Robust face tracking using colour Dempster-Shafer fusion and particle filter

F. Faux F. Luthon[1]
[1] Laboratoire LIUPPA (EA 3000)

Computer Science Dpt, University of Pau, France
faux@iutbayonne.univ-pau.fr

Abstract

This paper describes a real time face detection and tracking system. The method consists in modelling the skin face by a pixel fusion process of three colour sources within the framework of the Demster-Shafer theory.The algorithm is composed of two phases. In a simple and fast initialising stage, the user selects successively in an image, a shadowy, an overexposed and a zone of mean intensity of the face. Then the fusion process models the face skin colour.
Next, on the video sequence, a tracking phase uses the key idea that the face exterior edges are well approximated as an ellipse including the skin colour blob resulting from the fusion process. As ellipse detection gets easily disturbed in cluttered environnements by edges caused by non-face objects, a simple and fast efficient least squares method for ellipse fitting is used. The ellipse parameters are taken into account by a stochastic algorithm using a particle filter in order to realise a robust face tracking in position, size and pose.
The originality of the method consists in modelling the face skin by a pixel fusion process of three independant cognitive colour sources. Moreover, mass sets are determined from a priori models taking into account contextual variables specific to the face under study. Hence, the face specificity which is to present shadowy (neck) and overexposed zones (nose, front) is considered, so that sensitivity to lighting conditions decreases.
Results of face skin modelling, fusion, ellipse fitting and tracking are illustrated and discussed in this paper. The limits of the method and future work are also commented in conclusion.

Keywords

Face detection, Dempster-Shafer, colour fusion, face tracking, condensation, skin hue.

1 Introduction

Face detection and tracking in a video sequence is useful in many applications such as videoconference, HMI, telesurveillance or robotics.
Above 150 detection methods [1, 2], as low level methods using cues such as texture, color [3, 4], edges, up to high level methods such as appearance models, neural networks or Support Vector Machines are proposed in the literature [5, 6, 7, 8].
In these approaches, skin color is often used as a first localisation and segmentation primitive in order to decrease the research zone. Because of its specificity, skin colour cue is very pertinent [9] and allows rapid size and orientation invariant algorithms.
However it remains an important difficulty to accommodate unsupervised varying lighting conditions. Normalised RGB, HLS colour spaces are the most used. But these colour transforms remain quite sensitive to lighting conditions and very noisy in shadowy zones. In order to overcome these drawbacks, increase robustness and contrast, skin colour cues are mapped into the logarithmic colour space LUX [10]. Moreover, face is a non-rigid object which pecularity is to present shadowy (neck) and overexposed zones (nose, front). Their time varying localisation because of movement in the video sequence yields complex modelisation. So contextual variables are included in the process to deal with this aspect.
Pixel-based skin colour detection techniques can be gathered in explicit methods, which use empirical rules of decision, and statistics [11]. However, information to model is never perfect because of image acquisition principle (3D space towards 2D plan transformation). There exists different forms of artefact including ambiguity, imprecision or inadequacy. We ought to take into account these difficulties to improve modelisation. Classical probabilistic methods (Bayesian inference rule) loose performance when early vision learning stage is not representative of real measurements. That is the reason why a pixel fusion process of skin colour cues within the framework of the Dempster Shafer theory is used in this paper [12, 13].
A simple and fast initialising stage takes into account ground truth as the user selects successively on one image, a shadowy, an overexposed and a zone of mean intensity. For each zone, mass sets are determined from a priori models of three colour sources (in the LUX colour space), contextual variables and source confidence degrees. Then mass sets are fused and blobs are computed.
Next, on the video sequence, a tracking phase uses the key idea that the face exterior edges are well approximated as an ellipse including the skin colour blob resulting from the fusion process [14, 15]. As ellipse detection gets easily disturbed in cluttered environments by edges caused by non face objects, a simple and fast efficient least squares method for ellipse fitting is used [16]. The ellipse parameters (center, minor axis, major axis and orientation) are taken into account by a stochastic algorithm using a particle filter in order to realise a robust face tracking in position, size and pose.
The remainder of the paper is organised as follows. Section 2 develops the modelisation and fusion processes. Modelisation results are illustrated and commented in section

3. Section 4 explains the tracking algorithm and shows results. Finally, section 5 summarises the contribution and opens the suggestions for future work.

2 Method description

2.1 Sources and discernment field

Skin colour cues are mapped into the logarithmic colour space LUX [10]. The expression of the LUX components from the RGB colour space coded on 3×8 bits are :

$$L = (R+1)^{0.3}(G+1)^{0.6}(B+1)^{0.1} - 1$$

$$U = \begin{cases} 128\left(\frac{L+1}{R+1}\right) \text{for} R > L \\ 256 - 128\left(\frac{R+1}{L+1}\right) \text{ otherwise} \end{cases}$$

$$X = \begin{cases} 128\left(\frac{L+1}{B+1}\right) \text{for} B > L \\ 256 - 128\left(\frac{B+1}{L+1}\right) \text{ otherwise} \end{cases}$$

The fusion process uses 3 sources (colour "sensors") called S_j, $(j = 1, 2, 3)$ such as :

$$S_1 = U,\ S_2 = X \text{ et } S_3 = 0.5(L+U) = W \qquad (1)$$

For the source S_3, instead of L we add to S_1 the luminance component L, very rich from the semantic point of view, in order to better characterise the skin colour variations due to lighting conditions. Each source S_j provides a measurement noted M_j.
The discernment field Ω is defined by 2 assumptions such as : $\Omega = \{\omega_1, \omega_2\}$, where ω_1 represents the face assumption and ω_2, complement of ω_1, symbolises the background ($\omega_2 = \bar{\omega}_1$).

2.2 A priori model

In order to determine the a priori model, in an initialising stage, the user selects successively on an image three characteristic zones of the face : a shadowy zone, a zone of mean intensity and an overexposed zone. Three contextual variables z_i , $(i = 1, 2, 3)$ are taken into account : shadowy zone z_1 ; mean zone z_2 ; overexposed zone z_3.
The histograms calculated on each selected zone and for each measurement M_j allows to determine the conditional probability densities $p(M_j/\omega_1, z_i)$. These are approximated by Gaussian distributions $N_{ij}(\mu_{ij}, \sigma_{ij})$ (Fig. 3) where μ_{ij} and σ_{ij} are respectively the mean and the standard deviation of the measurement M_j on the zone z_i. In addition to statistical information on the zone, spatial data distribution (M_1, M_2, M_3) in the colour space (U, X, W)(Eq. 1), that is to say the colour domain D_i contributes to modelisation. For each context z_i, 3 pairs of coefficients α_{ij} and β_{ij} maps 6 straight-line segments ($Min_{ij} = \mu_{ij} + \beta_{ij} - \alpha_{ij}\sigma_{ij}; Max_{ij} = \mu_{ij} + \beta_{ij} + \alpha_{ij}\sigma_{ij}$) which synthesise a parallelepiped. These coefficients are adjusted such as the parallelepiped encompasses at best the colour domain D_i (Fig. 4). Then, the function $skin_{ij}$ defines the a priori model such as :

$$skin_{ij} = \begin{cases} \frac{N_{ij}(\mu_{ij},\sigma_{ij})}{\max(N_{ij}(\mu_{ij},\sigma_{ij}))} \text{ if } Min_{ij} \leq M_j \leq Max_{ij} \\ 0 \quad \text{else} \end{cases} \qquad (2)$$

2.3 Mass sets

Appriou [17] suggested to introduce each a priori density of probability $p(M_j/\omega_1, z_i)$ and its corresponding confidence degree d_{ij} in a mass set $m_{ij}(.)$. This set is defined by an axiomatic approach in the discernment field Ω. In the method developed here the densities of probability are replaced by the functions $skin_{ij}$ which implicitly take into account the spatial data distribution in the colour space (U, X, W).
Focal elements associated to $m_{ij}(.)$ are ω_1, ω_2 and Ω.
Mass sets are defined by :

$$\begin{aligned} m_{ij}(\omega_1) &= \frac{d_{ij} R_i skin_{ij}}{1 + R_i skin_{ij}} \\ m_{ij}(\omega_2) &= \frac{d_{ij}}{1 + R_i skin_{ij}} \\ m_{ij}(\Omega) &= 1 - d_{ij} \end{aligned} \qquad (3)$$

The weighting coefficients R_i take into account the data $skin_{ij}$ for each zone i.

2.4 Confidence degrees

For $S_1 = U$ and $S_2 = X$ sources, the ambiguity between classes is weak. Consequently a probabilistic modelling is used $d_{i1} = d_{i2} \approx 1$ ($m(\Omega) = 0$). On the other hand the S_3 source depends on brightness. The reliability of this source for the face class is maximum ($d_{i3} = 1$) only for the average grey level (μ_{i3}) of the modelled zone. Ambiguity between classes grows ($m(\Omega) > 0$) when M_3 deviates from μ_{i3}. Therefore a fuzzy function (Eq. 4) is used to characterise the reliability of the S_3 source in the context z_i (Fig. 1).
As $\mu_{13} \leq \mu_{23} \leq \mu_{33}$ we obtain :

$$\text{for } i = 1 : \quad d_{13} = \begin{cases} \frac{M_3 + 2\mu_{23} - 3\mu_{13}}{2\mu_{23} - 2\mu_{13}} & \text{if } 3\mu_{13} - 2\mu_{23} \leq M_3 \leq \mu_{13} \\ \frac{-(M_3 - \mu_{23})}{\mu_{23} - \mu_{13}} & \text{if } \mu_{13} \leq M_3 \leq \mu_{23} \\ 0 \quad \text{else} \end{cases}$$

$$\text{for } i = 2 : \quad d_{23} = \begin{cases} \frac{M_3 - \mu_{13}}{\mu_{23} - \mu_{13}} & \text{if } \mu_{13} \leq M_3 \leq \mu_{23} \\ \frac{-(M_3 - \mu_{33})}{\mu_{33} - \mu_{23}} & \text{if } \mu_{23} \leq M_3 \leq \mu_{33} \\ 0 \quad \text{else} \end{cases}$$

$$\text{for } i = 3 : \quad d_{33} = \begin{cases} \frac{M_3 - \mu_{23}}{\mu_{33} - \mu_{23}} & \text{if } \mu_{23} \leq M_3 \leq \mu_{33} \\ \frac{-(M_3 + 2\mu_{23} - 3\mu_{33})}{2\mu_{33} - 2\mu_{23}} & \text{if } \mu_{33} \leq M_3 \leq 3\mu_{33} - 2\mu_{23} \\ 0 \quad \text{else} \end{cases} \qquad (4)$$

2.5 Decision

Three contextual masses $m_i(s)$ are associated at each pixel $s(x, y)$ of colour components (M_1; M_2; M_3) by the orthogonal normalised combination of Dempster-Shafer :

$$m_i(s) = \oplus m_{ij}(s) \qquad (5)$$

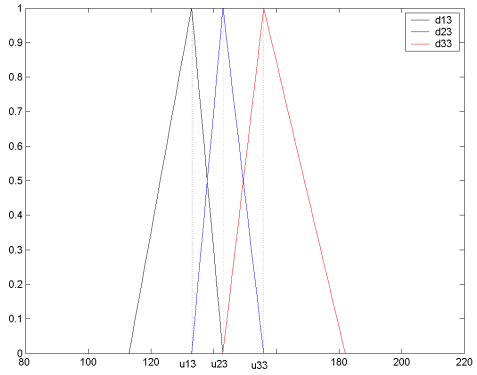


Figure 1: Confidence degrees d_{i3} according to M_3

The disjunctive fusion rule combines the contextual masses $m_i(s)$ in order to associate a single mass $m(s)$ to each pixel.
We obtain :

$$m(s) = m_1(s) \oplus_{\cup} m_2(s) \oplus_{\cup} m_3(s) \quad (6)$$

3 Modelisation results

3.1 Shadowed zone a priori model

As the initialisation stage is the same for each zone, only the modelling of the shadowed zone z_1 is presented in detail hereafter. First, the user selects on an image a shadowed zone of the face (Fig. 2). The probability densities

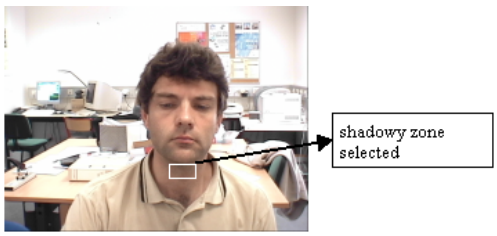
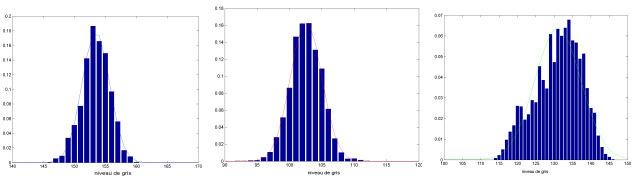


Figure 2: Phase of training shadowy face zone

are approximated for each measurement M_j by a Gaussian distribution $N_{1j}(\mu_{1j}, \sigma_{1j})$ (Fig. 3).



a) b) c)

Figure 3: Probability densities for the shadowy zone z_1: a) $p(M_1/\omega_1, z_1)$; b) $p(M_2/\omega_1, z_1)$; c) $p(M_3/\omega_1, z_1)$.

Straight line segments ($Min_{1j} = \mu_{1j} + \beta_{1j} - \alpha_{1j}\sigma_{1j}$; $Max_{1j} = \mu_{1j} + \beta_{1j} + \alpha_{1j}\sigma_{1j}$) given by varying manually α_{1j} and β_{1j}, generate a parallelepiped which encompasses the colour field D_1. Here $\alpha_{11} = \alpha_{12} = 2$; $\alpha_{13} = 1.7$; $\beta_{11} = \beta_{12} = \beta_{13} = 0$; and $\sigma_{11} = 2.3$; $\sigma_{12} = 2.55$; $\sigma_{13} = 6.5$.

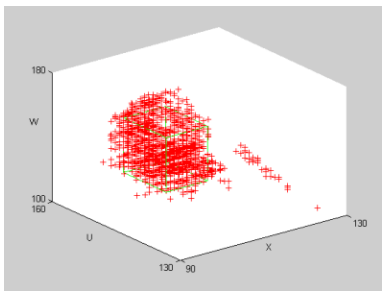
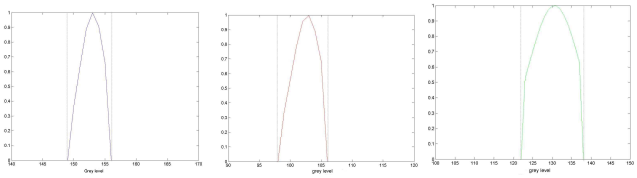


Figure 4: Colour pixel components in the colour space (U, X, W) (domain D_1) and parallelepipedic envelop associated.

These coefficients α_{1j} and β_{1j} limit the probability densities and determine the functions $skin_{1j}$ (Fig. 5).



a) b) c)

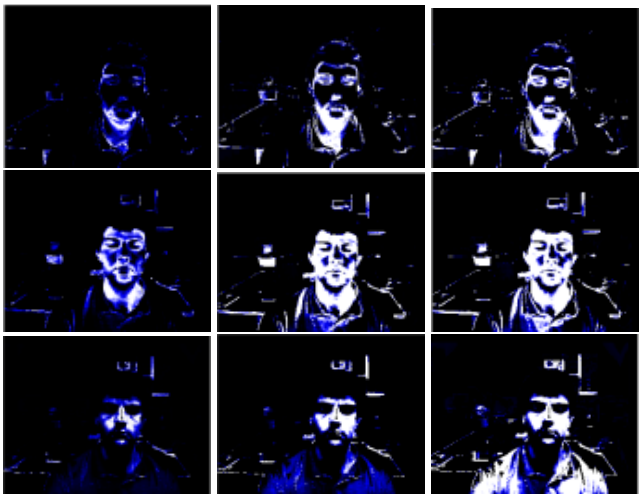
Figure 5: Functions $skin_{1j}$: a) $skin_{11}$; b) $skin_{12}$; c) $skin_{13}$

3.2 Shadowed zone detection

The mass sets $m_{1j}(\omega_1) = \frac{d_{1j} R_1 skin_{1j}}{1 + R_1 skin_{1j}}$ (Eq. 3) depend on the parameter R_1 which weights the data importance of $skin_{1j}$ characterising the face class.
Then, three masses $(m_{11}(s), m_{12}(s), m_{13}(s))$ are associated with each pixel in the site $s(x, y)$ in function of its colour components (M_1, M_2, M_3).
Finally a single mass $m_1(s)$ is affected to each pixel by the conjunctive fusion rule (Eq. 5).
The 3 top images of Fig. 6 show the experimental results for various values of R_1. For $R_1 = 1$ modelling is satisfactory as the shadowed skin colour (in white) present under the chin, the eyes contours and under the hair is detected. For $R_1 = 5$ or $R_1 = 10$ the modelling zone widens and false detections appear.

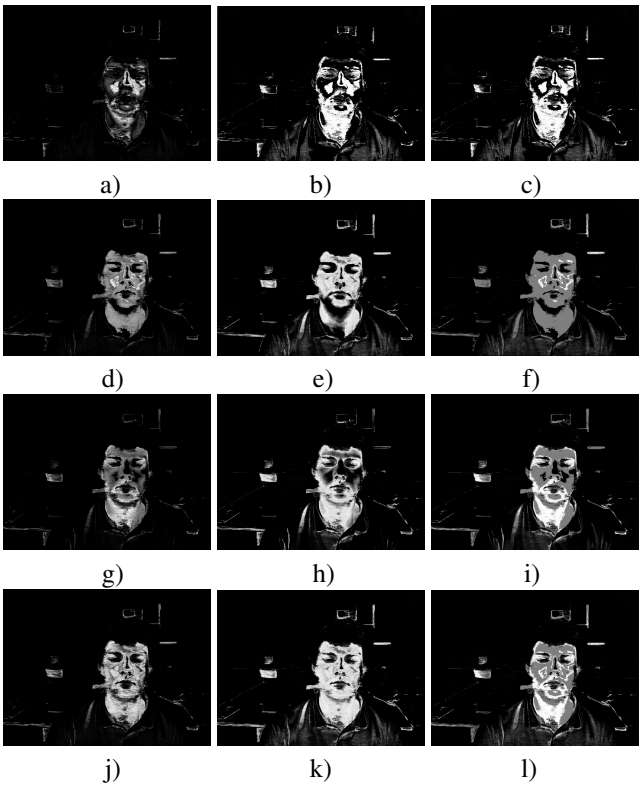
3.3 Modelisation synthesis

Fig. 6 (middle and bottom) shows the results of the conjunctive fusion for the mean intensity and overexposed zones, whose process is similar to those presented in sections 3.1 and 3.2 for the shadowed zone. Modelling detects accurately the shadowed and overexposed skin face zones (Fig. 6 top and bottom). Although more representative of the face, the average model (Fig. 6 middle), presents artefacts in the shadowy or overexposed zones. This is the reason why the disjunctive fusion (Eq. 6) is used to combine the various models and optimise the detection result (Fig. 7 left and center).
When two masses are combined, modelling is insufficient and presents defaults on mean intensity parts of the face (Fig. 7 a, b), on the shadowy parts of the face (Fig. 7 d, e) or on the overexposed parts of the face (Fig. 7 g, h).



a) b) c)

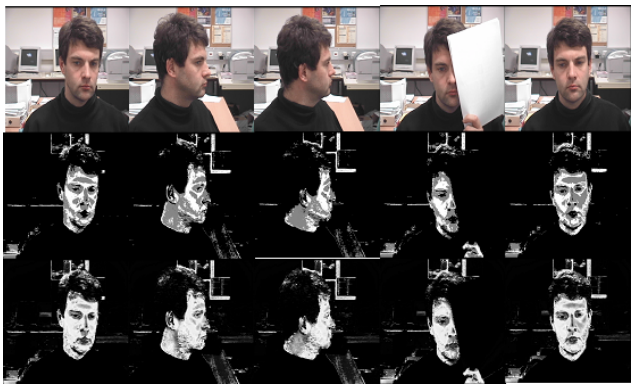
Figure 6: Detection realised from the mass sets m_i, $(i = 1, 2, 3)$ for various values of R_i : a) $R_i = 1$; b) $R_i = 5$; c) $R_i = 10$. Top: shadowy zone $(i = 1)$; Middle: average zone $(i = 2)$; Bottom: overexposed zone $(i = 3)$



a) b) c)
d) e) f)
g) h) i)
j) k) l)

Figure 7: Fusion results for different combinations of masses in the framework of DS evidence theory (left and center) and probabilistic (right): a) $R_1 = 1$, $R_2 = 0$, $R_3 = 1$; b)c) $R_1 = 5$, $R_2 = 0$, $R_3 = 5$; d) $R_1 = 0$, $R_2 = 1$, $R_3 = 1$; e)f) $R_1 = 0$, $R_2 = 5$, $R_3 = 5$; g) $R_1 = 1$, $R_2 = 1$, $R_3 = 0$; h)i) $R_1 = 5$, $R_2 = 5$, $R_3 = 0$; j) $R_1 = 1$, $R_2 = 1$, $R_3 = 1$; k)l) $R_1 = 5$, $R_2 = 5$, $R_3 = 5$;

Modelling is optimal when the 3 masses are combined $(R_1 = 1, R_2 = 1, R_3 = 1)$ (Fig. 7 j). However a too important weighting of parameter R_i deteriorates slightly the modelling quality (Fig. 7 k).
Within the framework of the Dempster-Shafer theory, in spite of a complex background made up of colour elements close to skin colour (table, sweater, poster), the fusion process ensures a good detection and proves to be robust to occlusions and pose variations (Fig. 8 bottom).
On the other hand, Fig. 7 right and Fig. 8 middle represent the result of modelling when $(d_{ij} = 1)$ i.e. when the reliability of the sensors is not taken into account any more. Consequently, $m_{ij}(\Omega) = 0$ (cf Eq. 3) and only the face (ω_1) and non-face (ω_2) classes are considered. Within this probabilistic framework, the detection result is of worse quality in the zones where the data inaccuracy is not negligible because of the less reliable learning stage model.



a) b) c) d) e)

Figure 8: a) image 10 ; b) image 14 ; c) image 18 ; d) image 24 ; e) image 30

4 Tracking by condensation algorithm

4.1 Particle filter formalism

The particle filter algorithm initially designed for signal processing problems [18], was applied in computer vision under the name of "condensation " algorithm [19]. In the case of a single object, the vector X_t representing the hidden state of the object of interest follows the law of evolution (Eq. 7) and is observed by the vector Y_t at discrete times according to Eq. 8 :

$$X_t = F_t(X_{t-1}, V_t) \quad (7)$$
$$Y_t = H_t(X_t, W_t) \quad (8)$$

No assumption is made on F_t and H_t functions and the processes V_t and W_t are two white noises not necessarily Gaussian, mutually independent and independent of the initial condition X_0.

4.2 Application to face tracking

In the framework of face tracking, where movement is not foreseeable and frequently changes in direction, particle filtering finds all its justification.
Moreover, the key idea that the face exterior edges are well approximated as an ellipse of center noted (x_{c_t}, y_{c_t}), of mi-

nor axis ℓ_t, large axis L_t and of orientation θ_t including the skin colour blob resulting from the fusion process is used. These parameters are gathered in the state vector $X_t = [X_{1_t}, X_{2_t}]$ where $X_{1_t} = [x_{c_t}, y_{c_t}]$ and $X_{2_t} = [\ell_t, L_t, \theta_t]$. After a phase of initialisation, the tracking method proceeds in two stages. First the algorithm determines the center then estimates the size and the orientation of the ellipse.

Initialisation. In order to select the face whatever its initial position in the image, the particles are distributed in position according to a uniform probability law whereas the parameters of size and orientation are fixed at a constant value ($\ell_0 = 20$; $L_0 = 25$; $\theta_0 = 0$) (Fig. 9). The weight of the particles is $q_0^n = 1/N$ where N is the number of particles (typ. $N = 150$).

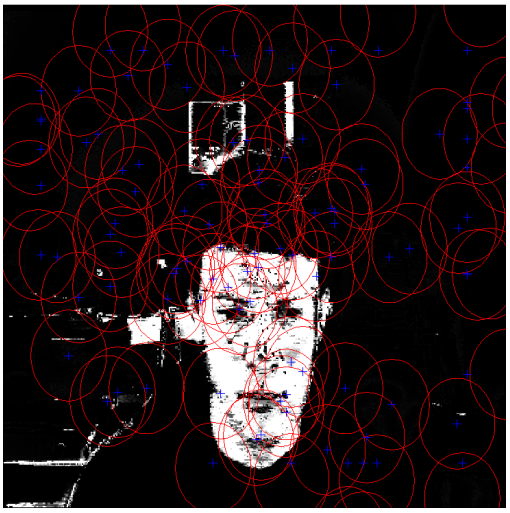


Figure 9: Particle initialisation

Model center estimation. In this stage, in order to determine the state vector position at time t size and orientation parameters are fixed. They are the components of X_{2_t} estimated at time $t-1$.
The dynamics of the vector X_{1_t} is described by the following model [20] :
$[p(X_{1_t}/X_{1_{t-1}}) = (1-\beta_u)N(X_{1_t}/X_{1_{t-1}}, \Sigma) + \beta_u U_\chi(X_t)\ ]$ where $N(./\mu, \Sigma)$ represents the Gaussian distribution of average μ and of covariance Σ.
$U_\chi(.)$ represents the uniform distribution on the unit χ.
The coefficient β_U , $0 \leq \beta_U \leq 1$ weights the uniform distribution and $\Sigma = diag({\sigma^2_{x_{c_t}}}, {y_{c_t}}^2)$ is the diagonal matrix made up of the variances of the X_{1_t} state vector components.
The introduction of an uniform component ($\beta_u = 0.1$) manages the rare erratic movements like jumps in the video sequence. It helps also the algorithm to be locked after one period of partial or total occlusion.
The model of measurement $Y_{1n,t}$ weights more or less strongly a particle n in function of the squared sum of masses inside the particle n. The probability of observation is consequently defined by :

$$p(Y_{1_t}/X_{1_t}) = \prod_n p(Y_{1n,t}/X_{1n,t}) \text{ where} Y_{1n,t} \propto \sum m(s)^2$$

The maximum of plausibility criterion allows selecting the most significant ellipse and its center defines the state vector position components. At the end of the initialisation stage, the estimated position (Fig. 10 b) does not correspond to face center because of incorrect face and orientation ellipse dimensioning. After some steps, these parameters being corrected (see hereafter), the algorithm estimates with a better accuracy the ellipse center (Fig. 10 d).

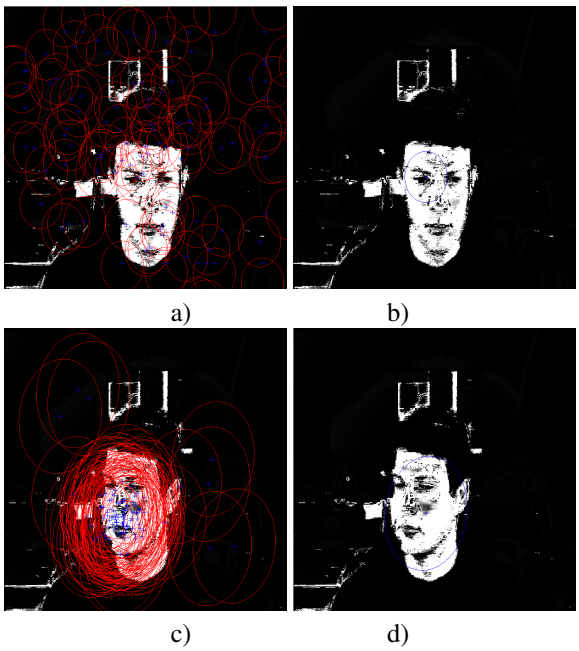


Figure 10: a) particles initialisation ; b) position image 1 ; c) particles image 13 ; d) position image 13

Size and pose estimation. From the ellipse center defined at the previous stage, a filling step then a binarisation is realised.
Moreover, the dynamics of the vector X_{2_t} is described by the following model : $[p(X_{2_t}/X_{2_{t-1}}) = N(X_{2_t}/X_{2_{t-1}}, \Sigma)]$ where $\Sigma = diag({\ell_t}^2, {L_t}^2, {\theta_t}^2)$ is the diagonal matrix made up of the variances of the X_{2_t} vector components.
Then face edges are extracted then approximated by an ellipse called measurement ellipse (Fig. 11 c,d). The Fitzgibbon al. [16] algorithm for ellipse fitting realises a compromise between speed and accuracy and is very robust to noise. It provides a measurement of ellipse parameters $\hat{X}_t = \left[\hat{x}_{c_t}, \hat{y}_{c_t}, \hat{\ell}_t, \hat{L}_t, \hat{\theta}_t\right]$.
Consequently the probability density is :

$$p(Y_{2_t}/X_{2_t}) = \prod_n p(Y_{2n,t}/X_{2n,t}) \text{ where} Y_{2n,t} \propto d^2$$

d is the eucledian distance between the predicted ellipse $X_{2_{n,t}}$ and the measured ellipse noted $\hat{X}_{2_t} = \left[\hat{\ell}_t, \hat{L}_t, \hat{\theta}_t\right]$.
The minimum of plausibility criterium enables to isolate the particle having the most probable form.
The results on the sequence (Fig. 12) exhibit the behaviour of the algorithm in the context of face viewed from aside (image 15) or partial occlusion (image 27).

5 Conclusion

The originality of the method consists in modelling the face skin by a Dempster-Shafer pixel fusion of three cognitive independant colour sources. Moreover, mass sets are determined from a priori models taking into account contextual

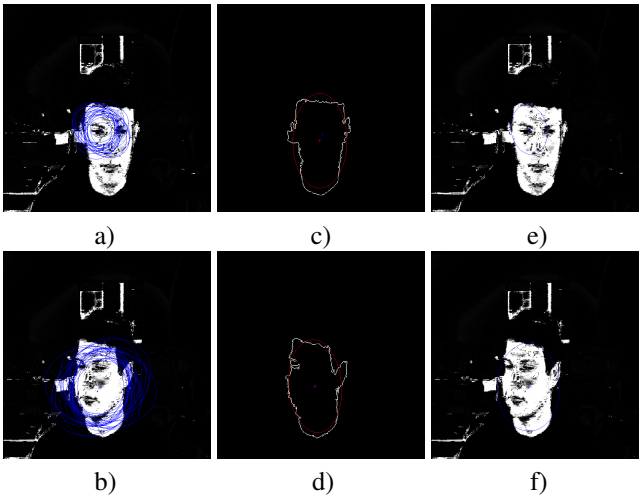


Figure 11: a,b) particles images 1 et 13 ; c,d) ellipse measurements images 1 et 13; e,f) results images 1 et 13.

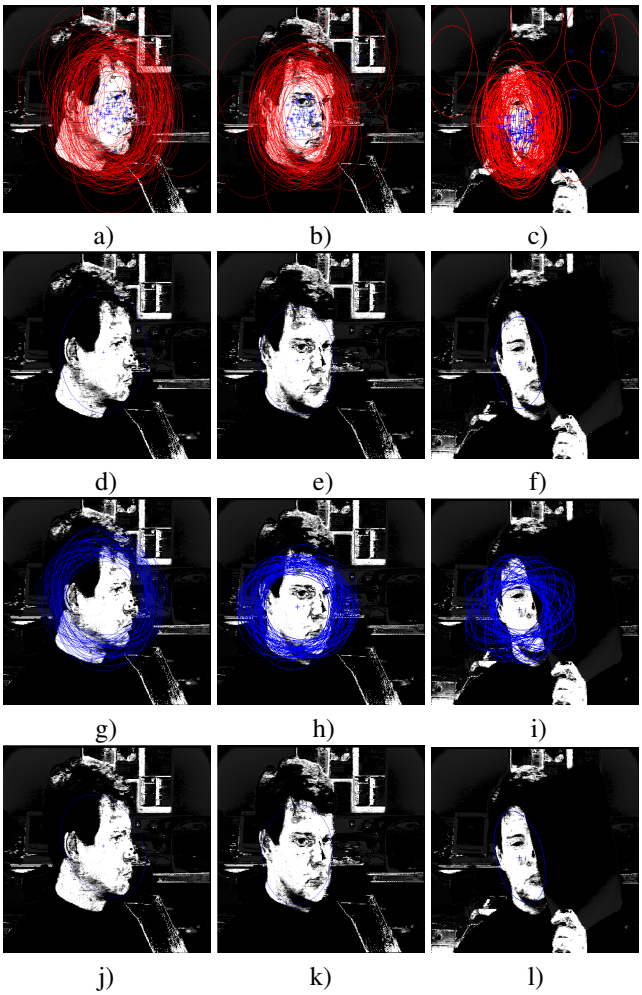


Figure 12: a,b,c) particles for position images 15,18,27 ; d,e,f) position images 15,18,27 ; g,h,i) particles for size and pose images 15,18,27; j,k,l) results images 15,18,27.

variables specific to the face under study. The shape criterion (ellipse) used in the particle filter algorithm removes modelisation artefacts and so improves segmentation and tracking.

However the fusion of only colour information is sometimes insufficient. The use of other independent cues such as the movement, texture or edges should improve detection. A feedback loop will be integrated in order to dynamically adapt the fusion process parameters and thus to optimise the information reliability (precision, uncertainty ...).

Finally, the objective will be to elaborate an active vision system where the dynamic collaboration of information and a feedback loop should contribute to increase segmentation and tracking robustness.

References

[1] E. Hjelmås and B.K. Low, Face detection: A survey, *Computer Vision and Image Understanding*, Vol. 83, No. 3, pp. 236-274, Sept. 2001.

[2] Ming-Hsuan Yang, David Kriegman and Narendra Ahuja, Detecting faces in images: A survey, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34-58, 2002.

[3] Prem Kuchi, Prasad Gabbur, P. Subbanna Bhat and Sumam David S., Human Face Detection and Tracking using Skin Color Modeling and Connected Component Operators, *IETE Journal of Research*, Vol. 38, No. 3&4, pp. 289-293, May-Aug 2002.

[4] S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, Blob Analysis of the Head and Hands: A Method for Deception Detection, *International Conference on System Science* (HICSS'05), Hawaii, Track 1, p. 20c, 2005.

[5] C.Garcia and M.Delakis. Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection, *IEEE trans. on Pattern analysis and machine intelligence*, Vol. 26, No. 11, pp. 1408-1423, Nov. 2004.

[6] F. Yang et M. Paindavoine. Implementation of an RBF neural network on embedded systems : real-time face tracking and identity verification. *IEEE Trans. on Neural Networks*, Vol.14, No.5, pp. 1162-1175, 2003.

[7] C.C. Chiang, W.N. Tai, M.T. Yang, Y.T. Huang, and C.J. Huang. A novel method for detecting lips, eyes and faces in real time. *Real-Time Imaging*, Vol. 9, No.4, pp. 277-287, Aug. 2003.

[8] P. Viola and M. Jones, Fast and Robust Classification Using Asymmetric AdaBoost and a Detector Cascade,*Advances in Neural Information Processing System* 14,MIT Press, pp. 1311-1318, Cambridge, MA, 2001.

[9] Michael J. Swain and Dana H. Ballard, Color indexing, *International Journal of Computer Vision*, Vol.7, No.1, pp. 11-32, Nov. 1991.

[10] M. Liévin, F. Luthon, Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of faces in video. *IEEE Trans. on Image Processing*, Vol. 13, No. 1, pp. 63-71, Jan. 2004.

[11] Vezhnevets V., Sazonov V., Andreeva A., A Survey on Pixel-Based Skin Color Detection Techniques, *Proc. Graphicon-2003*, pp. 85-92, Moscow, Russia, September 2003.

[12] Dempster, A. P., A generalisation of Bayesian inference, *Journal of the Royal Statistical Society*,Series B 30, pp. 205-247, 1968.

[13] Shafer, G. *A Mathematical Theory of Evidence.* Princeton University Press,1976.

[14] Stan Birchfield, Elliptical Head Tracking Using Intensity Gradients and Color Histograms, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, pp. 232-237, June 1998

[15] Nanda, H. and Fujimura, K., A robust elliptical head tracker, *Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, pp. 469-47, May 2004.

[16] A. Fitzgibbon, M. Pilu and R. Fisher, Direct Least Squares Fitting of Ellipses, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 5, pp.476-480, May 1999.

[17] A. Appriou, DTIM Multisensor signal processing in the framework of the theory of evidence, *NATO/RTO Lecture Series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems*, Châtillon, Nov 1999.

[18] N. Gordon, D. Salmond and A. Smith. Novel approach to non linear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2) :107-113,1993.

[19] M. Isard and A. Blake. CONDENSATION-conditional density propagation for visual tracking *Int. J. Computer Vision,* 29(1) : 5-28,1998.

[20] P. Pérez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. *Proceedings of IEEE*, 92(3):495-513, 2004.