



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2001

2D to pseudo-3D conversion of "head and shoulder" images using feature based parametric disparity maps

Chaminda Weerasinghe

Motorola Australian Research Centre

Philip Ogunbona

University of Wollongong, philipo@uow.edu.au

Wanqing Li

University of Wollongong, wanqing@uow.edu.au

Publication Details

Weerasinghe, C., Ogunbona, P. & Li, W. (2001). 2D to pseudo-3D conversion of "head and shoulder" images using feature based parametric disparity maps. *IEEE International Conference on Image Processing* (pp. 963-966).

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

2D to pseudo-3D conversion of "head and shoulder" images using feature based parametric disparity maps

Abstract

This paper presents a method of converting a 2D still photo containing the head & shoulders of a human (e.g. a passport photo) to pseudo-3D, so that the depth can be perceived via stereopsis. This technology has the potential to be included in self-serve photo booths and, also as an added accessory (i.e. software package) for digital still cameras and scanners. The basis of the algorithm is to exploit the ability of the human visual system in combining monoscopic and stereoscopic cues for depth perception. Common facial features are extracted from the 2D photograph, in order to create a parametric depth map that conforms to the available monoscopic depth cues. The original 2D photograph and the created depth map are used to generate left and right views for stereoscopic viewing. The algorithm is implemented in software, and promising results are obtained.

Keywords

conversion, feature, parametric, 3d, images, pseudo, 2d, disparity, shoulder, head, maps

Disciplines

Physical Sciences and Mathematics

Publication Details

Weerasinghe, C., Ogunbona, P. & Li, W. (2001). 2D to pseudo-3D conversion of "head and shoulder" images using feature based parametric disparity maps. IEEE International Conference on Image Processing (pp. 963-966).

2D TO PSEUDO-3D CONVERSION OF "HEAD AND SHOULDER" IMAGES USING FEATURE BASED PARAMETRIC DISPARITY MAPS

Chaminda Weerasinghe, Philip Ogunbona, Wanqing Li

Motorola Australian Research Centre

ABSTRACT

This paper presents a method of converting a 2D still photo containing the head & shoulders of a human (e.g. a passport photo) to pseudo-3D, so that the depth can be perceived via stereopsis. This technology has the potential to be included in self-serve photo booths and, also as an added accessory (i.e. software package) for digital still cameras and scanners. The basis of the algorithm is to exploit the ability of the human visual system in combining monoscopic and stereoscopic cues for depth perception. Common facial features are extracted from the 2D photograph, in order to create a parametric depth map that conforms to the available monoscopic depth cues. The original 2D photograph and the created depth map are used to generate left and right views for stereoscopic viewing. The algorithm is implemented in software, and promising results are obtained.

1. INTRODUCTION

Human vision is stereoscopic by nature. However, the majority of composed visual information, such as images and video content are presently conveyed in the two-dimensional (2D) or monoscopic format. Monoscopic images lack realism due to the absence of depth cues. Artificially inserting depth cues in the forms of shading and perspective views is common practice especially in three-dimensional (3D) graphics arena. However, decades of research indicate that stereopsis is a more effective cue for depth perception. Generating left and right views based on an artificially created disparity map, for stereoscopic viewing is generally known as pseudo-3D.

There are well-established methods for virtual view face image synthesis using 3D wire-frame [1] or 3D spring based face models [2]. However, these models require a generic 3D-face model to be deformed to match a set of known facial landmarks. Complex deformation of a 3D model is computationally intensive. Since the goal of pseudo-3D is limited to generating only 2 complementary synthesized views, use of a wire-frame with complex computations is unwarranted.

A simple image based rendering technique using a low-complexity disparity map is adequate to generate the required views. It is shown in literature that a low-accuracy disparity map has the ability to generate a visually acceptable complementary image of a given view, although it is difficult to generate an accurate disparity map from two given views [3]. The basis for this observation is the ability of the human visual system in combining monoscopic and stereoscopic cues for depth perception. Therefore, the depth does not have to be completely and accurately represented in the disparity map for visual purposes. However, it is also important not to produce conflicting depths from monoscopic cues [4] and stereoscopic cues (i.e. disparity).

In this paper, a method of converting a 2D still photo containing the head & shoulders of a human (e.g. a passport photo) to pseudo-3D is presented. The algorithm is based on facial feature based low-complexity parametric disparity map. This method exploits the ability of the human vision in combining monoscopic and stereoscopic cues for depth perception.

2. PROPOSED METHOD

Figure 1 illustrate the block diagram of the proposed method consisting of the following steps:

1. Locating the head and shoulders boundary of the subject to remove the existing background.
2. Parametric representation of head contour using multiple ellipsoids.
3. Locating main facial features (e.g. eyes, nose).
4. Parametric representation of eyes and nose shapes using simple geometry.
5. Facial feature based parametric depth map creation.
6. Estimation of left and right views, given the original photo of the subject, the parameterized depth map and any background image of choice.
7. Composition of the left and right views for stereoscopic viewing via an anaglyph.

In order to simplify the background removal and facial feature extraction, the following assumptions are imposed:

1. The original (supplied) photo contains a human head and shoulders.

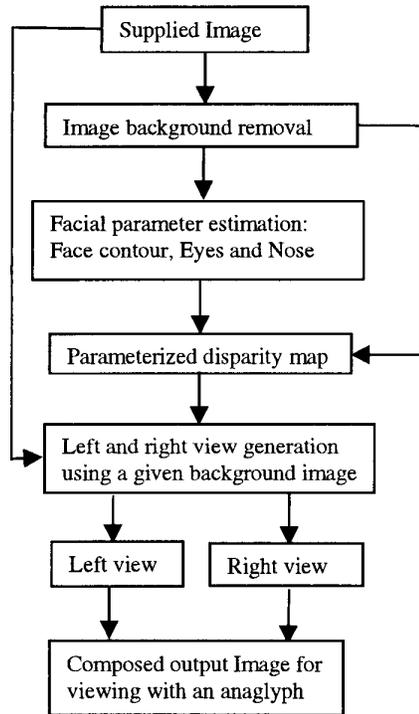


Fig. 1: Block diagram of the proposed 2D to pseudo-3D conversion system for head and shoulder photos.

2. The head orientation is such that the subject is looking directly towards the camera.
3. The photo should be taken with an uncluttered background.

It is implicitly known from the scope of this paper that, only a single still photo per subject is available and no camera parameters or range data are available.

3. HEAD AND SHOULDER MASK

Under the assumption of uncluttered background, the edge-image displays relatively low edge energy within the background. Therefore, a snake algorithm [5], which attracts to the edge energy, is chosen for the purpose of outlining the object mask of the “head and shoulder” images. The initial control points of the snake are assigned automatically. Each control point is moved inwards from the two vertical edges of the frame (**Fig. 2**) until an image edge is encountered. The separation of the initial points (δ) can be fixed as a percentage of the image frame height, and is defined *a priori*. For the test images used in this work, δ is defined as given in Equation 1.

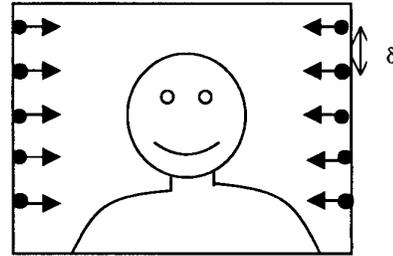


Fig. 2: Starting positions of search for snake control point initialization.

$$\delta = 1.5\% \times \text{FrameHeight} \quad (1)$$

Iterative minimization of snake energy moves the control points towards the nearest edge contours. The search is performed via greedy algorithm [5].

Since the converged contours for this particular application are similar in shape, the snake-parameters (i.e. $\alpha, \beta, \gamma, f_{db}$) [5] can be fixed for the entire system. For the experiments in this work, $\alpha = 0.5$, $\beta = 0.85$, $\gamma = 0.9$ and $f_{db} = 0.5$ are used. Improvements to the algorithm in terms of robustness is possible using the knowledge of the approximate shape of the final contour, which can be incorporated into the snake model.

It is also possible to implement the head and shoulder mask extraction stage to be interactive with an appropriate graphical user interface, which allows the user to adjust the final snake contour for improved accuracy.

Converged snake control points are then connected with lines using linear interpolation. Smoothness of the resulting contour can be ensured by using more snake control points (i.e. small value for δ), however, this will impose a computation time penalty. The final object mask is created by assigning pixel value I_{obj} ($= 255$) for the pixels enclosed by the snake contour and the bottom edge of the image frame, and assigning I_{bg} ($= 0$) to the remaining pixels. Although the head and shoulder mask do not have to match the shape of the head and shoulders exactly, a general shape conformity is assumed, as the parameter estimation for the depth map is based on this mask.

The following section describes the method of constructing the parametric depth map and parameter estimation based on extracted facial features.

4. PARAMETRIC DEPTH MAP MODEL

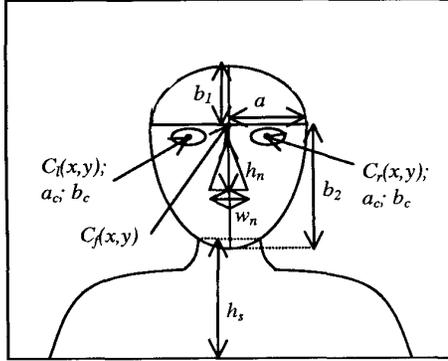


Fig. 3: Parameters of the facial feature based depth map model.

The following parameters are used for the feature based parametric depth map creation: Maximum disparity of ellipsoids for face (d_h), ellipse center (C_f), ellipse 1 / ellipse 2 minor axis half-length (a), ellipse 1 major axis half-length (b_1), ellipse 2 major axis half-length (b_2), height to the upper neck portion (h_s), maximum disparity of ellipsoids for both eyes (d_e), left eye ellipse center (C_l), right eye ellipse center (C_r), left / right minor axis half-length (a_c), left / right major axis half-length (b_c), maximum disparity of pyramid (3-sided) for nose (d_n), nose length (h_n) and nose width (w_n). Altogether there are 14 parameters assigned for the depth map.

4.1. Head and shoulders parameter estimation

The maximum depth of the head is determined using the proportion of the area of the mask (A) compared to the total image frame area ($H \times W$). Therefore, d_h is given by Equation 2.

$$d_h = \frac{r \times A}{H \times W} \times W = \frac{rA}{H} \quad (2)$$

Where $r=0.03$ is a scaling factor. The head contour is represented using two ellipsoids. The ellipse parameters are automatically estimated using horizontal projection width information of the head and shoulder mask. Let P_i represent the projection width at the i^{th} row, then a smooth contour (see Fig. 4) of the projection (P_{iA}) is obtained using Equation 3.

$$P_{iA} = \frac{1}{2q+1} \sum_{k=-q}^q P_{i+k} \quad (3)$$

Where $q=2$ is used in this work.

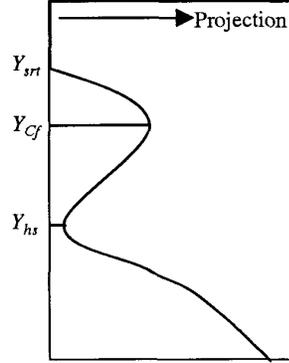


Fig. 4: Smooth projection contour of the head and shoulder mask.

Initially, the approximate neck position (Y_{hs}) is located using the projection width and its first derivative. The area above this point is searched for the maximum width of the head (W_h). Therefore C_f is computed using the mid-point of W_h and Y_{Cf} . If the width of the neck is W_{neck} , the required parameters are given by the following Equations.

$$a = 0.5 \times W_h \quad (4)$$

$$b_1 = Y_{Cf} - Y_{srt} \quad (5)$$

$$b_2 = \frac{(Y_{hs} - Y_{Cf})}{\sqrt{1 - ((W_{neck}/2)/a)^2}} \quad (6)$$

$$h_s = H - Y_{hs} \quad (7)$$

4.2. Eyes parameter estimation

Using the head parameters determined from Section 4.1, and a face template proposed by Hara et al. [6], the initial ellipses for the eye positions are located. An iterative algorithm, which maximizes the edge energy within the ellipse, computes the best fitting ellipse position and size (C_l, C_r, a_c, b_c). Disparity of the eyes (d_e) is 25% of d_h .

4.3. Nose parameter estimation

The location of the nose is determined using the distance between the centers of the eyes, the face template and nose contour extraction algorithm proposed by Hara et al. w_n and h_n are directly computed using the contour span and the distance from C_f . Disparity of the nose (d_n) is 25%

of d_h . In an interactive package, these parameters can easily be computed, if the user can select the facial features interactively using a graphical user interface.

5. RESULTS AND DISCUSSION

The following left and right views are generated from each given head and shoulder photo captured by a digital camera, using the proposed 2D to pseudo-3D conversion method.

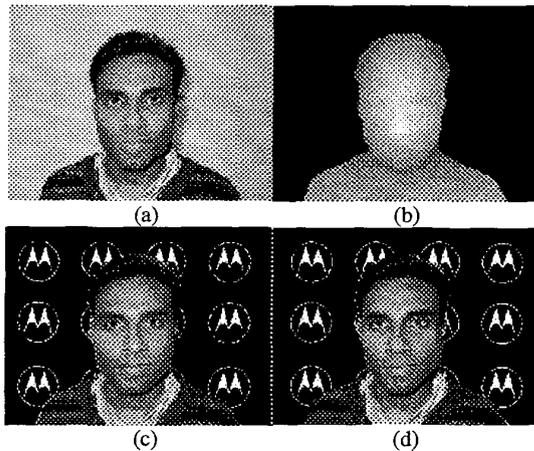


Fig. 4: Case 1 - A simple head and shoulder image: (a) Original photo; (b) Disparity map overlaid on object mask; (c) Generated left view and (d) Generated right view.

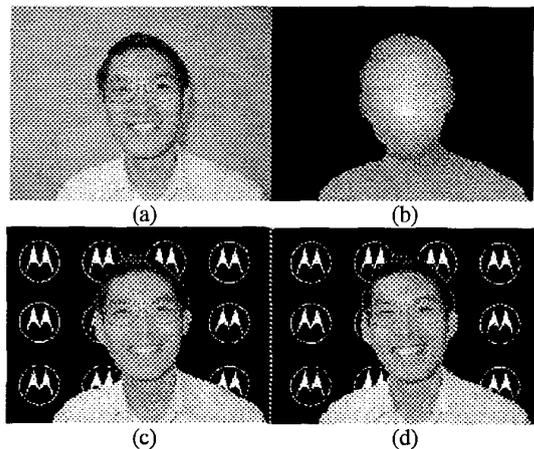


Fig. 5: Case 2 - Head and shoulder image with subject wearing spectacles: (a) Original photo; (b) Disparity map overlaid on object mask; (c) Generated left view and (d) Generated right view.

When the subject is having long hair, the algorithm failed to find the location of the neck. In this case, an additional mask was created by applying the snake energy minimization with higher feedback ($f_{db} = 1.1$) [5] until a strong edge is encountered.

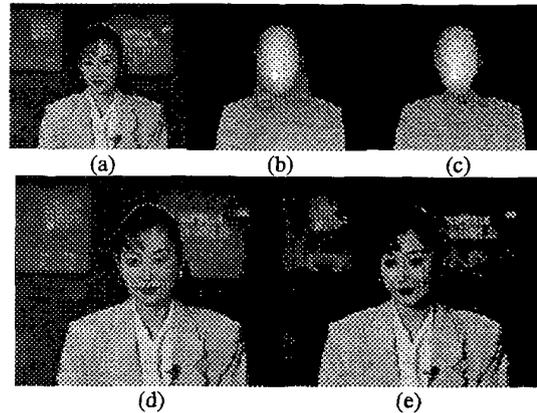


Fig. 6: Case 3 - Head and shoulder image with subject having long hair: (a) Original photo; (b) Depth map on object mask 1; (c) Depth map on object mask 2; (d) Generated left view and (e) Generated right view.

6. CONCLUSIONS

The acceptable quality of the results obtained using a simple depth map indicate that complex 3D face models are not required for conversion of 2D still photos in to pseudo-3D images. The proposed algorithm can easily be implemented in software as an interactive or non-interactive package. Use of an interactive graphical user interface will enable the constraints on image background to be released.

7. REFERENCES

- [1] M.W. Lee and S. Ranganath, "3D deformable face model for pose determination and face synthesis", Proc. Int. Conf. on Image Analysis and Processing, pp. 260-265, 1999.
- [2] G.C. Feng et al., "Virtual view face image synthesis using 3D spring-based face model from a single image", Proc. Int. Conf. on automatic face and gesture recognition, pp. 530-535, 2000.
- [3] G. Gimel'farb, "Experiments in probabilistic regularization of symmetric dynamic programming stereo", APRS/IEEE w/shop on stereo image and video processing, pp. 29-32, 2000.
- [4] S.B. Xu, "Qualitative depth from monoscopic cues", Int. Conf. on Image Proces. and its Applications, pp. 437-440, 1992.
- [5] L. Ji and H. Yan, "An Intelligent and attractable snake model for contour extraction", Proc. ICASSP'99, pp. 3309-3312, 1999.
- [6] F. Hara, K. Tanaka, H. Kobayashi and A. Tange, "Automatic feature extraction of facial organs and contour", IEEE Int. w/shop on robot and human communication, pp. 386-391, 1997.