Segmentation of a Head into Face, Ears, Neck and Hair for Knowledge–Based Analysis–Synthesis Coding of Videophone Sequences

Markus Kampmann
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 30167 Hannover, F.R.Germany
Email: kampmann@tnt.uni–hannover.de, WWW: http://www.tnt.uni–hannover.de/~kampmann

Abstract

Since in video telephony the image quality in the face is subjectively more important for a human observer than the image quality in other head parts like the hair, the neck and the ears, a knowledge–based analysis–synthesis coder should code different head parts with different qualities. For this, an automatic segmentation of a head into different head parts is necessary. In this contribution, an algorithm for automatic segmentation of a head into face, ears, neck and hair is presented. This segmentation is carried out using estimates for eye and mouth centers, chin and cheek contours, head silhouette and areas with skin color of the person. The proposed algorithm has been applied to the videophone sequences "Claire" and "Miss America". The segmentation of the heads into head parts is carried out with subjectively high accuracy.

1. Introduction

For coding of moving images at very low bit rates, an object–based analysis–synthesis coder (OBASC) has been introduced [1]. In an OBASC, real objects are described by model objects. A model object is defined by motion, shape and color parameters, which are estimated by an automatic image analysis. For the source model of moving 3D objects [2][3], the shape of a model object is represented by a 3D wireframe. The motion parameters describe translation and rotation of a model object, the color parameters denote luminance and chrominance reflectance of the object surface. In typical videophone sequences, head and shoulder of persons appear in the sequence. Using this a–priori knowledge about the sequence, an OBASC is extended in [4] to a knowledge–based analysis–synthesis coder (KBASC) by adaptation of the predefined 3D face model *Candide* [5]. First, eye and mouth center positions of the person in the sequence are estimated. Then, the face model *Candide* is adapted using these estimated center positions and is incorporated into the 3D wireframe of the person.

In videophone sequences, the image quality in the face is subjectively more important for a human observer than the image quality in other head parts like the hair, the neck and the ears. This can be exploited by a knowledge–based analysis–synthesis coder (KBASC) by coding different head parts with different qualities. For this, an automatic segmentation of a head into different head parts is necessary.

In the literature, some algorithms are proposed for segmentation of a head in videophone sequences. However, only the face is determined and no distinction between face, ears, neck and hair is carried out. In [6][7], the face is modelled by an ellipse. The parameters of this ellipse are estimated by searching for oval contours in the image. In [8][9][10][11], color segmentation is used for head segmentation.

In this contribution, a new algorithm for the segmentation of a head for a knowledge–based analysis–synthesis coder (KBASC) is presented. Whereas algorithms from the literature only estimate the face area, the proposed algorithm segments a head into face, ears, neck and hair. The algorithm requires as input data the eye and mouth center positions, the head silhouette and the chin and cheek contours of the person in the sequence. The eye and mouth center positions as well as the head silhouette are estimated by the algorithm in [4]. The chin and cheek contours are estimated with the algorithm in [12]. The new segmentation algorithm consists of five steps. In the first step, a color segmentation is carried out in order to estimate areas with skin color. This is described in Section 2. Using these estimated areas, the segmentation of the head into face, ears, neck and hair is carried out in the following four steps as described in Section 3.

2. Estimation of areas with skin color

In order to estimate areas with skin color, the color segmentation described in [13] is used. For this algorithm, the typical skin color of the specific person in the sequence has to be estimated first. For estimation of the typical skin color, areas on the person's cheeks are determined using the known eyes and mouth centers and chin and cheek contours of the person (Fig. 1).

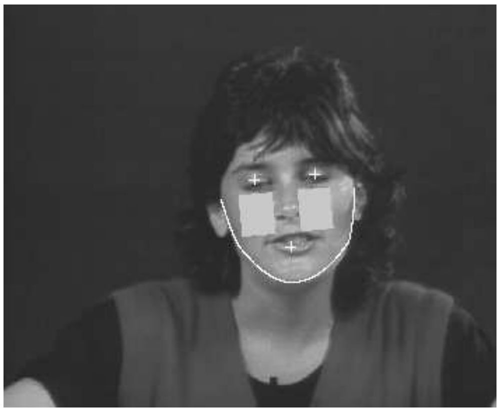

Fig. 1: Selected areas on the person's cheeks determined by estimated eyes/mouth centers and chin/cheek contours.

Inside these areas, the typical skin color of the person is estimated. The color value of each pel in these areas is represented in the YUV color space. For approximation of the typical skin color, a regression line through the origin of the YUV color space is calculated. Next, a cone is constructed around the estimated regression line. If a color value is inside the cone, it will be classified as skin color. If it is outside the cone, it will be classified as non skin color. Now, all pels of the image are examined and classified as skin color or non skin color. All pels classified as skin color are marked in a binary mask S. Fig. 2 shows an example of the mask S. Here, areas with skin color are marked like the face, the neck, the ears and the arm. Furthermore, parts of the hair are marked too. The eyebrows, the eyes, the mouth and shadows below the tip of the chin are not marked since they have a different color than the surrounding skin. Using the binary mask S, the face, the ears, the neck and the hair are estimated in the following steps.

Fig. 2: Binary mask *S* over original image. Pels with skin color are marked in *S*.

3. Estimation of face, ears, neck and hair

In a first step, the face of the person is estimated. For estimation, it is assumed that a human face has skin color and is bounded by the chin and cheek contours and by the hairline. For this reason, the largest area of S inside the estimated chin and cheek contours is selected. Subsequently, holes inside this selected part of S like eyes, eyebrows and mouth are filled. Fig. 3 shows an example of an estimated face.

Fig. 3: Estimated face.

Now, parts of the mask S outside of the chin and cheek contours and at the sides of the face are assumed to be ears (Fig. 4).

Fig. 4: Estimated ears.

For estimation of the neck, the part of the mask S bounded by the chin contour and below the face is selected first (Fig. 5).

Fig. 5: Selected part of the mask *S* bounded by the chin contour and below the face.

Holes inside the selected part like shadows below the tip of the chin are filled. Then, the length of the selected part is compared with the length of an average human neck. This length of an average human neck is determined using the distance between the estimated eye and mouth centers. If the selected part is shorter than an average human neck, the selected part is chosen as the neck. In the case of a neck–line showing more skin than the neck only, the selected part could be larger than an average human neck. In this case, the neck is calculated by additionally bounding the length of the selected part. This is done by using assumptions about the average length of a human neck (Fig. 6).

Fig. 6: Estimated neck after boundig the selected part of the mask *S* from Fig. 5.

Finally, the area of the hair is estimated. After estimation of the head silhouette using the algorithm in [4] (Fig. 7), areas of the head silhouette being neither face nor ears or neck are assumed to be hair (Fig. 8).

Fig. 7: Head silhouette.

Fig. 8: Estimated hair.

4. Experimental results

The described algorithm has been applied to the videophone sequences *Claire* and *Miss America* (CIF,10 Hz). Fig. 9 and Fig. 10 show representative results. Face and ears are determined with subjectively high accuracy. In the case of *Claire*, the length of the neck is bounded by the woman's blouse. In the case of *Miss America*, the length of the neck is bounded due to an assumption of the length of an average human neck. Because the hair region is restricted to the head silhouette in this algorithm whereas the hair of the persons in the videophone sequences also covers the shoulders, segmentation errors occur, especially at the transition from head to body.

5. Conclusions

An algorithm for segmentation of a head into face, ears, neck and hair for a knowledge–based analysis–synthesis coder (KBASC) has been presented. This algorithm requires the eye and mouth center positions, the head silhouette and the chin and cheek contours of the person in the sequence. Additionally, areas with skin color are estimated. Using all these informations, the segmentation into face, ears, neck and hair is carried out. The described algorithm has been applied to the videophone sequences *Claire* and *Miss America* (CIF,10 Hz). Face, ears and neck are estimated with subjectively high accuracy. Since the hair is restricted to the head silhouette, segmentation errors occur at the transition from head to body. Until now, no results from preceding images have been exploited for head segmentation of the current image. Taking into account these results, the described algorithm can be further improved. In the future, the described algorithm will be integrated in a KBASC and the different head parts will be coded with different qualities.

6. Acknowledgement

The author wishes to thank Dipl.–Ing. Ole Gerkensmeyer for software support and helpful discussions on this work.

7. References

[1] H.G. Musmann, M. Hötter, J. Ostermann, "Object–oriented analysis–synthesis coding of moving images", *Signal Processing: Image Communications*, Vol.3, No. 2, pp.117–138, November 1989.

[2] J. Ostermann, "Object–based analysis–synthesis Coding based on the source model of moving rigid 3D objects", *Signal Processing: Image Communications*, Vol.6, pp. 143–161, May 1994.

[3] J. Ostermann, "Object–based Analysis–Synthesis Coding (OBASC) based on the Source Model of Moving Flexible 3D Objects", *IEEE Trans. on Image Processing*, Vol.3, No.5, pp.705–711, September 1994.

[4] M. Kampmann, J. Ostermann, "Automatic Adaptation of a Face Model in a Layered Coder with an Object–based Analysis–Synthesis Layer and a Knowledge–based Layer", *Signal Processing: Image Communication*, Vol. 9, No. 3, pp. 201–220, March 1997.

[5] R. Rydfalk, "CANDIDE, A parameterised face", Internal Report Lith–ISY–I–0866, Linköping University, Linköping, Sweden, 1987.

[6] A. Eleftheriadis, A. Jacquin, "Automatic face location and tracking for model–assisted coding of video teleconferencing sequences at low bit–rates", *Signal Processing: Image Communications*, Vol.7, No. 3, pp. 231–248, September 1995.

[7] A. Eleftheriadis, A. Jacquin, "Automatic face location detection for model–assisted rate control in H.261–compatible coding of video", *Signal Processing:Image Communications*, Vol.7,Nos. 4–6, pp. 435–455, November 1995.

[8] E. Badiqué, "Knowledge–based facial area recognition and improved coding in a CCITT–compatible low–bitrate video–codec", *Picture Coding Symposium (PCS 90)*, Cambridge, Massachusetts, USA, No. 9.1, March 1990.

[9] D. Chai, K. Ngan, "Automatic Face Location for Videophone Images", *IEEE TENCON'96*, Perth, Australia, pp. 137–140, November 1996.

[10] T. Xie, Y. He, C. Weng, "A Layered Video Coding Scheme for Very Low Bit rate Videophone", *Picture Coding Symposium (PCS'97),* Berlin, Germany, pp. 343–347, September 1997.

[11] P. van Beek, M. Reinders, B. Sankur, J. van der Lubbe, "Semantic segmentation of videophone image sequences", *SPIE'92 Vol. 1818 Visual Communications and Image Processing '92*, Boston, Mass.,USA, pp.1182–1193, November 1992.

[12] M. Kampmann, "Estimation of the chin and cheek contours for precise face model adaptation ", *International Conference on Image Processing (ICIP–97)*, Santa Barbara, USA, pp. 300–303, October 1997.

[13] G. Klinker, S. Shafer, T. Kanade, "A Physical Approach to Color Image Understanding", *International Journal of Computer Vision*, Vol. 4, pp. 7–38, 1990.

Fig. 9: **Results of the proposed segmentation algorithm (videphone sequences *Claire*, *Miss America*): face and ears.**

Fig. 10: **Results of the proposed segmentation algorithm (videophone sequences *Claire*, *Miss America*): neck and hair.**