# AN EFFICIENT APPROACH FOR FINE-TUNING AND TRACKING OF FACE OBJECTS

*Raúl Medina Beltrán de Otálora, Stephan Herrmann, Paul Zuber, Walter Stechele*

*Raul.Medina@tum.de , Stephan.Herrmann@tum.de, Paul.Zuber@tum.de, Walter.Stechele@tum.de*
Lehrstuhl für Integrierte Systeme – TUM, Arcisstr. 21, 80290 Munich (Germany)

**ABSTRACT**

In this work we present a new face-tracking method based on fast segmentation and expansion plus relaxation processes. Markov Random Field (MRF) principles are considered in order to find the minimum energy configuration that fits to the segmented face. Active contours aspects are as well taken into account modifying the traditional conception of relaxation processes towards accelerating convergence. Pixel addressing optimisation is implemented in order to accelerate performance for recursive processing.

## 1. INTRODUCTION

Nowadays, the demand of video applications for detection and tracking of objects has been dramatically increased. As a particular case of this trend we find face tracking, which as paradigm is not very different to any other object-tracking problem. The main challenge of face tracking is that human faces are highly dynamic and non–rigid objects. Moreover, automatic object segmentation is an ill-posed problem. Automatic segmentation is therefore a problem without a general solution, at least at the current state-of-the-art. User-assisted segmentation offers an attractive solution by letting the user to introduce semantic aspects while keeping an important part of the process automatic. As explained before, face tracking could be thought of as a sub-set within the vast domain of object segmentation and tracking, therefore this premise clusters the problem in a smaller space.

This paper deals with the problem of first segmenting a face in an image to later on track it through consecutive frames based on both photometric and geometric considerations. Here, the term "photometric" is meant to group variables such as intensity and colour rather different to the "geometric" ones such as edges. For the sake of a better understanding of the work presented here some considerations must be taken into account. Face detection aims to answer the questions "is there a face in the picture? Well, where then?" Face recognition tries to answer the question: "do I know whose face is that?" we consider the problem of segmenting and tracking one or more faces within a video sequence, which is different.

For face detection two main approaches can be found in the literature; feature–based approach that extracts facial components such as eyes, nose, mouth, etc. versus the image–based approach that treats face detection as a recognition problem by training and learning. Moreover, most face trackers focus on modelling the object to track as an ellipse, a polygon, multiple blobs [1] or just track characteristic points such as eyes, nose, lips, etc. Our approach is to segment the face with the maximum possible accuracy and to track it (with deformations, occlusions, etc.) in a video sequence. Therefore, we assume to have somehow a rough estimation where a face could be in the image. Once this is accomplished, the first goal is to segment the face with maximum accuracy (only face pixels). The next step is to track every single face over the next frames keeping the segmentation accuracy. Finally, it has to be remarked that the face recognition problem is out of the scope of this paper. However, the results of this work could be applied for such purpose.

First, the segmentation problem has to be solved. In this case we take the image-based approach, see Section 2.1. Section 2.2 describes the fine-tuning of the mask by relaxation of the object boundary. The whole algorithm combining face-segmentation plus fine-tuning is presented in Section 3. Finally, conclusions and future work are presented in the last Section.

## 2. FORMULATION OF THE PROBLEM

As explained in the previous section we look at the tracking problem from a segmentation perspective. That is, taking segmentation as starting point. The goal of tracking a face can be formulated as finding the evolution of the segmentation mask through consecutive frames by means of expansion and relaxation processes of the contour, i.e., spatio-temporal fine-tuning of the segmentation mask.

### 2.1. Face segmentation

Face segmentation is currently a very active research area and the technology has come a long way since the survey of Chellappa et al. [2]. Skin modelling has become as well a very active field towards simplifying the decision rule that could discriminate between skin and non-skin pixels[3].

In this work the face segmentation stage aims to obtain a rough and fast estimation of face pixels to later on get a better segmentation mask applying expansion and relaxation processes. Pixels likely to belong to the face colour space are selected as a starting criterion.

Due to the nature of the object to segment, it is expected that not all pixels belonging to the face are properly detected, for example eyebrows, lips, eyes, etc. do have a

different colour. Thus, the colour histogram of the region of interest is analysed and three dominant colours are selected. It is expected to find some gaps or unlabelled pixels due to illumination, shadows, occlusions, etc.

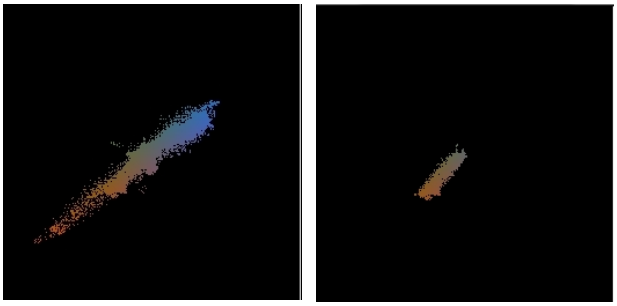

*Figure 1:Histogram of the U-V channels for the whole image (left) and for the region of interest (right)*

In order to collect as many valid pixels as possible a Euclidian distance is applied over these three dominant face colours and pixels with similar colour characteristics under a given threshold are selected. Figure 2 shows the colour histogram of a frame and the selected maxima. This similarity criterion produces three possible values, that is: a) labelled as face-pixel, b) labelled as not face-pixel, or c) unlabelled: not sure of this pixel is a face-pixel or not.

Once the most likely face-coloured pixels are collected following this elemental skin colour model, we have enough information to apply a watershed filter[4] in order to classify all the unlabeled pixels and produce the segmentation mask. This segmentation mask will be used later as initialisation for the fine-tuning.

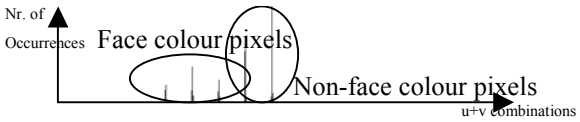


*Figure 2: Another representation of the colour histogram of the region of interest*

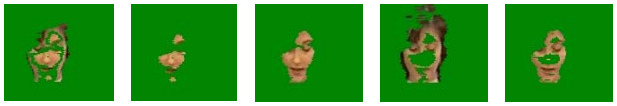

*Figure 3: rough estimation of face pixels depending on the distance to the supreme.*

Figure 3 shows the results of the face segmentation for different thresholds around the dominant colours (supreme values of the histogram). After applying the expansion process (see Figure 4) some pixels can be mistakenly taken as face pixels (hair, neck, surroundings of the face, etc). To solve this problem (we are only interested in the face) a relaxation-like process is applied to the segmented mask in order to have a fine-tuning of the contour, which is described in the next section.

*a)* *b)*

*Figure 4: a) Face after watershed filtering b) fine-tuning based on relaxation.*

### 2.2. Fine-tuning and tracking

Active contour models or snakes were introduced by Kass et al.[19]. They model curves and boundaries in images trying to minimise their associated energy function. The snake can be formally defined as a function of the arc length s by $v(s)=[x(s),y(s)]$ whit $s \in [0..1]$. The implied energy of a snake can by written in the following form:

$$E = \int E(v(s))ds = \int E_{\text{int}}(v(s)) + E_{image}(v(s))ds$$

Where $E_{int}$ represents the internal energy of the spline and $E_{image}$ is the energy introduced by the image itself driving the snake towards strong edges. Again here the formulation of the problem is well-known, traditional approaches model these two energy functions based in gradient minimisation. There have been many attempts aiming to solve the optimisation and minimisation of these energy functions including dynamic programming [20] and greedy optimisation [21] although they are still very sensitive to local minima caused by noise or initialisation.

On the other hand, Markov Random Fields (MRF) modelling has received a great deal of attention in the past decades [8][9][10][13]. This type of modelling, originally introduced in vision by Geman and Geman[14], has been widely used for edge detection[18], image restoration[12], stereovision, long-range motion and image classification [16].

Two main trends can be found in the literature, stochastic and deterministic relaxation. Stochastic relaxation approaches are based on Simulated Annealing[14][15]. These algorithms imply a high computational cost although assuring convergence in the global minimum. On the other hand, deterministic relaxation despite sub-optimal offers much faster results and there is a wide variety of algorithms (Graduated Non Convexity (GNC) [12], Iterated Conditional Mode (ICM) [11], Mean Field Annealing (MFA)[18], Modified Metropolis Dynamics (MMD) [17]).

Bearing in mind these two aspects; active contours and stochastic relaxation, we propose a new method able to accurately segment a face and track it over consecutive frames. Hence, the fine-tuning is viewed as a pixel-labelling problem and modelled as a pseudo-MRF, i.e., the image is viewed as a two-dimensional sequence of random variables as well as the active contour nature of the problem is considered. Two factors are taken into consideration for the labelling. On one hand neighbourhood influence (local statistics) through a MRF statistical dependence among the neighbouring pixels and on the other hand external influence (global statistics) due to the nature of the segment the pixel belongs to, proved to improve the behaviour of the algorithm for tracking of the contour. For a pixel, the energy function that we try to minimise is compound of three factors: local potential, external potential and an annealing temperature factor to converge to the global minimum.

We start from the assumption that there is a statistical dependency among neighbouring pixels, due to the assumption that the image is a Markov random field a local maximization of the probability also maximizes the global probability that the segmentation mask corresponds to the image content. The relaxation can change the label of a pixel and therefore the local probability that the pixel is a member of the corresponding segment is maximized. As local potential we compute the absolute distance (for the Y:U:V channels) to its eight neighbours i.e., photometric similarity constraints are considered as a minimising potential factor. If a neighbour has a different label a penalty is applied instead of the absolute difference for each different pixel/label in that given configuration. Geometric aspects are as well taken into account, straight lines are rewarded with lower potentials, and any other different configuration is penalized. The external potential is obtained by the absolute differences to the segment mean value and the dominant colours. This labelling scheme is performed later on for the tracking of the segmented faces.
Experiments on this labelling scheme have offered very good results for noisy TV sequences (see Figure 4 b). The basics for this fine-tuning of the contour are founded on a colour segmentation algorithm[5], in this particular case the fine-tuning had only local statistics into account

## 3. THE TRACKING ALGORITHM

In the previous section the fundaments of the tracking algorithm were briefly presented. Although there has been already work done on stochastic relaxation optimising active contours [22] the novelty of our work lies both on the extension of the MRF principles as well as its implementation. *Figure 5* shows how the problem of segmenting or tracking is addressed both based on the same fine-tuning scheme. First a face list is checked to know if there were already faces in the previous frame. For the first frame it is obvious that the way to proceed is to first detect and segment all the new faces (see Section 2.1).

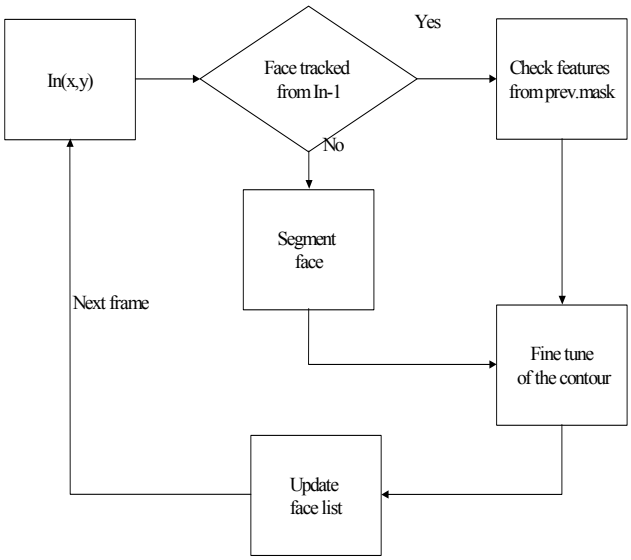


*Figure 5: Flow diagram of the tracking algorithm.*

Once the face is roughly segmented the fine-tuning is performed and the face list updated. For every face in the list a colour analysis is performed. Based on the assumption that the shape variation of the face between two consecutive frames is relatively low, see *Figure 6*, when processing a frame with already tracked faces; the fine-tuning is directly applied. Here the information from the immediately previous frame stored in the face list will be used as the external potential factor.

An important aspect of this implementation is that most of the computational complexity is concentrated in the fine-tuning. The segmentation phase is only carried out for initialisation purposes previous to the fine-tuning. Besides, the implementation of the relaxation is done using a recursive addressing scheme and implemented with a dedicated software library as described in the work done by Herrmann[6] for pixel addressing optimisation. First, only the pixels of the border are processed. And second, after applying the relaxation to a pixel, if its label does not change, that pixel will not be processed again. In the opposite scenario, if the pixel label is changed by the relaxation, it will be automatically queued up for the next iteration saving a lot of processing power.

*Figure 6: an illustration of the algorithm. The solid line indicates the initial position of the contour. The contour will flow to the dashed line, which minimises the energy function of the contour for that position of the face.*

## 4. CONCLUSIONS

In this paper we have presented a new algorithm for fine-tuning of the segmentation mask and its tracking applied to face objects. We combined two well-know techniques: active contours and stochastic relaxation. Experimental results have shown fast contour convergence mainly due to the extension of the MRF principles by introducing global statistics for the labelling decision. As well, it has been described that this relaxation scheme works both for segmentation and tracking as well as its implementation based on pixel addressing optimisation. The limitations of this algorithm are situations when there is no overlap between the segmentation mask of the previous frame and the face in the actual frame i.e., in *Figure 6* when the solid and the dashed line do not have an intersection. Then it is obvious that the relaxation will not succeed. In this case, running the segmentation phase again will solve the problem and provide a mask for the fine-tuning. Future work of this approach is to apply other external influences in the labelling stage such as global motion estimation; especially in scenarios where the photometric information is too fuzzy i.e., segments within the object are very different and therefore photometric similarity could fail. However, if the motion were too big even a block-matching algorithm would succeed but at a very high computational cost. Quantitative comparison against other segmentation/tracking solutions is still an ongoing task expecting promising results.

## 5. ACKNOWLEDGMENTS

This material is based upon work supported by the IST program of the EU in the project IST-2000-32795 SCHEMA (http://www.iti.gr/schema)

## 6. REFERENCES

[1] K. Toyama, “Prolegomena for Robust Face Tracking”, MSR Technical Report, MSR-TR-98-65, November 1998

[2] R. Chellappa, C. L.Wilson and S. Sirohey, “Human and machine recognition of faces: A survey”, Proc. IEEE, vol. 83, no. 5,1995.V.

[3] Vezhnevets, V. Sazonov and A. Andreeva, “A Survey on Pixel-Based Skin Color Detection Techniques”, Graphicon 2003, Moscow, Russia, Sep 2003.

[4] L. Vincent and P. Soille, “Watershed in Digital Spaces: An Efficient Algorithm based on Immersion Simulations”, IEEE Transaction on Patern Analyisis and Machine Intelligence, vol 13, pp. 583-598, 1991.

[5] S. Herrmann, H. Mooshofer, H. Dietrich, W. Stechele:"An Automatic - Semiautomatic Image and Video Segmentation Algorithm for Hierarchical Object Representation", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No. 8, Dec. 1999, pp. 1204-1215.

[6] S. Herrmann, H. Mooshofer, R. Medina Beltrán de Otálora, J.M. Martinez-Ibanez, W. Stechele: ”Image Processing in the MPEG-7 Reference Software using the AddressLib”WIAMIS 03, London,April 2003

[7] Klein, R. W. and Dubes, R. C. (1989) ”Experiments in projection and clustering by simulated annealing Pattern Recognition, 22(2): 213-220

[8] Azencott, R. (1987) “Markov fields and image analysis” Proc. AFCET, Antibes, 1987

[9] Azencott, R. (1992) “Parallel Simulated Annealing: Parallellization Techniques” Wiley.

[10] Bader, D.A. Jala, J., & Chellappa, R. (1995) “Scalable data parallel algorithms for texture synthesis using Gibbs random fields”. IEEE Tr. Image Processing, 4: 1456-1460.

[11] Besag, J. (1996) “On the statistical analysis of dirty pictures” Journal of Royal Statist. Soc. B-68: 259-302.

[12] Blake, A. (1989) “Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction” IEEE Trans. on Pattern Analysis and Machine Intelligence, 11: 2-12.

[13] Bouman, C.A. & Shapiro, M. (1994) “A Multiscale Random Field Model for Bayesian Image Segmentation”. IEEE Trans. Image Proc. 3: 162-177.

[14] Geman, S. & Geman, D. (1984) “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images”. IEEE Trans. Pattern Analysis and Machine Intelligence 6: 721-741.

[15] Kirkpatrick, S., Gellatt, C. & Vecchi, M. (1983) “Optimisation by simulated annealing”. Science 220: 671-690.

[16] Kato, Z., Zerubia, J. & Berthod, M. (1996) “A Hierarchical Markov Random Field Model and Multitemperature Annealing for Parallel Classification. Graphical Models and Image Processing”, 58(1): 18-37.

[17] Kato, Z., Zerubia, J. & Berthod, M. (1992) “Satellite image classification using a modified Metropolis dynamics”. Proc. of ICASSP, San Francisco, 1992. Circuits and Systems, 40(3 (II.)): 163-173.

[18] Zerubia, J. & Chellappa, R. (1993) “Mean Field approximation using Compound Gauss-Markov Random Field for edge detection and image estimation”. IEEE Trans. Neural

[19] M. Kass, A. Witkin, and D. Terzopoulos. “Snakes: Active contour models”. In Proceedings of the First International Conference on Computer Vision,1987.

[20] A. Amini, T. E. Weymouth, and R. C. Jain. “Using dynamic programming for solving variational problems in vision”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(9):855{867, September 1990.

[21] D. J. Williams and M. Shah. A fast algorithm for active contours and curvature estimation. CVGIP: Image Understanding, 55(1):14{26, 1992.

[22] D.Rueckert and P. Burger, "Contour fitting using stochastic and probabilistic relaxation for Cine MR Images". Computer Assisted Radiology 1995 (CAR 95), pp. 137-142, Berlin, 21-24 June 1995.