A reliable skin mole localization scheme

Taeg Sang Cho, William T Freeman
CSAIL
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

taegsang@mit.edu, billf@mit.edu

Hensin Tsao
Dept. of Dermatology
Harvard Medical School
Wellman Center for Photomedicine

htsao@partners.org

Abstract

Mole pattern changes are important cues in detecting melanoma at an early stage. As a first step to automatically register mole pattern changes from skin images, this paper presents a framework to detect and label moles on skin images in the presence of clutter, occlusions, and varying imaging conditions. The input image is processed with cascaded blocks to successively discard non-mole pixels. Our method first searches the entire input image for skin regions using a non-parametric skin detection scheme, and the detected skin regions are further processed using a difference of Gaussian (DoG) filter to find possible mole candidates of varying sizes. Mole candidates are classified as moles in the final stage using a trained support vector machine. To increase the mole classification accuracy, hair is removed if present on the skin image using steerable filters and a graphical model. The performance of the designed system is evaluated with 28 test images, and the experimental results demonstrate the effectiveness of the proposed mole localization scheme.

1. Introduction

Mole pattern changes are important cues in detecting early signs of melanoma, a deadly skin cancer [4]. Early detection is especially important for melanoma because, while advanced cases are not curable, the disease can be cured if detected early [16]. However, a principled system to register mole pattern changes is currently lacking. In fact, a major burden on the dermatological workforce is in manual surveillance of pigmented lesions, which is both time-consuming and prone to human error. In this paper, a computer vision-based mole localization system is introduced, which can potentially be valuable in registering mole patterns automatically.

While image processing techniques are extensively used in classifying moles as either malignant or benign [4, 19],
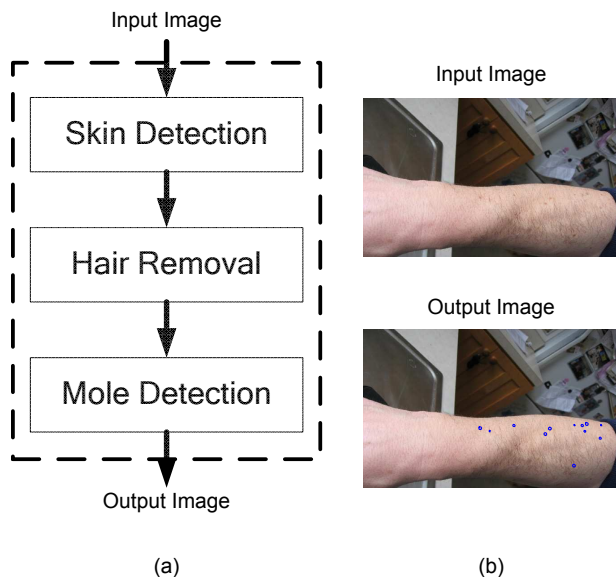

Figure 1. (a) A skin region is detected from the input image to reduce the computation and increase the reliability of the system. If any hair is present in the detected skin region, steerable filters are used to eliminate hair patterns. The hair removed image is further processed with the mole detection block to localize the moles. (b) Input - output image pair of the proposed system.

localizing moles in a larger skin image has received less attention. In Lee *et al.* [9], moles from back torso images, taken under constrained imaging conditions, are localized using meanshift clustering and heuristic classification schemes. A related work by Pierrard and Vetter [12] is aimed at detecting skin irregularities, such as moles, for face recognition. Pierrard and Vetter first detected possible mole candidates by using a Laplacian of Gaussian filter, and classified mole candidates as moles by using a normalized cross correlation and a saliency measure. While their approach is similar to ours in using scale-space filters to locate mole candidates, they register only salient skin irregularities that are isolated from other irregularities. This aspect differs
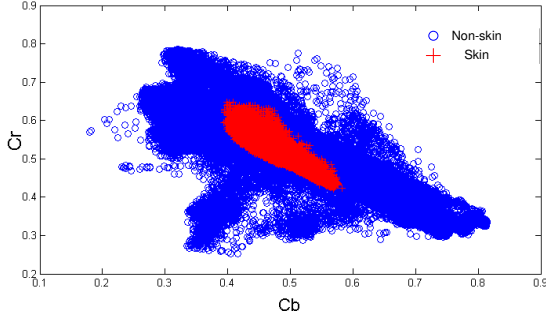
Figure 2. Skin color can be compactly localized in YCbCr color space [6]. This figure shows the skin and non-skin color distribution in CbCr space; the distributions serve to discriminate skin and non-skin pixels.

from our goal in that we are interested in locating all moles present in the image, not only the salient ones.

Our work goes beyond Lee *et al*. to detect moles in skin images other than that of back torso under varying imaging conditions. A rich descriptor is used to classify moles using a support vector machine. Section 2 describes the developed system, and the purpose of each block. Each block is discussed more in detail with implementation issues in three subsequent sections, and Section 6 will quantify the performance of the whole system.

2. Proposed system

Figure 1 (a) shows a block diagram of the designed system. The system is comprised of three stages: a skin detection stage, a hair removal stage, and a mole detection stage. Since the input image can contain objects other than the skin surface, skin is detected to focus further image processing only to skin regions. If the detected skin region has hair patterns (user-specified), a hair removal scheme is used to eliminate the hair patterns. The skin detection and hair removal steps reduce the computation in later stages and improve the mole detection performance.

Moles are modeled as circular regions of lower intensity than the surround: scale-space difference-of-Gaussian (DoG) filters [11] are used to locate possible mole candidates of varying sizes. Detected mole candidates are further classified as moles and non-moles using a trained support vector machine (SVM). Figure 1 (b) shows a typical input-output pair of the skin image processed by the proposed system.

3. Non-parametric skin region detection

Many skin region detection schemes have been proposed in literature [6, 7]; Vezhnevets *et al*. [17] provides an excellent review on methods using skin color as cues. An exact segmentation of skin using color as an only source of in-

formation is a hard problem in the presence of changing lighting conditions, color differences between people, etc. However, an exact segmentation of skin is not needed for our application: subsequent image processing stages can be used to discard data from non-skin regions. Thus, our goal is to classify skin regions, erring on the side of classifying ambiguous regions as skin. Once the skin regions are detected, the subsequent stages focus only on the detected skin region represented with the skin mask (Figure 3.)

Among many proposed skin detection methods, a non-parametric approach is used since it's hard to find an adequate parametric representation of skin color data. YCbCr color space is used to represent skin colors since it's known that CbCr space localizes skin colors [6]. Figure 2 shows the distribution of skin color in our dataset, along with non-skin colors: skin color is confined in an elliptical distribution. Note that it's important to have a representative skin color dataset: we collected skin images from the web to capture different skin colors under varying lighting conditions. This tends to broaden the skin color distribution (Figure 2) compared with that in [6].

The skin pixel classification is carried out using a Neyman-Pearson hypothesis test: for a given pixel value x and a threshold λ, classify the pixel as skin if the ratio of probability being skin to probability being non-skin is greater than λ. In equation,

$$\frac{P_{X/H_1}(x/H_1)}{P_{X/H_0}(x/H_0)} \geq \lambda \qquad (1)$$

where H_1 is a hypothesis that the pixel is skin and H_0 is a hypothesis that the pixel is non-skin. Note that $P_{X/H_1}(x/H_1)$ and $P_{X/H_0}(x/H_0)$ are the color distributions from the dataset. λ can be used to vary the operating point on the receiver operating characteristics (ROC) curve, and in this implementation, λ is fixed to be 1. Artifacts of using only the pixel values, and not incorporating connectivity information among the pixels, are salt-and-pepper type non-skin islands on inferred skin regions. These are removed by applying median filters on the detected skin regions, which results in a smoother skin mask. Figure 3 shows some skin detection results: note that false alarms exist.

4. Hair removal using steerable filters and a graphical model

Hair can hinder a reliable mole detection, but we cannot ask users to shave hair before using the mole localization system. In this section, an image processing technique to remove hair patterns on skin images is described. There have been a number of successful attempts to remove hair patterns using a morphological erosion operator [10, 13], but these algorithms require that the hair patterns are in-focus,
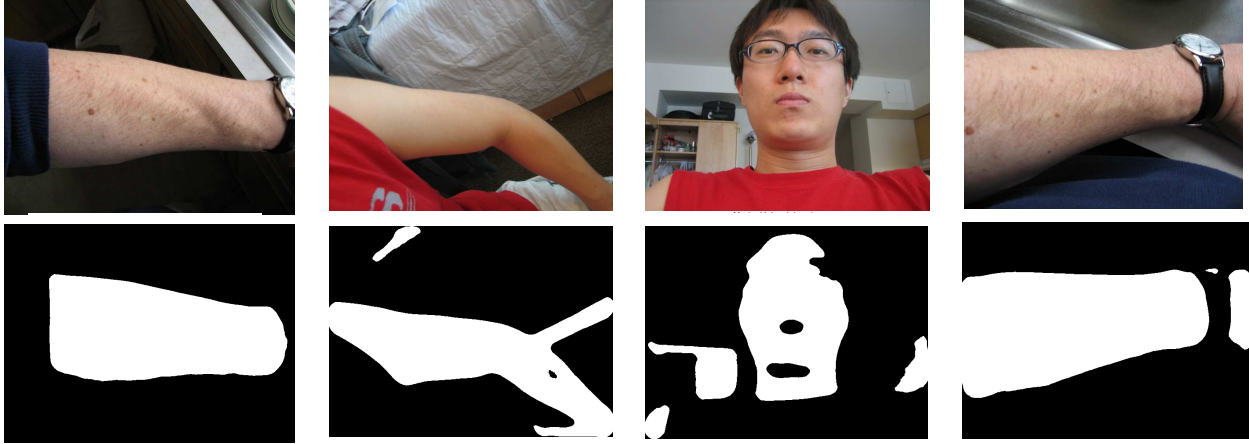
Figure 3. The skin region is detected using the skin color density distribution in Figure 2. Note that skin is not perfectly segmented, but this can be taken care of by the subsequent image processing stages.
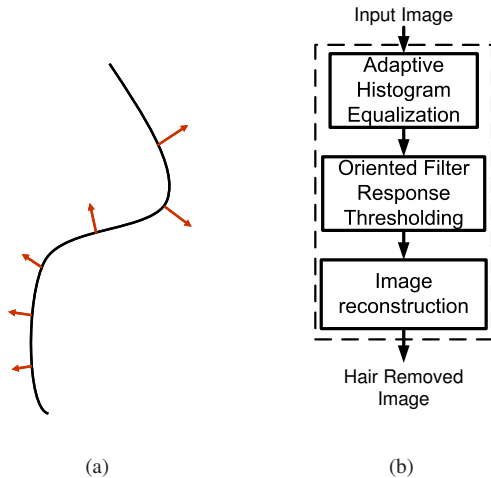

(a)   (b)

Figure 4. (a) The main idea behind the proposed hair removal scheme. Hair structures have a high derivative response at a direction normal to the hair orientation. Therefore, oriented filters can be used to extract out long hair-like structures. (b) The designed hair removal system. Hair is removed by thresholding the maximum magnitude of oriented derivative filters, and the skin image is reconstructed using a graphical model.

and the overlap among hair threads is minimal. Such constraints are not suited for our application since input images will be of larger skin regions with many hair threads overlapping. The proposed hair removal scheme solves this problem by using simple oriented filters.

The basic observation behind the proposed hair removal scheme is that hair patterns have high frequency components normal to the orientation (Figure 4 (a).) Since hair patterns can be oriented arbitrarily, oriented filters are used

to *search* for the maximum normal derivative response at a given neighborhood. Once the maximum derivative response is calculated across the whole image, pixels with normal derivative magnitude greater than a certain threshold ζ are classified as hair, and are discarded. The mathematical expression of the proposed operation is given as the following:

$$\log(\max_{\phi}(F_{\phi}(x))) \geq \zeta(im) \qquad (2)$$

where x is the intensity of the pixel, and F_{ϕ} is the magnitude response of derivative filter with orientation ϕ. Note that ζ is a function of the input image, and this will be explained more in depth.

If conventional oriented filters are used to find the normal direction, many (ideally infinite number of) oriented filters are needed. In our implementation, steerable filters [3] are used to reduce the number of oriented filters to 5. Filters introduced in [3] calculate the maximum derivative magnitude, as well as the normal direction (ϕ.) To increase the hair detection accuracy, input images are first adaptively histogram-equalized using a contrast-limited adaptive histogram equalization (CLAHE) scheme [20] to accentuate high frequency components.

The log of $\max_{\phi}(F_{\phi}(x))$ is rasterized into a histogram, denoted $H(x)$. Empirically, this histogram approximates a normal distribution, so we determine ζ in (2) by the first and second order statistics of $H(x)$:

$$\zeta(im) = mean(H(x)) - \alpha \times std(H(x)) \qquad (3)$$

Note that ζ is input image dependent. α is fixed at 1 in this work. To increase the hair removal performance, α should be adjusted to account for the amount of hair present in the image, but such an adaptive scheme was not considered in this work. The thresholded image will be missing many
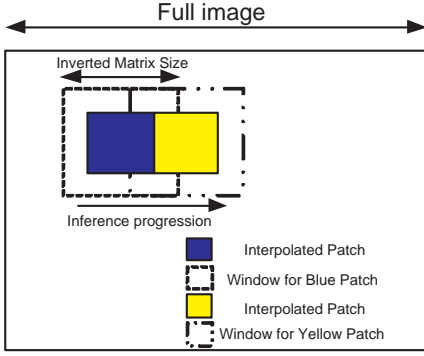
Figure 5. Only a windowed patch of J matrix is inverted for inference to reduce computational complexity. The blocky artifact of the proposed scheme is suppressed by using only the center portion of the interpolated $\hat{x}_{MAP}$.

pixel values (Figure 6 (b)); missing pixel values are recovered by using a graphical model with a smoothness prior.

A graphical model is a modular representation of how multiple random variables interact with one another [8]. Assuming a Gaussian Markov Random Field (GMRF) model, the *maximum a posteriori* (MAP) estimate is given by (4):

$$\hat{x}_{MAP} = \arg\max p(x|y) = J^{-1}h \qquad (4)$$

where J is an information matrix[1] and h is a potential vector [2] of the gaussian process. J and h encode how the state of a given node behaves as its neighbors and its measurement change, respectively. A thin-plate model [2] is used as a smoothness prior for x:

$$p(x) \propto \exp(-\alpha_2 \sum_{i \in V}(x_i - \frac{1}{|\mathcal{N}(x_i)|}\sum_{j \in \mathcal{N}(x_i)} x_j)^2) \quad (5)$$

A thin-plate penalizes the difference between the value at a given node and the average of its neighboring nodes.

In this work, every pixel is modeled as a node of the GMRF, and the image is represented as a mesh of nodes. Since the size of J matrix is the same as that of image, inverting J can be computationally expensive ($O(n^3)$, where n is the width of the matrix.) While successful inference algorithms exist to solve the inference problem without inverting the matrix [2, 18], we take a simpler approach to take an advantage of skin image characteristics.

The underlying assumption of this work is that the estimate of a given node will only depend on nodes within a patch: this is a locality assumption imposed at the patch-level. This assumption can be justified in case of skin images since a pixel in one corner of the image is likely to have small effect on a different pixel far away from itself. Therefore, we can crop the image into smaller windows, as

---

[1] J is also known as the inverse of a covariance matrix.
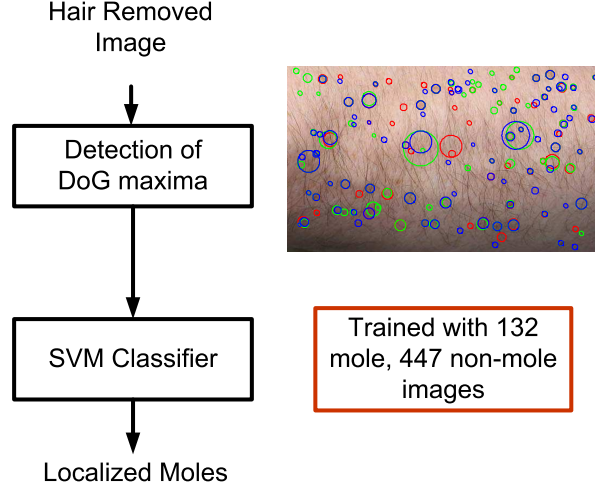[2] h is calculated as $J * mean(x)$.

Figure 8. The hair removed image is run through a difference-of-Gaussian (DoG) filter in the scale-space. The maxima of the DoG filter output are considered possible mole candidates. The possible mole candidates are further classified into moles with a trained SVM classifier.

shown in Figure 5, and compute the inverse J matrix of the cropped window. Since the cropped window is much smaller than the input image, the inversion of J matrix is computationally cheaper. Since we are inferring on blocks of image patches (i.e. ignoring pixels outside of the cropped window), the interpolated image will have blocky artifacts. Therefore, only part of $\hat{x}_{MAP}$ is used to interpolate the image, as shown in Figure 5.

Figure 6 shows how the image changes down the hair removal pipeline, and Figure 7 shows the result of the proposed hair-removal scheme. The proposed scheme leaves moles in tact, while removing the hair pattern. The size of moles may become a bit smaller as an artifact of using derivative filters. This is somewhat relieved by using a smoothness prior at the inference stage. The performance of the proposed hair removal scheme is compared with that of Dull Razor ([10]). Since Dull Razor is to remove hair in high resolution, in-focus images with less overlapped hair threads, it performs poorly on our dataset. In fact, Dull Razor misinterpretes moles as hair, and removes moles (Figure 7 (c)).

5. Mole localization using DoG filters and a support vector machine

The block diagram representation of the mole localization stage is shown in Figure 8. In this section, DoG scale-space filters and the designed support vector machine (SVM) classifier will be introduced. Since the size of moles can vary, moles should be searched in a multi-scale fashion. The SIFT feature detection scheme [11] selects the
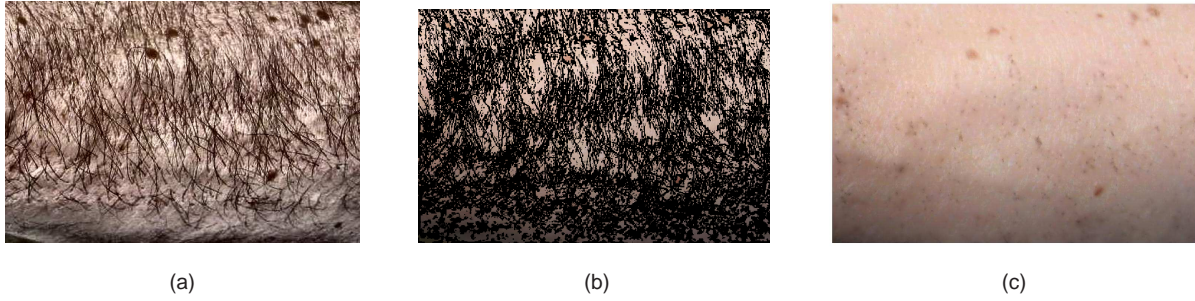
Figure 6. (a) The output of the image after adaptive histogram equalization. Note that the high frequency components are accentuated. (b) The thresholding operation removes long trails of hair patterns, while leaving the mole patterns in tact. (c) A graphical model is used to interpolate the missing pixel values. To reduce the computational complexity of graphical model inference, a patch-wise locality assumption is applied.
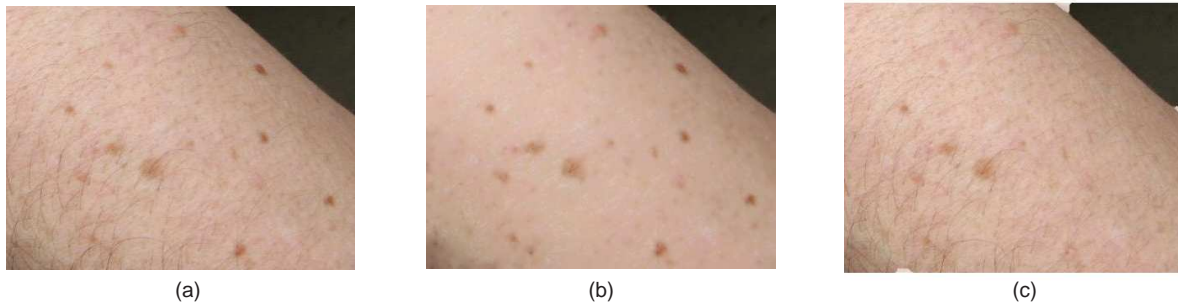

Figure 7. The input and output of the proposed system (a, b). The performance of the proposed scheme is compared with that of DullRazor [10].

proper scale by using difference-of-Gaussian (DoG) filters in the scale-space. The same idea is used here: the DoG filter is applied to RGB color channels separately, and the set union of the output maxima over scale in each channels are considered possible mole candidates. When combining the DoG maxima, any mole candidates occuring within a radius of another mole candidate is eliminated. Once mole candidates are localized, regions around mole candidates are cropped from the *hair-removed* image. The width of mole candidate patches is $2 \times \sqrt{2}$ times the radius of each DoG maximum.

Cropped mole candidates are classified as moles using a support vector machine (SVM) classifier. An SVM is a powerful tool to both generalize and classify objects: LIB-SVM [1] was used to build the SVM classifier. The feature vector to train the SVM closely resembles the image gist descriptor of Torralba *et al*. [15]: the gist feature is designed to describe texture patterns over space, as measured by oriented filter outputs at a variety of spatial scales. The feature vector extraction procedure used in this work is delineated in Figure 9. A mole candidate patch is first resized to 32 x 32 using a bicubic interpolation scheme, and is converted into a LA*B* color representation. LA*B* components are

then normalized as in (6) to increase the SVM classification performance.

$$ L_{Norm} = \frac{L}{100} \quad A_{Norm} = \frac{A}{256} \quad B_{Norm} = \frac{B}{256} \quad (6) $$

The L component of the input patch is steerable filtered into a 2-scale steerable pyramid using MatPyrTools [14]: each scale consists of 6 different oriented filter outputs. There are 16 images at hand (12 oriented filter outputs, the low and high frequency residue of L, A and B components of the input image), and they are each gridded into 4×4 squares. The values within the grid are averaged, and form an entry in the feature vector (i.e. each image generates 16 entries). The 16-entry vector from 16 images are rasterized into a single feature vector with 256 elements (Figure 9.) 132 mole patches and 447 non-mole patches are in the training set, and each mole patch is represented with a 256-entry feature vector described above. The training set is generated from a number of skin images *not* used in performance evaluation in Section 6. All mole patches come from *hair-removed* images, thus mole patches in the training set don't have any hair.
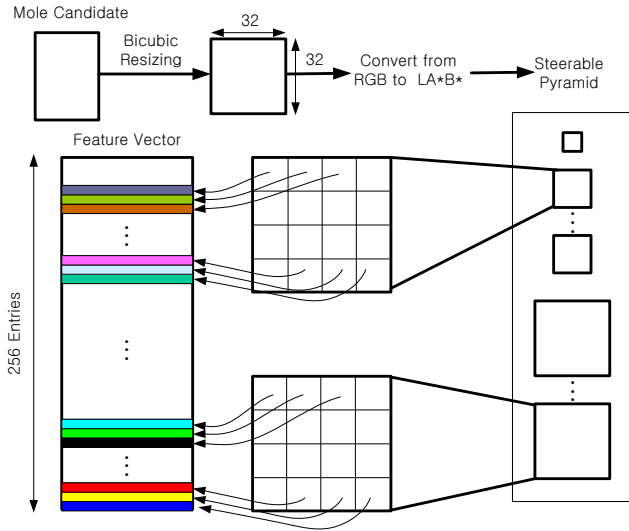
Figure 9. The mole candidate is resized into 32x32 patch image; the resized image is steerable filtered, and is represented as a 256-entry feature vector.

A standard dimensionality reduction scheme called Principal Component Analysis (PCA) is used to reduce the dimensionality of feature vectors. The number of principal components are chosen such that 98.5% of the data variance is conserved: 12 dimensions are sufficient with our dataset. The SVM training procedure is 10-fold cross-validated [5] to determine the optimal γ and C, which are two parameters of radial basis function-based SVM classifier.

6. Experimental results

To analyze how the hair removal scheme impacts the overall system performance, two sets of experiments are performed. In the first experiment, DoG filtering is applied to both the original input image and the hair removed image. Experiments show that the hair removal scheme reduces the number of false mole candidates, thus reduces the number of classifications needed in the SVM classifier stage. Figure 10 shows the maxima of DoG filter outputs when hair is present and removed, respectively.

The DoG filter in SIFT [11] discards maxima smaller than a certain threshold β. To make a fair comparison, β is adjusted in both test cases to get the fewest DoG maxima while retaining all moles. In some images, moles can be heavily occluded by hair such that the DoG cannot detect the mole pattern without hair removal. In such cases, the hair removal scheme can help locate a mole that would have been undetected otherwise.

In the second experiment, another type of SVM (called SVM 1 from here) is trained to quantify whether the hair removal scheme improves classification performance. SVM 1 is trained with mole patches that come from *skin surface*
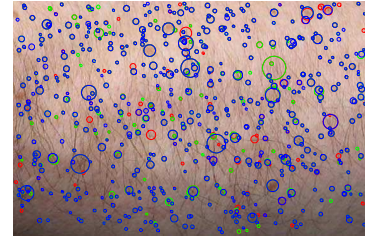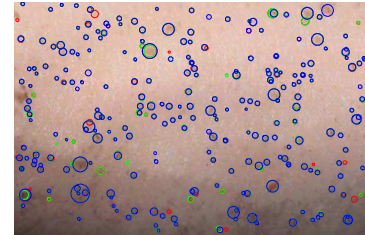

Figure 10. Hair removal scheme reduces the number of false mole candidates. The threshold of DoG response is adjusted in both images to get fewest DoG maxima while still retaining all moles.
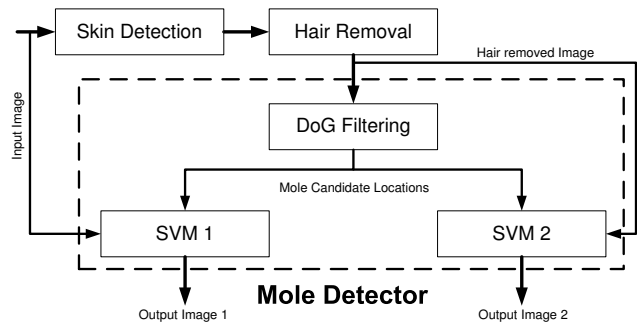

Figure 11. Another type of SVM is trained to verify the effectiveness of hair removal block in the system. While the mole candidate coordinates are the same for both SVMs, the SVM 1 crops the mole candidate patch from the original input image, while the SVM 2 crops the mole candidate patch from the hair removed image.

before hair removal. Thus, mole patches in this training set have hair on them. This differs from our original SVM (called SVM 2 from here) in that SVM 2 is trained with hair-removed mole patches. The experiment architecture is shown in Figure 11. Input to both SVMs are the list of mole candidate locations, and the image from which to crop the mole patches. The outputs of each SVMs are compared to characterize the classification performance. Both SVMs are trained to achieve the best classification rate. The test images for this experiment consist of 28 images capturing arms and the back torso of different people at different lighting conditions (and orientations in the case of arms).

Figure 12 shows the experimental results on a subset of images. As can be seen, the designed system works well
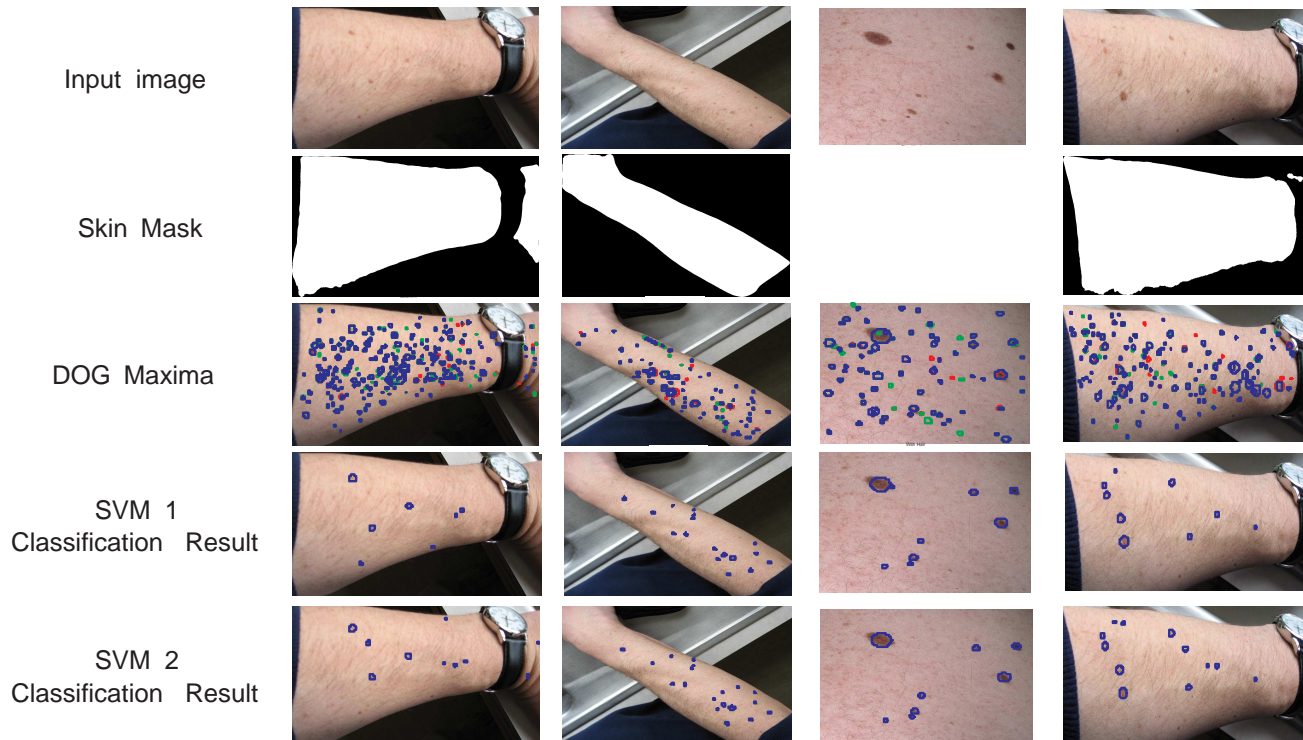
Figure 12. This figure shows the result of the experiment described in Figure 11. Interestingly, the mole localization performance in both SVMs are comparable.

for the given test images. While the skin detection method is not robust to objects with skin-like colors, subsequent stages can localize the moles reliably (i.e. at low false alarm rate.) Note that skin regions should not be discarded as non-skin regions in order to achieve high detection rate.

To quantify the mole detection performance, two measures are introduced [9]:

$$Sensitivity = \frac{TP}{TM}$$

$$DiagnosticAccuracy = \frac{TP}{TM + FP} \qquad (7)$$

where TP is the number of true moles detected, TM is the total number of moles in the image, and FP is the number of non-moles classified as moles. The detection result is summarized in Table 1. Note that classification performance is better for SVM 2, and this justifies our use of hair removal scheme in mole localization.

While both SVMs work well for most images, the system also has some failure modes. Figure 13 (a) is a crop of a back-torso image, and the false positive rate is high. This can be attributed to the fact that in high resolution images, hair roots and pores can be misinterpreted as moles

Table 1. A table comparing the performance of SVM 1 and SVM 2 (S: Sensitivity, DA: Diagnostic Accuracy)

SVM 1		SVM 2	
S	DA	S	DA
79.4%	76.5%	84.7%	79.3%

under certain lighting conditions. The false positive rate can be reduced with a user intervention: if the user specifies that any mole candidates with radius smaller than a certain threshold should be discarded, hair roots and pores will not be detected. In some cases, moles can be misclassified as non-moles if images are blurry, as shown in Figure 13 (b). Therefore, to increase the reliability of the system, the input image should be in-focus.

7. Conclusion

This paper presented a novel mole localization scheme that makes use of multiple cascaded filters. The explicit description of mole appearance using an introduced feature vector allows us to increase the mole detection accuracy at low false alarm rate. The failure modes can be circumvented by dealing only with focused images, or by user intervention to exclude mole candidates smaller than some specified size.
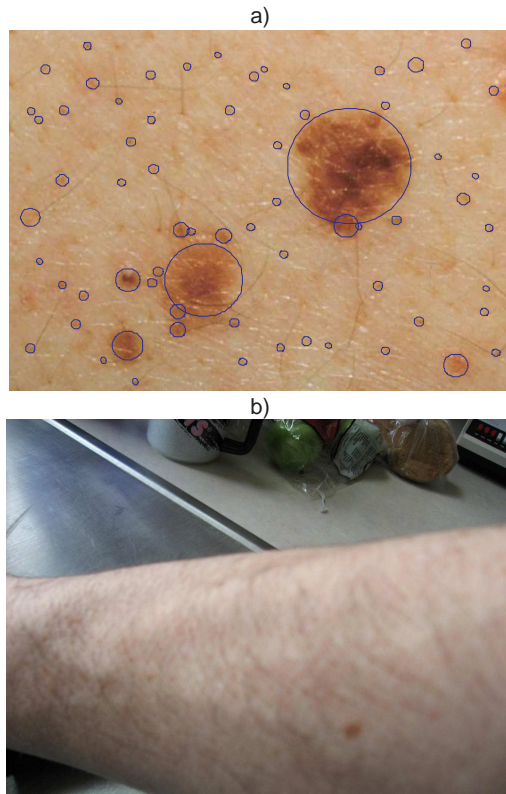
a)

b)

Figure 13. These two example images show the failure mode of the designed system. a) When the image is at high enough resolution to reliably visualize pores on skin, pores can be recognized as moles at certain lighting conditions. b) When mole patches are blurred, the SVM cannot recognize them as moles.

Acknowledgment

References

[1] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/˜cjlin/libsvm/, 2001.

[2] M. J. Choi. Multiscale gaussian graphical models and algorithms for large-scale inference. Master's thesis, Massachusetts Institute of Technology, 2007.

[3] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38:587 – 607, 1991.

[4] H. Ganster, A. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, 20:233 – 239, 2001.

[5] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. http://www.csie.ntu.edu.tw/˜cjlin/papers/guide/guide.pdf.

[6] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:696, 2002.

[7] M. Jones and J. Rehg. Statistical color models with application to skin detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1999.

[8] M. Jordan and Y. Weiss. Graphical models: probabilistic inference. In M. Arbib, editor, *Handbook of Neural Networks and Brain Theory 2nd edition*. MIT Press, 2002.

[9] T. Lee, S. Atkins, M. King, S. Lau, and D. McLean. Counting moles automatically from back images. *IEEE Transactions on Biomedical Engineering*, 52:1966 – 1969, 2005.

[10] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean. Dullrazor: A software approach to hair removal from images. *Computers in Biology and Medicine*, 27:533–543, 1997.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 2:91 – 110, 2004.

[12] J.-S. Pierrard and T. Vetter. Skin detail analysis for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[13] P. Schmid-Saugeon, J. Guillod, and J.-P. Thiran. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*, 27:65 – 78, 2003.

[14] E. Simoncelli. Matlab pyramid tools. http://www.cns.nyu.edu/˜lcv/software.html.

[15] A. Torralba, A. Oliva, M. Castelhano, and J. M. Henderson. Contextual guidance of attention in natural scenes: The role of global features on object search. *Psychological Review*, 113(4):766 – 786, 2006.

[16] H. Tsao, M. Atkins, and A. Sober. Management of cutaneous melanoma. *The New England Journal of Medicine*, 351(10):998–1012, 2004.

[17] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Graphicon*, 2003.

[18] M. Wainwright and M. Jordan. Graphical models, exponential families, and variational inference. Technical report, University of California, Berkeley, 2003.

[19] Z. Zhang, W. Stoecker, and R. Moss. Border detection on digitized skin tumor images. *IEEE Transactions on Medical Imaging*, 19:1128 – 1143, 2000.

[20] K. Zuiderveld. Contrast limited adaptive histogram equalization. In P. Heckbert, editor, *Graphics Gems IV*, pages 474–485. Academic Press, Boston, 1994.