Contributed paper

OPTO-ELECTRONICS REVIEW 11(3), 247–251 (2003)

Face detection in colour images using fuzzy Hough transform

A. PIETROWCEW*

Institute of Radioelectronics, Faculty of Electronics and Information Technology,
Warsaw University of Technology,
15/19 Nowowiejska Str., 00-665 Warsaw, Poland

Novel, fast, accurate, and low complexity face localisation algorithm is presented. It is robust for hard scene conditions, such as complex background, different lighting, and face poses. It exploits spatial masking of face colour areas in normalised perceptually plausible tint-saturation-luminance (TSL) space where Canny filter is applied. Then, facial areas are approximated by ellipses of constrained proportions using fuzzy Hough transform on detected edges. Finally eye localisation technique is applied in Cr-Cb colour space in order to determine face pose, confirm face presence, and it enables us normalisation of face recognition.

Keyword: face localisation, Hough transform, TSL colour space.

1. Introduction

The general problem of human face localisation in digital images has very great importance for used today, highly automated and autonomous digital systems [1,2]. Effective methods of face detection and localisation are especially needed in human identification and verification systems which are more and more commonly installed in public places, such as, airports, railway and subway stations, and market centres. Precise and accurate determination of face area and also face features, e.g., eyes, mouth and nose, plays key role in this kind of applications, because it is initial step in face recognition. Face or generally speaking human tracking is also a very desirable element of city monitoring and surveillance systems.

Therefore a great need exists for defining reliable algorithms which allow effective localisation of all faces present in the image, independently of their size, pose, expression, and lighting conditions. The problem is very challenging because human faces are very complex objects of great variability. Additionally, faces in images are in many situations partially occluded by glasses, headgears, chins or other obstacles. All this factors have important impact on efficiency of face localisation methods.

The problem of face localisation, and more generally pattern detection, has very extensive literature and many diverse techniques are known which can be used for less or more effective finding of required objects. Localisation methods based on different psychovisual grounds are related to human perception of faces and distinct mathematical foundations. Some of them rely on methodology of local feature analysis. The first step in these solutions is establishing the groups of features characteristic for human faces and then building a mathematical model with the best fit. The model is built using a training data set. For instance, Sung and Poggio exploit Gaussian clustering to model distribution of face and non-face patterns [3]. Then, for each location during the image scanning a feature vector is constructed, as a difference between image pattern and face model. Based on its values, trained classifier determines whether face is present at a current position or not. Colmenarez and Huang use a family of discrete Markov processes in order to model face and background patterns and to estimate probability functions [4]. The detection process is carried out by computing the likelihood ratio between two classes of patterns using learned model parameters. As in the previous work, they need training data to properly construct their model.

The method based on neural networks has been proposed by Rowley, Baluja, and Kanade [5]. They used a multilayer perceptron which receives, at its input, the image blocks of size 20×20 pixels. This block is divided into segments and then processed independently by different branches in the network. The purpose of this is to appoint local face features and memorise them in the network structure. The processed blocks are extracted at various image resolutions in order to capture faces with different sizes. A slightly distinct technique, but based on the same idea, is presented by Osuna, Freud, and Girosi [6], which make use of support vector machines in face localisation problem.

2. Face detection algorithm

The method presented in this article, due to effective exploitation of colour information, allows for fast and robust

*e-mail: adamp@pbi.pl

localisation of human faces in digital images. The minimal height of the face image should not be less then 30 pixels. Technique is independent of illumination and to a large extent of spatial face pose. It gives very precise approximation of face area and is resistant to complex background conditions. Because of this features it can be directly used as initial stage for face recognition algorithms, without further preprocessing and complicated normalisation steps. In the consecutive steps of the algorithm, skin colour regions are selected, then edge detection is done and visible contours are masked by selected regions. Later, Hough transform is applied to find face ovals. At the last step, eyes positions are located.

2.1. Skin region selection

In the digital images, colour information presents important source of information for initial preprocessing stages. It provides fast image segmentation and is relatively robust to varying lighting conditions, changes in pose, scale or shading. Using colour information, it is possible to separate the objects of interest from even very complex background as opposed to grey level only images. However, in order to obtain adequate distinction between skin and non-skin regions, colour transformation effectively separating luminance from chrominance is needed at first. Then, plausible model of skin chrominance should be build for a threshold purpose. Many authors proposed very different colour spaces for the task of skin colour modelling. Most commonly used are NRGB (normalised rgb), HSI, CIEXYZ, YCbCr, YUV.

There are researchers, for example Shin *et al.* [7] arguing that colour space transformations, in general, do not improve face and non-face classes separability. Statistical properties of spaces remain almost unchanged. Terrillon and Akamatsu [8] showed, from the other side, that face skin colour distributions in TSL and normalized r-g spaces are most compact and yield the best fit to the unimodal Gaussian model.

In the presented work, there were tested three-colour spaces, normalised r-g, YCbCr and TSL. For each of them we have built a histogram on the basis of approximately 4000 photos, of Caucasian subjects, taken at different lighting conditions and poses. During the experiments, it was confirmed that the best results in skin colour selection are achieved for the TSL space. This space was further used for building of the lookup table for a segmentation process. The transformation from the RGB to TSL space was done using the following formulas

$$S = \left[\frac{9}{5}(r^2 + g^2)\right]^{1/2}, T = \begin{cases} \frac{1}{2\pi}\arctan\left(\frac{r}{g}\right) + \frac{1}{4}, & \text{if } g > 0, \\ \frac{1}{2\pi}\arctan\left(\frac{r}{g}\right) + \frac{3}{4}, & \text{if } g < 0, \\ 0, & \text{if } g > 0 \end{cases} \tag{1}$$

$$L = 0.299R + 0.587G + 0.114B$$

The image preprocessing step of the described algorithm was made of skin colour regions selection which were further treated as binary masks for removing contours, not directly related to human faces. Of course, to preserve face shape contours connectivity, binary masks were dilated using 9×9 rounded filters. Figure 1 presents the results of colour masking in different colour spaces.

2.2. Hough transform

The method proposed by Hough in 1962 and called from his name the Hough transform, exemplify the universal tool, allowing for localisation of specific shapes in images, on the basis of objects contours present in them. Shape detection is carried out by analysis of edge points found in the image. The two dimensional image point space, described by the image matrix $I_{m \times n}$, is thus mapped to the parameterised feature space Ω. Object representation in the new space Ω is of course more compact, then in the origi-

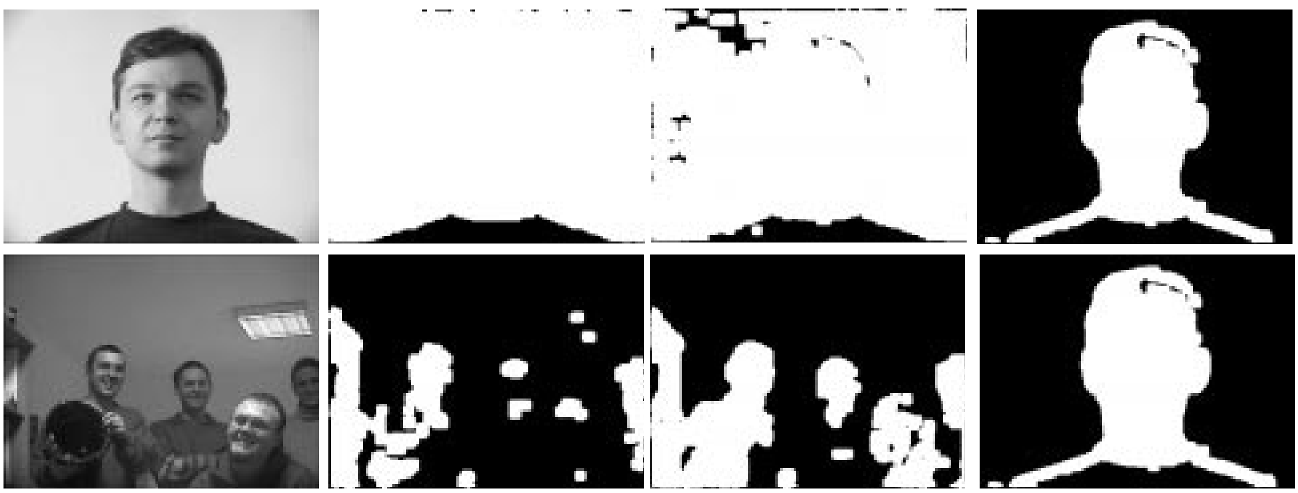

Fig. 1. Colour masks: left original image, next in CbCr, normalised r-g, and T-S space.

nal one defined on XY plane. Because of this, it is possible to localise objects in the image that covers large areas in the viewed scene.

There exist many forms of the Hough transform. For a comprehensive review refer to Illingworth and Kittler [9] or Leavers [10]. The simple form of it, standard Hough transform (SHT), increments values in the parameters space, which corresponds to the given edge point. In the case of generalised Hough transform (GHT), during the calculation of values in the parameter space, the normal determined on basis of gradients, at the given edge point is taken into account, additionally. This refinement reduces significantly the number of calculations necessary to fix the values in the feature space Ω. For the weighted Hough transform, a vote value in the parameter space is directly proportional to the gradient amplitude at the given edge pixel. Interesting solution, called weighted generalised Hough transform, was proposed by Bhnadrarkar [11]. In this technique, votes in the parameter space are inversely proportional to the distance between the given edge pixel and approximated shape. Application of this formula introduces blurring of votes to neighbouring locations in the parameter space Ω.

2.3. Fuzzy Hough transform

Because of high complexity of face localisation problem, due to large variability of conditions at which we capture the images and diversity of human faces, we exploited fuzzy methodology. It allows to better adjust the votes in the parameter space to human face shape, taking in account additional information present in the image.

In the proposed algorithm, shape of a human face is modelled as ellipse with constrained proportions, similarly to Seguier *et al*. [12]. For each dimension of ellipse axis, there is assigned one accumulator $A_i(a_i,b_i)$ which describes all possible ellipses in the image with a constrained size. All the above ellipses cover the image as in Fig. 2. Values of the parameters a_i and b_i are the subject to quantisation process in order to reduce a number of the analysed accumulators and to accelerate the voting. Having the gradient values Δx_P and Δy_P at the edge point P, we can determine coordinates of ellipse centre, passing through the point P. The values (x_i^c, y_i^c) are calculated for each accumulator $A_i(a_i,b_i)$ independently, according to the following equations

$$
\begin{aligned}
x_i^c &= x_P \pm \operatorname{sgn}(\Delta x_P)\frac{a_i}{\sqrt{1+\frac{b_i^2}{a_i^2}\left(\frac{\Delta y_P}{\Delta x_P}\right)^2}}, \\
y_i^c &= y_P \pm \operatorname{sgn}(\Delta x_P)\frac{b_i}{\sqrt{1+\frac{a_i^2}{b_i^2}\left(\frac{\Delta x_P}{\Delta y_P}\right)^2}}.
\end{aligned}
\tag{2}
$$

Because human heads in general are rather longer than wider, we can limit the proportions of the ellipse axes a_i and b_i in the considered accumulators. Without loosing the quality of localisation process we restrict, in the conducted experiments, the range of a_i to [0.6, 0.75] of b_i.

Of course, in reality, shape of the human face in the image, does not respond to ellipse too precisely, in many cases. Therefore in the applied method weighted fuzzy Hough transform was used. It means, that vote weight in the parameter space is directly proportional to the gradient amplitude at the given edge point. Moreover, an estimated vote weight is propagated to neighbouring ellipses in the parameter space, inversely proportional to the distance d of their centres from the calculated one, according to

$$
w_d = \begin{cases} \alpha_d e^{-\left(\frac{d}{\sigma_d}\right)^2}, & d \le \sigma_d \\ 0, & d > \sigma_d \end{cases} \tag{3}
$$

where σ_d determines the range of fuzzyfication and α_d characterises the level of influence on neighbours. Edge pixels, at which gradient amplitude is below threshold T_A, are not taken into account during image scanning.

In order to very closely approximate face oval, it was necessary to add additional parameterisation during Hough transform calculation. It is natural, that the pixel inside el-

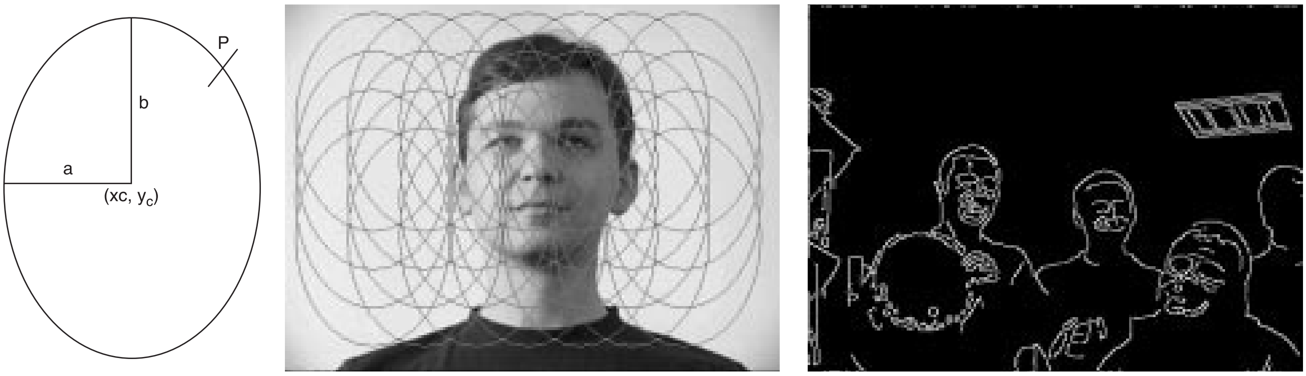

Fig. 2. Hough transform: ellipse parameterisation, sample accumulator, and extracted contours before masking.

lipse should have colour of face and outside, any other one. Assuming this, the new weight was constructed in such a way, that pixel values along normal at the point P, located at some small distance from P (dependent on the size of axis a_i, b_i in the given accumulator), were checked. The distances d_{c-}, d_{c+}, describing how far away the pixel colour is from T-S skin colour distribution, were computed and further used for votes weighting, in similar way as in Eq. (3). This assumption also allowed directly rejecting or filtering such points and ellipse votes for which the colours take opposite values along the normal, e.g., colour outside facial and inside non-facial.

Hough transform, in some situations, mistakenly generated ellipses which were large enough to generate many votes but which only had visible one side contours. These ellipses were not a result of face poses but rather scene cluttering by other object. To resolve this particular problem, each cell in Hough transform parameter space is split into four sub-cells, corresponding to ellipses quarters. The votes, for each the edge point P, are therefore distributed to appropriate ellipse sub-cells. Then, during a normalisation phase of a parameter space, the ellipses with equally distributed votes in all quarters are accepted, according to $v = i_b[(v_0 + v_1 + v_2 + v_3)/4]^2$, where i_b is the number of quarters with the vote weight above the threshold T_V .

Values in Hough transform parameter space are calculated only for edge pixels. In order to get them the following steps were applied:

- the luminance component of the image is blurred by Gaussian filter,
- gradients along OX and OY axis are calculated using 3×3 Sobel filter,
- Canny edge detector is applied to receive shape outlines,
- colour masking in T-S space is done.

The Canny method was used for edge detection as it gives continuous contours, even in the complex scenes and at low contrast.

2.4. Ellipse selection

According to the conditions described previously, the votes in the parameters space are the subject to weighting and fuzzification. As a result of accumulation process, taking place during scanning of all edge points in the image, we obtain, in the space Ω, local maxima peaks which correspond to potential faces present in the image. At the end, we choose among them, these which gathered most votes. For example, if we know that only one face should be present in the image we can choose one ellipse among all accumulators with maximum value of votes. However, because during the voting process the values are weighted and propagated to nearest neighbourhood, we rather get maxima clusters than single peaks. So, in the case of many faces, the analysis of only peaks can give selection of closely placed ellipses, corresponding to only one face. To overcome this drawback consolidation step is performed:

- for each accumulator the votes average T_{Pavg} and the variance T_{Pvar} are determined,
- only those ellipses are selected for which a vote value satisfies the condition $f > (T_{Pave} + T_{Pvar}/2)$,
- ellipses with similar sizes and near each other are merged, new vote value are calculated as average of 25% of ellipses with greatest votes weights,
- list of accumulated ellipses is sorted by votes weights,
- k first ellipses are picked from the beginning of the list for which the accumulated votes weights fulfil the condition $f_{accum} > 0.75Ta_{max}$, where Ta_{max} is the vote weight of the first element in the list.

After this, we get locations of potential faces in the image. To further confirm face presence, eyes detection is performed.

2.5. Eyes detection

Implemented approach to eyes localisation is similar to the one proposed by Hsu *et al.* [13], but we only analyse chrominance components during map building. In the hu-

Fig. 3. Face localisation results on sample images.

man face images, represented in Cb component only, the regions around eyes typically have high values, while in Cr component low ones. Having this, we construct eye map following to the formula

$$map[i, j] = (255 + Cb_{i,j} - Cr_{i,j}).$$

The morphological operations of dilation and erosion were used later to emphasis eyes and dismiss small disruptions.

3. Experimental results

The method was mainly tested on the source of Altkom facial database used by MPEG-7 group, which contains approximately 8000 colour photos of men and women, taken at different poses (from left to right half-profile, up and down, all intermediate poses) and lighting conditions. The proposed method allowed very precise localisation of face ovals in almost all images for this database. The obtained detection rate achieved 95.47%. Other test, conducted on the images acquired at our laboratory for different persons and groups of them also confirm high accuracy of the algorithm. The proposed method, in general, does not generate false positives. But in some rare situations, it omits the faces for which contours were not so clear, because of low contrast and skin colour masking effect. Masking process removes too many edge pixels on some faces and this causes that votes weights for ellipses approximating them were too small to choose them as possible face candidates.

4. Conclusions and future work

The proposed solution to face detection problem characterises very high accuracy overcoming many other face detection methods. It is also not complicated from computational standpoint and allows for very fast localisation of human faces in digital images. It was used in real time system for processing of about 5 frames per second on PIII 1.2 GHz CPU. Use of colour masking and fuzzy Hough transform with weighting and filtering, results thus in good, precise and fast detection system, which is robust to cluttered background. The future plans for face localisation include incorporation of motion estimation for human tracking in live video sequences.

References

1. X. Chen, L. Gu, S.Z. Li, and H.J. Zhang, "Learning representative local features for face detection", *Proc. IEEE CVPR*, 1126–1131 (2001).
2. M.H. Yang, N. Ahuja, and D. Kriegman, "Face detection using mixtures of linear subspaces", *Proc. IEEE ICAFGR'00*, 70–76 (2000).
3. K.K. Sung and T. Poggio, "Example-based learning for view-based human face detection", *IEEE TPAMI* **20**, 39–51 (1998).
4. A.J. Colmenarez and T.S. Huang, "Face detection with information-based maximum discrimination", *Proc. IEEE CVPR*, 782–787 (1997).
5. H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection", *IEEE TPAMI* **20**, 23–38 (1998).
6. E. Osuna, R. Freund, and F. Girosi, "Trainig support vector machines: an application to face detection", *Proc. IEEE CVPR*, 130–136 (1997).
7. M.C. Shin, K.I. Chang, and L.V. Tsap, "Does colour space transformation make any difference on skin detection?", *Proc. IEEE WACV'02* (2002).
8. J.C. Terrilon and S. Akamatsu, "Comparative performance of different chrominance spaces for colour segmentation and detection of faces in complex scene images", *Proc. ICAFGR'00*, 54–61 (2000).
9. J. Illingworth and J. Kittler, "A survey of the Hough transform", *Computer Vision, Graphics and Image Processing*, (1988).
10. V.F. Leavers, "Survey, which Hough transform?", *Computer Vision*, *Graphics and Image Processing* **58**, No 2 (1993).
11. S.M. Bhandarkar, "A fuzzy probabilistic model for the generalised Hough transform", *IEEE TSMC* (1994).
12. R. Sguier, A. Le Glaunec, and B. Loriferne, "Human face detection and tracking in video sequence", *Proc. RECPAD'95* (1995).
13. R.L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in colour images", *IEEE TPAMI* **24**, 696–706 (2002).

Handbook of Thin Film Devices

Academic Press

Editor-in-Chief: Maurice H. Francombe

Georgia State University, Atlanta

www.academicpress.com/thinfilms

The cutting-edge *Handbook of Thin Film Devices* provides the first multi-topical coverage of the semiconductor, optical, superconductor, magnetic, and ferroelectric devices and technologies responding to industrial needs. This superb reference comprises five topical volumes edited by world authorities in their fields, including semiconductor junction devices, semiconductor optics, superconducting film devices, magnetic film devices, and ferroelectric film devices. These leading experts cover recent progress in aspects ranging from deposition and fabrication to device modeling, measurements, and new cutting-edge design for improved performance. The versatile Handbook combines broad critical overviews with research level analysis and technical details.

***VOLUME ONE*:**
Hetero-Structures for High Performance Devices

- Epitaxial Film Growth and Characterization
- Processing of Epitaxial Heterostructure Devices
- Field Effect Transistors: FETs and HEMTs
- Heterojunction Bipolar Transistors
- GaN-Based Modulation-Doped Field-Effect Transistors and Ultraviolet Detectors
- Silicon Germanium Alloy Devices
- Silicon Carbide Power Devices
- GaN-Based Pyroelectronics and Piezoelectronics

***VOLUME TWO*:**
Semiconductor Optical and Electro-Optical Devices

- HgCdTe Infrared Detectors
- Antimony-Based Infrared Materials and Devices
- Quantum Well Infrared Photodetectors (QWIPs)
- An Introduction to the Physics of Quantum Well Infrared Photodetectors and Other Related New Devices
- Semiconductor Photoemissive Structures for Far Infrared Detection
- Interband III-V Mid-IR Semiconductor Lasers
- High-Performance Quantum Cascade Laser
- InGaN-Based UV/Blue/Green LED and LD Structure
- Plasma Displays
- Liquid Crystal Displays
- Recent Advances in Thin Film Solar Cells

***VOLUME THREE*:**
Superconducting Film Devices

- Synthesis and Characterization of Superconducting Thin Films
- Fabrication of Superconducting Devices and Circuits
- Digital Concepts and Applications
- Superconducting Microwave Device Properties and Applications
- Detectors for Infrared, Optical and X-Ray Photons
- SQUID Magnetometers

***VOLUME FOUR*:**
Magnetic Thin Film Devices

- Permanent Magnetic Films for Applications
- Magneto-Optical Recording Thin Films
- Epitaxial Garnet Films for Nonreciprocal Magneto-Optic Devices
- Microwave Magnetic Film Devices
- Magnetic Films for Planar Inductive Components and Devices

***VOLUME FIVE*:**
Ferroelectric Film Devices

- Ferroelectric Thin Films: Preparation and Characterization
- Ferroelectrical Random Access Memory (FeRAM) Technology
- Status of Metal-Ferroelectric-Semiconductor Field Effect Transistors (MFSFETs) and Related Devices
- Integration Aspects of Advanced Ferroelectric Thin-Film Memories
- Tunable Dielectric Materials and Devices for Broadband Wireless Communications
- Pyroelectric Devices and Applications
- Ferroelectric Devices for Microelectromechanical Systems (MEMS) Applications

ISBN: 0-12-265320-3
List Price: $1 800

