MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

HUMAN FACE ANALYSIS BASED ON DISTRIBUTED 2D APPEARANCE MODELS

Yasushi SUMI Yuichi OHTA

Institute of Information Sciences and Electronics
University of Tsukuba, Ibaraki, 305, Japan

ABSTRACT

We propose a new framework, called DCTOPS (Distributed Concurrent TOp-down Processing Scheme), for a computer vision system which is suitable in a parallel processing environment. A set of multiple top-down analyses are performed concurrently and distributively. Each of the analyses is based on a different 2D model corresponding to a different appearance of a 3D object. Using the framework it is possible to make a flexible and robust analysis mechanism without complex control structures. A human face analysis system has been developed in the framework.

INTRODUCTION

Locating a human face and its facial components in a scene is a basic process for various applications in computer vision and graphics which handle human faces. The fully automatic analysis of human faces, however, still remains as a challenging problem. In many researches, such as facial expression analysis, caricature generation and intelligent image coding, the locations of facial features are often extracted by hand. Researches of human face identification, which require fully automatic extraction of facial features, often uses a scheme of principal component analysis regarding the whole face image without locating the components on face[1][2]. The difficulty of the face analysis is caused by the huge variety of the appearances of a human face. In general it is very difficult to construct a computer vision algorithm which can cope with the variety of objects flexibly.

We have proposed a scheme called VIC (Vision module Integration in Cooperation and Concurrency)[3][4][5] as a basic concept on the integration of multiple vision algorithms in a parallel or distributed environment. The algorithm integration is indispensable for a vision system with flexibility and robustness. In the VIC, multiple algorithms each of which is a specialist to analyze only a specific target are constructed as independent agents. They work concurrently, they cooperate with each other and they try to achieve a shared goal.

In this paper we propose a new framework based on the VIC, called DCTOPS (Distributed Concurrent TOp-down Processing Scheme). The purpose of this framework is to realize a vision system which can analyze the structure of a 3D object in a scene although it presents various appearances in images. In the DCTOPS, multiple 2D appearance models are arranged in considering the variation of the 2D appearances of a 3D object. Multiple top-down analyses based on the 2D models are applied to an input image simultaneously. The mechanism which must be implemented in order to realize each of the analyses is nothing but a verifying process for the existence of features consistent with the model at a specified area in the input image. Therefore the structure of the mechanism is quite simple and the organization of the whole system is brief. The system, however, can cope with the variety of appearances flexibly owing to the redundant trials based on multiple 2D models.

A human face analysis system has been developed in the framework to demonstrate its feasibility. The pictorial structure of a human face may drastically change because of the 3D orientation of the head, the motion of facial components, and the existence of attachments such as glasses. Therefore it is quite difficult for an ordinary top down analysis scheme to cope with human face images without a tight restriction such as "front view" and "without glasses", etc.

In the following sections, we will describe the modeling scheme of a face in the DCTOPS and the implementation of the pilot system.

DISTRIBUTED 2D APPEARANCE MODELS

MULTIPLE APPEARANCE MODELS

A basic problem in computer vision is to recover the information of 3D world from 2D image data. This is an ill-posed inverse problem. Therefore it is necessary to reformulate it as a well-posed problem by incorporating a hypothesis on the scene to limit the solution space of the 3D world. A vision algorithm can be constructed under the hypothesis. This means, however, that every vision system is tuned to a single specific situation. It is true that in some environments it is relatively easy to establish an effective hypothesis on the scene. For example, in a factory environment, where the information on illuminant is known and the geometric parameters of the camera and the target objects can be easily obtained, many vision systems for manufacturing and inspection, are working in

practical situations. When the scene has variety, however, it is very difficult to establish a single hypothesis which covers the variety and is effective to build a vision system.

The DCTOPS proposed in this paper provides a framework to analyze objects which present various appearances. The variety of 2D appearances of a 3D object can be limited by setting an appropriate condition on the scene. For example, under a condition that the location and orientation of an object is relatively fixed to a camera, the appearance will be limited. Then we set multiple different conditions on the scene and establish a set of hypotheses $\boldsymbol{H}$.

$$\boldsymbol{H} = \{H_{c1}, H_{c2}, H_{c3}, \ldots, H_{cn}\},$$

where c_k is a condition set on the scene and H_{ck} is a hypothesis established under the condition c_k in order to derive a vision algorithm. When all the algorithms are carried out concurrently, a correct output will be obtained from the algorithm fitting to the input scene. Therefore if $\boldsymbol{H}$ is complete, it is possible to cope with the variety of all of appearances of the object.

In case that an object is represented by a geometric model such as polyhedron, the appearances can be classified into a finite number of clusters according to the visibility of each facet [7]. A human head, however, is a typical example which cannot be represented by a simple geometric model. It is difficult to classify the appearances mechanically according to geometrical conditions. However, we should realize that the purpose of the classification is not for a topological interest but for the analysis of image. That is, it is possible to classify the appearances into a finite number of clusters according to the applicability of the image analysis algorithms.

Appearance Model of Human Face

In this section we describe the distributed 2D appearance modeling of a human face for the extraction of facial components. Basic strategy for the modeling is to set conditions limiting the appearances of a human face in order to construct simple top-down algorithms.

We suppose that the appearance of a human face changes with two factors. Therefore the modeling is done at two levels as shown in Figure 1.

Facial component model: The first factor is the variation of facial components. For example, the appearance of a right eye can be limited by two conditions, the orientation $\{Front, Side\}$ and the state $\{Open, Close\}$. Then a set of hypotheses $\boldsymbol{H}_{RightEye}$ which specify the appearance of a right eye is established.

$$\boldsymbol{H}_{RightEye} = \{H_{FrontOpen}, H_{FrontClose}, H_{SideOpen}, H_{SideClose}\}$$

It is possible to construct an algorithm for extracting a right eye under a hypothesis $H_c \in \boldsymbol{H}_{RightEye}$. Each algorithm is a top-down process to verify the existence of a particular right eye. It is performed by extracting

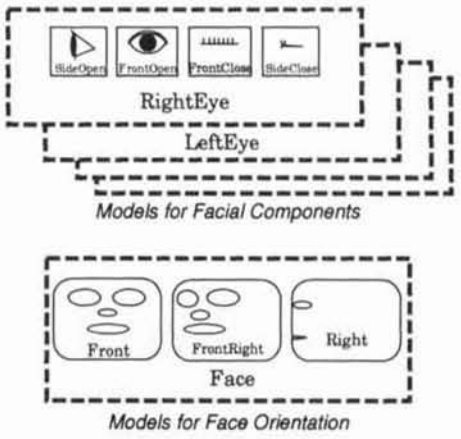

Figure 1: Appearance model of human face

evidences supporting the hypothesis H_c. Various feature extraction techniques can be employed to extract the evidences according to the pictorial structures. Basically, an algorithm is only effective to a specific appearance of right eye. However, the system is able to verify all the possible appearances of right eye based on the set $\boldsymbol{H}_{RightEye}$.

The set of hypotheses can be established in a similar way for each of the facial components. We denote the generic name of these sets as $\boldsymbol{H}_{FacialCompo}$.

Face model: The second and the more significant factor which affects the appearance of a face is the orientation of whole face. The change of face orientation distorts the relative location of facial components. We establish a set of hypotheses $\boldsymbol{H}_{Face}$.

$$\boldsymbol{H}_{Face} = \{H_{Right}, H_{FrontRight}, H_{Front}, \ldots, H_{Left}\}$$

The $H_f \in \boldsymbol{H}_{Face}$ is a hypothesis on the relative location of facial components and it is free from the variety of appearances of facial components. The algorithm to verify the H_f is a process to select facial components which agree to the hypothesis from the candidates extracted by facial component models. This is not only to make a good combination of the components in a bottom up way but also to update the confidence of facial components extracted by facial component models. The latter is realized by activating facial component models in order to extract lacking components based on the locations of known components. This means that face models control the behavior of facial component models in a top down way. It is effective to prune the useless search for facial components. When $\boldsymbol{H}_{Face}$ is complete, the system will be able to cope with all possible face orientations.

It is natural to regard that the two levels in the modeling of face appearance are in a hierarchical relationship. The face models are in higher level than the facial component models. In other words, $H_f \in \boldsymbol{H}_{Face}$ is a higher hypothesis than $H_c \in \boldsymbol{H}_{FacialCompo}$. A hypothesis which infers the location of a face in the scene is higher than H_f. This

kind of hypothesis will be used effectively if *a priori* information on the position of a person can be obtained. A hypothesis for the detail of a face, such as eyelid or lip, is lower than the H_c.

The DCTOPS is robust against the existence of attachments such as glasses. The attachments can significantly change the appearance of a face, so it has been difficult to cope with them by a simple analysis scheme. Especially this is serious for an analysis scheme which tries to extract features from the whole pattern of face without extracting facial components. On the other hand, the DCTOPS allows the existence of any attachments unless they do not affect the appearance of individual facial components. Attachments such as glasses or a beard can be modeled and extracted as facial components. In addition even though a face is partially invisible due to occlusion or noise, the DCTOPS can work to the rest of the face.

EVALUATION OF THE MODEL

In order to realize the mechanism described above, each model is implemented as an agent. We call the agent based on $H_f \in \boldsymbol{H}_{Face}$ as a face agent and the one based on $H_c \in \boldsymbol{H}_{FacialCompo}$ as a facial component agent, respectively. An agent carries out a top-down analysis by applying vision algorithms to appropriate regions on the input image based on its own hypothesis. Each agent is independent and autonomous. It can attempt to search the candidates of facial components by itself even when it cannot communicate with other agents.

An agent must have a function to evaluate the matching degree as a confidence value between its hypothesis and the evidence extracted on the input image. Confidence values from facial component agents are important information for face agents to verify their hypotheses.

The evaluation of confidence values in the DCTOPS is performed as follows. The vision algorithm in a facial component agent can be regarded as a mechanism to extract a set of evidences

$$\boldsymbol{E} = \{E_1,\ E_2,\ \ldots, E_n\}$$

for the verification of H_c. Where, n is the number of vision algorithms derived from H_c. By executing the algorithms to an input image, a set of actual features $\boldsymbol{e}$ is extracted.

$$\boldsymbol{e} = \{e_1,\ e_2,\ \ldots, e_n\}$$

Then a value $cf(\sim H_c \mid e_k)$, which indicates the negative confidence on H_c under the observation of e_k, is calculated as follows:

$$\begin{aligned} cf(\sim H_c \mid e_k) \ &= \ cf(E_k \mid H_c) \cdot cf(\sim E_k \mid e_k) \\ &\quad + cf(\sim E_k \mid H_c) \cdot cf(E_k \mid e_k), \end{aligned}$$

where $k = 1, \ldots, n$. The confidence values are handled in the framework of the Dempster and Shafer probability model[6]. Every $cf(\sim H_c \mid e_k)$ is combined into $cf(\sim H_c|\boldsymbol{e})$ according to the Dempster's rule of combination.

The hypotheses can be organized in a network illustrated in Figure 2. Confidence values obtained from the observation e_k are propagated through the arcs indicated as broken arrows and combined at a higher hypothesis H_f. In this way, the confidence value for H_f is evaluated based on the actual observation $\boldsymbol{e}$.

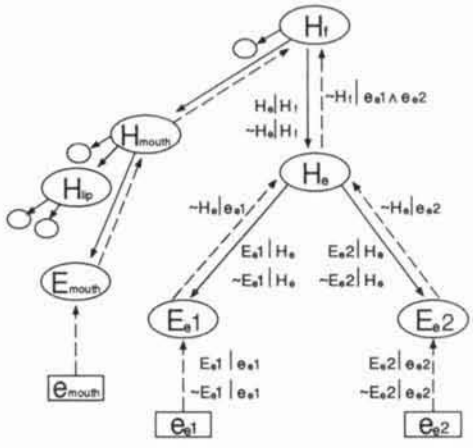

Figure 2: Hypothesis network

IMPLEMENTATION

Figure 3 illustrates the configuration of the system. Image data is accessed through image processing objects. An image processing object is a package of a basic image processing procedure such as smoothing, thresholding and so on. The facial component agent activates evidence extraction methods formed by combining multiple methods in image processing objects in order to verify the hypothesis. The facial component agents create a facial component objects from the extracted facial component for the face agents.

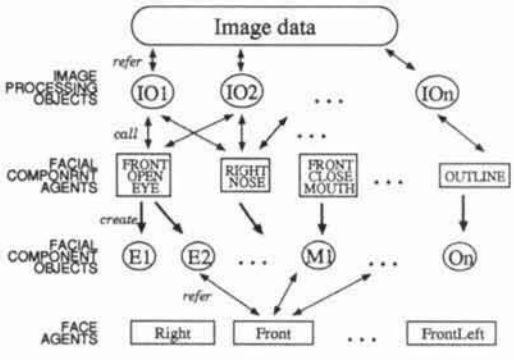

Figure 3: Configuration of the system

In the present system, the two kinds of methods for evidence extraction are implemented, edge based method and correlation based method. The edge based method uses three kinds of image processing objects, second order differentiation, thresholding, and binary shape analysis. By activating the three objects on an input image in that order, it is possible to obtain image features such as shape,

size, axis of moment. This method, however, is rather weak to noise. The correlation based method uses two kinds of image processing objects, template matching and thresholding. The template matching object decides a probable location of a facial component by calculating the correlation coefficient between the template and the input image. This is relatively tolerant to noise, but it is less flexible than the edge based method.

EXPERIMENTAL RESULTS

Examples of experimental results are shown in figure 4 and 5. The input image has 483×512 pixels and 256 intensity levels. The image data was obtained in a room with a CCD camera. The face position in the scene was not fixed strictly and the illuminating environment and the camera parameters were not cared for at all. Figure 4 is an input image on which the output from the agent based on the hypothesis $H_{FrontLeft}$ is superimposed. The agent had the best confidence value but the agent based on the hypothesis H_{Front} also output a fairly good result. Figure 5 shows that our scheme can adapt to a face wearing glasses. In this case the glasses are extracted as one of the facial components

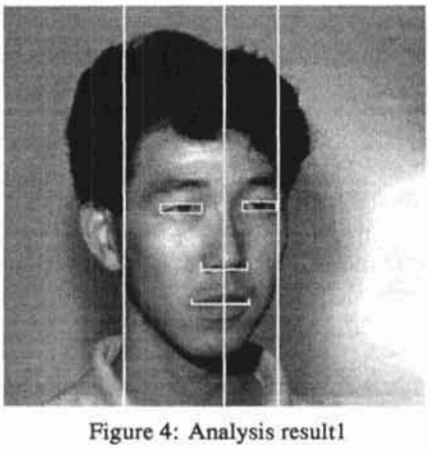

Figure 4: Analysis result1

The results shown here were obtained by a pilot system. The pilot system has been developed on a single CPU and simulates the behavior of the DCTOPS. At this point, It includes only three face agents, Front agent, FrontRight agent, and FrontLeft agent. Table 1 summarizes the score of analysis for 95 face images including the faces with glasses. It indicates that almost 90% of face images are analyzed successfully.

CONCLUSION

We proposed a framework to organize multiple 2D appearance models of a human face in a distributive way. The face analysis system developed within the framework can cope with various appearances of faces under the change of orientation and attachments and can successfully extract the locations of facial components.

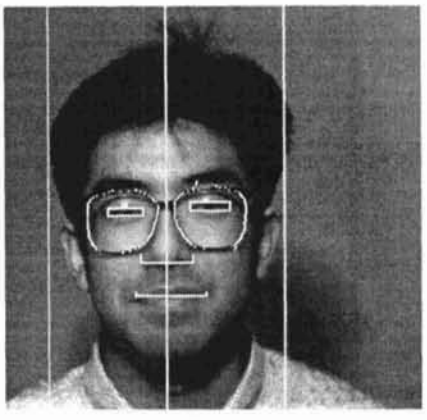

Figure 5: Analysis result2

The DCTOPS is suitable in a parallel processing environment. To realize a cooperative structure between agents on parallel processors is a future task.

This work is partly supported by the Grant in Aid for Scientific Research of the Ministry of Education, Science and Culture of Japan (Grant 04236106).

Table 1: Summary of results

face category	trials (glasses)	success (glasses)
right 15~30°	41 (19)	34 (14)
front	30 (15)	27 (13)
left 15~30°	24 (0)	24 (0)
	95 (34)	85 (27)

REFERENCES

[1] L. Sirovich, M. Kirby: "Low-dimensional procedure for the characterization of human faces", *J. Opt. Soc. Am. A*, **4**, 3, 519-524 (1987)

[2] A. Turk, A. Pentland: "Face recognition using eigenfaces", *Proc. CVPR*, 586-591 (1991)

[3] M. Watanabe, Y. Ohta: "Cooperative integration of multiple stereo algorithms", *Proc. ICCV*, 476-480 (1990)

[4] Y. Sumi, Y. Ohta: "A concurrent top-down analysis scheme and its application to human face images", *Proc. of International Conference on Automation, Robotics, and Computer Vision*, 1017-1021, Singapore (1990)

[5] Y. Ohta, *et al.*: "Approaches to Parallel Computer Vision", *IEICE trans.*, **E 74**, 2, 417-426 (1991)

[6] G. Shafer: *A mathematical theory of evidence*, Princeton Univ. Press (1976)

[7] K. Ikeuchi: "Generating an interpretation tree from a CAD model for 3D-object recognition in bin-picking tasks", *Int. J. Computer Vision*, **1**, 2, 145-165 (1987)