Multi-scale, Multi-feature Vector Flow Active Contours for Automatic Multiple Face Detection

Joanna Isabelle Olszewska

School of Computing and Engineering, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, U.K.

Keywords: Face Detection, Active Contours, Multiscale Snakes, Multi-feature Vector Flow, Unsupervised Segmentation.

Abstract: To automatically detect faces in real-world images presenting challenges such as complex background and multiple foregrounds, we propose a new method which is based on parametric active contours and which does not require any supervision, model nor training. The proposed face detection technique computes multi-scale representations of an input color image and based on them initializes the multi-feature vector flow active contours which, after their evolution, automatically delineate the faces. In this way, our computationally efficient system successfully detects faces in complex pictures with varying numbers of persons of diverse gender and origins and with different types of face views (front/profile) and variate face alignments (straight/oblique), as demonstrated in tests carried out on several datasets.

1   INTRODUCTION

Automatic face detection (Wood and Olszewska, 2012) is a cornerstone task in processes such as face recognition and analysis.

In particular, active contours are a suitable approach for face location and extraction in context of facial expression analysis (Lanitis et al., 2005) and facial animation generation (Hsu and Jain, 2003). Indeed, active contours are designed to automatically evolve from an initial position to the objects' boundaries by means of internal forces such as elasticity and rigidity and external forces computed based on image features (Olszewska, 2009). Therefore, active contours are continuous and smooth curves whose deformability enables them to precisely delineate the shape of foregrounds such as faces.

However, the current methods using active contours to detect faces are usually only dealing with images containing one single foreground such as in (Hanmin and Zhen, 2008), (Kim et al., 2010), (Zhou et al., 2010); a simple background (Sobottka and Pitas, 1996), (Gunn and Nixon, 1998), (Vatsa et al., 2003); frontal views of face(s) (Yokoyama et al., 1998), (Perlibakas, 2003), (Huang and Su, 2004); or they are time consuming as they require template learning (Lanitis et al., 2005), (Li et al., 2006) or prior training (Bing et al., 2004).

Another important issue for automatic face detection using active contours is their initialization.

Indeed, some approaches use manual initialization (Gunn and Nixon, 1998), (Vatsa et al., 2003), or quasi-automatic one which requires one user-defined point (Tauber et al., 2005) or two end-points (Neuenschwander et al., 1994), and thus are not fully automatic.

Automatic initialization techniques have been proposed in the literature and are usually based on the segmentation of the external force field as in the case of Center of Divergence (CoD) (Xingfei and Jie, 2002), (Charfi, 2010), Force Field Segmentation (FFS), or Poisson Inverse Gradient (PIG) initialization (Li and Acton, 2008). However, these approaches are limited to images with simple backgrounds. Other automatic methods use dense grids of boxes (Heiler and Schnoerr, 2005) or variations (Ohliger et al., 2010), but these techniques lead to the detection of groups of objects and not of distinct foregrounds. Some initialization techniques such as (Pluempitiwiriyawej and Sotthivirat, 2005) or (Olszewska, J. I. et al., 2007) are rather based on the object-of-interest movements, and therefore are not suitable in the case of static images.

Hence, some automatic initialization techniques more specific for face detection have been developed, and mainly rely on skin color information (Hanmin and Zhen, 2008), (Harper and Reilly, 2000), but they are not robust if other body parts are visible as well. Approaches using the elliptical shape detection (Huang and Su, 2004) fail in presence of other ellip-
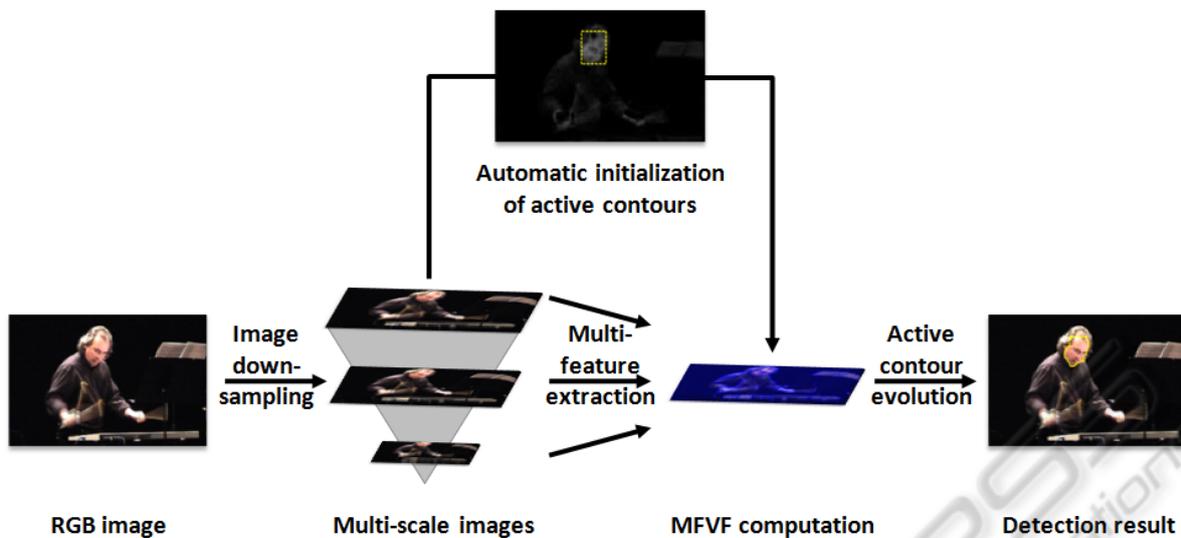
Figure 1: Overview of our automatic system of face detection based on multi-scale, multi-feature vector flow (MFVF) active contours.

tical foregrounds in an image, while those computing the interframe difference (Zhou et al., 2010), (Bajpai et al., 2011) are restricted to image sequences. Methods relying on the symmetry test (Yokoyama et al., 1998) or facial feature tests (Sobottka and Pitas, 1996) are only suitable for single-face images and are not adapted to profile views.

Thus, we propose in this paper to develop an active contour technique for fully automatic detection of multiple faces of people from different origins and with diverse gender, photographed under different views such as profile or frontal ones and whatever their alignment (straight/oblique), given a static picture.

For this purpose, we have applied the multi-feature vector flow method introduced by (Olszewska, J. I. et al., 2008) on multi-scale color images. In this way, the evolution of the active contours relies on features such as edges extracted from the original image and regions computed based on the scaled images.

We propose also an original automatic initialization technique based on color and area-based criteria, whose major advantages are:

- robustness even if other body parts are visible, unlike skin-color techniques;

- robustness when other elliptical foregrounds are present in an image, unlike methods using elliptical tests;

- compatibility with static images, unlike motion-driven approaches;

- robustness in case of complex backgrounds, unlike CoD techniques;

- robustness when faces are close to each other or occluded, unlike grid techniques;

- suitability for profile views of faces, unlike methods relying on facial symmetry detection;

- online compatibility since no training is required such as in Viola-Jones-based approaches.

Our multi-scale, multi-feature vector flow active contour method embeds the multi-target approach described in (Olszewska, 2012) to efficiently detect multiple faces in complex-background images, while our approach developed for the automatic initialization of the parametric active contours involves the use of RGB color space representation of real-world pictures.

In summary, the contribution of this paper is twofold. On one hand, we present a new active contour approach which involves a multi-scale, multi-feature vector flow leading to the computation of an efficient external field for a fast, robust, accurate, and automatic delineation of foregrounds. On the other hand, we introduce a new method for the automatic initialization of active contours. Our new multi-scale initialization approach uses color and areas criteria and outperforms other state-of-the-art active contour initialization methods in the case of multiple face detection.

The paper is structured as follows. In Section 2, we describe our automatic face detection method (see Fig. 1) which uses multi-scale representations of a color image in order to initialize and compute the multi-scale multi-feature vector flow active contours. The resulting system that automatically delineates the multiple faces present in the processed im-

Figure 2: Our system's performance in the case of multiple face detection in images with (a) mixed profile and frontal views; (b) variate elliptical foregrounds.

age has been successfully tested on real-world pictures characterized by complex background, varying numbers of persons of diverse gender/origins and presenting frontal as well as slanting views with variate alignments, as reported and discussed in Section 3. Conclusions are drawn up in Section 4.

2 MULTIPLE FACE DETECTION USING MULTI-SCALE ACTIVE CONTOURS

The multi-scale, multi-feature vector flow active contour approach consists first in their automatic initialization (Fig. 3 (d)) based on multi-scale representations (Figs. 3 (a),(b),(c)) of the given image (Fig. 3 (a)) as explained in Section 2.1, while their evolution is guided by the Multi-Feature Vector Flow (MFVF) (Fig. 3 (e)) built on features such as edges and regions extracted from the original and the downsampled images, respectively, as detailed in Section 2.2. This method leads to the accurate delineation of each of the faces present in a given color image (Fig. 3 (f)), as described in Section 2.3.

2.1 Multi-scale Active Contour Initialization

Let us consider a color image $I(x,y)$ with M and N, its width and height, respectively, and RGB, its color space. Firstly, this image is n-times downsampled by applying n ($n \in \mathbb{N}$) different ratios d_n ($0 < d_n \leq 1$) to get a set of n downscaled images $I_D = \{I_{d_n}\}$.

Secondly, in order to take into account the skin color to pre-detect faces at each scale, the corresponding R_n and G_n channels are subtracted to compute n

images I_{RGn} as follows:

$$\forall I_{d_n}(R_n, G_n, B_n), \ I_{RGn} = R_n - G_n. \quad (1)$$

Then, an average image I_{RG} is computed based on the rescaled I_{RGn} images. In order to compute the initialization mask I_{MASK}, I_{RG} is binarized using T as the threshold, leading to

$$I_B(x,y) = \begin{cases} 1 & \text{if } I_{RG}(x,y) > T, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

and treated by morphological mathematical operation as follows:

$$I_{MASK}(x,y) = I_B(x,y) \bullet S, \quad (3)$$

where $\bullet$ is the *closing* operation and S is a 3x3 square structuring element.

Next, b bounding boxes (B_b) are computed by taking the minimum and maximum values in x an y, respectively, of all the b ($b \in \mathbb{N}$) candidate regions $R_b(x,y) \subset I_{MASK}(x,y)$, such as $R_b(x,y) \neq 0$.

Finally, m initial active contours are initialized using the m bounding boxes B_m (with $m \leq b$) validated by the following area proportion criterion:

$$\forall B_b(x,y), \ B_m = B_b \ if \ A_{BOX} > \frac{A_{IMAGE}}{24}$$
$$\& \ A_{BOX} < \frac{A_{IMAGE}}{12}, \quad (4)$$

with $A_{IMAGE} = M \times N$ and $A_{BOX} = |B_b(x_{max}) - B_b(x_{min})| \times |B_b(y_{max}) - B_b(y_{min})|$.

Thus, our approach shows better face detection rates than the state-of-the-art ones as discussed in Section 3.

Figure 3: Our system's performance in the case of multiple face detection in an image of a group consisting of several persons with diverse gender and origins: (a) original image; (b)-(c) downsampled images; (d) initial active contours; (e) multi-scale, multi-feature vector flow field; (f) final active contour results.

2.2 Multi-scale Multi-feature Vector Flow Active Contours

In this work, the selected features are, on one hand, the edge map f_1 computed as follows

$$f_1(x,y) = |\nabla(G_{\sigma_e}(x,y) * I(x,y))|^2, \quad (5)$$

where

$$G_{\sigma_e} = \frac{1}{2\pi\sigma_e^2} \exp\left(-\frac{x^2+y^2}{2\sigma_e^2}\right), \quad (6)$$

with G_{σ_e}, an isotropic two-dimensional Gaussian function with standard deviation σ_e.

On the other hand, the multi-scale I_{RGn} images as computed by (1) constitute the second type of features (f_2) used in this work.

Once these features are selected, their association is ensured by the generic algorithm providing a unique multi-feature vector flow (MFVF) $\mathbf{\Xi}(x,y) = [\xi_u(x,y), \xi_v(x,y)]$ vectorial field, and defined as a weighted combination of the N_F feature vector flow $\mathbf{\Xi}_j(x,y)$ fields (Olszewska, 2009).

Each feature vector flow $\mathbf{\Xi}_j(x,y) = [\xi_{uj}(x,y), \xi_{vj}(x,y)]$ is generated by minimizing the following functional

$$\varepsilon_j = \iint \mu_j(\xi_{uxj}^2 + \xi_{uyj}^2 + \xi_{vxj}^2 + \xi_{vyj}^2) + (f_{xj}^2 + f_{yj}^2)((\xi_{uj} - f_{xj})^2 + (\xi_{vj} - f_{yj})^2)\,dxdy, \quad (7)$$

where μ_j is the diffusion parameter and f_j is corresponding to the jth adopted feature.

Hence, each of the multi-feature vector flow active contours, which is a parametric curve $\mathbf{C}(s) : [0,1] \rightarrow \mathbb{R}^2$, evolves from its initial position computed in Section 2.1 to its final position, guided by internal and external forces as follows

$$\mathbf{C}_t(s,t) = \alpha\,\mathbf{C}_{ss}(s,t) - \beta\,\mathbf{C}_{ssss}(s,t) + \mathbf{\Xi}, \quad (8)$$

where $\mathbf{C}_{ss}$ and $\mathbf{C}_{ssss}$ are respectively the second and the fourth derivative with respect to the curve parameter s; α is the elasticity; β is the rigidity; and $\mathbf{\Xi}$ is the multi-feature vector flow (MFVF).

2.3 Multiple Face Detection System based on Parametric Active Contours

The overall, fully automatic multiple face detection system such as presented in Fig. 1 uses each of these resulting MFVF active contours initialized and computed as described in Sections 2.1 and 2.2, respectively, in order to automatically delineate each of the faces. This proposed system is compatible with on-line applications since it does not require any training or model learning phases and since the developed active contour approach is computationally efficient. Moreover, the MFVF mechanism leads to very robust active contours that could coexist without collapsing nor merging, even when multiple targets are

close/occluding each other as proven in (Olszewska, 2012). Hence, the automatic and simultaneous detection of all the faces is done quickly and precisely as demonstrated in Section 3.

3 EXPERIMENTS AND DISCUSSION

We have tested our system by carrying out several experiments on image databases such as *Music Ensemble* and *Group*, running the MatLab software on a computer with an Intel (R) Core (TM)2 Duo CPU T9300 2.5 GHz processor, 2 Gb RAM, 32-bit OS.

The dataset called *Music Ensemble* consists of 407 color images of different music ensembles ranging from *solo* to *octet* ones, with c. 50 images per class. These images have an average resolution of 320x480 pixels and present real-world backgrounds as well as multiple foregrounds.

The dataset *Group* contains 593 color images presenting different groups of people from any gender and origins. Image resolution is around 500x330 pixels. This dataset presents challenges such as multiple faces per view, mixed face views (frontal and profile ones) and varying (straight or oblique) face alignments.

Examples of the results obtained using our method to detect multiple faces in images of *Music Ensemble* and *Group* datasets are presented in Figs. 2 and 3, respectively.

In the first experiment, we have assessed the performance of our automatic initialization method by computing the face detection rate defined as

$$detection\ rate = \frac{TP}{TP + FN}, \quad (9)$$

with TP, true positive and FN, false negative. The obtained average face detection rate by our method has been reported and compared to state-of-art techniques in Table 1.

Table 1: Face detection rates.

(Harper and Reilly, 2000)	(Huang and Su, 2004)	our method
89%	91%	99%

We can observe that our method outperforms the state-of-art ones. Indeed, techniques like (Harper and Reilly, 2000) applying skin-color tests usually over-segment images, such as Fig. 2 (a), containing many visible body parts which are not only faces. Thus, these state-of-art approaches do not detect the exact number of faces, unlike our method. On the other hand, methods such as (Huang and Su, 2004), which

use the elliptical validation of the detected faces, fail e.g. in situations depicted in Fig. 2 (b), where elliptical objects of interest other than faces are present in the processed image, whereas in that case, our method successfully detects all the faces without being disturbed by any other round foreground such as the tambourine.

In the second experiment, we have measured the precision of our multi-scale, multi-feature vector flow approach in delineating the faces in *Music Ensemble* and *Group* image datasets, using the following MPEG-4 segmentation error (Se) metric

$$Se = \frac{\sum\limits_{k=1}^{N_{fn}} d_{fn}^k + \sum\limits_{l=1}^{N_{fp}} d_{fp}^l}{card(M_r)}, \quad (10)$$

where $card(M_r)$ is the number of pixels of the reference mask defined by the reference curve C_r (groundtruth); d_{fn}^k is the distance of the kth false negative pixels from the computed active contour to the reference curve with N_{fn}, the number of false negative pixels; and d_{fp}^l is the distances of the lth false positive pixels from the computed contour to the reference curve with N_{fp}, the number of false positive pixels.

The value of Se is in the range of $[0, \infty[$, and the smaller, the better. Our approach results as well as those of the state-of-art methods are reported in Table 2.

Table 2: Face delineation error.

(Vatsa et al., 2003)	(Perlibakas, 2003)	our method
0.11	0.10	0.01

It appears that the segmentation error of our method is much lower than those of other state-of-the-art methods. Hence, our method is accurate in both detecting and delineating multiple faces in color images.

Moreover, our method is ten to twenty-five times faster than (Vatsa et al., 2003) and (Perlibakas, 2003), respectively, and could thus be used to detect faces in image sequences as well as in static images, unlike state-of-art methods such as presented in (Bajpai et al., 2011).

4 CONCLUSIONS

In this work, we have proposed a computationally efficient and robust multiple-face detection system that is entirely automatic and could handle with real-world images of persons with diverse gender and origins, whose faces are taken under different views such as

frontal or profile ones and could have varying alignments from straight to oblique ones.

Our developed system is based on parametric active contours whose innovative automatic initialization is based on the set of downscaled images and applies new validation criteria involving skin color and area information. The evolution of these active contours is guided by the multi-scale, multi-feature vector flow mechanism which uses the original combination of edges and regions extracted from the multi-scale representations of the processed color image.

REFERENCES

Bajpai, S., Singh, A., and Karthik, K. V. (2011). An experimental comparison of face detection algorithms. In *Proceedings of the IEEE International Workshop on Conference on Machine Vision*, pages 196–200.

Bing, X., Wei, Y., and Charoensak, C. (2004). Face contour extraction using snake. In *Proceedings of the IEEE International Workshop on Biomedical Circuits and Systems*, pages S3.2.5–8.

Charfi, M. (2010). Using the GGVF for automatic initialization and splitting snake model. In *Proceedings of the IEEE International Symposium on I/V Communications and Mobile Network*, pages 1–4.

Gunn, S. R. and Nixon, M. S. (1998). Global and local active contours for head boundary extraction. *International Journal of Computer Vision*, 30(1):43–54.

Hanmin, H. and Zhen, J. (2008). Application of an improved snake model in face location based on skin color. In *Proceedings of the IEEE World Congress on Intelligent Control and Automation*, pages 6897–6901.

Harper, P. and Reilly, R. B. (2000). Color based video segmentation using level sets. In *Proceedings of the IEEE International Conference on Image Processing*, pages 480–483.

Heiler, M. and Schnoerr, C. (2005). Natural image statistics for natural image segmentation. *International Journal of Computer Vision*, 63(1):5–19.

Hsu, R.-L. and Jain, A. K. (2003). Generating discriminating cartoon faces using interacting snakes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1388–1398.

Huang, F. and Su, J. (2004). Multiple face contour detection based on geometric active contours. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 385–390.

Kim, G., Suhr, J. K., Jung, H. G., and Kim, J. (2010). Face occlusion detection by using B-spline active contour and skin color information. In *Proceedings of the IEEE International Conference on Control, Automation, Robotics and Vision*, pages 627–632.

Lanitis, A., Taylor, C. J., and Cootes, T. F. (2005). Automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5):393–401.

Li, B. and Acton, S. T. (2008). Automatic active model initialization via Poisson inverse gradient. *IEEE Transactions on Image Processing*, 17(8):1406–1420.

Li, Y., Lai, J. H., and Yuen, P. C. (2006). Multi-template ASM method for feature points detection of facial image with diverse expressions. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 435–440.

Neuenschwander, W., Fua, P., Szekeley, G., and Kubler, O. (1994). Initializing snakes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 658–663.

Ohliger, K., Edeler, T., Condurache, A. P., and Mertins, A. (2010). A novel approach of initializing region-based active contours in noisy images by means of unimodality analysis. In *Proceedings of the IEEE International Conference on Signal Processing*, pages 885–888.

Olszewska, J. I. (2009). *Unified Framework for Multi-Feature Active Contours*. PhD thesis, UCL.

Olszewska, J. I. (2012). Multi-target parametric active contours to support ontological domain representation. In *Proceedings of the Conference on Shape Recognition and Artificial Intelligence (RFIA 2012)*, pages 779–784.

Olszewska, J. I. et al. (2007). Non-rigid object tracker based on a robust combination of parametric active contour and point distribution model. In *Proceedings of the SPIE International Conference on Visual Communications and Image Processing*, pages 6508–2A.

Olszewska, J. I. et al. (2008). Multi-feature vector flow for active contour tracking. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 721–724.

Perlibakas, V. (2003). Automatical detection of face features and exact face contour. *Pattern Recognition Letters*, 24(16):2977–2985.

Pluempitiwiriyawej, C. and Sotthivirat, S. (2005). Active contours with automatic initialization for myocardial perfusion analysis. In *Proceedings of the IEEE International Conference on Engineering in Medicine and Biology*, pages 3332–3335.

Sobottka, K. and Pitas, I. (1996). Segmentation and tracking of faces in color images. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 236–241.

Tauber, C., Batatia, H., and Ayache, A. (2005). A general quasi-automatic initialization for snakes: Application to ultrasound images. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 806–809.

Vatsa, M., Singh, R., and Gupta, P. (2003). Face detection using gradient vector flow. In *Proceedings of the IEEE International Conference on Machine Learning and Cybernetics*, pages 3259–3263.

Wood, R. and Olszewska, J. I. (2012). Lighting-variable AdaBoost based-on system for robust face detection. In *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 494–497.

Xingfei, G. and Jie, T. (2002). An automatic active contour model for multiple objects. In *Proceedings of the IEEE International Conference on Pattern Recognition*, volume 2, pages 881–884.

Yokoyama, T., Yagi, Y., and Yachida, M. (1998). Active contour model for extracting human faces. In *Proceedings of the IEEE International Conference on Pattern Recognition*, volume 1, pages 673–676.

Zhou, Z., Yang, W., Zhang, J., Zheng, L., and Lei, J. (2010). Facial contour extraction based on level set and grey prediction in sequence images. *Journal of Computational Information Systems*, 6(6):1975–1981.