33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

VirtualShave: automated hair removal from digital dermatoscopic images

M. Fiorese, E. Peserico, A. Silletti – Univ. Padova, Italy – melanoma@dei.unipd.it

Abstract— **VirtualShave is a novel tool to remove hair from digital dermatoscopic images. First, individual hairs are identified using a top-hat filter followed by morphological postprocessing. Then, they are replaced through PDE-based inpainting with an estimate of the underlying occluded skin. VirtualShave's performance is comparable to that of a human operator removing hair manually, and the resulting images are almost indistinguishable from those of hair-free skin.**

I. Introduction

The presence of hair represents a common obstacle in the dermatoscopic analysis of small skin lesions such as melanocytic naevi, both in the case of visual inspection by a dermatologist through a dermatoscope and in the case of automated analysis of a digital image. This is true even when hair is relatively sparse and occludes only a small fraction of the lesion's surface. In the case of visual inspection sparse hair disrupts the evaluation "at a glance" of the lesion's texture patterns. In the case of automated systems it can drastically alter the measurements of the lesion's size, shape, colour, and texture, completely ruining the analysis [9].

Shaving is an obvious solution – but it has drawbacks. First, it is time consuming, often far more than the lesion's visual inspection and/or digital image acquisition. Second, shaving tends to irritate the skin, creating a diffuse alteration of the colour of the *whole* lesion which may disrupt its analysis to an even greater extent than the initial hair.

An alternative solution that can be more effective when operating on digital images with sparse hair (whether the image is to be evaluated by a dermatologist or by an automated system) is to replace the hair pixels with a reasonable estimate of the underlying skin - an operation named *inpainting*. The two main tasks are then a) hair segmentation – i.e. classification of every pixel of the image as either hair or skin and b) inpainting – i.e. replacement of hair pixels with an estimate the underlying occluded skin. A crucial, but often disregarded, point is that what constitutes a "good" inpainting strongly depends on the application: if the image is to be examined by a human operator the ideal choice is usually an attempt to match the colour and texture of the neighbouring non-occluded skin, while sophisticated automated analysis tools may fare better if occluded pixels are correctly flagged as "unknown". This paper presents VirtualShave, an automated hair detection and removal tool, and an evaluation of its effectiveness.

Supported in part by P.A.T. and I.N.F.N. under proj. AuroraScience, by M.I.U.R. under P.R.I.N. "AlgoDeep", and by Univ. Padova under strategic proj. A.A.C.S.E. and under proj. "Naevi in silico".

II. Related Work

A number of hair-removal tools to process digital images can be found in the literature. DullRazor [9] uses a top-hat operator [19] to identify dark hair structures. Hair is then replaced with non hairy pixels. [16] performs a top-hat to retrieve the hair from grayscale images and morphological operations to remove false positives. Image inpainting is based on Perona's and Malik's diffusion PDE [16]. [20] instead uses Schmid's algorithm [12] to detect hair and then a RANSAC-based curve-fitting procedure [7] to refine the results and join the broken hair segments. Inpainting is example-based [4]. [1] recently (at the time of this writing the paper is still in press) proposed a Derivative of Gaussian (DoG) to detect hair, combining it with the fast marching image inpainting technique proposed by [3].

One serious deficiency most of these papers share is the lack of a rigorous quantitative evaluation of the results. As mentioned in Section I, this involves evaluating the effectiveness in both hair segmentation and inpainting. Segmentation is typically (not only in this context) evaluated in terms of the number of misclassified pixels compared to a "ground truth" provided by a human operator. The only two papers providing a quantitative evaluation of segmentation are [16] and [1]. The first normalises this quantity dividing it by the number of pixels classified as hair by the ground truth (the assessment is not entirely positive). The second instead normalises dividing by the number of pixels classified as hair either by the ground truth or the hair-removal software under test. As we shall see in Section IV in the case of hair this methodology suffers from a serious deficiency. [1] is the only paper providing a quantitative evaluation of inpainting quality – a somewhat artificial measure based on the inpainting's distortion of several image parameters such as contrast and entropy.

III. VirtualShave

VirtualShave segmentation module (Subsection III-A) first applies a top-hat operator to identify areas of high contrast as hair candidates; then filters out through a decision tree those candidates that do not satisfy certain morphological conditions; and finally attempts to remove "breaks" between multiple hair segments where hair almost match the colour of the surrounding skin and thus went unrecognized by the top-hat operator. VirtualShave inpainting module (Subsection III-B) is PDE-based and aims to maximize the quality of visual inspection (see Section I above); alternatively it has the option to mark all hair pixels with the same, "unknown"-flagged colour. Figure 1 depicts the overall procedure.

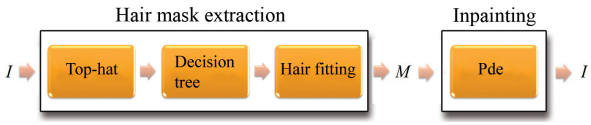

Fig. 1. VirtualShave segmentation module applies a top-hat operator, filters candidates through a morphology decision tree, and finally attempts to repair gaps in hair. It then applies PDE based inpainting.

A. Segmentation

a) Top-hat: The top-hat operator detects contrasted objects on non-uniform backgrounds [14]. Closure-based top-hat subtracts from the original image its closure [19]:

$$T = |I_{bw} - \{I_{bw} \cdot E\}| \tag{1}$$

where I_{bw} is a grayscale image, E is the structuring element and $\cdot$ denotes the closure operation.

VirtualShave utilizes a modified closure-based top-hat operator, where 8 structuring elements produce an equivalent number of closed images which are then merged in a unique image. The structuring elements (fig. 2) are 13 pixel-long straight lines. Each line has a direction (0°, 45°, 90°, 135°) and a thickness (1 or 2 pixels).

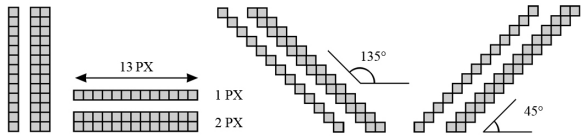

Fig. 2. 8 structuring elements for closure operation

T_r, the closure-based top-hat image of the red channel is computed as:

$$T_r = \left| I_r - \max_{\substack{d \in \{0°, 45°, 90°, 135°\} \\ t \in \{1, 2\}}} (I_r \cdot E_{d,t}) \right| \tag{2}$$

where I_r and $E_{d,t}$ are respectively the red channel of the image I and the structuring element with direction d and thickness t. A binary image M_r is then obtained by thresholding:

$$M_r(x, y) = \begin{cases} 1 & \text{if} \quad T_r(x, y) < th \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where th is the Otzu threshold, minimizing the histogram intra-class variance between bright and dark pixels [10],[17] (in a nutshell, we optimally "split" the bi-modal histogram of the grayscale image T_r into two classes). Equations 2 and 3 can be applied to the green and blue channels as well, obtaining respectively M_g and M_b. The final hair mask M (fig. 3) is the logical intersection of the 3 masks:

$$M = M_r \cap M_g \cap M_b \tag{4}$$

Unfortunately, top-hat relies only on *local* contrast [14] and thus often misclassifies pixels. Dark areas of skin can be classified as hair, while thin long hairs can be sometimes

Fig. 3. Raw mask M as computed in 3 and 4 by thresholding the modified top-hat operator. Small spurious hairs are detected, while long hairs are heavily fragmented.

classified as skin when overlapping e.g. a melanocytic lesion. The next two steps attempt to correct this problem.

b) Decision tree – false positive removal: A connected region $\mathcal{R}$ of M is a set of connected hair pixels. Ideally, a region $\mathcal{R}$ would represent a single hair or perhaps a few, partially overlapping hairs. Denoting for each region by A and P its area and perimeter, and by Ac and Pc those of its convex hull, we define the *Density*, *Sphericity* and *Convex Hull Sphericity* of the region respectively as:

$$D = \frac{A}{A_c} \qquad S = \frac{A}{4\pi P^2} \qquad S_c = \frac{A_c}{4\pi P_c^2} \tag{5}$$

D measures the "fullness" of the region, S and S_c its roundness. Informally, hair regions will be long and thin and thus will sport recognizable values of D, S and S_c.

Fig. 4. Region on the left has a high D value since $A \approx A_c$. The region on the right, on the opposite, has a very low D value since $A \leq A_c$. Long and curved hairs such as the one on the right are thus easily identified.

Each region traverses the decision tree depicted in figure 5. The tree rejects regions whose shape is geometrically incompatible with hair, and removes them from M.

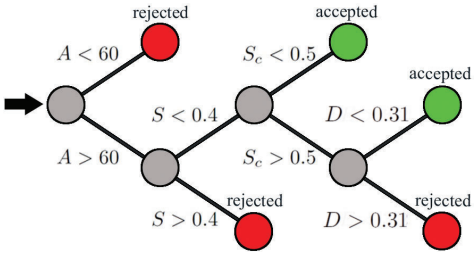

Fig. 5. The Decision Tree imposes geometrical coherence on hair regions. The tree is an effective yet easy to implement tool to remove false negative regions such as dots, superficial blood vessels and small pigmented areas. The order of the nodes is important. For instance, a decision on D is not reliable in the first stages since it could reject long and straight hair, which has a low D value. The thresholds are determined empirically.

c) Hair fitting – false negative removal: Thin and light hair often results in a number of small regions, rather than a long connected one. Figure 6 shows a frequent scenario, where a long hair has been fragmented into 5 pieces.

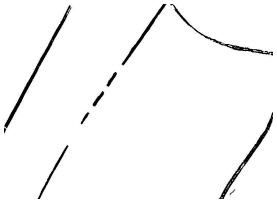

Fig. 6. Dealing with thin bright hairs is difficult since their local contrast is low and they can be confused with skin ripples.

For each region $\mathcal{R}$, we compute the set of endpoints $\{p_1, p_2...p_n\}$. For each endpoint i, we then compute the ρ_i and θ_i parameters of the approximation line (see the Appendix). Each region is then described as:

$$\mathcal{R} : \{\{p_1, p_2...p_n\}, \{\rho_1, \rho_2...\rho_n\}, \{\theta_1, \theta_2...\theta_n\}\} \quad (6)$$

For each pair of endpoints of distinct regions $\mathcal{R}^A$ and $\mathcal{R}^B$, we compute a "compatibility" value defined as:

$$C_{i,j}^{A,B} = k_d \cdot \mathcal{N}(0, \sigma_d)(||p_i^A - p_j^B||) + k_\theta \cdot \mathcal{N}(0, \sigma_\theta)(||\theta_i^A - \theta_j^B||) + k_\rho \cdot (0, \sigma_\rho)(||\rho_i^A - \rho_j^B||) \quad (7)$$

where k_x, $x \in \{d, \theta, \rho\}$, is the Gaussian normalization factor $1/\sqrt{2\pi\sigma_x^2}$; p_i^A and p_j^B are respectively the i^{th} endpoint of $\mathcal{R}^A$ and the j^{th} endpoint of $\mathcal{R}^B$; and ρ_i^A, ρ_j^B, θ_i^A, θ_j^B are respectively the i^{th} and j^{th} ρ, θ parameters of $\mathcal{R}^A$ and $\mathcal{R}^B$.

A high value of $C_{i,j}^{A,B}$ means that $\mathcal{R}^A$ and $\mathcal{R}^B$ are likely to be connected through p_i^A and p_j^B, and thus if $C_{i,j}^{A,B} > (\frac{k_d+k_\theta+k_\rho}{3})$ M is adjusted by interpolating the connecting region. In our experiments, $\sigma_d = 40$, $\sigma_\theta = 15$, $\sigma_\rho = 20$.

B. Inpainting

PDE-based inpainting [2] fills-in those hair occluded regions extracted in M. The gray-scale value $u(x)$ at position x is interpolated according to Perona's and Malik's non-linear diffusion equation [11], [15]:

$$\frac{\partial u(x,t)}{\partial t} = div(c(x,t)\nabla u(x,t)) \quad (8)$$

where div and ∇ are the divergence and gradient operators, and $c(x)$ is the scalar diffusivity at x:

$$c(x,t) = \frac{1}{1 + (\|\nabla u(x,t)\|/K)^2} \quad (9)$$

with K set empirically to 2.5. [16] provides a compact, discrete iterative approximation for $u(x,t)$ as:

$$u(x,t+1) = u(x,t) + \frac{\lambda}{|G(x)|} \sum_{x' \in G(x)} c(x',t)\Delta u(x',t) \quad (10)$$

where $G(x)$ is the set of pixel positions neighbouring x, λ is a scalar empirically set to 0.75, and Δ denotes the Laplacian operator. $u(x,0)$ is simply the original pixel value, and we stop after 50 iterations.

IV. Experimental evaluation

We tested VirtualShave on a set of 20 images (768 by 576 pixels, 24 bit rgb colour) acquired using Fotofinder [8] at 20X optical zoom. Each image depicts a melanocytic lesion surrounded by healthy skin, in the presence of sparse hair. VirtualShave required less than 3s per image on a cheap PC.

For the segmentation module, "ground truth" hair was provided by a human operator who manually marked pixels of the image as hair. VirtualShave misclassified pixels represented a total of 15.6% of all pixels classified as hair by either VirtualShave or the ground truth. In contrast, a simple implementation of DullRazor [9] misclassified more than 3 times as many pixels (47.1%).

These error values may appear high, considering the good "apparent" quality of the inpainted images. In fact, they were considerably higher than the numbers reported in [1] (even those for the same DullRazor algorithm). However, one must take into account that in images at this resolution a hair is only a few pixels wide, and thus exhibits a very large ratio of frontier pixels (hair pixels adjacent to non-hair ones) to internal pixels (hair pixels surrounded by other hair pixels). Thus, segmenting hair just a single pixel thicker or thinner than the ground truth can produce a considerable "error" even if the hair appears removed perfectly well.

To address this issue, we adopted the same method used in [13] to assess the lowest error achievable in a particular segmentation task: we had a *second* human operator segment hair on each image, and evaluated his error against the ground truth provided by the first operator. The second human operator actually misclassified slightly *more* pixels than VirtualShave. Thus, the error of VirtualShave's segmentation module is below the noise threshold introduced by the particular morphology of hair – a high threshold that casts doubts on the applicability of this standard evaluation methodology when applied to the segmentation of hair.

Thus, we resorted to an alternative methodology, which simultaneously allowed us to evaluate the quality of the inpainting. We used a second set of 20 images of melanocytic lesions, obtained through the same equipment, with the same resolution, zoom etc. but lacking visible hair. Then, we had six human operators, three of them dermatologists, identify each as the 40 images as "naturally hairless" or "shaved virtually". The three dermatologists misclassified 13, 14 and 19 images. The remaining three operators misclassified 13, 17 and 22 images. Thus, removing hair through VirtualShave produces images almost indistinguishable to the human eye from unprocessed images of skin without visible hair.

V. Conclusions

VirtualShave is a fast, accurate tool to detect and remove hair from medium resolution dermatoscopic images of melanocytic lesions. The best validation of its effectiveness is that the images it produces are almost indistinguishable, even by trained dermatologists, from those of naturally hairless skin. It would be interesting to test VirtualShave on larger collections of images – including very high resolution images and images of other types of lesions.

Fig. 7. The 20 test lesions, before and after hair removal.

REFERENCES

[1] Q. Abbas, M.E. Celebi, and I.F. García. Hair removal methods: A comparative study for dermoscopy images. *Biom. Signal Proc. and Control*, In Press., 2011.

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. *Proc. of SIGGRAPH Conf.*, pages 417–424, 2000.

[3] F. Bornemann and T. Mrz. Fast image inpainting based on coherence transport. *J. of Math. Imaging and Vision*, 28:259–278, 2007.

[4] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. *Proc. of IEEE CVPR Conf.*, 2:721–728, 2003.

[5] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. of the ACM*, 15:11–15, 1972.

[6] K.C. Fan, D.F. Chen, and M.G. Wen. Skeletonization of binary images with nonuniform width via block decomposition and contour vector matching. *Pattern Recognition*, 31(7):823–838, 1998.

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981.

[8] Inc. FotoFinder Systems. FotoFinder dermoscope. `http://www.fotofinder.de/en.html`.

[9] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean. Dullrazor: A software approach to hair removal from images. *Comp. Biol. Med.*, 27(6):533–43, 1997.

[10] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man and Cyb.*, 9(1):62–66, 1979.

[11] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Patt. Analysis and Machine Intell.*, 12:629–639, 1990.

[12] P. Schmid-Saugeon, J. Guillod, and J.P. Thiran. Towards a computer-aided diagnosis system for pigmented skin lesions. *Comp. Med. Imaging and Graphics*, 27(1):65–78, 2003.

[13] A. Silletti, E. Peserico, A. Mantovan, E. Zattra, A. Peserico, and A. Belloni Fortina. Variability in human and automatic segmentation of melanocytic lesions. *Proc. of IEEE EMBS Conf.*, 1(1):5789–92, 2009.

[14] I. R. Terol-Villalobos. Morphological connected contrast mappings based on top-hat criteria: A multiscale contrast approach. *Proc. of SPIE Conf.*, 43(7):1577–1595, 2004.

[15] G.W. Wei. Generalized perona-malik equation for image restoration. *IEEE Signal Proc. Lett.*, 6:165–167, 1999.

[16] F.Y. Xie, S.Y. Qin, Z.G. Jiang, and R.S. Meng. PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma. *Comp. Med. Imaging and Graphics*, 33(4):275–82, 2009.

[17] F. Yue, W.M. Zuo, and K.Q. Wang. Decomposition based two-dimensional threshold algorithm for gray images. *Acta Automatica Sinica*, 35:1022–1027, 2009.

[18] J.X. Zeng, G.M. Zhang, J. Chu, and Y.M. Lu. The application of hough transform in the detection of exponent function curve. *J. of Image Graph.*, 10(2):236–40, 2005.

[19] M. Zeng and J.X. Li. Optimized design of morphological improved top-hat filter based on improved genetic algorithms. *Acta Optica Sinica*, 26(4):510–5, 2006.

[20] H. Zhou, M. Chen, R. Gass, J.M. Rehg, L. Ferris, J. Ho, and L. Drogowski. Feature-preserving artifact removal from dermoscopy images. *Proc. of SPIE Conf.*, 6914, 2008.

APPENDIX – COMPUTING ENDPOINTS, θ AND ρ

Let $\mathcal{R}_s$ be the *skeleton* of a generic region $\mathcal{R}$ [6]. A pixel $p \in \mathcal{R}_s$ is a *candidate endpoint* if its eight-pixel-neighbourhood comprises at most one other pixel in $\mathcal{R}_s$ – otherwise p is an *internal* point. The set of candidate endpoints is then refined. For each candidate, consider all pixels in $\mathcal{R}_s$ within a distance of 15 pixels; if each of these pixels has at most two pixels belonging to $\mathcal{R}_s$ in its eight-pixel-neighbourhood, the candidate is an *endpoint*. The procedure finally returns all the pixels marked as endpoints.

Given a region $\mathcal{R}$ and an endpoint $p \in \mathcal{R}$, $\mathcal{R}_p \subseteq \mathcal{R}$ is the set of points whose distance from p is less than 20 pixels. In a nutshell, $\mathcal{R}_p$ is the final "portion" of $\mathcal{R}$. Let $H_{\mathcal{R}_p}(\theta, \rho)$ be the Hough transform [18] [5] of $\mathcal{R}_p$, and let Θ and P be the sets of the local maxima of $H_{\mathcal{R}_p}$:

$$(\Theta, P) = \arg\max_{\theta,\rho} \left\{ H_{\mathcal{R}_p} \right\} \tag{11}$$

$\widehat{\theta}$, $\widehat{\rho}$ are the average values of those local maxima:

$$\widehat{\theta} = \frac{1}{|\Theta|} \sum_{\theta \in \Theta} \theta \qquad \widehat{\rho} = \frac{1}{|P|} \sum_{\rho \in P} \rho \tag{12}$$

$\widehat{\theta}$ and $\widehat{\rho}$ are the estimated parameters of the line passing through p. The procedure returns $\widehat{\theta}$ and $\widehat{\rho}$.