Automated Detection of New or Evolving Melanocytic Lesions Using a 3D Body Model

Federica Bogo[1,2], Javier Romero[1], Enoch Peserico[2], and Michael J. Black[1]

[1] Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2] Università degli Studi di Padova, Padova, Italy

Abstract. Detection of new or rapidly evolving melanocytic lesions is crucial for early diagnosis and treatment of melanoma. We propose a fully automated pre-screening system for detecting new lesions or changes in existing ones, on the order of $2-3$mm, over almost the entire body surface. Our solution is based on a multi-camera 3D stereo system. The system captures 3D textured scans of a subject at different times and then brings these scans into correspondence by aligning them with a learned, parametric, non-rigid 3D body model. This means that captured skin textures are in accurate alignment across scans, facilitating the detection of new or changing lesions. The integration of lesion segmentation with a deformable 3D body model is a key contribution that makes our approach robust to changes in illumination and subject pose.

1 Introduction

Malignant melanoma is an aggressive form of skin cancer, the incidence of which is rapidly increasing worldwide. Early detection promptly followed by excision is the key to a favorable prognosis [2]. Unfortunately, in its early phases, a melanoma is often indistinguishable from a benign melanocytic lesion (a common mole). A sensitive sign of a malignant melanocytic lesion is its *evolution*; the appearance of a new lesion or changes in an existing one suggest an increased probability of a melanoma [2]. Digital imaging systems allow a dermatologist to compare pictures of a patient's body taken at different times [4]. However, manual comparison remains challenging (due to changes in the pose or illumination between scanning sessions) and time-consuming when applied to the whole body (many patients have hundreds of lesions). To improve early detection and comprehensive skin surface analysis, we develop an automated image acquisition and analysis system that provides a first level of surveillance; putative changes can then be evaluated by a dermatologist. The system can also find use in the acquisition and analysis of data for epidemiological studies.

Specifically, relying on the framework introduced in [1], we propose a fully automated pre-screening system to detect new or changing melanocytic lesions using a learned, deformable, parametric 3D body shape model. The approach is summarized in Fig. 1. First, we capture a 3D triangulated mesh, or "scan", using 22 pairs of high-resolution stereo cameras that capture body shape and 22 color cameras that capture skin texture. Acquisition is rapid (a few milliseconds

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 593–600, 2014.

per scan) and the system does not require patients to accurately hold a specific pose. Given scans of a patient taken on different days there will be changes in pose, shape, lighting, hair and skin texture. Our novelty lies in using a 3D body model to accurately register (align) such scans across time, correct for lighting, and to build a "map" of the skin surface that can be used for analysis. This involves several key technologies. First, we use a learned statistical model of body shape variation, constructed from thousands of detailed 3D scans of different people. This model accounts for variations in body shape and pose between scans. Second, we define a method that uses the 3D shape information together with image texture information to accurately align scans with the model. This brings every scan of the patient into correspondence. Once scans are aligned, we compare them across time to identify changes. To that end, we define a basic segmentation algorithm that detects putative lesions in the scan images; such lesions are then compared across scans using the registration.

In a pilot study using synthetic lesions, the method detects new lesions or changes in existing ones on the order of 2–3mm. The system is robust to changes in body pose, illumination, presence or absence of sparse body hair etc. Our segmentation scheme takes advantage of multiple camera views to robustly detect lesions by using consistency across views; artifacts tend not to be consistent. This goes beyond previous work to use many cameras to see most of the body at once and to integrate all this information into a coherent 3D model of body shape and appearance that enables lesion detection over time.

Related work. Most previous work on lesion change detection addresses the problem in high-magnification images of small regions surrounding a lesion obtained with a dermatoscope (see [4] for a survey). Tracking multiple lesions, however, is a challenging problem that has received surprisingly little attention [4].

The task can be subdivided into two parts: segmentation/detection and registration/matching. Segmentation/detection approaches usually identify a set of lesion candidates using simple image processing methods [7, 9–11] and then filter the results using unsupervised [7, 9] or supervised [10] classification. Matching lesions in images taken at different times is challenging and approaches take many forms. The diameter of a lesion is generally small compared to its displacement between different images, making matching hard. One approach solves for a rigid 2D transformation between images given user-provided matches [8]. In [5], back torso images are mapped to a common 2D template. Pose variation and non-rigid changes in body shape cause non-linear, anisotropic deformations of the skin, further complicating matching. Other approaches focus on the topological relations between lesions and use graph-matching methods to find the relationship between images [3, 5, 6]. While able to produce robust matchings, these approaches have difficulty with large numbers of lesions.

Voigt and Classen [11] perform both segmentation and registration. Images of the patient's front and back torso are acquired with a single camera and a positioning framework for adjusting the patient's pose. Lesion borders are detected by thresholding the output of a Sobel operation; due to the large number of skin features easily mistaken for lesions, such as hair, this can lead to poor

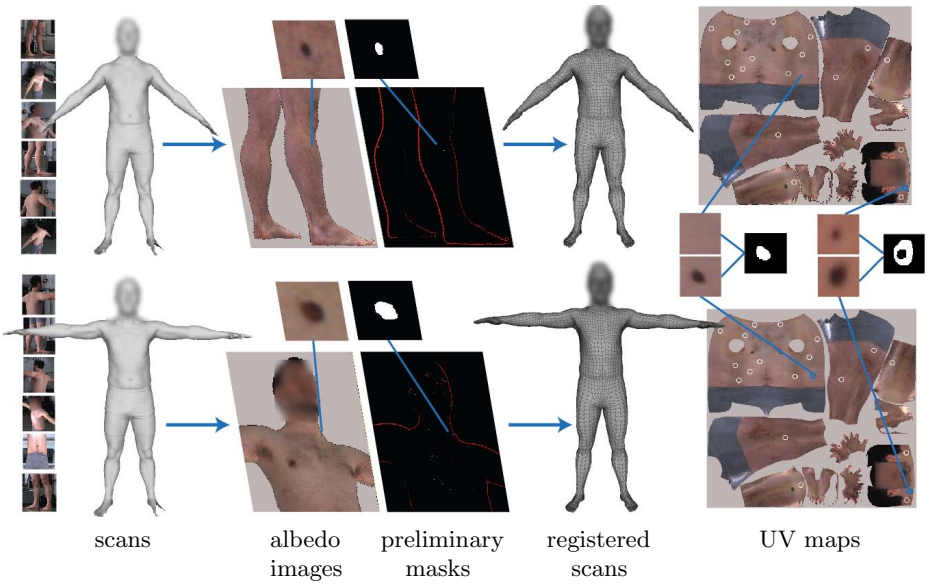

Fig. 1. Overview of our approach. After scan capture, we preprocess camera images in order to eliminate shadows and obtain a preliminary lesion segmentation. We bring scans into alignment with a 3D body model, that normalizes for pose and shape variations. Once scans are registered, we compare them and identify changes in the parameterized space defined by the model.

performance. The precise positioning of the patient attempts to remove the registration problem but, in practice, humans are deformable and there will always be non-rigid changes in pose and shape.

Virtually all previous segmentation and registration techniques are evaluated only on a small part of the body, commonly the back or front torso. These methods do not provide a solution to the full-body analysis/screening problem. In contrast, we consider the entire body surface (or most of it) at once, simplifying the acquisition process for both the patient and doctor.

2 Method

Our method proceeds in four steps (see Fig. 1): 1) acquisition of a textured 3D scan of the subject; 2) albedo extraction and preliminary lesion segmentation; 3) scan registration to a 3D body model (with coherent topology across scans and subjects); 4) segmentation refinement and skin surface tracking in the parameterized space defined by the model.

1. Scan acquisition. Our acquisition system is a full-body 3D stereo capture system (3dMD, Atlanta, GA) with 22 modular, medical standard scanning units. Each unit contains a pair of grayscale stereo cameras and one or two speckle projectors, plus a single RGB camera for texture extraction. A set of 20 flash

units illuminate the body during capture. Each scan results in a triangulated, high-resolution, non-watertight mesh. We process images at a resolution of 1224× 1024 pixels.

2. Albedo extraction and preliminary lesion segmentation. Automated lesion segmentation on the whole body may suffer from the presence of shadows and shading. To reduce these effects, we preprocess the camera images in order to discriminate between albedo and shading. As in [1], we model scene lighting as a combination of 9 Spherical Harmonic basis functions under the assumption of Lambertian skin reflectance. We assume light is constant across scan sessions, and simply precompute it (see [1] for details). The estimated light model is used for computing shadows, which are then removed from each original image I_j to produce the corresponding albedo image A_j (see Fig. 1).

In each image, we isolate skin from background and clothing by means of a simple thresholding of the hue. We choose a conservative threshold, since subsequent steps can deal with skin false positives.

We obtain an initial estimation of lesion borders using Laplacian-of-Gaussian (LoG) filtering [9, 10]. Since lesion radii can vary depending on the subject and camera viewpoint, the LoG filter is applied at five different scales. Linear Discriminant Analysis (LDA) is used to classify each pixel in A_j into a lesion binary mask M_j based on the output of the multi-scale LoG filter. LDA classification produces, for each albedo image A_j, a binary mask M_j, marking each pixel as lesional or non-lesional.

Facial features and occlusion boundaries, due to their high second-derivative response, may be erroneously identified as lesional (red pixels in preliminary masks, Fig. 1). However, these artifacts tend to be elongated, while lesions are spatially compact. We postprocess M_j in order to keep only compact connected components. For each connected component in M_j, we consider its minimum bounding box; if the ratio between its major and minor side is too high, or fewer than half of the pixels inside it are lesional, the component is discarded.

3. Scan registration. We register scans of the same subject captured in different sessions by aligning each scan S with a common, triangulated template mesh T^*. In this process, T^* is aligned (i.e. deformed) to S, giving a registered scan T. Our registration technique exploits both 3D shape and appearance information, and relies on a learned, statistical 3D human body model. It is similar, in many respects, to that described in [1]; we briefly review it here for completeness.

Our model factorizes body deformations into a set of pose-dependent transformations (parameterized by pose $\boldsymbol{\theta}$) and a set of identity-dependent transformations, D. The appearance of each subject is modeled through a high-resolution UV map, U. We learn the pose-dependent transformations from a corpus of registered scans of different people [1]. D and U are learned by registering an initial set of scans of the subject [1]; these scans need to be captured only once, during the first session (see Sec. 3 for details about the initialization).

The quality of the correspondence between T and S is measured in terms of an error with three components: E_S, E_U and E_C. E_S expresses how close the

mesh surfaces are in 3D space, while E_U enforces similarity between their color appearance; E_C encourages deformations that are consistent with the learned body model. Mathematically, we minimize the following energy function:

$$E(T, \boldsymbol{\theta}; S, U, D) = \lambda_S E_S(T; S) + \lambda_U E_U(T; U, \{A_j\}, \{M_j\}) + \lambda_C E_C(T, \boldsymbol{\theta}; S, D) \tag{1}$$

where λ_S, λ_U and λ_C represent weights assigned to the different terms.

With respect to the formulation provided in [1], we slightly modify the appearance term E_U to give more importance to appearance consistency around lesions. More precisely, given U, T and the calibration parameters of camera j, we render a synthetic image $\bar{A}_j$ (Fig. 2(b)). E_U encourages consistency between each albedo image A_j and the corresponding synthetic image $\bar{A}_j$:

$$E_U(T; U, \{A_j\}, \{M_j\}) = \sum_{\text{cams } j} \sum_{\text{pixels } \boldsymbol{y}} w_{M_j}(\boldsymbol{y}) (\Gamma_{\sigma_1,\sigma_2}(A_j)[\boldsymbol{y}] - \Gamma_{\sigma_1,\sigma_2}(\bar{A}_j)[\boldsymbol{y}])^2 \tag{2}$$

where $w_{M_j}(\boldsymbol{y})$ is a weighting function assigning higher weight to pixel $\boldsymbol{y}$ if $\boldsymbol{y}$ is marked as lesional in M_j, and $\Gamma_{\sigma_1,\sigma_2}$ defines a Ratio of Gaussians (RoG) of parameters σ_1 and σ_2.

4. Lesion segmentation refinement and change detection. The presence of sparse hair, small skin artifacts or generic image noise may affect the performance of the pre-segmentation described above, producing a high number of false positives. Using more restrictive classification thresholds or artifact removal algorithms (as in [10]) may produce false negatives, i.e. discard actual lesions. Crucially, these artifacts tend to be mistaken as lesions only from specific viewpoints. We exploit our multi-camera capture framework to filter out lesions that are not consistently detected by a number of relevant (i.e. with a good viewpoint) cameras. More formally, for any template surface point $\boldsymbol{x}$, denote by $uv(\boldsymbol{x})$ its mapping from 3D to UV space, and by $\pi_j(\boldsymbol{x})$ its projection onto the image plane defined by camera j. $M_j[\pi_j(\boldsymbol{x})]$ equals 1 if $\boldsymbol{x}$ is classified as lesional according to camera j, 0 otherwise. For each camera j, denote by $\omega_{\boldsymbol{x},j}$ the cosine of the angle between the surface normal at $\boldsymbol{x}$ and the ray from $\boldsymbol{x}$ to the camera's center. We denote the set of cameras for which $\boldsymbol{x}$ is visible by $J(\boldsymbol{x})$. Pixel $uv(\boldsymbol{x})$ is classified as lesional if and only if

$$\frac{\sum_{\text{cams } j \in J(\boldsymbol{x})} M_j[\pi_j(\boldsymbol{x})] \max(\omega_{\boldsymbol{x},j}, 0)}{\sum_{\text{cams } j \in J(\boldsymbol{x})} \max(\omega_{\boldsymbol{x},j}, 0)} > \delta \tag{3}$$

where δ is a system parameter. This corresponds to computing a weighted average of the classifications provided by different cameras – where the contribution of each camera is weighted according to the quality of its viewpoint. Figure 2 shows how the final segmentation varies depending on δ: artifacts like sparse hair tend not to be consistently detected across different cameras, and are therefore filtered out; lesions exhibit more consistency (see e.g. the bottom of the back and the right shoulder). We quantitatively evaluate the sensitivity of the system to the value of δ in Sec. 3.

Fig. 2. A real (a) and a synthetic albedo (b) image of a subject's back. Figures (c)-(e) show the final segmentation obtained by setting δ in Eq. 3 to 0, 0.2 and 0.5, respectively.

Fig. 3. (a) Skin patch exhibiting a synthetic lesion (large lesion towards upper left). (b) Scans of subjects, showing varied skin phenotype and pose.

Detected lesions are integrated into a full-body UV map (see Fig. 1). This greatly simplifies the tracking of changes in lesions compared to using multiple single images. Each UV map pixel is associated with the same template surface point, independently of subject pose and shape. UV maps from different times are therefore directly comparable. A detection that does not overlap with one in a previous map reveals a new lesion; a detection that does overlap, but comprises a higher number of pixels, is likely to reveal a lesion that has grown.

3 Experimental Evaluation

We evaluated our system on a set of 6 male and 6 female subjects of ages 23 to 44 years, height 160 to 186 cm, and weight 55 to 82 kg. There was considerable variation in terms of skin tone, number of melanocytic lesions, and presence of sparse body hair (Fig. 3(b)). We trained the LDA classifier (Sec. 2) on a set of 50 images of 10 different subjects, captured from different viewpoints; there is no overlap between the subjects used for evaluation and those used for training.

For this pilot study, we artificially created and altered lesions by drawing with a marker on the subjects' skin. Note that these synthetic lesions look realistic at the resolution of our images, as seen in Fig. 3(a).

Each subject was scanned in 2 poses, respectively with arms held horizontally, and pointing downwards at an angle (Fig. 3(b)). For each subject, we captured two initial scans in order to learn D and U (Sec. 2). After the initial scans, for

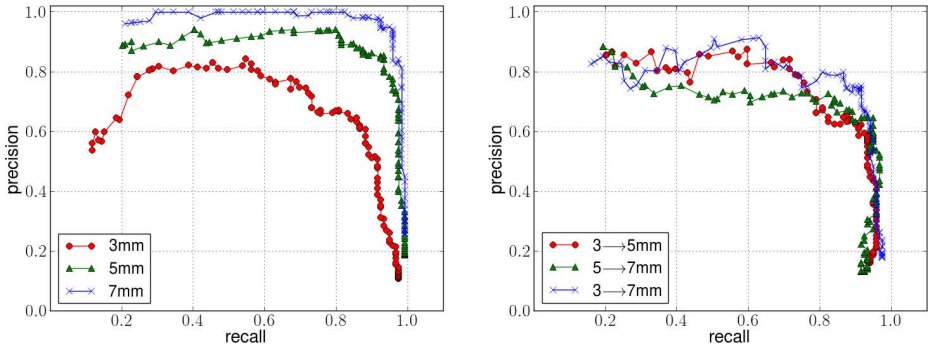

Fig. 4. Precision/recall curves for detecting new lesions (left) and increased lesion sizes (right), for 100 values of the parameter δ evenly spaced between 0 and 1. Precision and recall values are computed by aggregating the values obtained for all the subjects.

each subject we created 10 synthetic lesions with a diameter of 3mm, and re-scanned them in the 2 poses. We then expanded each synthetic lesion first to a diameter of 5mm, and then to a diameter of 7mm, re-scanning the subjects in the 2 poses each time. This yielded 4 timepoints with increasing lesion diameter (0, 3, 5 and 7mm): 3 pairs of timepoints ($0 \rightarrow 3$mm, $0 \rightarrow 5$mm, $0 \rightarrow 7$mm) correspond to the appearance of new lesions of different diameters, while the other 3 pairs ($3 \rightarrow 5$mm, $3 \rightarrow 7$mm, $5 \rightarrow 7$mm) correspond to changes in existing lesions. For each pose, and pair of timepoints, our system identifies a set of "suspect" lesions – lesions deemed either new or modified. For different values of δ (Sec. 2), our system yields different values of precision (the fraction of suspect lesions that were actually new or modified lesions) and recall (the fraction of new or modified lesions that were reported as suspect lesions).

Figure 4 reports the results for the "arms downward" pose, since it was the most comfortable for all subjects. The results for the other pose are almost identical. On average, a high recall ($> 90\%$) was achieved for all pairs of timepoints, with a precision $> 50\%$ in the case of small (3mm) new lesions, $> 80\%$ in the case of larger new lesions (5mm and 7mm), and $60 - 80\%$ in the case of changes in existing lesions. Note that, while high precision is desirable, high recall is more important since the consequences of missing a potential melanoma are much direr than those of a false alarm.

The acquisition of each scan requires a few milliseconds. Further processing (scan generation, alignment, UV map analysis) can be performed off-line; in our experiments, it required a few minutes per scan on a common desktop machine.

4 Conclusions

We have proposed a novel solution for "full-body" screening of melanocytic lesions. A multi-camera stereo system captures the 3D shape and skin texture of a subject. Given two such scans of the same subject, taken at different times,

we bring them into registration by aligning each scan with a learned, parametric 3D body model. Once scans are registered, we compare them across time and identify changes in skin lesions. In a pilot study, we show that our method automatically detects changes on the order of a few millimeters.

Based on our results, a longitudinal study of dermatological patients should be pursued. Future work should explore higher-resolution RGB imagery and the effect of varying number/resolution of cameras on detection. Another research line would explore less expensive scanning devices (e.g. the Kinect) for the acquisition of the 3D data and texture. Here the 3D body model could be exploited to integrate information from multiple poses (cf. [12]) and, given accurate alignment, image super-resolution could be used to obtain high-quality texture.

Acknowledgments. F. Bogo and E. Peserico were supported in part by MIUR proj. AMANDA, prot. 2012C4E3KT_001.

References

1. Bogo, F., Romero, J., Loper, M., Black, M.: FAUST: Dataset and evaluation for 3D mesh registration. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2014)
2. Dunki-Jacobs, E., Callender, G., McMasters, K.: Current management of melanoma. Current Problems in Surgery 50, 351–382 (2013)
3. Huang, H., Bergstresser, P.: A new hybrid technique for dermatological image registration. In: IEEE International Conference on BioInformatics and BioEngineering (BIBE), pp. 1163–1167 (2007)
4. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: A review. Artificial Intelligence in Medicine 56(2), 69–90 (2012)
5. Mirzaalian, H., Hamarneh, G., Lee, T.: A graph-based approach to skin mole matching incorporating template-normalized coordinates. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2152–2159 (2009)
6. Mirzaalian, H., Lee, T., Hamarneh, G.: Uncertainty-based feature learning for skin lesion matching using a high order MRF optimization framework. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 98–105. Springer, Heidelberg (2012)
7. Perednia, D., White, R., Schowengerdt, R.: Automated feature detection in digital images of skin. Computer Methods and Programs in Biomedicine 34, 41–60 (1991)
8. Perednia, D., White, R., Schowengerdt, R.: Automatic registration of multiple skin lesions by use of point pattern matching. Computerized Medical Imaging and Graphics 16, 205–216 (1991)
9. Pierrard, J., Vetter, T.: Skin detail analysis for face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2007)
10. Taeg, S., Freeman, W., Tsao, H.: A reliable skin mole localization scheme. In: IEEE International Conference on Computer Vision (ICCV), pp. 1–8 (2007)
11. Voigt, H., Classen, R.: Topodermatographic image analysis for melanoma screening and the quantitative assessment of tumor dimension parameters of the skin. Cancer 75, 981–988 (1995)
12. Weiss, A., Hirshberg, D., Black, M.: Home 3D body scans from noisy image and range data. In: IEEE International Conference on Computer Vision (ICCV), pp. 1951–1958 (2011)