Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Patch-guided facial image inpainting by shape propagation*

Yue-ting ZHUANG[1], Yu-shun WANG[1], Timothy K. SHIH[2], Nick C. TANG[3]

([1]*Institute of Artificial Intelligence, School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*)

([2]*Department of Computer Science, National Taipei University of Education, Taipei 10671, Taiwan, China*)

([3]*Department of Computer Science and Information Engineering, Tamkang University, Taipei 25137, Taiwan, China*)

E-mail: yzhuang@cs.zju.edu.cn; eawang@microsoft.com; tshih@cs.tku.edu.tw; nick.ctang@msa.hinet.net

Received Feb. 26, 2008; Revision accepted June 22, 2008; Crosschecked Dec. 26, 2008

Abstract: Images with human faces comprise an essential part in the imaging realm. Occlusion or damage in facial portions will bring a remarkable discomfort and information loss. We propose an algorithm that can repair occluded or damaged facial images automatically, named 'facial image inpainting'. Inpainting is a set of image processing methods to recover missing image portions. We extend the image inpainting methods by introducing facial domain knowledge. With the support of a face database, our approach propagates structural information, i.e., feature points and edge maps, from similar faces to the missing facial regions. Using the inferred structural information as guidance, an exemplar-based image inpainting algorithm is employed to copy patches in the same face from the source portion to the missing portion. This newly proposed concept of facial image inpainting outperforms the traditional inpainting methods by propagating the facial shapes from a face database, and avoids the problem of variations in imaging conditions from different images by inferring colors and textures from the same face image. Our system produces seamless faces that are hardly seen drawbacks.

Key words: Image inpainting, Face reconstruction, Feature point extraction
doi:10.1631/jzus.A0820138 **Document code:** A **CLC number:** TP391.41

INTRODUCTION

Images with human faces comprise an essential part of family photos and visual documentaries. As faces play the most substantial role in depicting human characters (Ersotelos and Dong, 2007), an occlusion (e.g., visible watermark) or damage in the facial part of an image will bring a remarkable information loss.

Image inpainting/image completion (Criminisi *et al*., 2004) is a technique to restore/complete the missing area of an occluded or damaged object, which is manually selected by the users. Compared with earlier approaches (Shih *et al*., 2003) that focus on removing only small areas on photos, the work reported in (Criminisi *et al*., 2004) produces fairly good results in general cases, especially when applied to large continuous areas. Image inpainting techniques can complete holes based on both spatial and frequency features. Structural properties such as edges of a house are extracted from spatial domain and used to complete an object with its structural property extended (Criminisi *et al*., 2004). On the other hand, textural information can be propagated from the surrounding areas toward the center of a hole such that a seamless natural scene can be recovered. Within the inpainting process, in general, the user has to select an object to be removed (and thus the hole is created). In addition, there are mechanisms suggesting that human intelligence can be incorporated to produce a better result (Nielsen and Nock, 2005; Sun *et al*., 2005). Nielsen and Nock (2005) used an interface to identify a source area, where texture information was used to inpaint the missing area. Sun *et al*.(2005) further suggested that most natural or artificial objects can be defined by a few main curves; by asking the user to draw a few curves, excellent image inpainting

* Project supported by the National Natural Science Foundation of China (No. 60525108) and the National Key Technology R & D Program of China (No. 2006BAH11B03-4)

results can be produced. In general, the problem of image completion can be defined as the following. Assume that the target image I is decomposed into two parts, $I=\Phi\cup\Omega$, where Ω is a missing area/hole manually identified by the user, and Φ is a source area with information to be used to complete Ω. And, there is no overlap between the missing area and the source area. These terms (i.e., I, Φ, and Ω) are used in most articles to discuss inpainting algorithms, as well as in this article.

In this paper, we try to incorporate specific domain knowledge into image inpainting, named 'facial image inpainting', which reconstructs partially damaged facial images with the support of a face database. By using a prior probabilistic distribution of facial structural information from a face database to guide the traditional inpainting process, seamless results can be achieved.

Hwang and Lee (2003) estimated optimal coefficients for linear combinations of reference face images, and used least-square minimization to reconstruct the damaged face image. Inspired by (Hwang and Lee, 2003), Mo *et al*.(2004) used human observation on correlations between different face regions and produced a monotonic density feature space, which is used for extrapolating estimated missing face regions. Using a reconstruction approach similar to that in (Hwang and Lee, 2003), Mo *et al*.(2004) produced reasonable good quality of reconstructed faces.

The experiments discussed in (Hwang and Lee, 2003; Mo *et al*., 2004) always remove a portion of the face completely. That is, the left eye, the right eye, the nose, the mouth, or their combinations are completely removed. The mechanisms obtain a missing face portion from a synthetic reference face image and paste the portion completely to the reconstructed face image. However, due to the various changes in lighting conditions and skin textures in real-life imaging, the linear combinations of reference facial images will not faithfully conform to the specific lighting and photographic conditions of the target image.

Compared with the previous methods (Hwang and Lee, 2003; Mo *et al*., 2004), this paper relates more closely to the traditional inpainting academy. Rather than synthesize the whole facial organ by a linear combination of reference facial images, we propagate the structural information (i.e., feature points and edge maps, called 'facial shape') from a face database, but reconstruct the textural information of the missing area by inferring from regions in the original image. Guided by the shape information from a face database, we use an updated exemplar-based image inpainting algorithm (Criminisi *et al*., 2004) to copy patches (i.e., small blocks of pixels) from the source face portion to the missing portion. Thus, the inpainted skin looks more natural.

PROBLEM DEFINITION AND PROCEDURE

Given a target face I, the user selects a missing area Ω. The target face is divided into two portions, $I=\Phi\cup\Omega$, where Φ is a source area. And, $\Phi\cap\Omega=\varnothing$. We define I_ε, Φ_ε and Ω_ε as the edge map of target face I, source area Φ and missing area Ω (which is missing before inpainting) respectively, and $I_\varepsilon=\Phi_\varepsilon\cup\Omega_\varepsilon$; define I_γ, Φ_γ and Ω_γ as a set of feature points of I, Φ and Ω (reconstructed) respectively, and $I_\gamma=\Phi_\gamma\cup\Omega_\gamma$.

We define a face database: $FDB=\{(I, I_\varepsilon, I_\gamma)^*\}$. FDB is a set of faces, edge maps, and feature point sets pre-computed. The procedure to inpaint missing face portions follows two main steps:

Step 1: Deducing feature point set and edge map.

The *FeatureSearch* function uses source feature points Φ_γ, and searches for similar faces I^x by comparing Φ_γ and $I_\gamma^x-\Omega_\gamma^x$. From the similar faces found, the edge map and the feature point set $(I_\varepsilon^x, I_\gamma^x)$ are computed. Therefore,

$$FeatureSearch(\Phi_\gamma, FDB)\rightarrow(I_\varepsilon^x, I_\gamma^x). \qquad (1)$$

Based on the relative position of Φ in I^x, we compute the deduced edge map $\tilde{\Omega}_\varepsilon$ and deduced feature point set $\tilde{\Omega}_\gamma$ of $\tilde{\Omega}$ (i.e., the final inpainting result) from $(I_\varepsilon^x, I_\gamma^x)$. The recovery of and searching for feature points can be conducted by the algorithm discussed in Section 3.

Step 2: Inpainting by patch guidance.

The inpainting procedure takes the position of Ω, and uses the deduced information from Step 1, to compute the final inpainting result. Thus,

$$Inpaint(\Omega, \tilde{\Omega}_\gamma, \tilde{\Omega}_\varepsilon)\rightarrow\tilde{\Omega}. \qquad (2)$$

The detailed inpainting algorithm is discussed in Section 4. Next, we illustrate how to search for faces with missing areas.

FACE SEARCHING WITH MISSING AREA

The process of matching a query face image to the most similar ones from a database mainly consists of two steps: (1) feature extraction; (2) distance measuring (Zhao *et al*., 2003).

Feature extraction is essential to face recognition and other pattern recognition issues (Zhao *et al*., 2003). The most famous and widely used feature extraction method is the principal component analysis (PCA) (Turk and Pentland, 1991), which represents the features as scores of principal components, i.e., eigenvectors with the largest eigenvalues. Low-level features, such as the face shape, the eyebrow tilt, and the eye shape (Oda and Kato, 1993), are regarded as salient features to define face images. Parts of the face as well as holistic features are used separately and weighted interactively by Zhang *et al*.(2006) to represent inter-personal face characteristics.

Distance measuring can be used to classify a query image into one or more categories in a database. Classification methods such as linear discriminant analysis (LDA) (Swets and Weng, 1996) and hidden Markov models (HMMs) (Martínez, 1999) are often used. Lu *et al*.(2006) used a boosting method to cascade several weak classifiers to form a strong classifier for face matching. Zhang *et al*.(2006) employed support vector machine (SVM) with radial basis function (RBF) kernels to train a similarity measuring function.

There have been a number of methods reported to identify, verify and retrieve faces (Turk and Pentland, 1991; Oda and Kato, 1993; Swets and Weng, 1996; Martínez, 1999; Zhao *et al*., 2003; Lu *et al*., 2006; Zhang *et al*., 2006). However, searching of faces with missing areas, which make the search more complicated, remains an open issue. In order to solve this problem, we introduce 'partial face-shape descriptors' to represent facial features. These partial face-shape descriptors are similar to that proposed in (Oda and Kato, 1993) by using facial feature points to represent facial shapes. If there are missing parts in the image, the lost information is composed into the descriptor by removing the corresponding landmarks. We implement the method proposed in (Gross *et al*., 2004) to extract facial feature points in images with missing areas. Before two images are compared, the two facial shapes are normalized into the same scale including size and position. We use L_2-norm between two facial shapes to measure the similarity between a target face and faces in a database. By ranking the measured distances, the top N faces with least distances will be obtained. As an example, Fig.1a is the query face, with the top 5 faces in Fig.1b selected from the face database.

(a)

(b)

Fig.1 Searching for similar faces
(a) Target face; (b) Top 5 faces found in the face database

After face searching, the next step is to propagate the structure information to the target face. We first compute the mean shape of the N faces that are returned by the face searching procedure: $P_{\text{mean}} = \underset{1 \le i \le N}{mean}(P_i)$. From feature points in the source area (i.e., Φ_γ), we can learn a mapping function from P_{mean} to the shape of the target face $Q_{\text{target}} = \{\boldsymbol{q}_i\}$. Constructing such a function can be regarded as an interpolation or approximation problem, which solves the problem of approximating a continuous multivariate function $f(\boldsymbol{x})$ by an approximate function $F(\boldsymbol{x}, \boldsymbol{c})$ with an appropriate choice of parameter set $\boldsymbol{c}$. The family of RBFs (Zhang *et al*., 2006) is well known for its ability to approximate high-dimensional smooth surfaces, and is often used in model fitting. The network of RBFs to recover the coordinates of missing feature points is

$$\boldsymbol{q}_j = \sum_{i=1}^{k} \boldsymbol{c}_{ji} \phi(\| \boldsymbol{p}_j - \boldsymbol{q}_i \|), \tag{3}$$

where $\boldsymbol{p}_j$ is the corresponding feature point in the mean shape of a missing point $\boldsymbol{q}_j$ on the target face, $\boldsymbol{q}_i$ is a source feature point on the target face, $\|\boldsymbol{p}_j - \boldsymbol{q}_i\|$ is the Euclidean distance, k is the number of source feature points on the target face, c_{ji} denotes the parameters to be learned, and $\phi(r)$ is a radial symmetric basis function. Plugging the Hardy basis function into the network of RBFs results in

$$\boldsymbol{q}_j = F_j(\boldsymbol{p}_j) = \sum_{i=1}^{k} \boldsymbol{c}_{ji} \sqrt{\| \boldsymbol{p}_j - \boldsymbol{q}_i \|^2 + s_i^2}, \qquad (4)$$

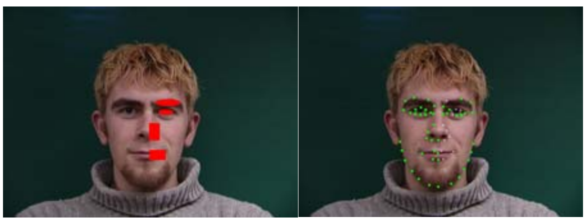
where $s_i = \min_j(\| \boldsymbol{p}_j - \boldsymbol{q}_i \|)$ is the stiffness coefficient.

Substituting the k pairs of source feature points in both mean shape and target face shape as training data ($\boldsymbol{p}_i$, $\boldsymbol{q}_i$) into Eq.(4) results in a linear system of k equations. Solving the linear system yields

$$\boldsymbol{c} = \boldsymbol{H}^{-1}\boldsymbol{Q}_{\text{source}}, \qquad (5a)$$

$$\boldsymbol{c} = (\boldsymbol{H} + \lambda \boldsymbol{I})^{-1}\boldsymbol{Q}_{\text{source}}, \qquad (5b)$$

where $\boldsymbol{Q}_{\text{source}}$ is a k×2 matrix, whose ith row is $\boldsymbol{q}_i$, and λ=0.0001 is a disturbing factor determined empirically to decrease the impact of noise. As an example, Fig.2a illustrates a target face (with missing areas in squares and ellipses), and Fig.2b shows the recovered feature points (in white crosses). The squared feature points in Fig.2b were extracted from the source area, and then used to search for similar faces in the face database. Edge maps can be deduced in a similar way by image warping (Wolberg, 1994) using the recovered feature points as control boundaries.



(a) (b)

Fig.2 Recovery of missing feature points
(a) Target face (with the missing areas given as squares and ellipses); (b) Recovered feature points (in white crosses)

PROPOSED FACIAL IMAGE INPAINTING FUNCTION

The face image inpainting algorithm we propose is based on the exemplar-based inpainting algorithm proposed in (Criminisi *et al*., 2004). However, our revised algorithm introduces the concept of 'patch guidance', which was not found in the previous literature. Two important techniques introduced in (Criminisi *et al*., 2004) for image inpainting are priority map $P(p)$ and confidence term $C(p)$. The priority map, $P(p)=C(p)D(p)$, is defined to be the product of a data term $D(p)$ and the confidence term $C(p)$. The data term is based on the product of the isophote and a unit vector orthogonal to the front of an inpainted area. The data term and the confidence term together result in a strategy to select the best patch to be inpainted. However, computing the isophote and the unit vector for each patch at each iteration can be time consuming.

The purpose of the priority map $P(p)$ is to find the patch to be inpainted. The confidence term $C(p)$ is to indicate how reliable the pixel p is inpainted. We integrate these two factors in the face inpainting algorithm, by considering the selection process of pixels, p, to be inpainted using the confidences of surrounding pixels of p. Thus, our revised inpainting algorithm is presented as follows.

We assign the initial confidence term of pixel p in I: $C(p)$=1.0 iff $p\in\Phi$, and $C(p)$=0 iff $p\in\Omega$.

Let $\delta\Omega$ be a front contour on Ω and adjacent to Φ. Let Ψ_p be a patch centered at pixel $p\in\delta\Omega$. We re-define the initial confidence term $C(p)$ for pixels in $\delta\Omega$,

$$\forall p \in \delta\Omega,\ C(p) = \frac{1}{|\Psi_p|} \sum_{q \in \Psi_p \cap (\Phi_\varepsilon \cup \tilde{\Omega}_\varepsilon)} C(q), \qquad (6)$$

where $|\Psi_p|$ is the area of Ψ_p. The area is equal to 9 pixels. That is, a 3×3 patch was used in our experiments. Essentially, the confidence term computes the percentage of useful pixels in a patch Ψ_p. Useful pixels mean the coverage of source pixels in Φ. Since the edge map of the source region is computed only once for the entire image, computation time is saved, as compared to computing the isophote and the unit vector for each patch at each iteration (Criminisi *et al*., 2004).

Next, we introduce the concept of 'patch guidance'. In most inpainting algorithms, a large percentage of computation time is used while a patch is searching for a similar patch to be inpainted. The concept of patch guidance hints the searching procedure to find the most reliable patches, by using reference features. We believe that similar human faces have similar structural information (e.g., the continuity of edge in a mouth boundary). There is no need to search for the entire face to find better patches (i.e., the mouth will not be similar to the eyes). These reference features are the deduced edge map $\tilde{\Omega}_\varepsilon$ and deduced feature point set $\tilde{\Omega}_\gamma$. The deduced edge map is used for the calculation of the confidence term $C(p)$ (using $\Phi_\varepsilon \cup \tilde{\Omega}_\varepsilon$). The deduced feature point set $\tilde{\Omega}_\gamma$ can be used as a starting position to locate a candidate patch within a range. Let Ψ_q be a candidate patch that can be used for inpainting. With respect to Ψ_q, there is a nearby feature point $f \in \tilde{\Omega}_\gamma$, which can be selected by computing the minimal Euclidean distance between points q and f. Define by $\Theta(\Psi_q, d)$ an area centered at f, within a distance of d (which was set to 5 pixels in our experiments). Thus, our proposed face inpainting algorithm is described as follows:

Inpaint(Ω, $\tilde{\Omega}_\gamma$, $\tilde{\Omega}_\varepsilon$), repeat until $\Omega=\varnothing$.

Step 1: Compute boundary $\delta\Omega$ and $\forall p\in\delta\Omega$, compute $C(p)$.

Step 2: Using the structure information obtained from the target face in the searching procedure as guidance to propagate texture informaion obtained from a nearby undamaged area.

Step 2.1: Find $\Psi_{\hat{p}}$ with $\max C(p)$, $\forall p\in\delta\Omega$.

Step 2.2: Find $\Psi_{\hat{q}} = \min\limits_{\Psi_q \subseteq \Theta(\Psi_q, d)} diff(\Psi_{\hat{p}}, \Psi_q)$.

Step 2.3: Copy $\Psi_{\hat{q}} \cap \Omega$ to $\Psi_{\hat{p}} \cap \Omega$.

Step 2.4: Update $C(p) = \dfrac{1}{|\Psi_p|} \sum\limits_{q \in \Psi_p \cap (\Phi_\varepsilon \cup \tilde{\Omega}_\varepsilon)} C(q)$.

The function $diff(\Psi_{\hat{p}}, \Psi_q)$ computes the sum of squared differences (SSDs) based on the CIE Lab color space for the difference between the two patches $\Psi_{\hat{p}}$ and Ψ_q.

EXPERIMENTAL RESULTS

We used the IMM Face Database, which comprises 240 annotated color images of 40 different human faces (Stegmann *et al.*, 2003). The gender distribution is 7 females and 33 males. Fifty-eight landmarks were labeled on the facial structures of the eyebrows, the eyes, the nose, the mouth, and the jaw. Each face was captured in a six-image group with different expressions, illuminations, and viewpoints.

Some results are shown in Fig.3. The left column contains the original faces, the middle column shows masks, and the right column shows the inpainted faces. Detailed inpainting results are shown in Fig.4, which is a comparison of results from our system against the results obtained by an exemplar-based image inpainting approach (Criminisi *et al.*, 2004). The exemplar-based method works well near the borderline of the occlusion. However, as the inpainting goes further towards the center of the occlusion, the results become less and less reliable, due to the lack of shape information in the inpainting process. With shape propagation from the face database, our approach achieved much better results. In Figs.5 and 6, the details are illustrated as a good and a bad example, respectively. Fig.6 contains the details of a failure case (see the squares). These results are inpainted automatically by shape propagation. It may be noted that, using only the symmetric property of human faces to manually copy missing areas will bring artifacts, for most faces are not in perfect symmetry. Furthermore, since patch search only obtains a similar patch against the original, in some extreme situation, an inpainted face may not look exactly like the original one.

CONCLUSION

Our facial image inpainting algorithm is new to image inpainting, by incorporating facial domain knowledge to enhance inpainting results. Partial face-shape descriptors are used to search faces under different circumstances of missing areas. It is also the first time that patch guidance has been used in image inpainting. Patch guidance is computed from a face database by propagating facial shapes. We demonstrate that the proposed algorithm works on faces with different skin colors. The result is visually pleasant.

Our approach also has limitations. For instance, if the original face has an abnormal appearance (e.g., a scabbed face), which is partially covered by a mask, the inpainted face may not show the scab. In addition, we assume that the face has an ordinary distribution of luminous intensity; that is, the light source for taking the face photo is fixed in front of the person. How to overcome these restrictions is considered as our future work.

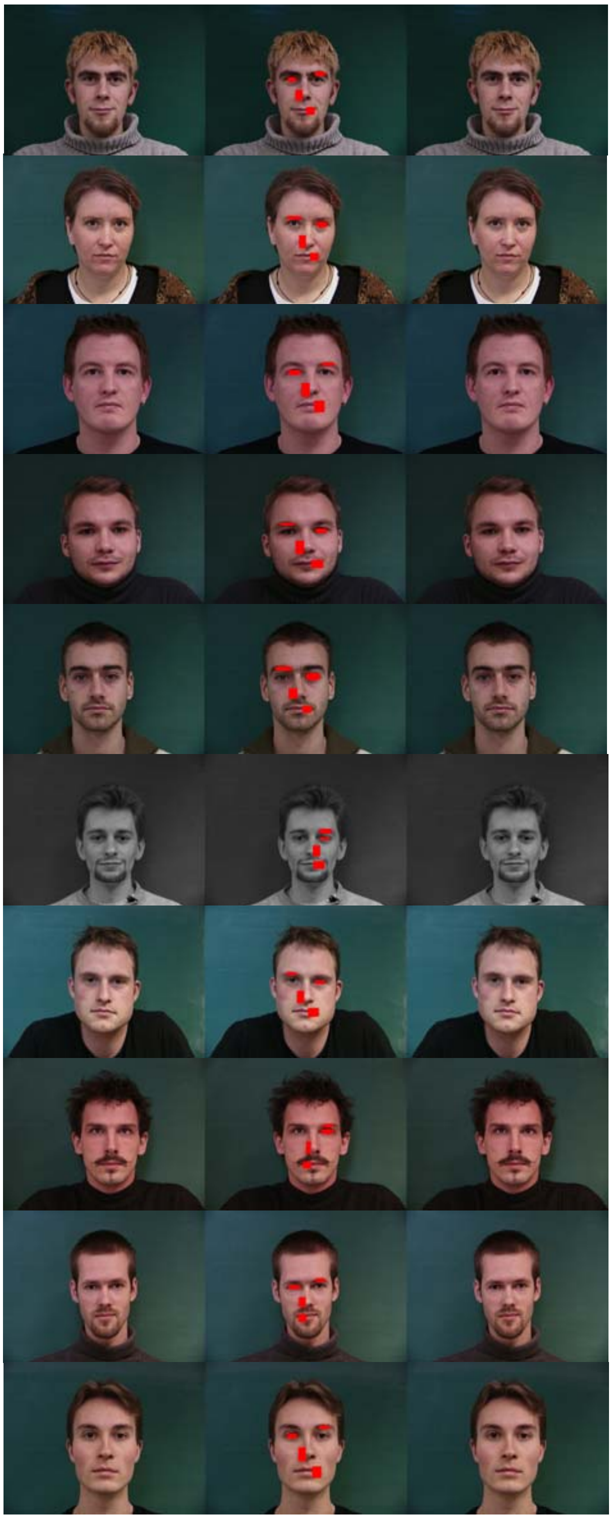


Fig.3 Face inpainting results
Left: Original faces; Middle: Masks; Right: Inpainted faces

(a) (b)

(c) (d)

Fig.4 Comparison of detailed inpainting results
(a) and (c): inpainting results by exemplar-based inpainting without shape propagation; (b) and (d): inpainting results by our approach

(a)

(b)

Fig.5 Details of a successful face inpainting
(a) Original; (b) Inpainted

(a)

(b)

Fig.6 Details of a face inpainting failure
(a) Original; (b) Inpainted

References

Criminisi, A., Perez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Processing*, **13**(9):1200-1212. [doi:10.1109/TIP.2004.833105]

Ersotelos, N., Dong, F., 2007. Building highly realistic facial modeling and animation: a survey. *The Visual Computer*, **24**(1):13-30. [doi:10.1007/s00371-007-0175-y]

Gross, R., Matthews, I., Baker, S., 2004. Constructing and Fitting Active Appearance Models with Occlusion. Conf. on Computer Vision and Pattern Recognition Workshop, p.72. [doi:10.1109/CVPR.2004.317]

Hwang, B.W., Lee, S.W., 2003. Reconstruction of partially damaged face images based on a morphable face model. *IEEE Trans. Pattern Anal. Mach. Intell.*, **25**(3):365-372. [doi:10.1109/TPAMI.2003.1182099]

Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N., Li, S.Z., 2006. Ensemble-based discriminant learning with boosting for face recognition. *IEEE Trans. Neural Networks*, **17**(1):166-178. [doi:10.1109/TNN.2005.860853]

Martínez, A., 1999. Face Image Retrieval Using HMMs. Proc. IEEE Workshop on Content-based Access of Image and Video Libraries, p.35-39. [doi:10.1109/IVL.1999.781120]

Mo, Z.Y., Lewis, J.P., Neumann, U., 2004. Face Inpainting with Local Linear Representations. British Machine Vision Conf., London, UK.

Nielsen, F., Nock, R., 2005. ClickRemoval: Interactive Pinpoint Image Object Removal. Proc. 13th Annual ACM Int. Conf. on Multimedia, p.315-318. [doi:10.1145/1101149.1101214]

Oda, M., Kato, T., 1993. What Kinds of Facial Features Are Used in Face Retrieval. Proc. 2nd IEEE Int. Workshop on Robot and Human Communication, p.265-270. [doi:10.1109/ROMAN.1993.367710]

Shih, T.K., Lu, L.C., Chang, R.C., 2003. Multi-resolution Image Inpainting. Proc. IEEE Int. Conf. on Multimedia & Expo, Baltimore, USA, p.485-488.

Stegmann, M.B., Ersboll, B.K., Larsen, R., 2003. FAME—a flexible appearance modeling environment. *IEEE Trans. Med. Imag.*, **22**(10):1319-1331. [doi:10.1109/TMI.2003.817780]

Sun, J., Yuan, L., Jia, J., Shum, H.Y., 2005. Image completion with structure propagation. *ACM Trans. Graph.*, **24**(3):861-868. [doi:10.1145/1073204.1073274]

Swets, D.L., Weng, J., 1996. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, **18**(8):831-836. [doi:10.1109/34.531802]

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *J. Cogn. Neurosci.*, **3**(1):71-86. [doi:10.1162/jocn.1991.3.1.71]

Wolberg, G., 1994. Digital Image Warping (1st Ed.). IEEE Computer Society Press, Los Alamitos, CA, USA.

Zhang, L.Z., Yang, Q., Bao, T., Vronay, D., Tang, X., 2006. ImLooking: Image-based Face Retrieval in Online Dating Profile Search. ACM SIG CHI, Canada, p.1577-1582.

Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 2003. Face recognition: a literature survey. *ACM Comput. Surv.*, **35**(4):399-458. [doi:10.1145/954339.954342]