An And-Or Graph Model for Face Representation, Sketching and Aging

Feng Min[1,2], Jinli Suo[2,4] and Song-Chun Zhu[2,3]

[1]Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, China.
[2]Lotus Hill Institute for Computer Vision and Information Science, China.
[3]Departments of Statistics and Computer Science, University of California, Los Angeles.
[4]Graduate University of Chinese Acadamy of Sciences, China.
fmin.lhi, jlsuo.lhi@gmail.com sczhu@stat.ucla.edu

Synonyms

And-Or Graph model

Definition

For face modeling, an And-Or graph model was first proposed in [9] as a compositional representation for high resolution face images. In an And-Or graph, the And nodes represent coarse-to-fine decompositions and the Or-nodes represent alternative components for diversity. The And-Or graph face model, as Fig.1 illustrates, has three levels: the first level describes the general appearance of global face and hair; the second level refines the representation of the facial components (eyes, eye brows, nose, mouth) by modeling the variations of their shapes and subtle appearance; and the third level provides further details of the face components and divides the face skin into 9 zones where the wrinkles and speckles are represented. The And-Or graph provides a expressive model for face diversity and details, and thus is found to be especially efficient for applications in **face sketching** generation and **face aging** simulation.

Main Body Text

Introduction

Human faces have been extensively studied in computer vision and graphics for its wide applications: detection, recognition, tracking, expression recognition, non-photorealistic rendering (NPR), and many face models have been proposed, for example, EigenFace[1], FisherFace[2], Laplacianfaces[3] and their variants[4], deformable templates[5], the active shape models and active appearance models [6, 7, 8]. Most of these models are mainly used for face detection, localization, tracking, and recognition.

Although these face models have achieved reasonable successes in face detection, recognition, tracking, they use templates of fixed dimensions at certain low-middle resolutions, and thus are limited by their expressive powers in describing facial details in higher resolutions, for example, subtle details in the different types of eyes, nose, mouths, eyebrows, eyelids, muscle relaxations due to aging, skin marks, motes, and speckles. Consequently, these models are less applicable to applications that entail high precision, such as face sketch generation and face aging simulation. For the latter tasks, Xu et al[9] proposed a compositional And-Or graph representation for high resolution face images. Adopting a coarse-to-fine hierarchy with Or nodes representing the alternatives , the And-Or graph can represent a large diversity of human faces at different resolutions.

Compositional And-Or graph representation for faces

A compositional And-Or graph describes all types of faces collectively at low, medium, and high resolutions, as Fig.1 shows. There are three types of nodes in the And-Or graph – And-nodes, Or-nodes and leaf-nodes. An And-node either represents a

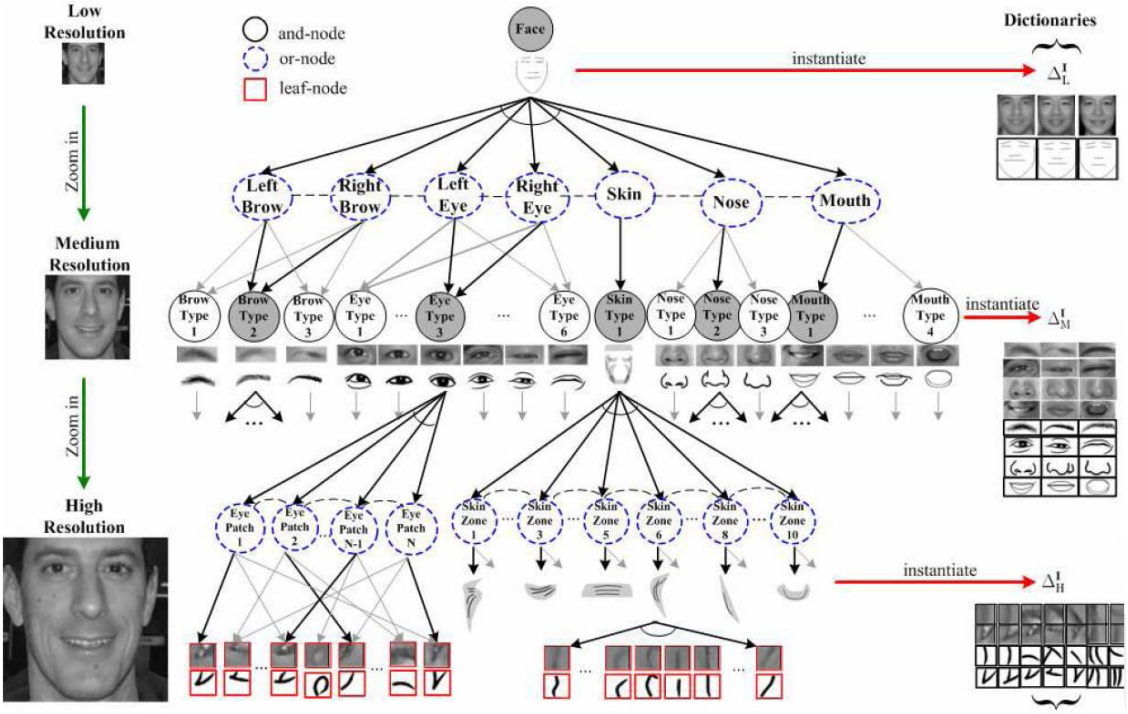

Fig. 1. An illustration of the compositional And-Or graph representation of human face. Left column is a face image at three resolutions. All face images are collectively modeled by a three level And-Or graph in the middle column. The And nodes represent decomposition and Or nodes represent alternatives. Spatial relations and constraints are represented by the horizontal links between nodes at the same level. By selection of alternatives, the And-Or graph turns into a *parse graph* for a face instance. The right column represents the dictionaries at three scales: Δ_H^I, Δ_M^I, Δ_L^I. *From Xu et al.[9].*

way for decomposition at higher resolution or terminates in a Leaf-node at lower resolution. An Or-node stands for a switch pointing to a number of alternatives components. For example, an Or-node of eye could point to different types of eyes. A leaf-node is an image patch or image primitive with shape and appearance attributes. The And-Or graph has horizontal lines (see dashed) to specify the spatial relations and constraints among the nodes at the same level. By choosing the alternatives at Or nodes, the And-Or graph turns into an And-graph representing a face instance, which is called a *parse graph*. Thus the And-Or graph is like a "mother template" which produces a set of valid face configurations, each of which is a composition of the image patches or primitives at its leaf nodes.

At the low resolution, face is represented as a traditional Active Appearance Model(AAM)[7], which describe the general face shape, skin color etc. At medium resolution, the face node expands to a number of Or-nodes for facial components (eyebrows, eyes, nose and mouth) and skin zone. For each component, a number of AAM models are used as the alternatives of the Or node. At high resolution, the nodes of facial component and skin zone further expand into a number of Or-nodes describing the local structure of components and free curves (wrinkles, marks, etc.) in details.

Model computation

For an input high resolution face image, the algorithm computes a parse graph in a Bayesian framework in three levels from coarse-to-fine.

At the first level, face image is down-sampled, and the algorithm computes the AAM-like representation W_L with global transform T, geometrical deformation α_{geo} and photometric appearance β_{pht} by maximizing the posterior probability,

$$W_L = \arg\max\ p(I_L^{\text{obs}}|W_L;\Delta_I)p(W_L) \qquad W_L = (T, \alpha_{\text{geo}}, \beta_{\text{pht}}). \tag{1}$$

At the second level, a number of AAM-like models are trained for each facial component. The algorithm takes a down-sampled medium resolution face image and W_L as the input and conducts a constrained search for W_M conditioned on W_L. The variables are computed by maximizing the posterior probability,

$$W_M = \arg\max\ p(I_M^{\text{obs}}|W_L, W_M; \Delta_I, \Delta_I^{\text{CP}})p(W_M|W_L) \qquad W_M = (l^i, \alpha^i_{l^i,\text{geo}}, \beta^i_{l^i,\text{pht}})_{i=1}^6 \tag{2}$$

At the third level, the face area is decomposed into zones, which refine the sketches of local structures based on the searching results at medium resolution level. The variables at this layer are inferred by maximizing the posterior,

$$W_M = \arg\max\ p(I_H^{\text{obs}}|W_M, W_H; \Delta_I^{\text{CP}}, \Delta_I^{\text{SK}})p(W_H|W_M) \qquad W_H = (K, \{(l_k, t_k, \alpha_k) : k = 1, 2, ..., K\}) \quad (3)$$

Applications

The And-Or graph face model has been applied to two applications – automatic face sketch and portraiture generation in [10] and face aging simulation in [11].

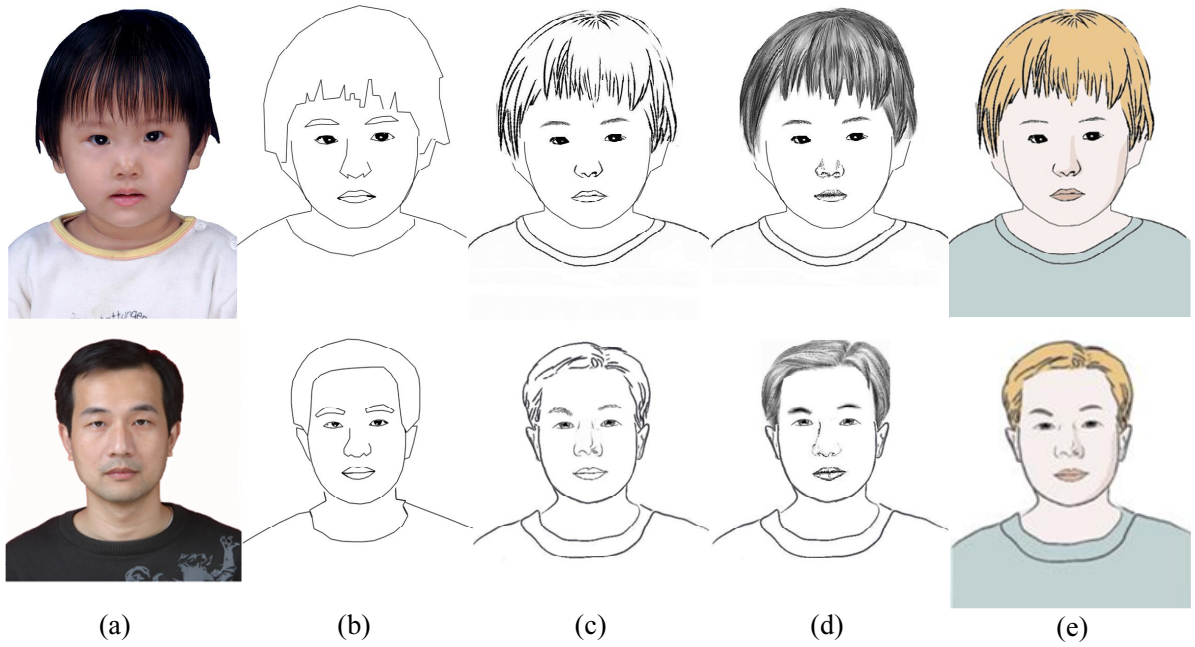

Fig. 2. The result of applying compositional And-Or graph model to portraiture generation. (a) is an input frontal face image, (b) is a draft sketch obtained by image processing methods based on AAM search result of face contour and face component, (c), (d), (e) are separately three rendered results by the sketch dictionaries in literary, pencil, and colored style. *From Min et al.[10].*

Min[10] developed an automatic human portrait system based on the And-Or graph representation. The system can automatically generate a set of life-like portraits in different artistic styles from a frontal face image as shown in Fig.2. The And-Or graph is adopted to account for the variabilities of portraits, including variations in the structures, curves, and drawing style. Given a frontal face image, a local AAM search is performed for each facial component, based on the search result, the hair and collar contours can be inferred. Then using predefined distances, a template matching step finds the best matching template from sketch dictionaries for each portrait component. Finally the strokes of specific style will render each component into stylistic results. Making good use of the large sketch dictionaries in different styles, it can conveniently generate realistic portraits with detailed face feature of different styles.

Suo et al[11] augmented the compositional face model [9] with aging and hair features. This augmented model integrates three most prominent aspects related to aging changes: global appearance changes in hair style and shape, deformations and aging effects of facial components, and wrinkles appearance at various facial zones. Then face aging is modeled as a dynamic Markov process on this graph representation which is learned from a large dataset. Given an input image, the aging approach firstly computes the parse graph representation, and then samples the graph structures over various age groups according to the learned dynamic model. Finally the sampled graphs generate face images together with the dictionaries. Fig.3 is an illustration of the dynamic model for aging over the parse graphs. I_1 is an input young face image and G_1 is its graph representation. I_2 to I_5 are four synthesized aged images in four consecutive age groups generated by G_2 to G_5. $\{G_1, G_2, ..., G_5\}$ is a chain of parse graphs describing face aging procedure.

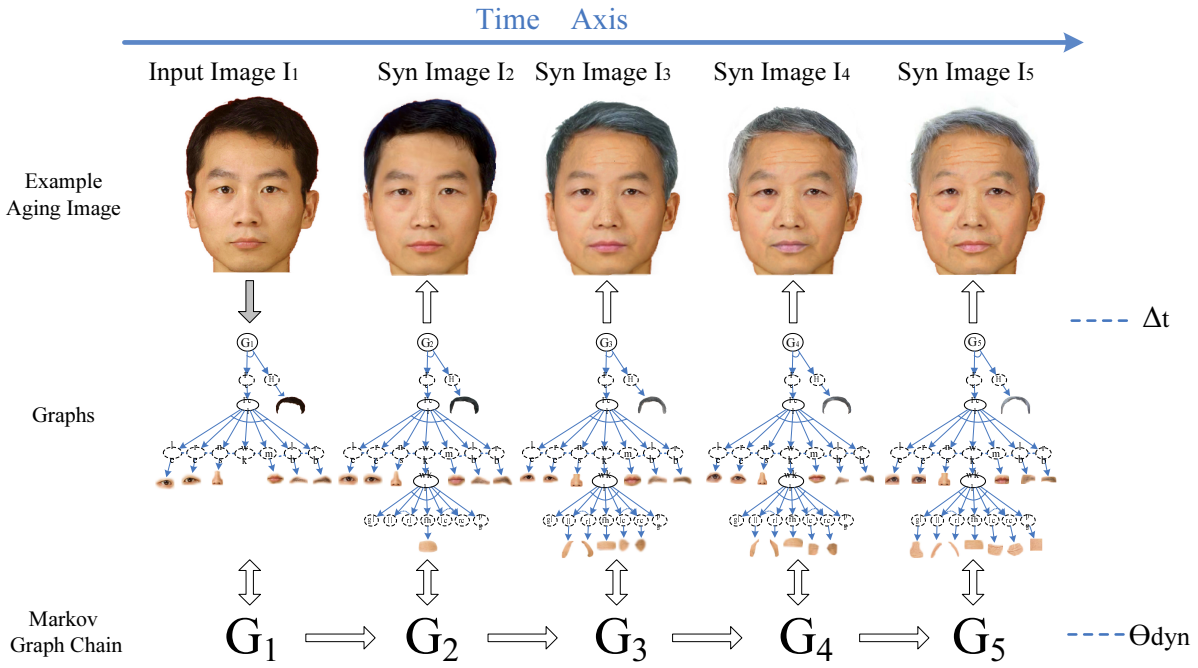

Fig. 3. An aging process can be modeled by a Markov Chain on the parse graphs G_t where t is an age period. The first row is a aging sequence of face, I_1 is the input image, and the other four are simulated aged images. The second row is the graph representations of the image sequence. Third row is the corresponding parse graphs G_t, which form a Markov Chain. $\Theta_{img,t}$ includes the parameters for generating the images from G_t and Θ_{dyn} the parameters for aging progression. *From Suo et al.[11].*

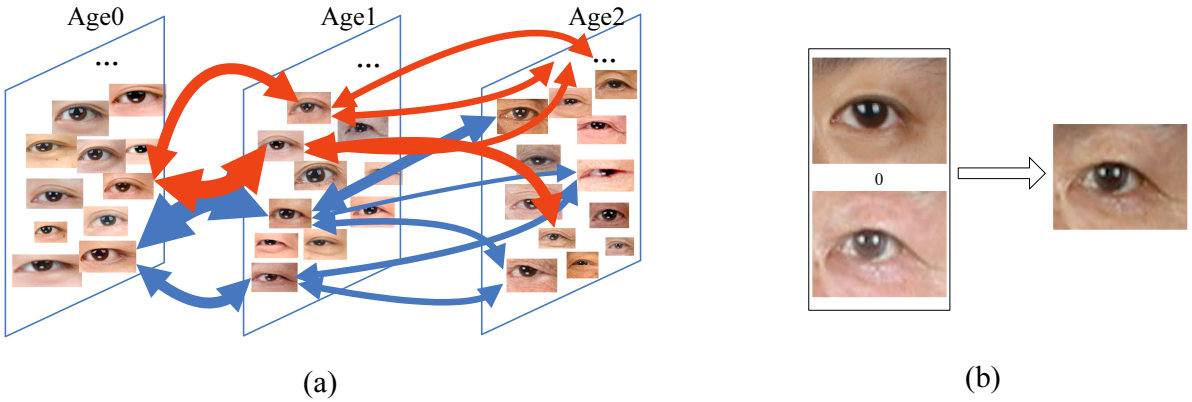

Fig. 4. (a) is a training subset for dynamic learning of face aging. (b) is one simulated result of eye aging. *From Suo et al[11].*

The compositional model decomposes face into parts, this strategy provides the potential of learning the statistics of each node separately. In Suo et al[11], aging dynamics are learned from similar parts cropped from different persons, Fig.4 (a) gives a training subset of eye aging and (b) is the aging results. Human experiments have validated that this aging process is perceptually plausible.

Summary

The compositional And-Or graph model is an expressive representation of high resolution human face. With selection of alternatives at Or nodes, the And-Or graph can model the large diversity of different faces as well as the artistic styles. The decomposition allows learning of parts and the spatial constraints, and alleviates the difficulty of training set collection. The model has been applied to automatic portrait generation, face aging simulation. We argue that the model should also improve other applications, such as to face recognition and expression analysis.

Related Entries

And-Or graph; Face sketching; Face aging

References

1. M. Turk and A. Pentland "Eigenfaces for recognition", *CVPR*, 1991.
2. P. Belhumeur, J. Hespanda, and D. Kriegman "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *PAMI*, vol. 19, no. 7, pp. 711C720, 1997.
3. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang "Face recognition using Laplacianfaces", *PAMI*, vol. 27, no. 3, pp. 328C340, 2005.
4. J. Kim, J. Choi, J. Yi, and M. Turk "Effective representation using ICA for face recognition robust to local distortion and partial occlusion", *PAMI*, vol. 27, no. 12, pp. 1977C1981, 2005.
5. A. L. Yuille, D. Cohen and P. Hallinan, "Feature extraction from faces using deformable templates", *IJCV*, 8(2):99-111, 1992.
6. T. F. Cootes, C. J. Taylor, D. Cooper and J. Graham, "Active shape models-their training and application", *Computer Vison and Image Understanding*, 61(1):38-59, 1995.
7. T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models", *proceedings of ECCV*, 1998.
8. T. F. Cootes and C. J. Taylor "Constrained active appearance Models", *proceedings of ICCV*, 2001.
9. Z. J. Xu, H. Chen and S. C. Zhu. "A high resolution grammatical model for face representation and sketching". *CVPR*, 2005.
10. F. Min, J. L. Suo, S. C. Zhu and N. Sang. "An Automatic Portrait System Based on And-Or Graph Representation". *EMMCVPR*, 2007.
11. J. L. Suo, F. Min and S. C. Zhu. "A Multi-Resolution Dynamic Model for Face Aging Simulation". *CVPR*, 2007.

Definitional Entries

And-Or graph

An And-Or graph is a 6-tuple for representing an image grammar G.

$$G_{and-or} =< S, V_N, V_T, R, \Sigma, P > \tag{4}$$

S is a root node for a scene or object category, V_N is a set of non-terminal nodes including an And-node set and an Or-node set, V_T is a set of terminal nodes for primitives, parts and objects, R is a number of relations between the nodes, Σ is the set of all valid configurations derivable from the grammar, and P is the probability model defined on the And-Or graph.

Face sketching

A face sketching is a parsimonious yet expressive representation of face. It depicts concise sketches of face that captures the most essential perceptual information with a number of strokes.

Face aging

Face aging is to predict the future appearance of human face by learning the aging patterns, child growth and adult aging are two type of aging.