Quality-Driven Face Occlusion Detection and Recovery

Dahua Lin[1]
[1]Dept. of Information Engineering
The Chinese University of Hong Kong
dhlin@ie.cuhk.edu.hk

Xiaoou Tang[1,2]
[2]Microsoft Research Asia
Beijing, China
xitang@microsoft.com

Abstract

This paper presents a framework to automatically detect and recover the occluded facial region. We first derive a Bayesian formulation unifying the occlusion detection and recovery stages. Then a quality assessment model is developed to drive both the detection and recovery processes, which captures the face priors in both global correlation and local patterns. Based on this formulation, we further propose GraphCut-based Detection and Confidence-Oriented Sampling to attain optimal detection and recovery respectively. Compared to traditional works in image repairing, our approach is distinct in three aspects: (1) it frees the user from marking the occlusion area by incorporating an automatic occlusion detector; (2) it learns a face quality model as a criterion to guide the whole procedure; (3) it couples the detection and occlusion stages to simultaneously achieve two goals: accurate occlusion detection and high quality recovery. The comparative experiments show that our method can recover the occluded faces with both the global coherence and local details well preserved.

1. Introduction

Face occlusion that often occurs in practical face recognition systems may incur performance degradation. Thus a technique that can automatically recover the occluded parts of the faces is important for these systems.

A related topic called image inpainting[5][6][3][4]. has been extensively studied in recent years. Most of the current inpainting methods share the same principle: to make use of the consistency between neighboring pixels or textured patches to infer the target area. Hence, their success relies on the condition that the colors or textures in the target region should be basically homogeneous and similar to the surrounding parts. Consequently, they will encounter great difficulty in recovering heterogeneous structures, especially when some components are completely occluded.

In face recovery, the capability of restoring unseen structures is necessary due to the structural nature of faces.
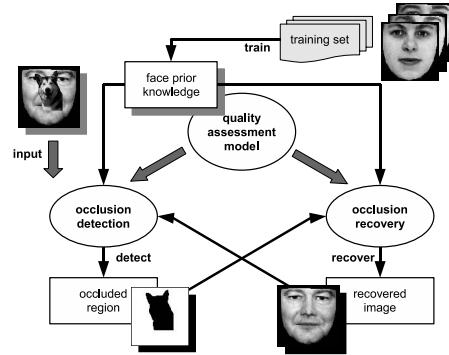

Figure 1. The relations between components in the framework.

Generic image inpainting methods would fail in this case, due to the fundamental limitation that they lack the prior knowledge of face appearance. Hence, utilizing the face appearance priors is essential to recover occluded faces.

Recently, some algorithms are proposed to recover partially occluded faces using examplar-based face priors. Hwang and Lee[2] presented a method to recover damaged faces by linear combination of faces. Mo et al[12] further enhance the reconstruction quality by enforcing the local convex constraint in coefficient estimation. Though these methods are feasible approaches, they suffer from the following drawbacks: (1) linear combination of holistic appearance often lead to blurred results; (2) the errors in coefficient estimation may break the overall coherence; (3) they need user input to determine the occluded region.

This paper addresses the face recovery problem with two targets: detecting the occlusion automatically and recovering the occluded parts in high quality. We consider the formation of an occluded image as a generative process, and thus derive a Bayesian MAP formulation unifying both goals. The quality assessment model is first constructed by learning the prior knowledge from a set of face images, which is then used to guide the procedure of recovery. This model characterizes the face appearance in two aspects: the global correlation defining the overall appearance structure, and the local patterns characterizing the textured details.
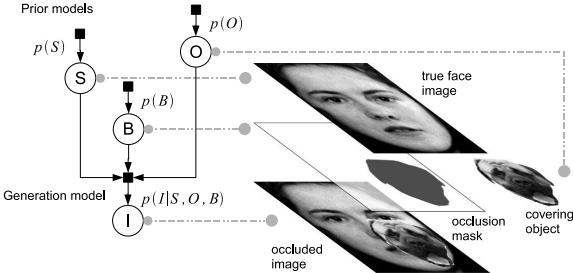
Figure 2. The generative model of occluded images

They are integrated in our probabilistic formulation.

The framework is illustrated in fig.1. When a new image is input, the occlusion detector localizes the occluded region by solving a belief map using convex quadratic programming. Based on the detection results, the recovery stage then exploits both the face priors and the information in the reliable parts to restore the occluded parts using confidence-oriented sampling. Subsequently, the initially recovered image is fed back to the detector to refine the localization of the occlusion by GraphCut. In this progressive procedure, the detection stage works in cooperation with the recovery stage to enhance the quality of the reconstructed image.

In sum, the contributions of our work lie in the following aspects: (1) the probabilistic unification of detection and recovery; (2) a quality assessment model coupling both global correlation and local patterns to drive the recovery process; (3) an automatic occlusion detector; (4) new techniques to efficiently achieve optimal solutions: GraphCut-based detection and Confidence-oriented sampling.

2. Unified Probabilistic Formulation

In our generative model as illustrated in fig.2, an occluded face image I is a masked combination of a true face image S and a covering object O. The occlusion locations are indicated by a binary mask B, in which a value 1 means that the corresponding pixel is non-occluded. In addition, we use m to denote the number of pixels in each image, j to index the pixels in the images, and s_j, b_j, o_j, and v_j to denote the j-th pixel in S, B, O, and I.

Suppose that the true face image, the covering object, and the occlusion locations are mutually independent, then the joint probability can be written as

$$p(S, O, B, I) = p(S)p(O)p(B)p(I|S, O, B). \quad (1)$$

Here, $p(S)$, $p(O)$, and $p(B)$ respectively capture the prior knowledges of faces, covering objects, and the occlusion locations; while $p(I|S, O, B)$ reflects the relation between these elements in the generative process. $p(S)$ is given by the quality assessment model introduced in next section. Here, we first analyze the other three factors.

The prior of the covering object $p(O)$ depends on what we know about the occlusion. In a general case, we have little information about the occlusion, and thus simply use an independent uniform distribution as

$$p(O) = \prod_{j=1}^{M} p(o_j); \quad (2)$$

$$p(o_j) = \begin{cases} c & o_j \text{ is in a valid range,} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The prior of the occlusion locations $p(B)$ is determined by the shape of the occlusion. Generally, we can set the prior by considering the probability of occlusion and the connectivity of the occluded parts. Then $p(B)$ can be given in the following form as an Markov Random Field.

$$p(B) = \frac{1}{Z_B} \prod_{j=1}^{M} \eta^{(1-b_j)} \prod_{l \in \mathcal{N}_a(j)} a(j,l)^{|b_j - b_l|}, \quad (4)$$

$$a(j,l) = \exp\left(-\frac{1}{\sigma_a^2}((x_j - x_l)^2 + (y_j - y_l)^2).\right) \quad (5)$$

Here, η is an estimation on the tendency to occlusion, $\mathcal{N}_a(i)$ is pixel j's neighborhood, and $a(j,l)$ is the spatial affinity between two pixels. The factor $\eta^{(1-b_i)}$ reflects the occlusion potential, $a(j,l)^{|b_j - b_l|}$ reflects the spatial continuity of the occluded parts. We can rewrite the Eq.(4) into the energy form as $p(B) = \exp(-E_B(B) - \log Z_B)$:

$$E_B(B) = \sum_{j=1}^{M} (1 - b_j) \log \frac{1}{\eta} + \frac{1}{\sigma_a^2} \sum_{j=1}^{M} \sum_{l \in \mathcal{N}(j)} |b_j - b_l| d^2(j,l). \quad (6)$$

The conditional pdf of the observation $p(I|S, O, B)$ describes the relationship between the variables. Assume that the generation of each pixel can be formulated as the following function:

$$v_j = \begin{cases} s_j + \varepsilon_j & (b_j = 1), \\ o_j + \varepsilon_j & (b_j = 0), \end{cases} \quad (7)$$

and the noise term ε_j independently satisfy normal distributions: $\varepsilon_j \sim N(0, \sigma_v^2)$. Then we have

$$p(I|S, O, B) = \prod_{j=1}^{m} p(v_j|s_j, o_j, b_j) =$$

$$\frac{1}{Z_I} \prod_{j=1}^{m} \exp\left(\frac{b_j(v_j - s_j)^2}{\sigma_v^2}\right) \exp\left(\frac{(1 - b_j)(v_j - o_j)^2}{\sigma_v^2}\right). \quad (8)$$

Furthermore, if $p(O)$ satisfies Eq.(2), we have

$$p(I|S, B) = \prod_{j=1}^{m} p(v_j|s_j, b_j), \qquad (9)$$

$$p(v_j|s_j, b_j) = (p(v_j|s_j, 1))^{b_j} (\int_o p(v_j|o, 0)do)^{1-b_j}. \qquad (10)$$

Accordingly, the pdf can be written into an energy form $p(I|S, B) = \exp(-E(I|S, B) - \log Z_I)$:

$$E(I|S, B) = \sum_{j=1}^{m} b_j E_v(v_j|s_j, 1) + (1 - b_j)E_v(v_j|0), \qquad (11)$$

$$E_v(v_j|s_j, 1) = \frac{1}{\sigma_v^2}(v_j - s_j)^2, \qquad (12)$$

When $p(o_j)$ is given by Eq.(3), $E_v(v_j|0) = E_c$.

Bayesian MAP Formulation Since the occlusion detection and recovery are respectively to infer the true image S, and the occlusion locations B, they can be formulated as a Bayesian MAP problem to maximize $p(S, B|I)$ as follows:

$$p(S, B|I) \propto p(S)p(B)p(I|S, B). \qquad (13)$$

We adopt the alternate optimization strategy here, and solve the problem by iterating the following two updating steps:

$$S^{(t+1)} = \underset{S}{\arg\max}\, p(S)p(I|S, B^{(t)}); \qquad (14)$$

$$B^{(t+1)} = \underset{B}{\arg\max}\, p(B)p(I|S^{(t+1)}, B). \qquad (15)$$

Eq.(14) and Eq.(15) respectively correspond to the detection and recovery stages. They are coupled in the Bayesian formulation.

3. Quality Assessment Model

The perceptual quality of a face has two aspects: global correlation and local patterns.

3.1. Global Correlation

In the paper, we describe the global correlation by a Gaussian subspace model. The model considers the image S as an m-dimensional vector of pixel values $\mathbf{s} = \{s_1, s_2, \ldots, s_m\}$. Since there exist strong correlation between the pixels, we can assume that its major components reside on a subspace of much lower dimension r:

$$\mathbf{s} = \boldsymbol{\mu} + \mathbf{W}\mathbf{x} + \boldsymbol{\varepsilon}. \qquad (16)$$

Here, $\boldsymbol{\mu}$ is an $m \times 1$ mean vector, $\mathbf{W}$ is an $m \times r$ matrix, the intrinsic factor $\mathbf{x}$ and the noise term $\boldsymbol{\varepsilon}$ are $r \times 1$ and $m \times 1$ vectors, independently satisfying Gaussian distributions:

$$\mathbf{x} \sim N(\mathbf{0}_{r \times 1}, \mathbf{I}_{r \times r}); \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}_{m \times 1}, \sigma_g^2 \mathbf{I}_{m \times m}). \qquad (17)$$

Thus, $\mathbf{s}$ also satisfies a Gaussian distribution

$$\mathbf{s} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}): \quad \boldsymbol{\Sigma} = \mathbf{W}\mathbf{W}^T + \sigma_g^2 \mathbf{I}_m. \qquad (18)$$

It is essentially the probabilistic PCA model[7]. According to PPCA, with the optimal solution of $\mathbf{W}$ and σ_g^2, we have

$$\boldsymbol{\Sigma} = \mathbf{U}_r \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \sigma_g^2 \mathbf{I}_{m-r} \end{pmatrix} \mathbf{U}_r^T. \qquad (19)$$

Here, $\mathbf{D}$ is an $r \times r$ diagonal matrix containing the r largest eigenvalues of covariance matrix $\mathbf{C}$, $\mathbf{U}_r$ is an $m \times r$ matrix composed of the corresponding eigenvectors, σ_g is the mean of the remaining eigenvalues. With the Gaussian model, the prior on global correlation can be written in an energy term: $p_g(S) = p_g(\mathbf{s}) = \exp(-E_g(S) - \frac{1}{2}\log(2\pi|\boldsymbol{\Sigma}|))$, with

$$E_g(S) = (\mathbf{s} - \boldsymbol{\mu})^T \mathbf{G}(\mathbf{s} - \boldsymbol{\mu}). \qquad (20)$$

Here, $\mathbf{G} = \frac{1}{2}\boldsymbol{\Sigma}^{-1}$ encodes the connections between pixels.

3.2. Local Patterns

We analyze the local patterns based on patches of radius r_p. The patch centered at pixel j is notated by $\mathcal{P}(j)$ as

$$\mathcal{P}(j) = \{q \mid (x_q - x_j)^2 + (y_q - y_j)^2 \leq r_p^2\}. \qquad (21)$$

In addition, we use $\mathcal{R}(j)$ to denote the set of non-centered values in $\mathcal{P}(j)$: $\mathcal{R}(q) = \{s_q | q \in \mathcal{P}(j),\ q \neq j\}$.

In this paper, we characterize a local pattern using the dependency of the center pixel on its surrounding pixels in a patch, which is modeled by belief-based multiple hypothesis integration. For the j-th pixel in S, a set of referenced patches are collected from all training samples in the corresponding position and neighboring positions. The center pixel values of these patches are denoted by $s'_1, \ldots, s'_K$, and the set of non-centered pixel values by $\mathcal{R}'_1, \ldots, \mathcal{R}'_K$. Each referenced patch gives an evaluation, which are then combined together with a belief value

$$\phi_j(\mathcal{P}(j)) = \sum_{k=1}^{K} bel(\mathcal{R}'_k|\mathcal{R}) \exp\left(-\frac{(s_j - s'_j)^2}{\sigma_l^2}\right). \qquad (22)$$

The belief values are determined based on the similarity of non-centered values in the patch,

$$bel(\mathcal{R}'_k|\mathcal{R}) = \frac{\exp(-d^2(\mathcal{R}, \mathcal{R}'_k)/\sigma_l^2)}{\sum_{k=1}^{K} \exp(-d^2(\mathcal{R}, \mathcal{R}'_k)/\sigma_l^2)}; \qquad (23)$$

$$d^2(\mathcal{R}, \mathcal{R}'_k) = \sum_{q \in \mathcal{P}(j) - \{j\}} (s_q - s'_q)^2. \qquad (24)$$

Then the joint probability of all patches in the image is

$$p_l(S) = \frac{1}{Z_l} \prod_{j=1}^{m} \phi_j(\mathcal{P}(j)). \qquad (25)$$

Table 1. The procedure of occlusion recovery

1. Initialize the confidence map CM: set the confidence of the known pixel to 1, and that of the unknown pixel to 0.

2. Initialize the parent confidence map PCM based on initial CM. The parent confidence of a pixel j is the average confidence of the other pixels in $\mathcal{P}(j)$.

3. Repeat the following cycles until convergence:

    3.a: Initialize the set of pixels to be updated: $\mathcal{U} = \{j \mid b_j = 0\}$;

    3.b: Find the pixel in $\mathcal{U}$ with highest parent confidence, denote its index by j_c;

    3.c: Select a set of candidate values for pixel j_c: $\mathcal{C}(j_c)$.

    3.d: Compute the conditional posteriori of the candidate values, and draw the value s_{new} from the candidates according to the posteriori (Gibbs sampling);

    3.e: Compute the new confidence: $c_{new} = PCM(j_c) \times$ the posteriori of s_{new};

    3.f: If $c_{new} > CM(j_c)$, accept the update: set s_{j_c} to s_{new}; set $CM(j_c)$ to c_{new}; and update all affected parent confidences;

    3.g: Remove j_c from $\mathcal{U}$: $\mathcal{U} = \mathcal{U} - \{j_c\}$;

    3.h: Repeat the steps from 3.b to 3.g until $\mathcal{U} = \emptyset$.

It can be written in an energy form as $p_l(S) = \exp(-E_l(S) - \log Z_l)$:

$$E_l(S) = \sum_{j=1}^{m} E_{l(j)}(\mathcal{P}(j)) = \sum_{j=1}^{m}(-\log \phi_j(\mathcal{P}(j))). \quad (26)$$

The energy E_l actually measures the deviation of all local patches from the expected patterns.

4. Optimization Algorithms

4.1. Recovering Occluded Parts

Occlusion recovery is based on Eq.(14), which has two factors: $p(S)$ and $p(I|S, B^{(t)})$. Since the occlusion recovery only concerns the values of s_j with $b_j = 0$, which do not affect the latter factor $p(I|S, B^{(t)})$, the objective can therefore be simplified to maximize $p(S)$ by pursuing the optimal values of s_j in the occluded parts.

We develop the *Confidence-oriented sampling* method for recovery. Given the occluded region B, it restores the occluded pixel values by Gibbs sampling, and simultaneously maintains a confidence map to direct the process. The procedure of the algorithm is specified in table.1. Its strategy can be briefly explained in two points:
(1) It updates the pixels in an order from high confidence to low confidence, such that information is flowed from high

confidence regions to low confidence ones.
(2) Each pixel is updated by Gibbs sampling: drawing its value from conditional distribution.

In this procedure, we have to address two issues: (1) how to select the candidate values for each target pixel; (2) how to compute their conditional posteriori. They are both related to the quality assessment model. As discussed above, the quality model comprises two aspects: global correlation and local patterns, respectively measured by the energy E_g in Eq.(20) and E_l in Eq.(26). Thus, the sum of the global correlation terms related to s_j is

$$E_{g(j)}(s_j) = 2\sum_{i \neq j} g_{ij}(s_i - \mu_i)(s_j - \mu_j) + g_{jj}(s_j - \mu_j)^2, \quad (27)$$

here $g_{ij} = g_{ji} = \mathbf{G}(i,j)$ ($\mathbf{G}$ is symmetric). While the relevant local pattern term is $\phi_j(\mathcal{P}(j))$ given in Eq.(22).

With other pixel values fixed, we have

$$s_{j(gopt)} = \underset{s_j}{\operatorname{argmin}} E_{g(j)}(s_j) = \mu_j - \frac{1}{g_{jj}}\sum_{i \neq j} g_{ij}(s_i - \mu_i). \quad (28)$$

During the recovery process, the pixels in target region is uncertain, so we use the confidences in the computation as

$$s_{j(gopt)} = \mu_j - \frac{1}{g_{jj}}\sum_{i \neq j} CM(i)g_{ij}(s_i - \mu_i). \quad (29)$$

Then we can construct the candidate set $\mathcal{C}(j)$ by selecting two types of candidates: (1) the value given by optimizing the global correlation: $s_{j(opt)}$; (2) the pixel values hypothesized by the referenced patches for local patterns: $s'_1, s'_2, \ldots, s'_K$. After the candidates are selected, we compute the relevant quality energy for each candidate value s based on the quality assessment model:

$$E_{j(rele)}(s) = E_{g(j)}(s) - \log \phi_j(\mathcal{R}(j) \cup \{s\}). \quad (30)$$

Then the probability to set value s for pixel j is

$$p(s|j) = \frac{\exp(-E_{j(rele)}(s))}{\sum_{s \in \mathcal{C}(j)} \exp(-E_{j(rele)}(s))}. \quad (31)$$

4.2. Detecting Occlusion by GraphCut

Under the Bayesian formulation, with the true image S given, occlusion detection can be accomplished by optimizing B as in Eq.(15), which can be considered as an energy minimization problem to minimize $E_D(B)$ as

$$E_D(B) = E_B(B) + E(I|S, B). \quad (32)$$

Here, $E_B(B)$ corresponds to the connectivity of the occluded regions, while $E(I|S, B)$ measures how the detection results match the generative model. According to

Eq.(6) and Eq.(11), it can be rewritten as

$$E_D(B) = \sum_{j=1}^{m} E_{D(j)}(b_j) + \sum_{j=1}^{m}\sum_{j'=1}^{m} E_{D(j,j')}(b_j, b_{j'}); \quad (33)$$

$$E_{D(j)}(b_j) = \begin{cases} E_v(v_j|s_j, 1) & (b_j = 1) \\ E_v(v_j|0) + \log(\frac{1}{\eta}) & (b_j = 0) \end{cases}, \quad (34)$$

$$E_{D(j,j')}(b_j, b_{j'}) = \begin{cases} \frac{1}{\sigma_a^2} d^2(j, j') & (b_j \neq b_{j'}) \\ 0 & (b_j = b_{j'}) \end{cases}. \quad (35)$$

We can see that in this formulation, the interaction terms satisfy the sub-modular condition $E(0,0) + E(1,1) < E(0,1) + E(1,0)$, thus it can be formulated as a binary GraphCut problem[10]: the connections from each node to the source 0 and sink 1 respectively indicate the cost of assigning a pixel to be occluded or non-occluded; while the connections between different nodes reflect their spatial affinity. Hence, we can readily obtain the globally optimal solution to this problem by GraphCut[11].

4.3. Initialize the Detection

Initially, both the true image S and the occlusion locations B are unknown. To activate the alternate updating procedure, we first initialize B. It is accomplished by coupling the prior of B and the quality assessment, so that we can attain a balance between the spatial continuity of B and the image quality in the non-occluded region. The problem can be formulated as an optimization problem as

$$B^{(0)} = \underset{B}{\arg\min} E_B(B) + E_{g(B)}(B) + E_{l(B)}(B). \quad (36)$$

Here, $E_{g(B)}$ is the energy of global correlation in the region with $b_j = 1$, while $E_{l(B)}$ is the energy of local patterns in that region. For convenience of derivation, we introduce an m-dimensional binary vector $\mathbf{b} = (b_1, \ldots, b_m)^T$. Then,

$$E_B(B) = 2\mathbf{b}^T(\mathbf{D}_A - \mathbf{A})\mathbf{b} + \log\frac{1}{\eta}(\mathbf{1} - \mathbf{b}), \quad (37)$$

where $\mathbf{A}$ is an $m \times m$ matrix with $\mathbf{A}(i,j) = a(i,j)$, $\mathbf{D}_A$ is an $m \times m$ diagonal matrix with $\mathbf{D}_A(i,i) = \sum_{j=1}^{m} a(i,j)$. According to Eq.(20), we have

$$E_{g(B)}(B) = \sum_{i=1}^{m}\sum_{j=1}^{m} b_i b_j g_{ij}(s_i - \mu_i)(s_j - \mu_j) = \mathbf{b}^T\mathbf{M}_g\mathbf{b}, \quad (38)$$

where $\mathbf{M}_g$ is an $m \times m$ matrix with $\mathbf{M}_g(i,j) = g_{ij}(s_i - \mu_i)(s_j - \mu_j)$. In addition, from Eq.(26), we have

$$E_{l(B)}(B) = \mathbf{h}^T\mathbf{b}, \quad (39)$$

in which $\mathbf{l}$ is an $m \times 1$ vector with $\mathbf{h}(j) = E_{l(j)}(\mathcal{P}(j)) = -\log\phi_j(\mathcal{P}(j))$. Integrating Eq.(37), Eq.(38), and Eq.(39),

we arrive at the following expression of the objective

$$\mathbf{b}^T(2(\mathbf{D}_A - \mathbf{A}) + \mathbf{M}_g)\mathbf{b} + (\mathbf{h} - \log\frac{1}{\eta})^T\mathbf{b}. \quad (40)$$

By relaxing the value of b_j to be real value in $[0, 1]$, we can solve the global optima of the relaxed problem using convex quadratic programming. After that, the binary value of $b_j^{(0)}$ can be obtained by thresholding the real values.

5. Experiments

5.1. Data Preparation and Experiment Settings

We conduct the experiments on a mixed dataset from FERET[8] and Purdue AR[1] databases, which consists of 1400 face images, and is divided into two disjoint sets: a training set with 700 face images, and a testing set with the other 700 images. In the preprocessing stage, we first use an alignment algorithm[9] to locate the key points on a face, and thereby normalize the face shape. Then we normalize the pixel values of each image, such that their mean is 0, and the variance is 1. The occluded images are obtained by superposing objects on the face images. To test the algorithms under different circumstances, we use different objects to simulate the occlusion in different cases: black and white circular masks, dogs, flowers, oranges, and text.

For the global correlation model, the mean vector $\boldsymbol{\mu}$ and the matrix $\mathbf{G}$ are estimated on the training set. The dimension of the principal subspace of r is determined such that 98% of the energy is preserved in the space. While for local pattern model, the training images are used as source of referenced patches for evaluating the local energy, and the pixel variance σ_l is set to 0.01. In addition, there are several design parameters in the prior formulation. For the occlusion prior given in Eq.(6), we set $\eta = 0.3$, and $\sigma_a = 2.0$. For the conditional pdf of observations in Eq.(8), we set $\sigma_v = \sigma_l = 0.01$. In the experiments, we find that the selection of the parameter values does not notably affect the final results if they are in a reasonable range.

5.2. Testing the Quality Assessment Model

We first test the quality assessment model, which is composed of a global correlation model and a local pattern model. In order to investigate how the proportion of occlusion influence the quality measurement, we generate different sets of occluded images by scaling the covering objects to different sizes such that they occlude specific ratios of the target faces. The testing results acquired by averaging the quality energies E_g and E_l (divided by the number of pixels) in different cases are shown in fig.3.

From the results, we can see that the quality energy values are acute to occlusion: the average value is below 1.0 when there is no occlusion, while it drastically rises when
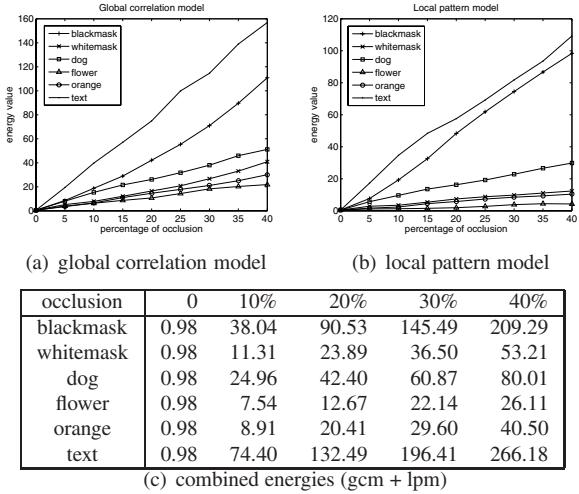
(a) global correlation model   (b) local pattern model

occlusion	0	10%	20%	30%	40%
blackmask	0.98	38.04	90.53	145.49	209.29
whitemask	0.98	11.31	23.89	36.50	53.21
dog	0.98	24.96	42.40	60.87	80.01
flower	0.98	7.54	12.67	22.14	26.11
orange	0.98	8.91	20.41	29.60	40.50
text	0.98	74.40	132.49	196.41	266.18

(c) combined energies (gcm + lpm)

Figure 3. The average quality measure vs. occlusion proportion

Table 2. The table of recovery evaluation by SNR (dB)

occlusion	Hwang's	Mo et al.'s	gcm	lpm	gcm + lpm
blackmask	50.95	53.76	59.38	51.14	**64.15**
whitemask	53.18	55.80	63.67	54.16	**69.35**
dog(small)	40.09	45.30	48.67	40.36	**53.46**
dog(large)	36.43	40.29	43.84	36.14	**48.60**
flower	40.96	43.86	55.12	41.27	**55.74**
orange	42.76	48.75	49.69	42.37	**54.78**
text(small)	51.87	56.58	88.16	63.33	**90.92**
text(large)	42.67	56.21	81.48	57.62	**83.43**

Note: gcm = global correlation model, lpm = local pattern model.

only a small part of the face image is occluded. Moreover, it continuously increase when the proportion of occlusion increases. This clearly shows that the our quality assessment model is a good criterion to evaluate the occlusion.

5.3. Testing Occlusion Detection and Recovery

We compare our method with Hwang's method[2] and Mo et al.'s method[12]. Since their works do not address the occlusion detection problem, we launch their recovery methods based on the initial detection results yielded by our occlusion detection algorithm. Fig.6 illustrates the results.

The first two rows show the recovery from occlusion by circle masks. In both these cases, most of the algorithms can work out satisfactory results, except that Hwang's method lead to some artifacts in the masked region. The 3rd to 6th rows show the recovery from occlusion by real objects. which are more difficult cases. The results produced by our method (column (h)) is remarkably better than those yielded by Hwang's (column (d)) and Mo et al's (column (e)). Their approaches cause over-blurring and notable artifacts in the recovered region and loss some subtle details. While our method preserves both the global coherence and the local details in a high quality.

We also compare the results yielded by the global model (column (f)), the local model (column (g)), and their combination (column(h)). Notice that the results acquired by only using the local pattern model have some artifacts when the eye is partially occluded. It is due to that the local model does not encode the information of global relation, thus it is weak in maintaining global coherence. With the global model incorporated, we can see that the faces recovered by the combined model are near perfect.

In addition to subjective observation, we compare the algorithms by quantitative evaluations. The recovered images are compared against the ground truth, and the signal-noise ratio (SNR) is computed by

$$SNR = 10\log\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(rp_{ij} - gp_{ij})^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}gp_j^2}\ (\mathrm{dB}),\quad (41)$$

here rp_{ij} and gp_{ij} are respectively the recovered value and ground truth value of the j-th pixel on the i-th image. The evaluation results are shown in table 2. The table shows that our algorithm achieves SNR $10\mathrm{dB}$ higher than other algorithms. It means that recover error energy is reduced to $1/10$ compared to others, which is a significant improvement. Moreover, the integration of the global and local models leads to further gain in SNR (about $2 - 6\mathrm{dB}$) compared to only using the global model.

6. Conclusion

This paper presented a new algorithm to detect and recover the occluded parts in face images, which is guided by a quality assessment model evaluating both global coherence and local details. It has been sufficiently manifested by the experimental results that our algorithm consistently outperforms others in recovery quality.

Acknowledgment

References

[1] A. M. Martinez and R. Benavente. The ar face database. CVC Technical Report 24, Purdue Univ., 2000.

[2] B. Hwang and S. Lee. Reconstruction of partially damaged face images on a morphable face model. *IEEE Trans. PAMI*, 25(3):365–372, 2003.

[3] I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. In *Proc. of ACM SIGGRAPH'03*, 2003.

[4] J. Sun, L. Yuan, J. Jia, and H-Y. Shum. 2005. In *Proc. of ACM SIGGRAPH'05*, 2005.

[5] M. Bertalmio and G. Sapiro. Image inpainting. In *Prof. of ACM SIGGRAPH'00*, 2000.

[6] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Trans. Imag. Proc.*, 12(8):882–889, 2003.

Figure 4. The experimental results of face occlusion and face recovery. The figure has 9 columns: (a) the input occluded image, (b) the belief map given by initialization, (c) the mask of initial detection results by thresholding the belief map, (d) the results of Hwang's method, (e) the results of Mo et al's method, (f) our results by using only global correlation, (g) our results by using only local patterns, (h) our results by using both global and local models, (i) the ground truth of non-occluded image.

[7] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *J. Roy. Stat. Soc.*, 61, 1999.

[8] P.J. Phillips, H. Moon, P.J. Rauss, and S. Rizvi. The feret evaluation methodology for face recognition algorithms. *IEEE Trans. PAMI*, 22(10), 2000.

[9] S. Yan, M. Li, H. Zhang, and Q. Cheng. Ranking prior likelihood distributions for bayesian shape localization framework. In *Proc. of ICCV' 03*, 2003.

[10] V. Kolmogorov and R. Zabih. What energy functions can be min-imized via graph cuts? *IEEE Trans. on PAMI*, 26(2):147–159, 2 2004.

[11] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization. *IEEE Trans. on PAMI*, 26(9):1124–1137, 9 2004.

[12] Z. Mo, J.P. Lewis, and U. Neumann. Face inpainting with local linear representations. In *Proc. of BMVC'04*, 2004.