Isomorphic Manifold Inference for Hair Segmentation

Dan Wang[1,2], Shiguang Shan[1], Hongming Zhang[3], Wei Zeng[3], Xilin Chen[1]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing, China
[2]Graduate University of CAS, Beijing, China
[3]NEC Laboratories China, Beijing, China
{dan.wang, shiguang.shan, xilin.chen }@vipl.ict.ac.cn; { zhang_hongming, zeng_wei}@nec.cn

***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 method, named Isomorphic Manifold Inference (IMI). Given a head-shoulder image, a Coarse Hair Probability Map (Coarse HPM), each element of which represents the probability of the pixel being hair, is initially calculated by exploring hair location and color priors. Then, based on an observation that similar Coarse HPMs imply similar segmentations, we formulate Coarse HPM and corresponding ground segmentation (Optimal HPM) as a pair of isomorphic manifolds. Under this formulation, final hair segmentation is inferred from the Coarse HPM with regression techniques. In this way, the IMI implicitly exploits the hair-specific prior embodied in the training set. Extensive experimental comparisons are conducted and the results strongly encourage the method. The generality of IMI to other class-specific image segmentation is also discussed.**

Keywords-Hair segmentation; Shape prior; Isomorphic Manifold Inference; Graph Cuts

I. INTRODUCTION

Hair segmentation has attracted increasing interest, since it can benefit face retrieval [1], gender classification [2], head detection [3], hair re-colorization [4], skin segmentation [5] and glasses trying on [6]. Besides the above applications, hair segmentation can also contribute to computer graphics and virtual reality, such as hair styling [7] and animation [8]. Especially, hairstyle trying on 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.1. First, hair color is very diverse and often non-uniform especially for young people, which implies building a universal prior color model 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.

A. 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 [9-15], 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

Figure 1: Challenging examples for hair segmentation[1]. The first row gives original images. Red parts of the second row show segmentation results with single hair location model while blue parts of the third row are results with our proposed method.

[11][12][14][15] extract the specific class by using image fragments, which is suitable merely for structured objects. Veksler [16] 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 [17] and Active Appearance Model [18], 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. [19]. Later Kampmann [20] adopts a skin color model and facial location features to segment a head into face, ears, neck and hair regions. Wang et al. [4] proposed an exemplar-based model, which cannot deal with multi-pose cases. Moreover, most recently proposed methods [1][5][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 neighboring regions are similar to hair regions in color or texture, e.g. clothes with similar colors to hair or trees as shown in Figure 1.

To solve the problem, a natural idea is to explore hair-specific priors to reject false regions. The method in [6] 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 [5], the methods are apt to apply average shape models on all images, which neglect the

[1]Some faces are blurred in this paper for privacy issue.

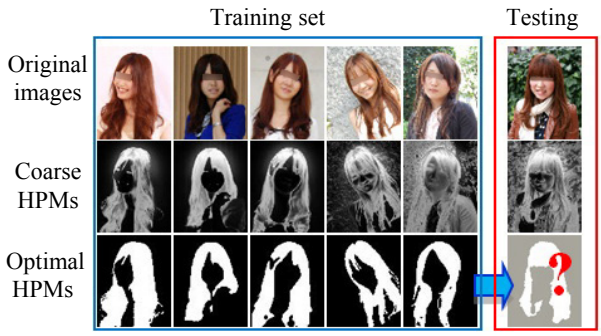

Figure 2: Description of our goal in this paper. The Coarse HPMs are calculated with the method elaborated in Sec. II.A

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 Cut [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.

B. *This paper*

As shown in Fig.2, given a training set containing head images and ground truth, our goal is to infer the segmentation for a test image by implicitly or explicitly exploiting the hair-specific prior in it. First, for 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. We are motivated by the observation: ***Coarse HPMs and ground truth segmentations have similar structures***, which can be seen from the examples in Fig.2. Based on this observation, the Coarse HPM and Optimal HPM (i.e., ground truth segmentation) are formulated 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 distribution.

Under this isomorphic manifold formulation, we propose Isomorphic Manifold Inference (IMI) method to compute a more accurate HPM (**Refined HPM**), by employing some regression techniques [25]. In some sense, our IMI is a 'noise-filtering' procedure, which exploits the hair prior implied in the training set to refine the Coarse HPMs. In IMI, appropriate constraints are imposed on regression models, which can assure that only the training samples, the shapes of which are related to the input, are utilized.

A recent class-specific segmentation method similar to ours is [26], which, however, adopts a different manner by using kernel principal component analysis

The rest of the paper is organized as follows: Sec.0 presents the formulation and implementation of the proposed method. Then Sec.III presents experiments to validate our method and Sec.IV concludes the paper.

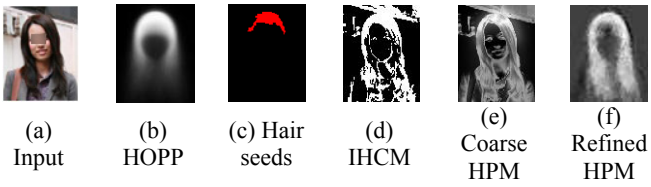

Figure **3**: An example of hair segmentation procedure.

II. 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 calculat the Coarse HPM and then give the detailed description of the IMI method.

A. *Automatic generation of the Coarse HPM*

The main procedure, as shown in Fig.3, includes the following steps: 1) perform face/head detection and coarsely normalize the image (Fig.3(a)); 2) choose seed pixels that reliably belong to hair (Fig.3(c)); 3) learn the image-specific hair color model based on the seeds and predict the Coarse HPM (Fig.3(e)). The steps are elaborated as follows:

1) *Hair seed selection using Bayesian model*

To choose seed pixels which are surely hair pixels, the hair color and location knowledge are both used by using a Bayesian model, similar to [22]. The location prior information is represented by location prior by Hair Occurrence Prior Probability (HOPP) for each pixel. Fig.3(b) illustrates the visualization of HOPP which is learned from 800 normalized images with labeled ground truth. Moreover, we learn a Generic Hair Color Model (GHCM) towards RGB values from a labeled set. The GHCM is represented by Gaussian Mixture Model (GMM) distribution and learnt using Expectation-Maximization (EM) algorithm. Thirteen Gaussian components are preserved for GHCM in our implementation.

After learning the GHCM and HOPP, we combine the two models with Bayesian framework. The weights of color and location model are set to be 0.6 and 0.4 respectively. Then certain pixels that are surely hair pixels can be chosen. Unlike [22], which performs pixel-wise choosing, our model operates in region-wise mode based on the over-segmented regions. The Mean Shift [27] method is adopted to perform over-segmentation. Hair seed regions are shown in red in Fig.3(c).

2) *Online learning of hair color and the HPM predicting*

With selected hair seed regions, an Image-specific Hair Color Model (IHCM) can be learnt with EM algorithm. The hair color is modeled by five Gaussian components. Fig.3(d) shows the hair probability map, predicted IHCM. Combining this model with HOPP, the Coarse HPM is calculated in pixel-wise mode. The weight of the image-specific color model is set to be 0.7 when calculating Coarse HPMs. The Coarse HPM is shown in Fig.3(e). Additionally, Fig.3 (f) is the Refined HPM, produced by the IMI method, elaborated in the next subsection.

B. *Isomorphic Manifold Inference for hair segmentation*

Our idea is to learn the relationship between the Coarse HPM of a test image and that of training ones, and then apply this relationship to pursuit the desired Optimal HPM. Our assumption is that the Coarse HPMs and Optimal HPMs have the same distribution. Specifically, we formulate the Coarse HPM and the Optimal HPM as a couple of isomorphic manifolds, as shown in Fig.4. The method includes two steps: 1) ***learn*** the relationship between the Coarse HPM of a test image and that of the training images; 2) ***transfer*** the relationship to the Optimal HPMs to produce an approximation of 'Optimal HPM' for the test image.

1) *Learning*

We discuss how to learn the relationship under different assumptions of the manifold geometry: 1) the data points on the manifold are locally linear and each point can be linearly reconstructed by its neighbors; 2) the data points on the manifold are globally linear and each point can be reconstructed by all samples.

Each HPM, denoted as a feature vector, represents a point in one of the two manifolds, shown in Fig.4. For an image ***I***, We define x, y as two *p*-dimensional column vectors. x is a Coarse HPM and y is an Optimal HPM of ***I***. For clarity, let x_s, y_s be the Coarse HPM and Optimal HPM vector of a *training* image and x_t, y_t be vectors of a *test* image. Let X_S be a $p \times N$ matrix, with its column being the training Coarse HPMs x_s .Then the object functions and their solutions, corresponding to different geometry properties of the HPM manifolds, are discussed as follows:

a) Under the locally 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:

$$\hat{\beta}^{kNN} = \underset{\beta}{\operatorname{argmin}} \left\| \mathrm{x}_t - \sum_{\mathrm{x}_s \in \mathcal{N}(\mathrm{x}_t)} \beta_s \mathrm{x}_s \right\|^2, \qquad (1)$$

where $\mathcal{N}$ 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.

b) Under the global linear assumption, each point on the manifold can be reconstructed by all *N* samples in the training set. The most popular estimation method of the reconstruction weights is least squares (**LS**) [25]. In the approach, the coefficients β can minimize the residual sum of squares

$$RSS(\beta) = \sum \left\| \mathrm{x}_t - \sum_{\mathrm{x}_s \in X_S} \beta_s \mathrm{x}_s \right\|^2 . \qquad (2)$$

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

$$\hat{\beta}^{ls} = \left({X_S}^T X_S \right)^{-1} {X_S}^T \mathrm{x}_t , \qquad (3)$$

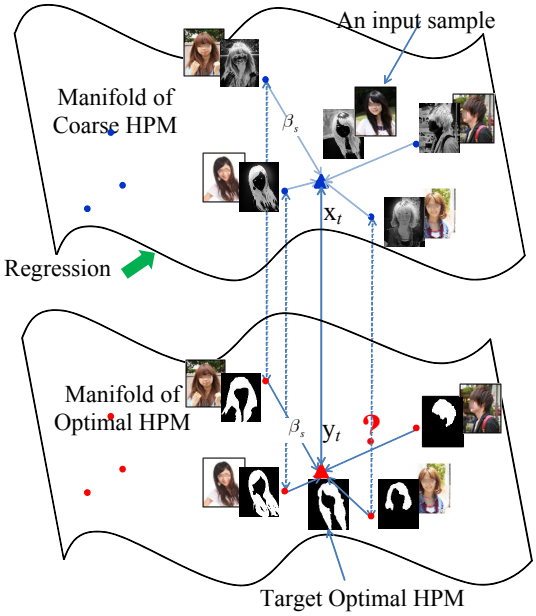

Figure 4: Illustration of our IMI method.

To achieve a bias-variance tradeoff, we add a regularization term, i.e.

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \left\| \mathrm{x}_t - \sum_{\mathrm{x}_s \in X_S} \beta_s \mathrm{x}_s \right\|^2 + \lambda_{ridge} \sum_{\mathrm{x}_s \in X_S} {\beta_s}^2 \right\}, \qquad (4)$$

which is called Ridge regression (RR) [28]. The corresponding solution is

$$\hat{\beta}^{ridge} = \left({X_S}^T X_S + \lambda_{ridge} I \right)^{-1} {X_S}^T \mathrm{x}_t . \qquad (5)$$

Likewise, if we replace the L2-norm regularization term in Eq.(4) with L1-norm, then the target function becomes the Lasso [29]:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left(\left\| \mathrm{x}_t - \sum_{\mathrm{x}_s \in X_S} \beta_s \mathrm{x}_s \right\|^2 + \lambda_{lasso} \sum_{\mathrm{x}_s \in X_S} |\beta_s| \right). \qquad (6)$$

An equivalent way to write the problem is

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left(\left\| \mathrm{x}_t - \sum_{\mathrm{x}_s \in X_S} \beta_s \mathrm{x}_s \right\|^2 \right), \quad \text{subject to} \sum_{\mathrm{x}_s \in X_S} |\beta_s| \le t. \qquad (7)$$

For small value of the bound *t*, Eq.(7) 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) *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 weights can be transferred to Optimal HPM manifold and explored to approximately infer the Optimal HPM:

$$\hat{y}_t = \sum_{x_s \in X_S} \hat{\beta}_s y_s. \quad (8)$$

Note that $\hat{y}_t$ is not a hard map, but a soft map with real values in the range of [0, 1]. We call $\hat{y}_t$ the Refined HPM, which is more accurate than the Coarse HPM x_t. That is because $\hat{y}_t$ is a linear combination of the ground segmentations, which can reduce some noise effects in the Coarse HPM. An example can be seen from Fig.3(e) and Fig.3(f), in which the Refined HPM suppress the noise of some confusing backgrounds.

Based on the Refined HPM obtained, one way to get the final segmentation is to set a threshold (**strategy (a)**). However, we exploit the Graph Cuts algorithm instead (**strategy (b)**), in order to make the final segmentation more smooth and make the fore/background regions more continuous. These two different strategies will be compared in Sec.III.D.

III. EXPERIMENTS

A. Database and measurement

1) Database

We collect two challenging databases from Internet with no overlapping with each other. One is for performance evaluation while the other is used to train hair color and location models (GHCM and HOPP). The evaluation database includes 3800 images, with a large range of variations in head-shoulder poses, hair styles, hair colors, illustration and backgrounds. Among these, there are 2900 near-frontal images and 900 non-frontal ones. The training set contains 1000 head-shoulder images. Since non-frontal images are really not easy to collect, the training images are near-frontal or with little pose variations.

In the experiments, all head-shoulder-images are aligned to the size of 80x100 pixels according to head rectangle position and the ground truth of hair regions are labeled manually. No other preprocessing is conducted. Some examples of normalized images have been shown in Fig.1 and Fig.2.

2) Measurement

Performances are evaluated by assessing the consistency of automatic segmentation and the manually labeled ground truth in terms of 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.

B. Experimental set-up

To investigate the performance of the proposed method, two experiments are conducted: 1) Sec.III.C compares the performances of different regression techniques for IMI; 2) Sec.III.D compares our method with other existing ones.

Specifically, we randomly select 2000 images as the basis (training set) and the rest 1800 are for evaluation. Then

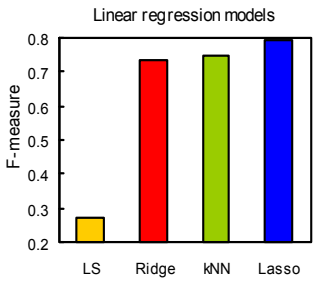

Figure 5: Performance comparison of different regression models.

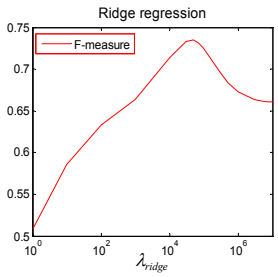

Figure 6: F-measure of Ridge varying with λ_{ridge}.

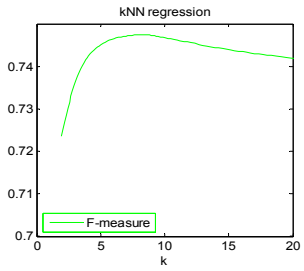

Figure 7: F-measure of kNN regression varying with *k*.

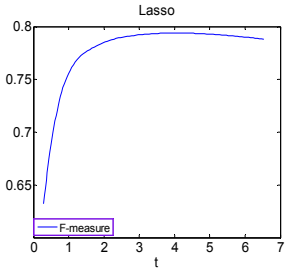

Figure 8: F-measure of Lasso varying with *t*.

according to results in experiment 1), we choose an appropriate regression technique for IMI and compare it with other existing segmentation method in experiment 2).

C. 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: kNN regression, least square, Ridge regression and the Lasso. Specifically, the LARS 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.(8)) is cut by threshold 0.5 to produce the segmentation.

Fig.5 shows the results, 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 lease square, since it can trade a little bias on the Corse HPM for a large reduction of the error influences. Nevertheless, neither of the two solutions is satisfactory. The reason maybe is that the globally linear assumption on the data points on manifolds seems not so 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

Table 1: Performance comparison of different methods

Algorithm	F- score	Precision/Recall
Bayesian method [22]	**0.619**	**0.612/0.714**
Graph Cut	**0.683**	**0.705/0.776**
Automatic version of MSRM [30]	**0.702**	**0.811/0.700**
***Proposed* with strategy (a)**	**0.794**	**0.830/0.789**
***Proposed* with strategy (b)**	**0.804**	**0.883/0.764**

applies the sparse constraints and can adaptively cause some coefficients to be exactly zero. Note that in Fig.5 the Lasso performs better than kNN regression.

Additionally, Fig.9 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.10 gives the corresponding samples with the largest Lasso coefficients (plotted in Fig.9). Note that most of these hair styles appear similar to the style of the input image. This indicates that Lasso provides a compact representation of the HPM, while rejecting error effects from many other samples.

As for the effects of the parameters, Fig.6 gives the F-measure with the varying λ_{ridge} in Eq.(4) for Ridge regression. Higher λ_{ridge} means more shrinking of coefficients, which also means the higher variances of the data distribution. Moreover, Fig.7 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 plots the F-measure as the t changes when using the Lasso. 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.

D. Comparison with other methods

In this section, we compare our method with latest methods: the Baysian method in [22], Graph Cut and the maximal similarity based region merging (MSRM) method [30].

According to Sec.III.C, 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.(7), is set to be 4.0, according to cross validation. After obtaining a Refined HPM, the two strategies (a) and (b) are evaluated respectively. We set the threshold T of strategy (a) to be 0.5. Bayesian method, including the paramter settings, is implemented exactly as the literature [22] presents. For Graph Cut, it is essentially similar to the method in [5] 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 Sec.II.A. Based

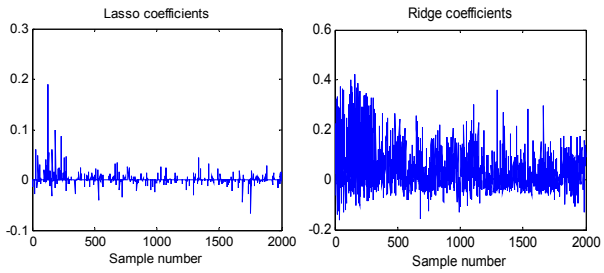

Figure 9: Lasso and Ridge coefficients for the image in Fig.10.

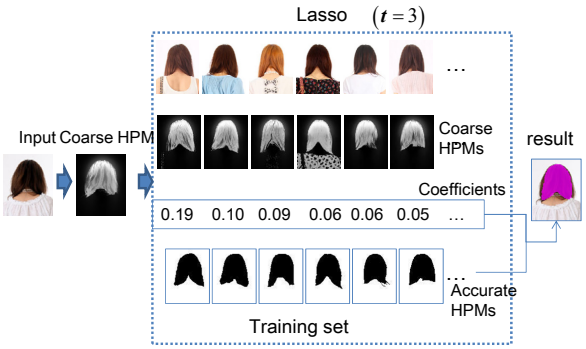

Figure 10: Examples of the images with the largest coefficients in Lasso. The decimals in the figure are the Lasso coefficients of the corresponding images in the same column. The right side is the labeled result (see color image) resulted from the weighted linear combination of the training segmentation ground truth.

on these configurations, Table 1 tabulates performance comparisons of different methods. It can be easily seen that both of the two strategies of the proposed method achieve superior performances to the other methods.

Some examples of segmentation results for different methods are shown in Fig.11. For our methods, we show the results with strategy (b). It can be seen that other than the bottom row in Fig.11, the rest are successful segmentation examples. The hair segmentation result can often be prevented from being influenced by the similar clothes (Row 1, 2, 4, 5,) and other confusing backgrounds (Row 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 supress the noise in the Coarse HPM, just like a 'noise filter'. Therefore, the resulte will not be largely beyond normal hair regions. The Bayesian method and Graph Cuts method often fails because of inproper parameters for the images, while the merging strategy is not always effective due to unreasonalbe fore/background models.

IV. 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 calculated 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. Generalizing the method to other specific class is our future work.

ACKNOWLEDGMENT

This paper is partially supported by National Basic Research Program of China (973 Program) under contract 2009CB320902, and Natural Science Foundation of China under contracts No. 61025010 and No. 61222211.

REFERENCES

[1] Y. Yacoob and L. S. Davis, "Detection and analysis of Hair," T-PAMI, 28(7):1164-1169, 2006.
[2] K. Ueki, H. Komatsu , S. Imaizumi, K. Kaneko, N. Sekine, J. Katto, and T. Kobayashi, "A method of gender classification by integrating facial, hairstyle, and clothing images," In ICIP, pp. 446-449, 2004.
[3] Z. Zhang, H. Gunes, and M. Piccardi, "An accurate algorithm for head detection based on XYZ and HSV hair and skin color models," In ICIP, pp. 1644-1647, 2008.
[4] N. Wang, H. Ai and S. Lao, "A Compositional Exemplar-based Model for Hair Segmentation," In ACCV, pp.171-184, 2010.
[5] K.C. Lee, D. Anguelov, B. Sumengen, and B. Gokturk, "Markov random field models for hair and face segmentation," In FG, pp. 1-6, 2008.
[6] P. Julian, C. Dehais, F. Lauze, V. Charvillat, A. Bartoli, and A. Choukroun, "Automatic hair detection in the wild," In ICPR, pp. 4617-4620, 2010.
[7] F. Bertails, B. Audoly, B. Querleux, F. Leroy, J. L. Leveque, and M. P. Cani, "Predicting natural hair shapes by solving the statics of flexible rods," Eurographics, 2005.
[8] F. Bertails, B. Audoly, M. P. Cani, B. Querleux, F. Leroy, and J. L, "Lévêque. Super-helices for predicting the dynamics of natural hair," SIGGRAPH, pp.1180-1187, 2006.
[9] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for Image Understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," IJCV, 81(1):2-23, 2009.
[10] J. Winn and N. Jojic, "LOCUS: Learning object classes with unsupervised segmentation," In ICCV, pp. 756-763, 2005.
[11] E. Borenstein and S. Ullman, "Class-specific, top-down segmentation," In ECCV, pp. 639-641, 2002.
[12] E. Borenstein and S. Ullman, "Learning to segment," In ECCV, pp. 315-328, 2004.
[13] X. He, R. S. Zemel, and D. Ray, "Learning and incorporating top-down cues in image segmentation," In ECCV, pp. 338-351, 2006.
[14] A. Levinand and Y. Weiss, "Learning to combine bottom-up and top-down segmentation," IJCV, 81(1):105-118, 2009.
[15] B. Alexe, T. Deselaers, and V. Ferrari, "ClassCut for unsupervised class segmentation," In ECCV, pp. 380-393, 2010.
[16] O. Veksler, "Star shape prior for Graph-Cut image segmentation" In ECCV, pp. 454-467, 2008.
[17] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models- their training and application," CVIU, 61(1): 38–59, 1995.
[18] T. F. Cootes, G. J. Edwards, and C. J. TAYLOR, "Active appearance models," T-PAMI, 23(6):681–685, 2001.
[19] Z. Liu, J. Guo, and L. Bruton, "A knowledge-based system for hair region segmentation," In ISSPA, vol. 2, pp. 575-576, 1996.

Figure 11: Results comparison. (a) Original image; (b) Bayesian method; (c) Graph Cuts (d) MSRM; (e) Proposed method.

[20] M. Kampmann, "Segmentation of a head into face, ears, neck and hair for knowledge-based analysis-synthesis coding of videophone sequences," In ICIP, 1998.
[21] C. Rousset and P.Y. Coulon, "Frequential and color analysis for hair mask segmentation," In ICIP, pp. 2276-2279, 2008.
[22] D. Wang, S. Shan, W. Zeng, H Zhang, and X. Chen, "A novel two-tier Bayesian based method for hair segmentation," In ICIP, pp. 2401-2404, 2009.
[23] T. F. Chan and L. A. Vese, "Active contours without edges," TIP, 10(2):266–277, 2001.
[24] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," T-PAMI, 23(11):1222-1239, 2001.
[25] T. Hastie and R.Tibshirani and J. Friedman. The elements of statistical learning. 2nd edition, pp. 43-93, 2008.
[26] S. Dambreville, Y. Rathi, and A. Tannenbaum, "Shape-based approach to robust image segmentation using kernel PCA," In CVPR, pp. 977 – 984, 2006.
[27] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," T-PAMI, 28(7):603–619, 2002.
[28] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," Technometrics 12(1):55-67, 1970.
[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. Royal. Statist. Soc B., 58(1): 267-288, 1996.
[30] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging," Pattern Recognition, 43(2):445-456, 2010.