



Data-driven hair segmentation with isomorphic manifold inference[☆]



Dan Wang^{a,b}, Shiguang Shan^{b,*}, Hongming Zhang^c, Wei Zeng^c, Xilin Chen^b

^a University of Chinese Academy of Sciences, Beijing, China

^b Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China

^c NEC Labs China, Beijing, China

ARTICLE INFO

Article history:

Received 8 June 2013

Received in revised form 28 November 2013

Accepted 25 February 2014

Available online 12 March 2014

Keywords:

Hair segmentation

Data driven

Shape model

Isomorphic manifold inference

ABSTRACT

Hair segmentation is challenging due to the diverse appearance, irregular region boundary and the influence of complex background. To deal with this problem, we propose a novel data-driven method, named Isomorphic Manifold Inference (IMI). The IMI method assumes the coarse probability map and the binary segmentation map as a couple of isomorphic manifolds and tries to learn hair specific priors from manually labeled training images. For an input image, firstly, the method calculates a coarse probability map. Then it exploits regression techniques to obtain the relationship between the coarse probability map of the test image and those of training images. Finally, this relationship, i.e., a coefficient set, is transferred to the binary segmentation maps and a soft segmentation of the test image will be achieved by a linear combination of those binary maps. Further, we employ this soft segmentation as a shape cue and integrate it with color and texture cues into a unified segmentation framework. A better segmentation is achieved by the Graph Cuts optimization. Extensive experiments are conducted to validate effectiveness of the IMI method, compare contributions of different cues and investigate the generalization of IMI method. The results strongly encourage our method.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Hair segmentation has attracted increasing interest, since it can benefit face retrieval [1], gender classification [2], head detection [3], skin segmentation [4] and glasses trying on [5]. Besides the above applications, hair segmentation can also contribute to computer graphics and virtual reality, such as hairstyle synthesis [6] and animation [7]. Especially, hairstyle trying on (See Fig. 1) and retrieval of pedestrians with specific hair color or style in video surveillance are two interesting applications. However, hair segmentation is a very challenging problem, as shown in Fig. 2. First, hair color is very diverse and often non-uniform especially for young people, which implies building a universal prior color model is very hard. Second, texture of hair region is hard to model since the hair may be complicated with many crimps. What's more, in case of low resolution, texture feature might disappear. Third, as a kind of flexible object, the outer contour of the hair region may be very irregular. Finally, the background or the clothes might be quite similar to the hair region in color or texture, which occurs very often in video surveillance.

1.1. Related previous work

Generally speaking, hair segmentation belongs to a special type of image segmentation, i.e., class-specific object segmentation. There have been many researchers addressing the problem [8–14], by using either supervised or unsupervised methods. In these works, class-specific shape prior has proven to be valuable for segmentation. Some of these methods [10,11,13,14] extract the specific class by using image fragments, which is suitable merely for structured objects. Veksler [15] introduces a prior specific to the “star shape” objects, which additionally requires knowing the shape center. Other representative approaches to learn object shape prior include Active Shape Model [16] and Active Appearance Model [17], which learn point distribution model of object landmarks. However, it is infeasible to define semantic landmarks for hair contour due to its irregularity and diversity.

As an instance of class-specific segmentation problems, hair segmentation is first posed, to our knowledge, by Liu et al. [18]. Later Kampmann [19] adopts a skin color model and facial location features to segment a head into face, ears, neck and hair regions. Wang et al. [20] proposed an exemplar-based model, which cannot deal with multi-pose cases. Moreover, most recently proposed methods [1,4,21,22] perform hair segmentation via three steps: 1) Finding initial hair pixels (seeds) that are surely in hair regions; 2) Building image-specific hair color model and predict the probability of each pixel being hair/background; 3) Performing segmentation based on the probability. Nevertheless, the above methods often fail to work when the

[☆] This paper has been recommended for acceptance by Lijun Yin.

* Corresponding author at: 6 Kexueyuan South Road, Zhongguancun, Haidian, Beijing, 100190, China. Tel.: +86 10 6260 0554.

E-mail address: sgshan@ict.ac.cn (S. Shan).



Fig. 1. An example of hairstyle trying-on.

neighboring regions are similar to hair regions in color or texture, e.g. clothes with similar colors to hair or trees as shown in Fig. 2.

To solve the problem, our idea is to explore hair-specific shape priors to reject 'impossible' hair shapes and false regions. The method in [5] employs active contour [23] to fit the upper hair shape model, but the landmark definitions are not so accurate and it can only get the upper hair parts. In [22] and [4], the methods are apt to apply average shape models on all images, which neglect the peculiarity of different instance shapes. Although they can reject some 'impossible hair shapes', it may also reject some true hair shapes occurring not frequently.

In all, the study on hair segmentation is still in an exploratory stage and current methods cannot work well in complex cases. Although image-based (bottom-up) methods like Graph Cuts [24] have been employed, very few works on learning hair-specific prior, like 'how to infer segmentation', from a set of training samples, have been addressed.

1.2. This paper

For an input image, if we firstly calculate a coarse Hair Probability Map (see Fig. 3) based on low level cues (e.g., color and location), human can

easily obtain the ground truth segmentation according to their knowledge. That is because human have seen many data about what is hair like and have the prior knowledge of hair shapes. Therefore, perceptually, the segmentation can be inferred from coarse HPM by seeing many training samples. Our target is to automatically learn the segmentation priors from training images and their segmentations. Specifically, we propose the Isomorphic Manifold Inference (IMI) method to learn the prior.

As shown in Fig. 3, given a training set containing head-shoulder images and ground truth, we want to infer the segmentation for a test image by implicitly or explicitly exploiting the hair-specific prior in the training set. First, for an input image and each sample in the training set, we calculate a coarse hair probability map (Coarse HPM) by exploring the information of the detected face and hair color model. Then the problem can be converted to how to infer segmentation from the Coarse HPM. Motivated by the observation that *Similar Coarse HPMs correspond to similar ground truth segmentations*, we formulate the Coarse HPM and Optimal HPM (i.e., ground truth segmentation) as a couple of manifolds that are isomorphic. The manifold formulation is valid because the hair contour is irregular and even unkempt, thus forms a non-linear



Fig. 2. Challenging examples for hair segmentation. The first row gives original images. Red parts of the second row show segmentation results with hair color model while blue parts of the third row are results of our proposed method.

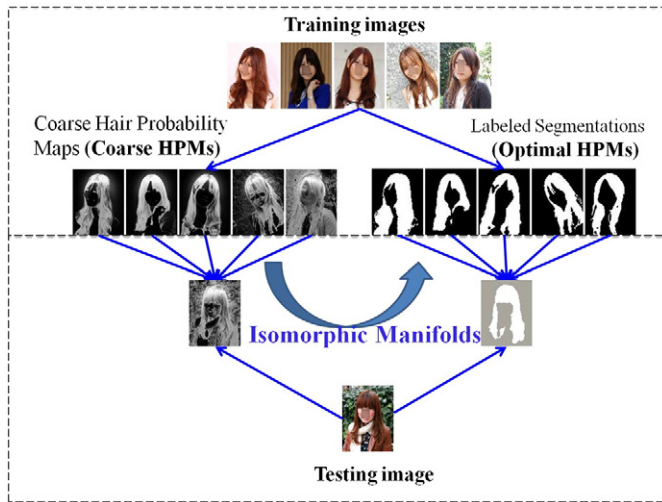


Fig. 3. Basic idea of this paper.

distribution. This allows us to exploit the relationships between Coarse HPMs to infer relationships between ground truth segmentations.

Specifically, the IMI method first employs some regression techniques [25] to pursue the relationship between the Coarse HPM of the test image and those of the training images. Then, it transfers this relationship, i.e., a set of coefficients, to the desirable segmentations of the test image and those of training images. Finally, a Refined HPM of test image can be obtained by linearly combining the segmentations of the training samples.

Since this Refined HPM is a linear combination of ground truth segmentations, it can assure the segmentation shape is globally hair-like. Thus we say, in some sense, our IMI method is a ‘noise-filtering’ procedure and the Refined HPM can be seen as a kind of shape cue.

In this paper, we further integrate the shape cue with color and texture cues into a unified framework. The final segmentation can be achieved by Graph Cuts optimization.

A preliminary version of this work appeared as a conference paper in [26]. The conference paper proposed the method of Isomorphic Manifold Inference (IMI) for hair segmentation. In this journal version, differences fall in three main folds. Firstly, texton-based features are further exploited to learn the hair texture prior, which is integrated into a unified energy function together with the shape prior learnt by our IMI method and the color prior. Corresponding experiments are also added to validate the effectiveness of texture prior and compare contributions of different priors. Secondly, to validate the generalizability of the proposed approach, we further apply it to horse segmentation by evaluating on a public horse dataset. Thirdly, experimental comparisons with two recent hair-specific methods in [20,27] are added and our method shows remarkable superiority.

The rest of the paper is organized as follows: Section 2 presents the IMI method and Section 3 introduces texton-based hair model and formulates shape, texture and color model into a unified segmentation framework. Then Section 5 gives extensive experiments to validate the effectiveness and the generalizability of the proposed method. Finally, Section 5 concludes the paper.

2. Method

Given a head-shoulder image, a coarse hair probability map is first calculated. With this Coarse HPM as input, our goal is to infer the Optimal HPM, by learning from the training images with their Coarse HPMs

and Optimal HPMs. We will first describe how to calculate the Coarse HPM and then give the detailed description of the IMI method.

2.1. Generation of the coarse HPM

Similar to the segmentation method proposed in [22], we employ a two-tier Bayesian method to generate a coarse hair probability map. In the first tier, a Bayesian model is constructed by integrating hair occurrence prior probabilities (HOPP) with a generic hair color model (GHCM) to obtain some reliable hair seed pixels. In the second tier, all of these selected seeds are used to train an Image-specific Hair Color Model (IHCM), which is combined with HOPP to build the second Bayesian model to calculate the Coarse HPM. Some interim results are shown in Fig. 4.

2.1.1. Hair color model and occurrence prior probability

The statistics tell that most hair color values scatter within certain ranges and obey specified principles. It is assumed that the distribution of generic hair color values follows Gaussian Mixture Model (GMM). This Generic Hair Color Model (GHCM) is learnt by using Expectation–Maximization (EM) algorithm. Thirteen Gaussian components are preserved for GHCM in our implementation. The number of components is determined by cross-validation. Selecting the number from 5 to 15 does not influence the result too much. Then we can calculate the conditional probability of each pixel under the GHCM.

Hair generally surrounds the face. Therefore, the HOPP at each location is an important clue. Fig. 4(b) visualizes HOPP which is learned from 1000 normalized images with labeled ground truth.

2.1.2. Bayesian model for hair seeds selection

After learning the GHCM and HOPP, we combine the two models with Bayesian framework. Note that we select hair seeds toward the over-segmented regions, which are obtained by the Mean Shift algorithm [28]. In our implementation, we simplify the calculation of Bayesian conditional probability as the weighted sum of the probability under the GHCM and HOPP. The weights of color and location model are set to be 0.6 and 0.4 respectively. Then certain regions with higher Bayesian conditional probability will be chosen as seeds, which are surely hair pixels. An example of hair seed regions are shown in red in Fig. 4(c).

2.1.3. Bayesian model for pixel classification

With selected hair seed regions, an Image-specific Hair Color Model, which is also GMM, can be learnt with EM algorithm.

Fig. 4(d) shows the hair probability map under IHCM. Combining this model with HOPP, the Coarse HPM is calculated in pixel-wise mode. The Coarse HPM is shown in Fig. 4(e). Additionally, Fig. 4(f) is the Refined HPM, produced by the IMI method. The method will be elaborated in the next subsection.

2.2. Isomorphic manifold inference for hair segmentation

Given a test image, a coarse HPM, generated by the above procedure, is an initial estimation of hair segmentation. Although it contains some errors, it provides rich information. For example, human can easily guess the segmentation from the coarse HPM with their knowledge, because human have seen many hair samples and have the prior of hair shapes. Our target is to automatically learn the segmentation (shape) specific to a test image, from training images and their segmentations.

The basic idea of our method is that similar Coarse HPMs correspond to similar ground truth segmentations (Optimal HPMs). Specifically, we assume that the Coarse HPMs and ground truth segmentation have the same distribution and formulate them as a couple of isomorphic manifolds, as shown in Fig. 5. Therefore, we first learn the relationship between the Coarse HPM of a test image and those of training ones, and then apply this relationship to pursue the desired Optimal HPM. The relationship between the Coarse HPMs is built with regression

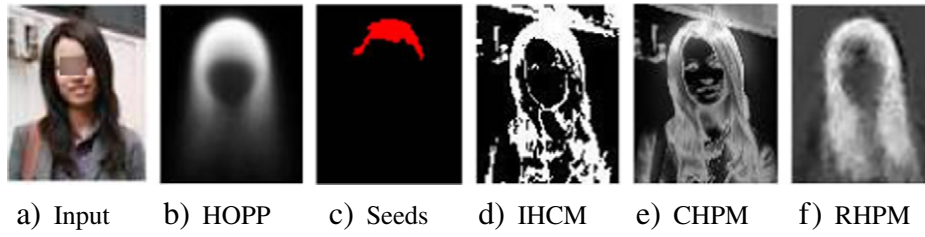


Fig. 4. An example of hair segmentation procedure. IHCM means image-specific hair color model. CHPM means coarse hair probability map and RHPM means refined hair probability map.

techniques and a set of coefficients is obtained as a solution. Then this relationship, i.e., the coefficient set, is applied to the ground truth segmentations. The whole framework of the IMI method is illustrated in Fig. 5. The method includes two steps: 1) *learn* the relationship between the Coarse HPM of a test image and those of the training images; 2) *transfer* the relationship to the Optimal HPMs to produce an approximation of 'Optimal HPM' for the test image.

2.2.1. Learning

We discuss how to learn the relationship under different assumptions of the manifold geometry: 1) each data point on the manifold can be linearly reconstructed by its k nearest neighbors; 2) each data point on the manifold can be reconstructed by a few samples under the sparse constraints.

Each HPM, denoted as a feature vector, represents a point in one of the two manifolds, shown in Fig. 5. For an image I with size $p = W \times H$, we define x, y as two p -dimensional column vectors and $x = [x_1, x_2, \dots, x_p]^T, y = [y_1, y_2, \dots, y_p]^T$. Assume x is the Coarse HPM and y is the Optimal HPM of I . For clarity, let x_s, y_s be the Coarse HPM and Optimal HPM (Ground Truth) vector of a training image and x_t, y_t be those of

a test image. We denote each element of x_t and y_t as x_{tv} and y_{tv} , for each location v . Note that each dimension of vectors x and y is normalized to be zero-mean. Let X_s, Y_s be two $p \times N$ matrix, with their columns being the training Coarse HPMs and Optimal HPMs respectively. Then the object functions and their solutions, corresponding to different geometry properties of the HPM manifolds, are discussed as follows:

1. Under the *local linear* assumption, we choose its k nearest neighbors (kNN) from the training set to represent the input Coarse HPM. The optimal weights can be achieved by:

$$\beta^{kNN} = \underset{\beta}{\operatorname{argmin}} \left\| x_t - \sum_{x_s \in N(x_t)} \beta_s x_s \right\|^2, \quad (1)$$

where $N(x_t)$ denotes the neighbors of x_t . In our implementation, the normalized similarity between x_s and x_t is utilized as the weight β_s . We call this case as 'kNN regression' briefly. Denote the similarity between x_t and its i th nearest neighbor x_{ki} as $S_{t,i}$, then $\beta_i^{kNN} = \frac{S_{t,i}}{\sum_{i=1}^k S_{t,i}}$.

2. Under the *sparse linear* assumption, each point on the manifold can be reconstructed by limited number of the N training samples.

HPM: Hair Probability Map

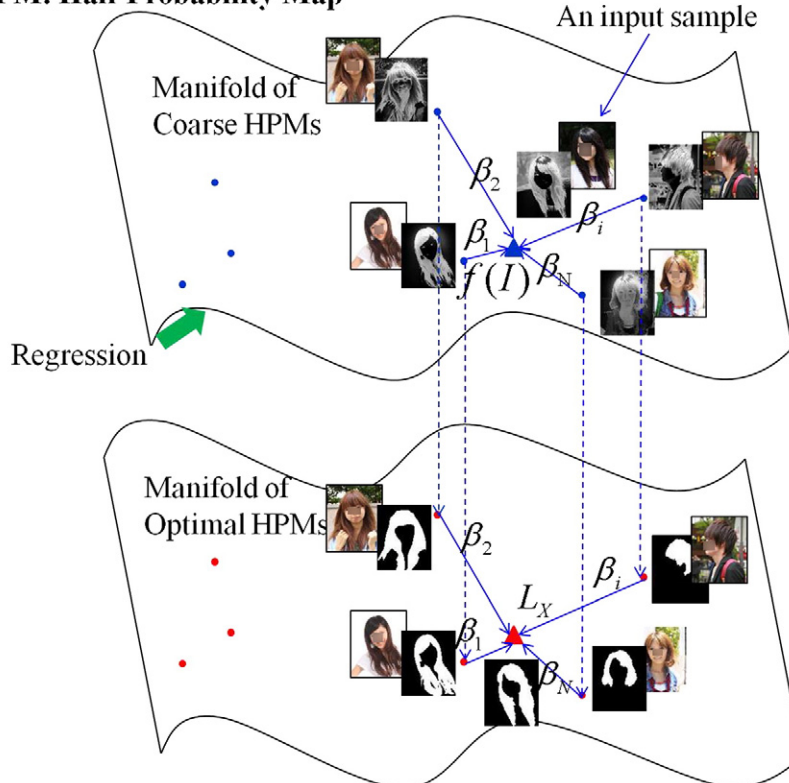


Fig. 5. Illustration of our IMI method.

Then it becomes a lasso [29] penalized least square problem and the target function becomes

$$\beta^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \left\| x_t - \sum_{x_s \in X_s} \beta_s x_s \right\|^2 + \lambda_{lasso} \sum_{x_s \in X_s} |\beta_s| \right\}. \quad (2)$$

An equivalent way to write the problem is

$$\begin{aligned} \beta^{lasso} = \underset{\beta}{\operatorname{argmin}} \sum \left\| x_t - \sum_{x_s \in X_s} \beta_s x_s \right\|^2 \\ \text{s.t.} \quad \sum_{x_s \in X_s} |\beta_s| \leq t. \end{aligned} \quad (3)$$

For small value of the bound t , Eq. (3) has a sparse solution, by causing some of the coefficients of x_s to be zero. The Lasso coefficients can be solved by a minor modification of the efficient Least Angle Regression (LARS) algorithm [29].

2.2.2. Transferring

We have presented how to learn the weights of an input HPM represented by others on the Coarse HPM manifold, under different assumptions of data distribution. Now the estimated weights can be transferred to Optimal HPM manifold and explored to get the Refined HPM:

$$y_t = \frac{1}{Z} \sum_{y_s \in Y_s} \hat{\beta}_s y_s \quad (4)$$

where $Z = 1/\sum_{i=1}^N \hat{\beta}_s$. It should be noted when using kNN regression, $\beta_s = 0$ for each $x_s \notin N(x_t)$. For each location v , we rewrite the Eq. (4) as the scalar form:

$$\hat{y}_{t_v} = \frac{1}{Z} \sum \hat{\beta}_s y_{s_v}. \quad (5)$$

where \hat{y}_{t_v} and y_{s_v} are the v th element of \hat{y}_t and y_s , respectively.

Note that y_t is not a hard segmentation map, but a soft map with real values in the range of [0,1]. We call y_t the Refined HPM, which is more accurate than the Coarse HPM x_t .

Since y_t is a linear combination of the ground segmentations, which can remove effects of noise in the Coarse HPM, it provide an important cues for hair shape. An example can be seen from Fig. 4(f), in which the Refined HPM suppress the noise of some confusing backgrounds.

3. Segmentation

As mentioned above, the Refined HPM can globally guide the segmentation to be hair-like, since it is a linear combination of ground truth segmentations. Therefore, the segmentation results, generated by thresholding on the Refined HPM have the advantage of being globally hair-like. However, it also causes the problem that hair boundaries cannot be well preserved due to the averaging of ground truth segmentations. To alleviate this problem, we introduce the texture information to help predict the hair regions and integrate the color and texture cues, with the shape cue to obtain the final segmentation result. In this section, we will first introduce how to exploit the texture information and then formulate texture, color and shape cues into a unified energy function and produce the segmentation by minimizing the function.

3.1. Hair texture cue

It has been realized texture information plays an important role in the representation of hair [30]. To capture hair specific texture characteristics, we employ the texton [31] based representation proposed by Varma and Zisserman [32]. Since the texton based feature is statistical, we over-segment the image into different regions with the Mean Shift algorithm [28] and extract the feature for each region. In the training process, all images are firstly convolved with 17 filters, which are suggested

by Winn et al. [33]. Then the responses are used to produce universal textons by using clustering techniques. Thus each pixel in an image can be given a texton index and finally each over-segmented region is described by a normalized texton index histogram. With these texton-based histograms we train Support Vector Machines (SVM) [34,35] to learn the model for the purpose of hair and non-hair region classification.

At run-time, the complete testing procedure is:

1. Segment the input image into different regions with the Mean Shift algorithm. The parameters of color resolution spatial resolution and the minimum region size are [color, spatial, min Region] = [2, 5, 100].
2. Convolve the input image with 17 filters and each pixel is classified into a cluster center. Then each pixel can be assigned a texton index.
3. Calculate the histogram of texton indices for each over-segmented region.
4. Run the hair/non-hair SVM and obtain the probability of each region being hair.

All pixels in the same region will be assigned with the same probability of being hair according to the prediction result of the SVM. This probability will be integrated with shape and color priors into an energy function to produce the final segmentation result.

3.2. Energy function

Let Λ be an image lattice (e.g. $W \times H$) and I_Λ a color image defined on Λ . For each location $v \in \Lambda$, $I_v = [R_v, G_v, B_v]^T$ is the pixel color value. The hair segmentation refers to seeking a labeling $L_\Lambda = (l_1, l_2, \dots, l_p)^T$, with $p = W \times H$ and

$$l_v = \begin{cases} 1, & \text{if pixel } v \text{ belongs to hair} \\ 0, & \text{if pixel } v \text{ belongs to background.} \end{cases} \quad (6)$$

The labeling L_Λ is also called ‘the desired shape’. Given an image I from a hair class, we determine optimal labeling L^* by minimizing the following energy function

$$E(L; I) = U(L; I) + \lambda_{gc} B(L; I). \quad (7)$$

This energy function integrates two terms and λ_{gc} is a weight to balance effects of them. The segmentation result will be achieved by performing the Graph Cuts algorithm [24]. The first term U is called data term, which reflects to what extent the pixel l_v fits into the observed data. It integrates two terms: the appearance and shape potentials and has the form of

$$U(L; I) = \omega_s \varphi^S(L; I) + \phi^A(L; I). \quad (8)$$

The factor ω_s is used to balance the importance of these two terms. The shape potential φ^S can globally guide the segmentation to be hair-like and is defined as a function of the Refined HPM in Eq. (5):

$$\varphi^S(L; I) = - \sum_{v \in \Lambda} \log \hat{y}_{t_v}. \quad (9)$$

On the contrary, the appearance potential ϕ^A focuses more on local fitness. In this study, it measures how well the low-level color and texture of the segmentation fit to the hair/background models. Formally, it can be written as:

$$\phi^A(L; I) = - \sum_v \left\{ \omega_\theta \log P(I_v, l_v | \theta_{gmm}) + \log P(\tau_v, l_v | \Theta) \right\} \quad (10)$$

where τ_v and l_v are texture and color feature vectors of location v in the image I . The parameter Θ denotes the hair-class texture model and θ_{gmm} is the Image-specific Hair Color Model, i.e., IHCM described in Section 2.1. The weight ω_θ is used to balance the importance of two terms.

The second term B on the right side of Eq. (7) is called the smoothness term, which is a penalty for a discontinuity between neighboring pixels. 8-neighbor configuration is adopted here. We choose a function related to the distance of RGB features and define the term as

$$B(L; I) = \sum_{(a,b) \in \text{Neighbors}} \delta(l_a \neq l_b) \cdot \gamma(a, b) \quad (11)$$

with

$$\gamma(a, b) = \exp\left(-\frac{\|I_a - I_b\|^2}{2\sigma^2}\right) \cdot \frac{1}{\text{dist}(a, b)} \quad (12)$$

and

$$\delta(l_a \neq l_b) = \begin{cases} 1, & \text{if } (l_a \neq l_b) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where (a, b) is any neighboring pixel pair, I_a and I_b are their RGB vectors, σ is the averaging discontinuity on the image and $\text{dist}(a, b)$ is their spatial distance.

For clarity of our method, Algorithm 1 describes the detailed procedure. Note that the final segmentation can be achieved by two strategies. Strategy (a) is to set threshold on the Refined HPM, i.e., only using the shape cue produced by IMI method. Strategy (b) is to combine color, texture and the shape prior into a unified energy function. The final segmentation is achieved by Graph Cuts optimization. The performance of Strategy (b) will be reported in the experiments, unless otherwise stated.

4. Experiments

To investigate the performance of the proposed method and validate its generalizability to other object class segmentations, four experiments are conducted: Section 4.2 compares the performances of different regression techniques for IMI; Section 4.3 compares our method with some state-of-the-art methods; Section 4.4 evaluates the effects of different cues, including the shape cue generated by the IMI, the color cue and the texture cue; Section 4.5 generalizes the proposed method to other object class segmentation and performs experimental comparisons on a public dataset.

Algorithm 1. IMI for data-driven hair segmentation

1. Given a training set of RGB images and the corresponding manually labeled ground truth YS.
2. Compute the corresponding Coarse HPM x_s for each training image, with the method in Section 2.1. These column vectors x_s consist the Coarse HPM matrix X_S .
3. For each testing image:
 1. Compute Coarse HPM.
 2. Obtain the estimated coefficients $\hat{\beta}$, by solving the Eqs. (1) or (2).
 3. Compute Refined HPM according to Eq. (4).
 4. Produce the final segmentation result in either of the following strategies.
 - (a) Setting threshold for \hat{y}_t .
 - (b) Integrate the color, texture cue with the shape cue into the energy function of Eq. (7) and perform Graph Cuts to get the final segmentation result.

4.1. Database and measurement

4.1.1. Database

We collect two challenging databases from internet with no overlapping with each other. The first is for training the hair color and location models (GHCM and HOPP). It includes 1000 images, which are near-frontal or with little pose variations. The second is the evaluation set, including 3800 images. Since the method is based on regression techniques, the evaluation set is randomly divided into two parts. One is the basis set, including 2000 images, and the other is testing set, including 1800 images. The evaluation set has a large range of variations in head-shoulder poses, hair styles, hair colors, illustration and backgrounds. Among these 3800 images, there are 2900 near-frontal images and 900 non-frontal ones.

In the experiments, all head-shoulder images are aligned to the size of 80×100 pixels according to head rectangle position. The ground truth of hair regions is manually labeled. Some examples of normalized images have been shown in Fig. 2 and Fig. 3.

4.1.2. Measurement

Performances are evaluated by assessing the consistency of automatic segmentation and the manually labeled ground truth in terms of the F-measure and accuracy. The F-measure defined as $2PR/(P + R)$. P stands for the precision which calculates the hair pixels of automatic segmentation overlapping with the ground truth and R stands for recall which measures hair pixels of the ground-truth overlapping with automatic segmentation. As F-measure is defined at all points on the precision-recall curve, the maximum F-measure score for each method is reported. The accuracy is defined as the percentage of correctly classified pixels in a whole image.

4.2. Evaluation of different regression methods

In this section, we will evaluate different regression techniques and analyze the distribution of the HPMs. The performances of the following regression methods will be compared: least square (LS) [25], Ridge regression (RR) [36], kNN regression and the Lasso. The LARS [29] is employed to solve lasso coefficients. Considering the speed of regression, all HPMs are downsampled to 40×50 pixels. It should be noted that the resulted Refined HPM (Eq. (4)) is cut by threshold 0.5 to produce the segmentation.

Note that the least square and Ridge regression is based on the global linear assumption of the data points. Thus we can investigate which assumption of the data distribution is more appropriate. For the least squares, the coefficients β can minimize the residual sum of squares, which is

$$RSS(\beta) = \sum \left\| x_t - \sum_{x_s \in X_S} \beta_s x_s \right\|^2. \quad (14)$$

If $X_S^T X_S$ is nonsingular, the solution of the minimization is

$$\hat{\beta}^{ls} = (X_S^T X_S)^{-1} X_S^T x_t. \quad (15)$$

For the Ridge regression, a regularization term is added:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \left\| x_t - \sum_{x_s \in X_S} \beta_s x_s \right\|^2 + \lambda_{\text{ridge}} \sum_{x_s \in X_S} \beta_s^2 \right\}. \quad (16)$$

The corresponding solution is

$$\hat{\beta}^{\text{ridge}} = (X_S^T X_S + \lambda_{\text{ridge}} I)^{-1} X_S^T x_t. \quad (17)$$

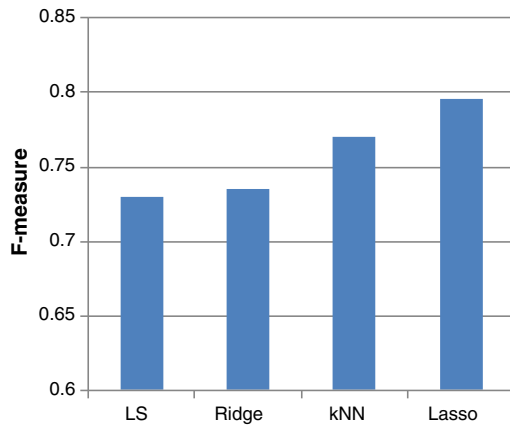


Fig. 6. Performance comparison of different regression models.

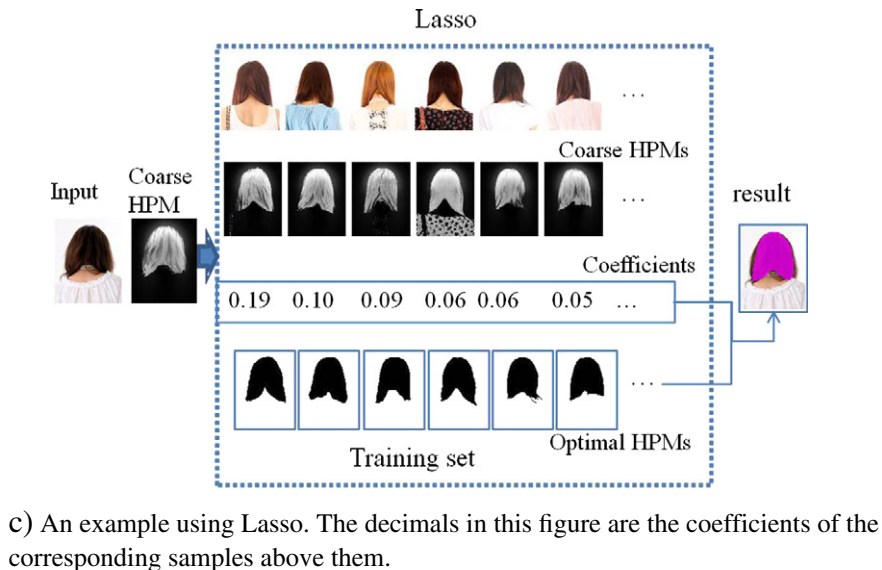
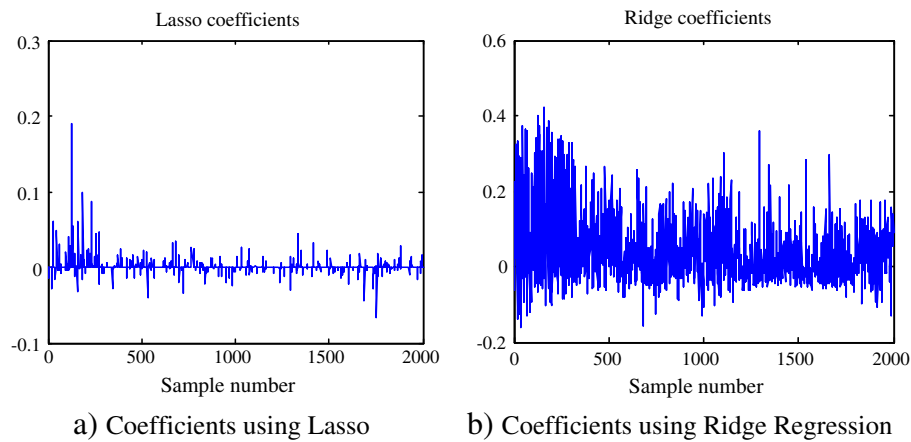
Fig. 6 shows the results comparison of different regression methods, from which it can be obviously seen that Ridge regression significantly outperforms least square. That is because least square considers all basis Coarse HPMs, which bring too many unpredictable errors. Ridge regression performs superior to least square, since it can trade a little

bias on the Coarse HPM for a large reduction of the error influences. Nevertheless, neither of the two solutions is satisfactory. The reason is that the globally linear assumption on the data points is not appropriate. In contrast, kNN regression, which considers only the neighboring similar samples, performs better. Based on the locally linear assumption, kNN regression can significantly depress the error influences from samples, which are far from the input. The experimental result supports more on this locally linear assumption.

As another point of view, kNN regression can be seen as a method of variable selection [25]. The problem is that setting the parameter k to be the same value for different images is not reasonable. The Lasso, fortunately, can solve the problem. It applies the sparse constraints and can adaptively cause some coefficients to be exactly zero. Note that in Fig. 6 the Lasso performs better than kNN regression.

Additionally, Fig. 7 plots the Ridge and Lasso solution coefficients for the input image. It can be seen that compared with Ridge regression, the Lasso coefficients are much more sparse. Fig. 7(c) gives the corresponding samples with the largest Lasso coefficients (plotted in Fig. 7(a)). Note that most of these hair styles are similar to the style of the input image. This indicates that Lasso provides a compact representation of the HPM, while rejecting effects from samples which are very different from the input.

As for the effects of the parameters, Fig. 8(a) gives the F-measure with the varying λ_{ridge} in Eq. (16) for Ridge regression. Higher λ_{ridge}



c) An example using Lasso. The decimals in this figure are the coefficients of the corresponding samples above them.

Fig. 7. An example using Lasso.

means more shrinking of coefficients, which also means the higher variances of the data distribution. Moreover, Fig. 8(b) gives kNN regression performances varying with the number of neighbors. The best result is obtained when only 8 nearest neighboring samples are preserved. Finally, Fig. 8(c) plots the F-measure as the t changes. The parameter t is the upper bound of the coefficients' L1-norm. The smaller the parameter t is, the sparser the coefficients are. We achieve the best performance when $t = 4$, which also demonstrates the sparse structure of data distribution.

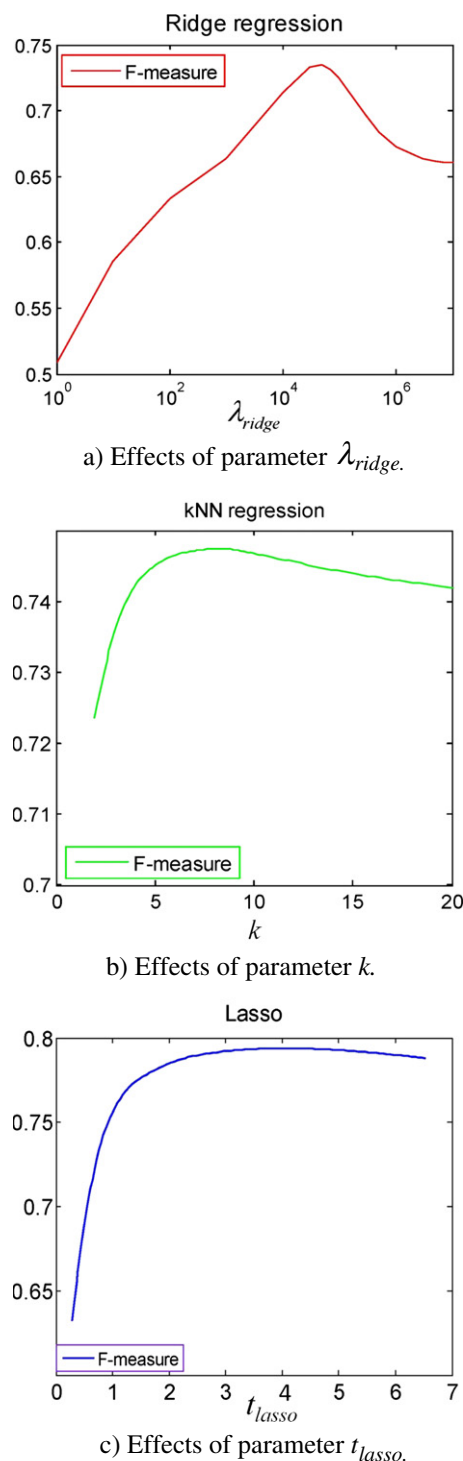


Fig. 8. Parameter effects on the performance.

Table 1

Performance comparison of different methods.

Algorithm	F-measure	Precision/recall
Bayesian method [22]	0.619	0.612/0.714
Graph cuts [4]	0.683	0.705/0.776
Automatic version of MSRM [37]	0.702	0.811/0.700
Proposed method	0.865	0.897/0.850

4.3. Comparison with other methods

On our database, the proposed method is compared with latest methods: the Bayesian method in [22], Graph Cuts [4] and the maximal similarity based region merging (MSRM) method [37].

According to Section 4.2, Lasso performs best for IMI to learn hair specific priors. Consequently, our experiments adopt Lasso in the learning step. The upper bound of L1-norm of coefficients, which is the parameter t in Eq. (3), is set to be 4.0, according to cross validation. After



Fig. 9. Comparison of results of different methods. (a) Original image; (b) Bayesian method; (c) graph cuts; (d) MSRM; (e) proposed method.

Table 2

Comparison with two recent hair-specific methods in [20] and [27] on the LFW dataset.

Algorithm	Accuracy
Exemplar-based model [20]	0.891
Hair shape modeling [27]	
Proposed method	0.914

obtaining a Refined HPM, we combine the shape, color and texture priors to obtain the final segmentations, by minimizing the Eq. (7). Bayesian method, including the parameter settings, is implemented exactly as the literature [22] presents. For Graph Cuts, it is essentially similar to the method in [4] which performs and updates the hair color model iteratively.

Additionally, the approach of MSRM needs human labeling for the fore-background seed regions, which is not fair for other automatic segmentation approaches. Therefore, in our implementation the interactive step is replaced by the automatic seed selection method, described in Section 2.1 Based on these configurations, Table 1 tabulates performance comparisons of different methods. It can be easily seen that the proposed method can achieve superior performances to the other methods.

Some examples of segmentation results for different methods are shown in Fig. 9. Note that other than the bottom row in Fig. 9, the rest are successful segmentation examples. The hair segmentation result can often be prevented from being influenced by the similar clothes (Rows 1, 2, 4, 5,) and other confusing backgrounds (Rows 3, 6). This is mainly due to the strong hair-specific prior by IMI, which learns the prior using the Lasso. Such IMI method tells what are abnormal ‘hair’ shapes and can suppress the noise in the Coarse HPM, just like a ‘noise filter’. Therefore, the result will not be largely beyond normal hair regions. The Bayesian method and Graph Cuts method often fail because of in proper parameters for the images, while the merging strategy is not always effective due to unreasonable fore/background models.

Additionally, we also evaluate our method on a public dataset, Labeled Faces in the Wild [38]. The manually labeled hair segmentation of the dataset is available on the website <http://media.cs.tsinghua.edu.cn/wangnan/>. The dataset contains 1046 images, covering large hair style variations, illuminations and clutter backgrounds. On the LFW dataset, the IMI method is compared with two recent hair-specific methods in [20] and [27]. The experimental settings are exactly the same as those of [20] and [27]. Half of these images are selected randomly as basis set and the other half for testing. Comparison results are shown in Table 2. The data tell that the proposed method can significantly outperform the method in [20] and [27]. Some segmentation results of our method are illustrated in Fig. 10. Note our proposed method shows satisfactory results, although the images contain different hair colors, styles and backgrounds.

4.4. Evaluation of different cues

To analyze the contribution of each term in Eqs. (7) and (10), we present some interim results in Fig. 11. Note that, the shape model produced by the IMI method can coarsely restrain the segmentation to be hair-like and suppress incorrect regions adjacent to the hair. Since the shape model is the sparse linear combination learning of training shapes, it can globally preserve some hair style cues. On the contrary, the color and texture cue can help to preserve more local details like the hair tail regions.

Fig. 12 illustrates more quantitative comparisons on hair database. It can be seen that each cue facilitates the segmentation performance and the combinations can further improve it. It should be pointed out that, the performance of ‘color + texture’ is about 0.79, and it can be upgraded to 0.865 when adding the shape model.

Fig. 13 also plots the curve of the segmentation performance varying with the weight ω_i in Eq. (7). It shows the value of F-measure increases with the increasing of the weight before 0.7 and decreases after that. This also demonstrates that the crucial role of the shape cue generated by the IMI method.

4.5. Generalization of the IMI method

The IMI method has been demonstrated effective for hair segmentation. It can also be generalized to other object class segmentation. In this section, we take the horse class as an example and evaluate its performance on a public horse database to validate the generalization ability of IMI method.

We adopt the Weizmann database [10] which contains 328 horse images and their ground truth segmentations, covering different poses and colors. All images are cropped to 120×96 according to the horse rectangle box, which is automatically detected by the deformable part model [39]. We randomly select one third images as the testing set and the rest as the basis set. All parameters in the horse experiments are set as those of the hair segmentation experiments.

We present some qualitative results from the Weizmann horse dataset in Fig. 14. Notice that, despite the presence of different colors and poses, the algorithm is able to reliably separate the horse from the background.

Table 3 summarizes the segmentation performance comparison with the Obj Cut [40], Auto-context [41] and the kernelized structural SVM method in [42] proposed by Bertelli et al. Note that in Obj Cut in [40] only reports the accuracy on a small subset of the horse database. The table shows we achieve comparable accuracy with the other three approaches. The PR-Curve comparison between our method and the Auto-context is also plotted in Fig. 15, where our curve is directly plotted on the figure cited from [41]. Note the performance of our method is comparable to the Auto-context.



Fig. 10. Some segmentation results on the LFW dataset. The proposed method shows satisfactory results, although the images contain different hair colors, styles and backgrounds.

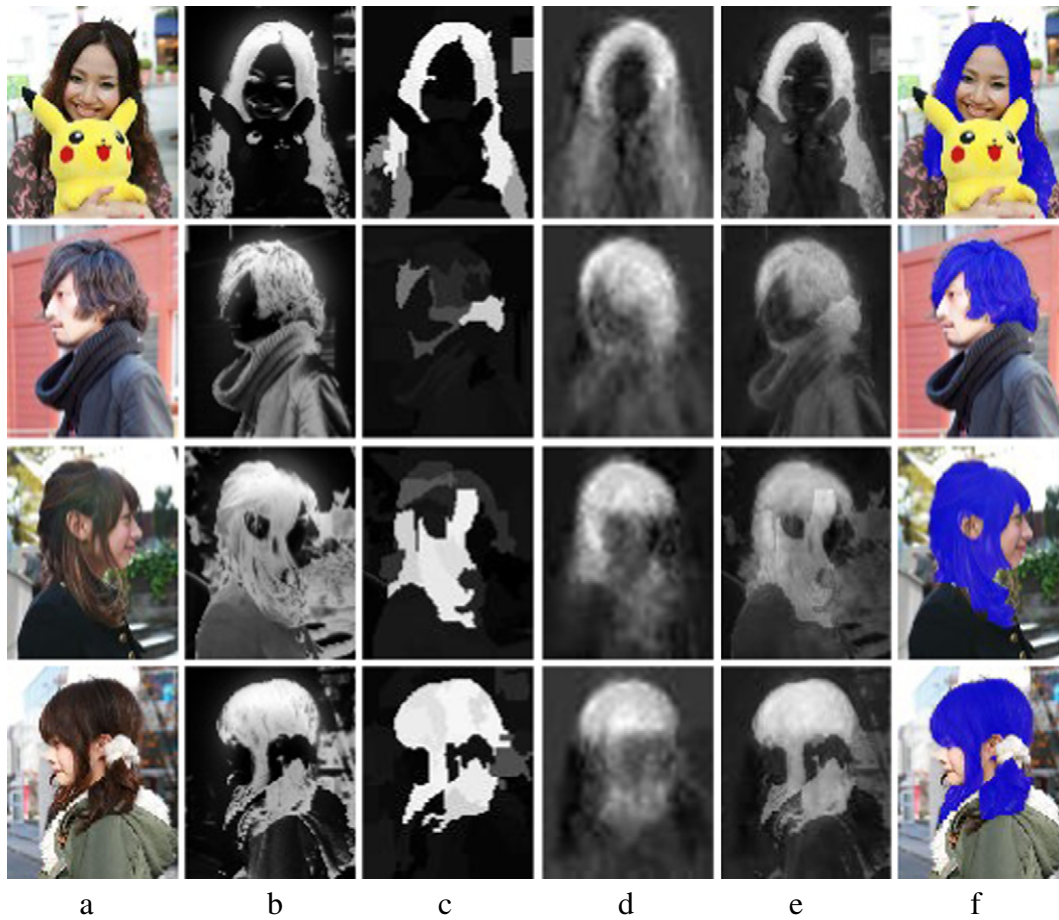


Fig. 11. Illustration of the contributions of different cues. (a) Cropped images; (b) color probability maps; (c) texture probability maps; (d) IMI maps (shape); (e) combination of color, texture and IMI maps; (f) segmentation results with the blue masks.

5. Discussion

In this paper, we propose isomorphic manifold inference method for hair segmentation. The IMI method implicitly learns hair-specific prior information from a set of training samples, by using some regression techniques. Specifically, given an input image, a normalized Coarse HPM will be first estimated based on the detected face and hair color information. Then a Refined HPM will be calculated via IMI, by imposing appropriate constraints on regression models. It should be noted that

the IMI method do not care how to generate a coarse probability map, but only provide a manner for learning class-specific priors and ‘filter’ some noise of the coarse hair probability map. Experimental results strongly encourage the proposed method.

We also extend the method to combine it with color and texture models for more accurate segmentation. The results show the shape cue generated by the IMI method can globally guide the segmentation to be more hair-like, while the color and texture models can facilitate preserving the hair boundaries.

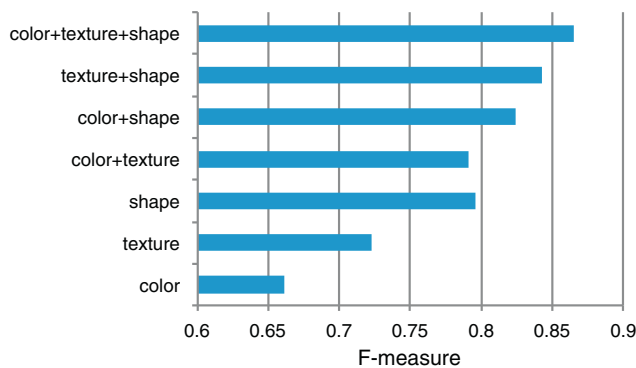


Fig. 12. Contribution comparison of different cues on the hair database. The notation ‘+’ means a combination of cues on both sides.

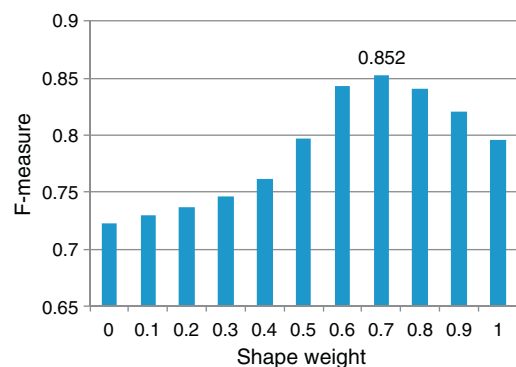


Fig. 13. Effects of the weight ω_θ of shape cues.



Fig. 14. Some segmentation results on the Weizmann horse dataset. Despite the presence of different colors and poses, the proposed algorithm shows promising performances.

Furthermore, we take the horse class as an example to discuss how to generalize the IMI method to other object class segmentation and obtain satisfactory results. Moreover, it can segment any specific class, which can be coarsely aligned. For example, it can also be used to segment head-shoulder regions, since the image can be normalized according to the detected face bounding box and a coarse head-shoulder probability map may be calculated in the similar way as described in Section 2.1. Then our IMI methods can be performed to infer the segmentation result. Similar case occurs on human body segmentation and car segmentation. The images can be normalized according to rect-angle provided by the pedestrian and car detectors.

Acknowledgment

This paper is partially supported by Natural Science Foundation of China under contracts No. 61025010, No. 61222211, and 61272321.

Table 3

Comparison with other state of the art algorithms on the horse dataset.

Algorithm	Accuracy	F-measure
Obj cut [40]	0.960	–
Auto-context [41]	–	0.835
Bertelli et al. [42]	0.946	–
Proposed method	0.931	0.839

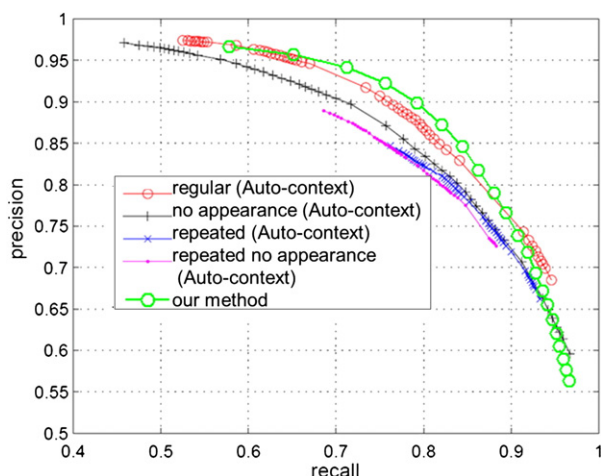


Fig. 15. Performance comparison of our method and auto-context.

References

- [1] Y. Yacoob, L. Davis, Detection and analysis of hair, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (7) (2006) 1164–1169.
- [2] K. Ueki, H. Komatsu, S. Imaizumi, K. Kaneko, N. Sekine, J. Katto, T. Kobayashi, A method of gender classification by integrating facial, hairstyle, and clothing images, *Proceedings of International Conference on Pattern Recognition*, vol. 4, 2004, pp. 446–449, (Cambridge, England, UK).
- [3] Z. Zhang, H. Gunes, M. Piccardi, An accurate algorithm for head detection based on xyz and hsv hair and skin color models, *Proceedings of IEEE International Conference on Image Processing*, 2008, pp. 1644–1647.
- [4] K. Lee, D. Anguelov, B. Sumengen, S. Gokturk, Markov random field models for hair and face segmentation, *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1–6.
- [5] P. Julian, C. Dehais, F. Lauze, V. Charvillat, A. Bartoli, A. Choukroun, Automatic hair detection in the wild, *Proceedings of International Conference on Pattern Recognition*, 2010, pp. 4617–4620.
- [6] F. Bertails, B. Audoly, B. Querleux, F. Leroy, J.-L. Lévêque, M.-P. Cani, Predicting natural hair shapes by solving the statics of flexible rods, *Eurographics Short Papers*, Dublin, Ireland, 2005.
- [7] F. Bertails, B. Audoly, M.-P. Cani, B. Querleux, F. Leroy, J.-L. Lévêque, Super-helices for predicting the dynamics of natural hair, *ACM Transactions on Graphics*, vol. 25, 2006, pp. 1180–1187.
- [8] J. Shotton, J. Winn, C. Rother, A. Criminisi, Textonboost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context, *Int. J. Comput. Vis.* 81 (1) (2009) 2–23.
- [9] J. Winn, N. Jojic, Locus: learning object classes with unsupervised segmentation, *Proceedings of International Conference on Computer Vision*, vol. 1, 2005, pp. 756–763.
- [10] E. Borenstein, S. Ullman, Class-specific, top-down segmentation, *Proceedings of European Conference on Computer Vision*, 2002, pp. 639–641.
- [11] E. Borenstein, S. Ullman, Learning to segment, *Proceedings of European Conference on Computer Vision*, vol. 3023, 2004, pp. 315–328.
- [12] X. He, R. Zemel, D. Ray, Learning and incorporating top-down cues in image segmentation, *Proceedings of European Conference on Computer Vision*, 2006, pp. 338–351.
- [13] A. Levin, Y. Weiss, Learning to combine bottom-up and top-down segmentation, *Proceedings of European Conference on Computer Vision*, 2006, pp. 581–594.
- [14] B. Alexe, T. Deselaers, V. Ferrari, Classcut for unsupervised class segmentation, *Proceedings of European Conference on Computer Vision*, 2010, pp. 380–393.
- [15] O. Veksler, Star shape prior for graph-cut image segmentation, *Proceedings of European Conference on Computer Vision*, vol. 5304, 2008, pp. 454–467.
- [16] T. Cootes, C. Taylor, D. Cooper, J. Graham, et al., Active shape models—their training and application, *Comput. Vis. Image Underst.* 61 (1) (1995) 38–59.
- [17] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (6) (2001) 681–685.
- [18] Z. Liu, J. Guo, L. Bruton, A knowledge-based system for hair region segmentation, *International Symposium on Signal Processing and Its Applications*, 1996, pp. 575–576.
- [19] M. Kampmann, Segmentation of a head into face, ears, neck and hair for knowledge-based analysis-synthesis coding of videophone sequences, *Proceedings of IEEE International Conference on Image Processing*, vol. 2, 1998, pp. 876–880.
- [20] N. Wang, H. Ai, S. Lao, A compositional exemplar-based model for hair segmentation, *Proceedings of Asian Conference on Computer Vision*, 2010, pp. 171–184.
- [21] C. Rousset, P. Coulon, Frequential and color analysis for hair mask segmentation, *Proceedings of IEEE International Conference on Image Processing*, 2008, pp. 2276–2279.
- [22] D. Wang, S. Shan, W. Zeng, H. Zhang, X. Chen, A novel two-tier Bayesian based method for hair segmentation, *Proceedings of IEEE International Conference on Image Processing*, 2009, pp. 2401–2404.
- [23] T. Chan, L. Vese, Active contours without edges, *IEEE Trans. Image Process.* 10 (2) (2001) 266–277.
- [24] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (11) (2001) 1222–1239.
- [25] J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning*, vol. 1, 2001.

- [26] D. Wang, S. Shan, H. Zhang, W. Zeng, X. Chen, Isomorphic manifold inference for hair segmentation, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2013.
- [27] N. Wang, H. Ai, F. Tang, What are good parts for hair shape modeling? *Proceedings of International Conference on Computer Vision and, Pattern Recognition*, 2012, pp. 662–669.
- [28] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (5) (2002) 603–619.
- [29] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B Methodol.* (1996) 267–288.
- [30] H. Chen, S.-C. Zhu, A generative sketch model for human hair analysis and synthesis, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (7) (2006) 1025–1040.
- [31] J. Malik, S. Belongie, T. Leung, J. Shi, Contour and texture analysis for image segmentation, *Int. J. Comput. Vis.* 43 (1) (2001) 7–27.
- [32] M. Varma, A. Zisserman, A statistical approach to texture classification from single images, *Int. J. Comput. Vis.* 62 (1–2) (2005) 61–81.
- [33] J. Winn, A. Criminisi, T. Minka, Object categorization by learned universal visual dictionary, *Proceedings of International Conference on Computer Vision*, vol. 2, 2005, pp. 1800–1807.
- [34] C. Cortes, V. Vapnik, Support vector machine, *Mach. Learn.* 20 (3) (1995) 273–297.
- [35] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 27.
- [36] A. Hoerl, R. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics* 12 (1) (1970) 55–67.
- [37] J. Ning, L. Zhang, D. Zhang, C. Wu, Interactive image segmentation by maximal similarity based region merging, *Pattern Recogn.* 43 (2) (2010) 445–456.
- [38] G. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, *Tech. rep. University of Massachusetts, Amherst*, 2007.
- [39] P. Felzenszwalb, D. McAllester, D. Ramanan, A discriminatively trained, multiscale, deformable part model, *Proceedings of International Conference on Computer Vision and, Pattern Recognition*, 2008, pp. 1–8.
- [40] M. Kumar, P. Ton, A. Zisserman, Obj cut, *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 18–25.
- [41] Z. Tu, X. Bai, Auto-context and its application to high-level vision tasks and 3D brain image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (10) (2010) 1744–1757.
- [42] L. Bertelli, T. Yu, D. Vu, B. Gokturk, Kernelized structural svm learning for supervised object segmentation, *Proceedings of International Conference on Computer Vision and, Pattern Recognition*, 2011, pp. 2153–2160.