CCITT SGXV
Working Party XV/1
Specialist Group on Coding
for Visual Telephony

Doc # 294
Jan 1988

Title: **Feature Identification in Facial Images**
Source: **UK**

Introduction

As the search for methods of compressing image data even further, new techniques are required for rendering reconstructed images of adequate quality. Techniques relying on artificial intelligence to identify image contents can be used to augment existing image data compression methods or open doors for the development of fundamentally new algorithms. This document addresses the problem of identifying significant facial features in a typical head and shoulders video phone scene. The process described here has two distinguishable stages. The first consisting of primary feature enhancement and the second of feature identification.

Primary feature enhancement

The features of interest are the eyes, nose and mouth. It is thought that these features are particularly important in visual perception and therefore important to identify from a machine vision point of view. It has been determined that the eyes, nose and mouth can be characterised as consisting of dark areas surrounded by lighter areas. A technique based of morphological processing has been developed which is capable of identifying small dark areas within an image. Applying this technique to a frame extracted from the 'Missa' sequence (Figure 1) yields the result in Figure 2. The eyes , nose and mouth features predominate the result. This data set is further simplified by representing the black clusters of pels by a single pel placed at the centre of gravity of the cluster and also removing any cluster, the size of which fell below a threshold. This simplified data set may now be used to attempt feature extraction.

Feature Extraction

A technique known as 'back-tracking' [BALL82] was used to attempt feature identification. Consider the cluster positions being represented by a set of points S = {s1, s2, s3.....sn}. Trees consisting of the set points are constructed starting with each set point in turn. Each tree is constructed by assuming the the first point in the tree represent the left eye. The next level of the tree is generated by including all set points which lie on a preset bearing from the set point at the level above. For example, as the first point is expected to be the left eye, the next required feature is the right eye and would be expected to lie to the right within +/- 10 degrees of the horizontal. The tree is generated until all possible routes are exhausted or a return path has been found to the left eye. Only a tree which has a complete route from the left eye, through the features as required and back to the left eye can be considered as a candidate identifying the features.

Results

The video sequence shown illustrates the process applied to a section of the 'Missa' sequence. In general the features are identified, although mistakes are made sometimes. Figures 3 and 4 illustrate 2 frames extracted from the sequence.

Conclusion

The method presented here for feature extraction shows significant promise and has been shown to work well on a section of the 'Missa' sequence.

[BALL82]
D. H. Ballard, C.M Brown, Computer Vision, Prentice Hall Inc, New Jersey, 1982.

/4/

Figure 1: 'Miss America' original

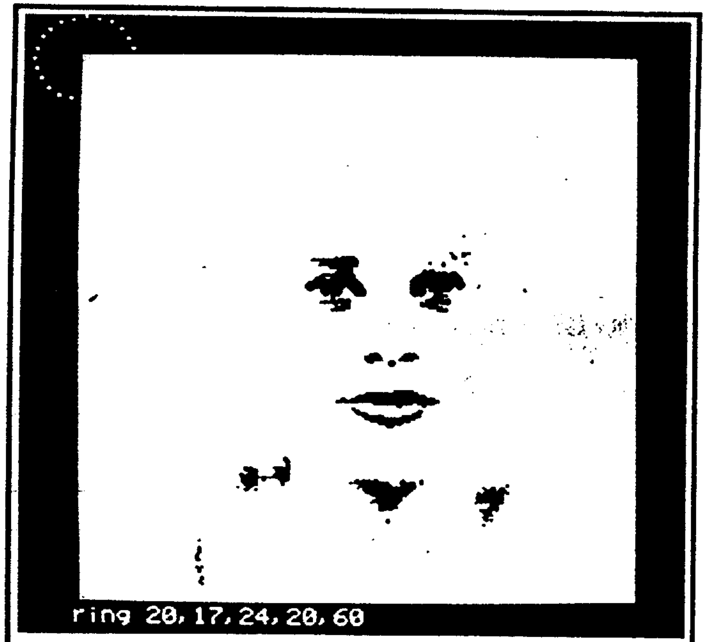

Figure 2: Dark region identification

Figure3 : 'Miss America', feature identification with error on right eye

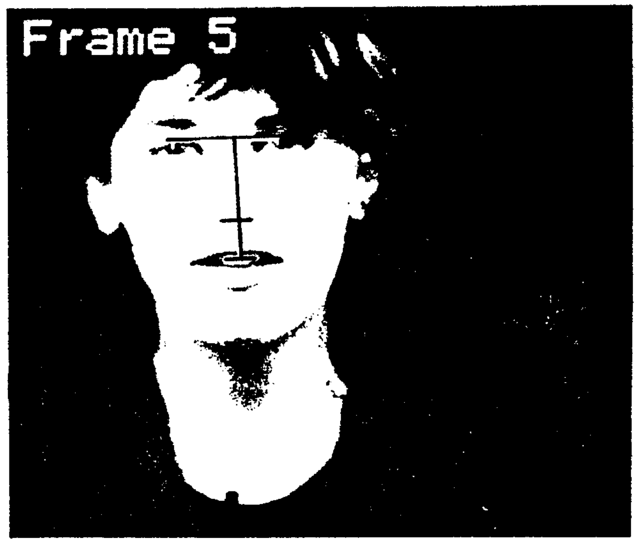

Figure 4: 'Miss America', features correctly identified