\mathbf{m}}(i)$ are positive parameters, calculated at each algorithm iteration.

Using the independency assumption of the hidden variables, and the above inequalities, we define a new quantity to be minimized w.r.t $q(\mathbf{x})$, $q(\mathbf{m})$ and $q(\mathbf{b})$, as well as w.r.t to $u_{\mathbf{x}}(i)$ and $u_{\mathbf{m}}(i)$, $\forall i$, and the parameters $\lambda_{\mathbf{x}}, \lambda_{\mathbf{m}}$. This is

$$L(q(\mathbf{x}), q(\mathbf{b}), q(\mathbf{m}), u_{\mathbf{x}}, u_{\mathbf{m}}, \lambda_{\mathbf{x}}, \lambda_{\mathbf{m}}) = \quad (6)$$

$$\int q(\mathbf{x})q(\mathbf{m})q(\mathbf{b}) \log \frac{p(\mathbf{z})q(\mathbf{x})q(\mathbf{m})q(\mathbf{b})}{p(\mathbf{x})\hat{p}(\mathbf{m})p(\mathbf{b})p(\mathbf{z}|\mathbf{x},\mathbf{m},\mathbf{b})} d\mathbf{x}d\mathbf{m}d\mathbf{b},$$

where the Bayes' rule has been used to express the posterior in terms of the joint distribution.

Next, due to space limitations, without loss of generality, $\mathbf{x}(i)$, $\mathbf{m}(i)$, their estimates and $\mathbf{z}_k^D(i)$ denote a single coordinate of the respective vector (e.g. $\mathbf{x}(i) = x_1(i)$). This does not hold in norms. The update equations at t-th iteration for the posterior estimates as well as the parameters are [6]:

$$q^t(\mathbf{x}) = N(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}}), \quad (7)$$

where $\mathbf{C}_{\mathbf{x}} = (\mathbf{B}_{\mathbf{x}} + \lambda_{\mathbf{x}}\mathbf{Q}^T\mathbf{U}_{\mathbf{x}}\mathbf{Q}), \boldsymbol{\mu}_{\mathbf{x}} = \lambda_{\mathbf{b}}\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{B}\mathbf{z}_1$. $\mathbf{B}_{\mathbf{x}}$ and $\mathbf{U}_{\mathbf{x}}$ are diagonal matrices with i-th elements:

$$\mathbf{B}_{\mathbf{x}}(i,i) = \lambda_{\mathbf{b}}\Big(\sum_{k}^{K_i^L} d_k^L(i)\hat{b}_k^L(i) + \sum_{k}^{K_i^R} d_k^R(i)\hat{b}_k^R(i)\Big),$$

$$\mathbf{U}_{\mathbf{x}}(i,i) = \frac{1}{\sqrt{u_{\mathbf{x}}(i)}},$$

and $\mathbf{Q}$ is the first order differences operator. Also,

$$u_{\mathbf{x}}^t(i) = \|\boldsymbol{\mu}_{\mathbf{x}}(i) - \boldsymbol{\mu}_{\mathbf{x}}(i-1)\|_2^2 + [\mathbf{Q}\mathbf{C}_{\mathbf{x}}\mathbf{Q}^T]_{i,i} \quad (8)$$

$$\hat{b}_k^L(i) = \frac{2}{1 + \|\mathbf{z}_k^L(i) - \boldsymbol{\mu}_{\mathbf{x}}(i)\|_2^2 + 4\mathbf{C}_{\mathbf{x}}(i,i)}, \quad (9)$$

$$\hat{b}_k^R(i) = \frac{2}{1 + d_k^L(i)\|\mathbf{z}_k^R(i) + \boldsymbol{\mu}_{\mathbf{m}}(i) - \boldsymbol{\mu}_{\mathbf{x}}(i)\|_2^2 + 4\mathbf{C}(i,i)}. \quad (10)$$

where $\mathbf{C} = \mathbf{C}_{\mathbf{x}} + \mathbf{C}_{\mathbf{m}}$. $\hat{b}_k^D(i)$ is the mean of $b_k^D(i)$ w.r.t to the estimated posterior distribution:

$$q^t(b_k^D(i)) = \text{Gamma}(2, c),\ \forall i, k, D = L, R. \quad (11)$$

where c is the denominator of (9) for $D = L$ and (10) for $D = R$. Also,

$$z_1(i) = \sum_{k=1}^{K_i^L} d_k^L(i)b_k^L(i)z_k^L(i) + \sum_{k=1}^{K_i^R} d_k^R(i)\hat{b}_k^R(i)(z_k^R(i) + \mu_{\mathbf{m}}(i)).$$

In a similar manner:

$$q^t(\mathbf{m}) = N(\boldsymbol{\mu}_{\mathbf{m}}, \mathbf{C}_{\mathbf{m}}) \quad (12)$$

is the update for the posterior of $\mathbf{m}$, where $\mathbf{C}_{\mathbf{m}} = (\mathbf{B}_{\mathbf{m}} + \lambda_{\mathbf{m}}\mathbf{Q}^T\mathbf{U}_{\mathbf{m}}\mathbf{Q}), \boldsymbol{\mu}_{\mathbf{m}} = \mathbf{C}_m^{-1}\mathbf{B}_m\mathbf{z}_2$ and $\mathbf{B}_{\mathbf{m}}$ and $\mathbf{U}_{\mathbf{m}}$ are diagonal matrices with elements:

$$\mathbf{B}_{\mathbf{m}}(i,i) = \lambda_{\mathbf{b}} \sum_{k}^{K_i^R} d_k^R(i)\hat{b}_k^R(i),\ \mathbf{U}_{\mathbf{m}}(i,i) = 1/\sqrt{u_{\mathbf{m}}(i)},$$

$$u_{\mathbf{m}}^t(i) = \|\boldsymbol{\mu}_{\mathbf{m}}(i) - \boldsymbol{\mu}_{\mathbf{m}}(i-1)\|_2^2 + [\mathbf{Q}\mathbf{C}_m\mathbf{Q}^T]_{i,i}, \quad (13)$$

$$z_2(i) = \sum_{k=1}^{K_i^R} d_k^R(i)\hat{b}_k^R(i)(\mathbf{z}_k^R(i) + \mu_{\mathbf{m}}(i)).$$

As for the $\lambda_{\mathbf{x}}$ and $\lambda_{\mathbf{m}}$ parameters, their update is:

$$\lambda_{\mathbf{x}}^t = \frac{4N}{\sum_{i=1}^N u_{\mathbf{x}}(i)},\ \lambda_{\mathbf{m}}^t = \frac{2N}{\sum_{i=1}^N u_{\mathbf{m}}(i)} \quad (14)$$

We have to mention that for $\mathbf{d}$ we give a fixed value:

$$d_k^R(i) = \frac{1}{2K_i^R}, d_k^L(i) = \frac{1}{2K_i^L}.$$

One algorithm iteration consists of (7), (8) , (9), (10), (12), (13) and (14). After convergence or the maximum iteration number is reached, $\boldsymbol{\mu}_{\mathbf{x}}$ and $\boldsymbol{\mu}_{\mathbf{m}}$ are the estimates of left and right channel ideal ROI coordinates, respectively.

5. EXPERIMENTS

In the first set of the experiments, we evaluated the performance of the algorithms FP and FBP in two sample single-view videos, where human bodies are tracked. The forward tracking algorithm (F) based on particle filters [4] was first applied. The first and middle ROIs depicting the tracked human was not localized by the tracking algorithm but was specified manually. In Table 1 we see an improvement in tracking accuracy in two sample videos (see next for the definition of the tracking accuracy metric used).

In the second set of the experiments we evaluate the performance of the proposed post-processing algorithm in stereo videos. We perform first single channel face tracking on two stereo videos (indepedently on the right and left channel), using a face detector [9] and a forward tracking procedure [4] based on particle filters. The face detection algorithm, used to periodically re-initialize the tracker, is based on Haar-like feature detection and, in parallel, exploits color skin information [9]. For each video, face detection was performed in varying frequency: every 20, 30 and 40 frames as can be seen in Table 2. The forward (F) tracking algorithm, whose results we try to improve with the proposed post-processing framework, works as follows. Face detection is performed periodically every n frames. The frames between successive face sophisticated prior for the ROIs coordinates. More accurate estimation of $\mathbf{m}$ will also be pursued.

Table 1. Algorithm performance ($\hat{a}$) in single view video

Video Name	F	FP	FBP
WalkByShop1front2	0.178	0.196	0.204
OneStopMoveEnter1cor4	0.522	0.531	0.539

Table 2. Algorithm performance ($\hat{a}$) using stereo videos. The results shown for FP and FBP concern the left channel, while for F and FBSP the results for the left and right channel are seperated by a slash.

Video	n	F	FP	FBP	FBSP
Vid.1	20	0.614/0.643	0.620	0.625	0.634/0.65
Vid.1	30	0.604/0.62	0.612	0.624	0.630/0.64
Vid.1	40	0.580/0.6	0.589	0.594	0.601/0.62
Vid.2	20	0.431/0.53	0.541	0.555	0.561/0.64
Vid.2	30	0.462/ 0.52	0.546	0.554	0.555/ 0.64
Vid.2	40	0.415/0.49	0.498	0.521	0.522/ 0.62

detections are extracted by the previously mentioned tracking algorithm. Tracking accuracy is evaluated in four scenarios: a) forward tracking (F) without post-processing and b) forward tracking c) forward-backward tracking and post processing (FBP), d) forward-backward tracking plus stereo post processing (FBSP). It must be noted that the parameter $\lambda_{\mathbf{b}}$ was set equal to 300 and kept fixed in every experiment. The Average Tracking Accuracy ($\hat{a}$) [8] metric was used to measure tracking accuracy:

$$\hat{a} = \frac{1}{N}\sum_{i=1}^{N}\frac{D_i \bigcap G_i}{D_i \bigcup G_i} \tag{15}$$

where D_i is the estimated ROI area, while G_i is the ideal (ground truth) ROI areas obtained by manual video annotation. For the FBSP algorithm, D_i corresponds to $\mu_{\mathbf{x}}(i)$ when assessing tracking accuracy of the left channel and to $\mu_{\mathbf{x}}(i) - \mu_{\mathbf{m}}(i)$ of the right channel. For FP and FBP, D_i corresponds to $\mu_{\mathbf{x}}(i)$, since there is only one channel and thus $\mathbf{m}$ is not estimated ($D = L$).

Table 2 depicts tracking accuracy in two sample stereo videos, video 1 and video 2, having 775 and 665 frames each, respectively, for face detection frequency $n = 20$, $n = 30$, $n = 40$. For both videos we see an increase of the tracking accuracy when the post-processing methodologies are employed.

6. CONCLUSIONS

We have presented a post-processing Bayesian methodology that refines the outputs of standard tracking algorithms. The results are promising and show that the information provided by a forward-backward tracking as well as stereo tracking can boost the tracking performance. In future, we plan to augment the proposed Bayesian methodology by using a more sophisticated prior for the ROIs coordinates. More accurate estimation of $\mathbf{m}$ will also be pursued.

7. REFERENCES

[1] N. Nikolaidis, M. Krinidis, E. Loutas, G.Stamou, and I.Pitas, "Motion tracking in Video", *in The Essential Guide to Video Processing*, 2nd edition, edited by: Al Bovik, Elsevier, 2009.

[2] A. Dore, M. Soto, and C. S. Regazzoni, "Bayesian Tracking for Video Analytics", *IEEE Signal Processing Magazine*, vol. 27 , no. 5, pp. 46 - 55, September 2010.

[3] I. Cherif, V.Solachidis and I.Pitas, "A Tracking Framework for Accurate Face Localization", *in Proc. of Int. Federation for Information Processing Conf. on Artificial Intelligence (IFIP AI 2006)* , Santiago, Chile, 21-24 August, 2006.

[4] Shaohua Kevin Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance adaptive models in particle filters," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1491-1506, November 2004.

[5] S. D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter estimation in tv image restoration using variational distribution approximation," *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 326–339, March 2008.

[6] G. Chantas N. Galatsanos R. Molina and A. K. Katsaggelos, "Variational bayesian image restoration with a spatially adaptive product of total variation image priors," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 351– 362, February 2010.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Verlag, 2006.

[8] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, " Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.31, no. 2, pp:319-336, 2009.

[9] G.Stamou, M.Krinidis, N.Nikolaidis and I.Pitas, "A monocular system for automatic face detection and tracking" in Proc. of Visual Communications and Image Processing (VCIP 2005), Beijing, China, 12-15 July, 2005.

[10] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1