Multimodal Data Acquisition for Dermatology using Google Glass

João Antunes
joao.guilherme.martins@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2014

Abstract

Melanoma is now the deadliest form of cancer. The best survivability rate is still attained with early detection. Several diagnosis methodologies have been developed specifically to optimize fast melanoma screening, like the ABCD rule and Menzies’ method. To complement this, several computer aided diagnosis systems were developed. These systems are plagued by the low availability of public labeled datasets, due to the resource intensity the task of creating a dataset entails. Also, state-of-the-art systems do not provide a way to locate magnified lesion images in space. This thesis proposes a hands-free non-invasive, wireless system using Google Glass that enables labeled dataset creation on-the-fly during a doctor’s appointment, without the need to have an expert clinician spending hundreds on man-hours labelling image sets. The proposed system also includes a module to, without any extra input from the user, locate the lesion in the patients body, resorting to identifiable tags.

Finally, a tele-dermatology framework using the proposed system is described. The proposed system worked as expected, but some limitations were encountered, mainly due to Google Glass’s immaturity as a product, and due to the fact that the detail camera used to obtain the lesions was a smartphone camera, as opposed to a dermoscope camera. Nevertheless, the proof-of-concept entailed was successful.

Keywords: Labelled dataset, Google Glass, Super-Resolution, Melanoma, Dermoscopy

1. Introduction

Over the years, the incidence of melanoma has shown a steady increase. Melanoma is defined as a malignant proliferation of melanocytes (melanin-producing cells located on the skin’s epidermis) that has the potential to metastasise. From all the skin cancers, melanoma is by far the deadliest form [28]. Despite all the advances in the treatment of melanoma, the ultimate goal for a dermatologist is still its early detection, for this route is the one with the highest survival rate. Fortunately, with an early prognosis, many melanomas can be detected at earlier stages with smaller, thinner and potentially curable lesions [2].

From the need to arrive at early prognosis for melanoma several clinical methodologies have been developed to aid the dermatologist. One of the most prevalent techniques is designated dermoscopy. This technique lies on examination of morphological structures not visible with the naked eye, through the use of a dermoscope. A dermoscope is an instrument composed of a magnification lens and a light source, coupled with an immersion fluid. By examining these structures the correct diagnosis of melanoma by experienced clinicians has shown significant improvement [18]. To guide the screening process, a few procedures have been developed like pattern analysis [2], the ABCD rule [29] and the 7-point checklist [2]. However, visual examination and interpretation of dermoscopic images is a time consuming task [19]. Furthermore, there is no easy way to do follow-up examination of skin lesions. Typically, dermoscopic images have a high magnification factor. Due to this, image-to-image identification of the same lesion, as well as lesion-to-image correspondence is currently a challenge for most dermoscopy systems. This is crucial, in order to facilitate identification of lesions undergoing rapid changes, another characteristic of melanoma. Furthermore, this methodology requires the presence of a skilled clinician on site.

Due to the time consuming nature of dermoscopy techniques (especially in patients with a high number of lesions to examine), several Computer Aided Diagnosis (CAD) systems have been developed. They can be divided into two main groups. One group is based on melanoma diagnostic methods, using image processing techniques to locate and identify the shape of the lesion, extract image parameters, and infer a diagnosis based on these parameters. This group tries to mimic the diagnosis procedures used by dermatologists. The other group is based on pattern recognition and machine learning techniques applied on standard image features (e.g.: color and texture) [22][4][13]. Such systems usually require the a priori existence of labeled datasets, which are used to apply a variety of machine learning techniques, supported by the knowledge base built by dermoscopy techniques. Although there are some datasets available, there
is no common standard unifying all the data available. Furthermore, most of those available datasets are not free [22]. Also, the manual labelling must be done by expert dermatologists, in order to ensure a reliable ground truth upon which to build and compare CAD systems. Due to these constraints, building a labeled dermoscopy dataset is a cumbersome task.

2. Dermoscopy Background

2.1. Introduction to dermoscopy

Dermoscopy, also known as epiluminiscence microscopy, is a non-invasive diagnosis technique for observation of in vivo Pigmented Skin Lesions (PSLs). This technique allows visualization of both skin surface and subsurface structures that are not discernible by the naked eye (e.g.: streaks or pigmented networks). This proves to be a valuable asset to experienced clinicians, leading to a significant increase in the accuracy of the diagnosis of melanoma [18].

This technique consists of inspecting a skin lesion by placing an immersion fluid (mineral oil/alcohol) on a skin lesion and examining the morphological structures (e.g.: globules, streaks or pigment networks) therein with a dermoscope. Typically, the magnification of a dermoscope ranges from 6x to 40x, although the most commonly used is 10x. The fluid placed directly on the lesion eliminates surface reflection, rendering the cornified layer translucent. This allows observation of the pigmented structures within the epidermis, the dermoepidermal junction and the superficial dermis [2].

Figure 1: Commercially available dermoscope using crosspolarized light produced by Dermlite®.

To assist in the screening process, a few diagnosis procedures have been developed, mainly the ABCD rule (see Section 2.2) and the 7-point checklist (see Section 2.3). However, it should be stated that these methods for melanoma screening are used to determine whether suspicious lesions could be a melanoma. The actual diagnosis is only carried out via biopsy.

Although the main goal is to classify a lesion in a binary fashion it is relevant to be acquainted with the most common melanocytic lesions in order to understand the difficulty of the process. Several lesions have a protein nature which make their identification trickier, and distinguishing them from melanoma is frequently a difficult task. Some of them will be briefly exposed [2]:

1. Melanoma: It is defined as a proliferation of melanocytes that has the potential to metastasize. A malignant melanoma causes the majority (75%) of deaths related to skin cancer [16][21]. Furthermore, in its advanced stages, (with signs of metastases), melanoma is at present incurable, with the available treatments being solely palliative [1]. The clinical features presented by melanoma are protean. They are, however, usually small, irregularly shaped and slightly elevated from the skin, being pink to dark-brown in color. Instances have appeared of melanomas that are skin-colored, being called amelanotic melanomas. Invasive melanomas exhibit a wider array of colors, exhibiting shades of brown, black, red, white (lighter than healthy skin) and blue. These melanomas are also usually papular (a circumscribed, solid elevation of skin with no visible fluid) or nodular, and often ulcerated.

2. Clark Nevus: These are the most common type of nevi, considered to be a precursor lesion of melanoma, although no scientific consensus on what criteria must be met for a given number of Clark nevi to develop into melanoma has been achieved [2]. The distinction between Clark nevi and melanoma is the biggest challenge in the realm of PSL classification. As is the case with melanoma, the clinical, dermoscopic and histopathologic features of Clark Nevi are protean. They are PSLs with a slight elevation, and are found mainly in skin that has been exposed to sunlight (the trunk and the extremities).

3. Blue Nevus: Blue nevi are usually very clear indicators of malignancy due to their morphological distinction. Usually, blue nevi are monotonous being either blue, gray-blue or gray black in pigmentation. They are circular with well defined borders. They exhibit a homogeneous pattern with absence of local features (no pigment network structures, globules or dots). This absence of local features coupled with the well defined borders are usually the distinction criteria between blue nevi and melanoma.

The two main melanoma screening procedures, the
ABCD rule and the 7-point checklist, will now be presented.

2.2. ABCD rule

The ABCD rule for melanoma diagnosis, introduced by Stolz et al. [29], is the most widely followed procedure for melanoma screening, mainly due to its ease-of-use [25][3]. This procedure is based on assessing four different criteria semiquantitatively and then combining them to determine a Total Dermoscopy Score (TDS). The four criteria to be assessed are Asymmetry, Color, Border and Differential Structures present. From these four criteria the acronym ABCD was coined. This information is presented in Table 1.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Score Range/Weight Factor</th>
<th>Description</th>
<th>Asymmetry</th>
<th>Border</th>
<th>Color</th>
<th>Differential Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetry</td>
<td>[0 - 2]/1.3</td>
<td>Color, contour and structure symmetry along the two axes with the least score. 1 point for each axis that exhibits asymmetry.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Border</td>
<td>[0 - 8]/0.1</td>
<td>Lesion is divided in 8 slices. 1 point for each slice that has a clear border.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>[1 - 6]/0.5</td>
<td>Presence of up to 6 colours.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differential</td>
<td>[1 - 5]/0.5</td>
<td>Presence of up to 5 differential structure types: pigment networks, structureless or homogeneous areas, streaks, dots and globules. 1 point for each type present.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: ABCD rule

The formula for calculating the TDS is:

\[ TDS = (A\text{Score} \times 1.3) + (B\text{Score} \times 0.1) + (C\text{Score} \times 0.5) + (D\text{Score} \times 0.5) \]  

(1)

Where A\text{Score}, B\text{Score}, C\text{Score}, D\text{Score} correspond, respectively, to the scores obtained in the Asymmetry, Border, Colour and Differential Structures sections.

Finally, the TDS score is used to make a prognosis according to Table 2.

<table>
<thead>
<tr>
<th>TDS</th>
<th>Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;4.75</td>
<td>Benign melanocytic lesion</td>
</tr>
<tr>
<td>4.8 - 5.45</td>
<td>Suspicious lesion</td>
</tr>
<tr>
<td>Highly suspicious</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: TDS score interpretation

In 2004, Abbasi et al. [3] proposed that the ABCD rule should be revised to ABCDE, to include an E standing for evolution. They stated that a lesion’s evolution over time (e.g.: an increase in the total area of the PSL) is a very good indicator to be used in early melanoma detection. Menzies et al. [23] conducted a study that demonstrated that the presence of change in a lesion is a reliable sign of a developing melanoma, even by itself.

2.3. 7-Point Checklist

This diagnostic procedure has its foundation on the incidence of 7 standard criteria, selected by their frequent association with melanoma [2]. These criteria are sorted by their diagnostic weight, yielding two types of criteria. The major criteria are: atypical pigment network, blue-whitish veil and atypical vascular pattern. The minor criteria are: Irregular streaks, irregular pigmentation, irregular dots/globules and regression structures. Major criteria are given a score of 2 and minor criteria are given a score of 1. By adding the individual scores of all the criteria present in the lesion the total score of the lesion is determined. If the total score of a given lesion is 3 or more, it is classified as a highly suspicious lesion.

2.4. Clinical Relevance

Although the described melanoma screening procedures have had an impact in how early melanoma can be diagnosed, for patients with a large number of PSLs these methods can still be very time consuming [19]. Due to this, Computer Aided Diagnosis (CAD) systems were developed.

3. CAD Systems

CAD systems of digital dermoscopic images have been developed to aid in the clinical evaluation of skin lesions. They make use of the available technology, jointly with machine learning techniques and computer vision knowledge. Usually, the images to be evaluated by the CAD are treated prior to being fed to the system (e.g.: hair removal or color calibration). Then a CAD system will conform to the following framework steps [7][5][26][14]: image segmentation (also known as border detection), feature extraction, and, finally, lesion classification. The features extracted vary from system to system.

CAD systems can be roughly divided into two main groups: one group attempts to extract the same features used by clinicians when diagnosing a PSL (e.g.: asymmetry of the lesion, what differential structures are present, how many colors the PSL exhibits), while the other group uses standard image features (e.g.: color and texture) to apply statistical pattern recognition and machine learning tools [22]. Processing the dermoscopic image prior to inputting it in the CAD system, has shown to have a significant impact on the system performance [19]. The framework mentioned above is now explained.

3.1. Border Detection

A wide array of segmentation methods has been developed [6]. Most methods are based on thresholding, region growing and color transformation. Some other techniques have been developed that lie on the knowledge provided by artificial intelligence methods like fuzzy borders and declarative knowledge [19][27]. Some semi-automatic methods have also been developed. These methods usually use contour detection approaches using classical edge detectors that produce a collection of edges, and allow the user to select the most appropriate one.
3.2. Feature Extraction
The feature extraction step is dependent on the method of diagnosis to be used, be it emulating the diagnosis procedures used by clinicians or machine learning techniques.

The most common feature descriptors extracted are asymmetry, border, color and texture descriptors. Other relevant features in dermoscopy like regression structures are rarely used in CAD systems due to their complexity [20].

3.3. Lesion Classification
The last step is to generate the output desired: the classification of the lesion depicted in the input image. Depending on the type of system, the output can take several shapes. Most commonly, the output falls in one of three categories: binary (e.g.: suspicious/non-suspicious, malignant/benign), ternary (e.g.: melanoma/dysplastic nevus/common nevus), or n-ary, classifying n different types of PSL’s [19]. Usually the result is presented as some variation of malignancy percentage probability.

The learning stage that the CAD system performs prior to its very first classification is made on a labeled dataset, meaning that there is a set of images that have a label associated with them (each image has a ground truth diagnosis attached to it). The system will perform the described operations to extract the desired feature descriptors from the image. Only then are the feature descriptors used to classify the lesions. To classify the lesions, various classification methods are used on the extracted feature descriptors. After the learning has been made on a ground truth dataset, the system can then classify any subsequent input image.

3.4. Challenges found when developing a CAD system
A CAD system needs a dataset for training, testing and evaluation purposes. This, however, poses a fundamental problem: there aren’t many datasets available. Also, dataset creation has one major problem associated with it as well: ground truth variability. For example, the border detection step exhibits high intra- and inter-observer variability among dermatologists [8][17][15]. The dataset creation step includes manual segmentation of each individual lesion by a trained clinician. However, the most common approach is to treat this information as a fusion of several manual segmentations [19].

Typically, when annotating a dataset, a trained clinician must: manually segment the lesion, make a detailed description of a few features (e.g.: setting the ground truth for border values, analyzing asymmetry), and perform the ground truth diagnosis. This is very time consuming. This poses problems for dataset creation, which in turn, makes their availability more limited. Also, the databases available are not normalized to each other (e.g.: different magnification used, different color calibration), making benchmarking different methods, that have been trained in different datasets, expect different inputs, and express results in varying outputs, a cumbersome task.

Furthermore, while research of computerised analysis had been steadily increasing, spanning hundreds of publications, comprehensive literature reviews are scarce. This has lead to the appearance of a significant overlap between researchers work: without comprehensive benchmarking of the systems and technologies, the natural selection of CAD systems has been impaired.

To allow follow-up examination to be feasible, a PSL must be identified and compared with the data collected from the same lesion months before. To do so, it is necessary to perform correct body-to-image and image-to-image correspondence. Establishing this correspondence is often a cumbersome task.

There are several CAD systems that enable follow-up examination, but this is usually limited to allowing simultaneous exhibition of two images of the same lesion months apart. Then the changes in the lesion must be visually assessed by the clinician, with no quantitative information regarding changes in shape, area or color [19].

4. Multimodal Data Acquisition System
The proposed system stemmed from the attempt to address the lack of freely available labelled datasets for dermoscopy. The contributions that were born from this system are threefold: a framework to create a labeled dataset without the requirement of intensive manual labelling of the images in a dataset, to create a simple way to do lesion follow-up, and proposing a tele-dermatology framework.

The main goals of the system can be categorised as follows:

1. Image Acquisition: The system should be able to acquire the images that compose the dataset to be created.

2. Image Identification: Each image should have information associated with it. For any given image, the patient should be easily identifiable, as well as where, in the patient, the lesion is located.

3. Information Management: All the information pertaining to every image in the dataset should be handled automatically, without extra workload imposed on the user.

4. Lesion Follow-up: Correspondence between to lesions examined in two different doctor’s appointments, spread out in time, should be easily established for the user.

5. Tele-dermatology: The system should be able to provide a framework with which tele-dermatology appointments can be carried out.
The way the system is organised is now described (see Figure 2).

The system is composed of Google Glass, a detail camera, and a server. Google Glass is used to interact with the whole system and to take a picture of the whole acquisition process. It was chosen to be used in the system for several reasons. It was crucial that the whole system was not intrusive in the patient-doctor relationship. Google Glass performed a crucial role here. Also, the way to interact with Google Glass was desirable, since only voice commands are needed to operate the whole system.

The detail camera is used to acquire high-quality images of the lesions to be examined, since Google Glass’s camera proved insufficient.

The server is used to manage all the information acquired and to display it in a user-friendly GUI. This shows how goals 1 and 3 are achieved.

In order to achieve goals 2 and 4 extra information is needed. To do so, identifier tags were used. One of these tags is placed on the patient at the start of an acquisition cycle, and one permanently fixed on the detail camera. Using these tags, not only is the patient identified, but also, the location of the lesions can be determined. Due to this, goals 2 and 4 were also achieved.

Finally, to propose a tele-dermatology framework, the communication between all the parts of the system must be made wirelessly. As such, the data acquiring parts of the system (Google Glass and the detail camera) are the only parts required to be close to the patient. A computer can be used remotely to have access to all the data being acquired. Tele-dermatology appointments can then be performed, with the expert clinician needing only a computer connected wirelessly to the system through the internet. So, goal 5 was also achieved.

Upon using the proposed system, the interaction is usually comprised of a cycle of acquisitions. Each of these cycles is very similar from the point of view of the user. During one acquisition cycle the following events take place:

1. The user wakes Google Glass with the prompt and says a keyword that corresponds to the diagnosis of the lesion currently being acquired
2. Google Glass takes a photo
3. The detail camera takes a photo
4. The acquired data is displayed on the Graphical User Interface (GUI)

This cycle corresponds to adding to the database all the information pertaining to a single lesion on a patient’s skin. Prior to looping in this cycle, the user should start an acquisition session with the "New Patient" command. Then repeat the cycle for each lesion to be examined. When finishing the current acquisition session the cycle can be ended with the "Finish" voice command. This comprises a full acquisition session.

4.1. Interaction with the system

Interaction with the system is made exclusively with Google Glass’s voice commands. Upon starting to use the system the clinician is faced with Google Glass’s prompt screen. When the clinician says "ok glass", Google Glass will then respond to a voice command. Each voice command triggers a different routine. There are 3 main tasks:

1. Start a new acquisition session: The first thing to be done is to identify the current patient with his unique identifier tag, and check whether or not he has a file in the database. If he has one, all the information acquired is displayed. If he doesn’t, a new entry is created in the database.

2. Acquire the information pertaining to one lesion: There are several commands in Google Glass that trigger this routine. Each command corresponds to a diagnosis. The behaviour all of them trigger is the following: an image is
captured by Google Glass’s camera to allow image registration for image-to-image correspondence (lesion-to-image matching as well). Then, the detail camera is instructed to take a high quality picture of the lesion being inserted in the database. This is done with no action from the clinician: all the steps in the routine are triggered in the succession after the voice command is received by Google Glass. Finally, all the information is committed to the database and displayed in the GUI.

3. End the current acquisition session: All loose ends are tied from a computing stand-point. All the new information in the database is committed to disk, and the GUI is still fully available. Changes to the patient’s file can still be made in the computer.

A diagram of a complete acquisition session is now explained (see Figure 4). At first, both Google Glass and the Server are in stand-by. When Google Glass is prompted to start a new acquisition session by examining a new patient an image is acquired and sent to the server, along with the "new patient" command. The server checks the database for the patient found in the picture. If information is available it is displayed. Regardless of the availability of information in the database, the acquired image is stored and displayed. The server is now waiting to receive more images pertaining to this patient. Then, a series of detail images are acquired and labeled. These acquisitions are done simply by operating Google Glass: a voice command corresponding to a label is detected by Glass and an image is acquired by Google Glass’s camera. Then this image is sent to the server. The server stores it, registers the position of the lesion in the lesion map, and instructs the detail camera to acquire an image of the lesion. This process is repeated once for each lesion to be examined. Finally, when the acquisition session is done, the "finish" command should be given to Google Glass, and the whole system is again in stand-by.

4. End the current acquisition session: All loose ends are tied from a computing stand-point. All the new information in the database is committed to disk, and the GUI is still fully available. Changes to the patient’s file can still be made in the computer.

A diagram of a complete acquisition session is now explained (see Figure 4). At first, both Google Glass and the Server are in stand-by. When Google Glass is prompted to start a new acquisition session by examining a new patient an image is acquired and sent to the server, along with the "new patient" command. The server checks the database for the patient found in the picture. If information is available it is displayed. Regardless of the availability of information in the database, the acquired image is stored and displayed. The server is now waiting to receive more images pertaining to this patient. Then, a series of detail images are acquired and labeled. These acquisitions are done simply by operating Google Glass: a voice command corresponding to a label is detected by Glass and an image is acquired by Google Glass’s camera. Then this image is sent to the server. The server stores it, registers the position of the lesion in the lesion map, and instructs the detail camera to acquire an image of the lesion. This process is repeated once for each lesion to be examined. Finally, when the acquisition session is done, the "finish" command should be given to Google Glass, and the whole system is again in stand-by.

4.2. Lesion Localization

First Image 5(b) is examined. The goal is to determine the coordinates of the planar tag that rests atop the detail camera in the referential that corresponds to the planar tag that is associated with the patient. The diagram shown in Figure 6, should be enlightening regarding how the registration process takes place.

To determine the rotation matrix that relates the coordinates of both the tags, only algebraic manipulation is necessary. For simplicity of notation, the tag in the detail camera will be denoted TagC, with rotation matrix $R_c$ and translation vector $T_c$, and the tag in the patient will be denoted TagP, with rotation matrix $R_p$ and translation vector $T_p$. The rotation matrices and the translation vectors described are used to denote the relationship between the tags’ referentials and the referential of Google Glass’s camera, as seen in Figure 6.

The coordinates of TagC are $T_c$. Projecting them to the referential of TagP we get the coordinates of the lesion in this referential, denoted by the vector $P$:
\[ P = R_p^T T_p - R_p^T T_p \]  
(2)

And thus in \( P \) we have the coordinates of TagC in \( T_p \)’s referential. In the mapping image shown in Figure 5(a) the same tag is present in the same position, although, given that it is a different image, it has a different pose, and, as such, a different rotation matrix and translation vector. Denoting these as \( R_F \) and \( T_F \), we can now recover the coordinates of the lesion, \( P_F \) in the camera referential pertaining to this image:

\[ P_F = R_F P + T_F \]  
(3)

Finally, to establish the correspondence between the camera referential and the pixel coordinates, the camera matrix \( K \) is used:

\[
\begin{bmatrix}
    a \\
    b \\
    c
\end{bmatrix} = K P_F; \quad x = \frac{a}{c}, \quad y = \frac{b}{c}
\]  
(4)

where \( x \) and \( y \) are the horizontal and vertical pixel coordinates of the lesion, respectively.

5. Super-Resolution

5.1. TI-DTV

Over the course of this thesis, the image acquisition was always plagued by the notion that the images from a dermoscope were of much higher quality than the images acquired by the detail camera used. To try to overcome this limitation, while still being bound by the same constraints, a super-resolution algorithm was employed to the acquired images.

The spatial resolution of a system is defined by how closely two neighboring atomic entities can be distinguished. Super-resolution is defined, in a broad sense, in techniques designed to enhance the resolution of a sensing system. Its interest stems from the fact that sensing systems have a limit to the maximum resolution they can achieve. The goal of super-resolution methods is to recover High Resolution (HR) details from one or more Low Resolution (LR) input images. The methods can be broadly divided in two main categories: Multi-image super-resolution, and single-image Super-Resolution. Due to imaging blur and noise this is an ill-posed problem [24]. As such, additional constraints are required to be able to arrive at a solution (e.g.: smoothness of the solution, low-rank approximations, sparse representations of matrices, edge-preserving priors).

In a LR image there are two main types of image degradation present: blurring and down-sampling. Both of these degradations are inserted by the sensing system, and are different for each system. Due to this, the super-resolution problem (as well as image restoration in general) belong to the class of inverse problems. Several methods try to directly invert the down-sampling and blurring by upsampling the LR image and applying a deblurring kernel. Some methods try to estimate the HR information present in the image from several sub-pixel shifted LR images.

Example-based super-resolution has been shown to exceed the limits encountered by multi-frame super-resolution. However, the hallucinated HR details are not necessarily correspondent to the true HR details [12][10]. This type of single-image super-resolution attempts to ”hallucinate” the HR details present in the image. So example-based super-resolution methods are unfit for this system, since it must be used in a medical setting. However, the acquisition of several images would severely slow down the pace of usage of the system. This is also undesirable.

C. Granda and E. Cand`es [9] published a paper describing a technique to include high-level structure features into a super-resolution framework. He leads off by presenting an example using a Directional Total Variation (DTV) regularizer; suppose that the image that needs to be super-resolved is a checkerboard with the edges of the squares aligned perfectly with the vertical and horizontal directions. In this case, the gradient of the image will not only be sparse, it will be group sparse, and a proposed regularizer is:

\[
\text{DTV}(I) = \sum_{x=1}^{N_1} \sum_{y=1}^{N_2-1} (I(x, y+1) - I(x, y))^2 + \\
\sum_{y=1}^{N_2} \sum_{x=1}^{N_1-1} (I(x+1, y) - I(x, y))^2
\]  
(5)

where \( I \) is an image with \( N_1 \) and \( N_2 \) horizontal and vertical sizes, respectively. This regularizer enforces that the super-resolved image will prioritize images with group sparse gradients, meaning that most of the edges of the image will be aligned with the main horizontal and vertical directions. This regularizer, however, has very little practical utility, since seldom will the majority of the edges of an image be perfectly aligned with the horizontal and vertical directions. The authors then state that "images with vertical and horizontal edges then to be approximately low-rank when viewed as a matrix". Based on this knowledge, they suggest the existence of a transform \( \tau \) such that a general image \( I \) can be transformed to have most of its edges vertically or horizontally by applying the transform to it: \( I \circ \tau \). Thus, they suggest the more general regularizer Transform Invariant - DTV (TI-DTV):

\[
\text{TI-DTV}(I) = \text{DTV}(I \circ \tau)
\]  
(6)

This transform must be learned \textit{a priori}. To learn the correct transform to use, they use the Transform Invariant Low-rank Textures (TILT) framework [30].
The observation model shown by Candès and Granda is now explained. Given a LR image $I_{LR}$ with dimensions $n_1$ and $n_2$ the goal is to retrieve a HR image with dimensions $N_1$ and $N_2$ with $N_1 > n_1$ and $N_2 > n_2$ such that:

$$I_{LR} \approx D(K \otimes I_{HR}) \quad (7)$$

with $D$ being a down-sampling operator from $N_1 \times N_2$ to $n_1 \times n_2$, and $K$ a blurring kernel. TILT is applied a priori to the LR image in order to learn the appropriate transform $\tau$. Then, the following optimization problem is solved:

$$\min_{I_{HR}} \|D(K \otimes \hat{I}_{HR}) - I_{LR}\|_2$$
$$+ \lambda \cdot DTV(A_{\tau} \hat{I}_{HR})$$
$$+ \beta \cdot TV(\hat{I}_{HR}) \quad (8)$$

where the first term of the sum is a data fidelity term, $A_{\tau}$ is a matrix that maps $\hat{I}_{HR}$ to the space where its gradient is group sparse (i.e.: most of its edges are aligned vertically or horizontally), making the second term of the sum enforce the TILT-DTV regularizer, and the third term simply enforces sparsity overall (again, enforcing the intuition that the pixels of an image are not as wild as completely random variables). This method approaches super-resolution in a new way, that neither requires multiple-frames of the same scene, nor does it hallucinate the HR details, which, in a medical setting, where fidelity is of the utmost importance, is essential. Due to these features, this was the method used to super-resolve the images acquired in the present thesis.

5.2. TI-DTV - Example
The picture of a lesion taken with a regular camera was taken (shown in Figure 7), and then super-resolved (shown in Figure 8). The corresponding lesion was also documented with a dermoscope (shown in Figure 9) for comparison.

Figure 7: Super-Resolution: Original Camera Image
As can be seen from Figures 7, 8 and 9, although the gain in resolution from the used super-resolution framework is significant, the image quality is still far from the one found in images acquired with a dermoscope.

6. Results
6.1. Example of the system in a real life scenario
The first example is of a patient with two lesions in his left arm. First, the MatLab GUI is started. Then, instructing Google Glass to insert the current patient in the dataset, the first global image is captured (see Figure 10) and shown in the GUI.

Figure 8: Super-Resolution: Super-Resolved Camera Image
Figure 9: Super-Resolution: Dermoscope Image

Figure 10: Example: Global Image

After that acquisition has ended, the first lesion can be captured and the GUI is updated (see Figure 11).
Finally, after that acquisition ends, the third acquisition can capture the second lesion (see Figure 12).

With all lesions acquired, the lesion map is created (see Figure 13).

7. Conclusions
During the course of this thesis, 3 main limitations were encountered in the proposed system:

1. Google Glass is still a prototype: This android platform has yet to reach the market in its final form. With the advantages of using new technology (i.e.: allowing hands-free manipulation with wireless communication capabilities, and incorporated keyword recognition) came several problems. The Operating System (OS) present in Google Glass underwent monthly overhauls. In fact, during the course of this thesis, the OS morphed from version XE7 to version XE22. These updates were not optional, being pushed automatically by Google, and sometimes entailed changes to the Application Programming Interface. This rendered code obsolete, at times, after hours of being written. Furthermore, the lag between updating the Google Glass’s OS and updating the documentation led to a significant increase in the development time of the Glassware created for this thesis. Also, on two separate occasions, the update pushed by Google was faulty. In total, the camera API, which was of the essence for this thesis, was unavailable for approximately 3 months.

2. Google Glass’s camera: Although high in resolution, the field-of-view of the camera is very large. This proved counterproductive in a medical setting since the amount of pixels exhibiting relevant information is reduced and the amount of non-useful pixels in a picture skyrocket. This proved especially relevant during the image registration process due to the difficulty in recognizing the ArUco [11]tags.

3. Detail Camera: The dedicated detail camera used was a very poor substitute for a dermoscope. A dermoscope, apart from usually having a degree of magnification (most commonly 10x magnification), is an imaging instrument that works in conjunction with an immersion fluid, being able to provide for inspection both skin surface and subsurface structures that aren’t discernible with the naked eye. Of course, none of this is possible with a smartphone camera. This is the most important upgrade that should be made to the proposed system.

From information acquired from direct contact with dermatologists it can be stated that a system capable of organizing the acquired data automatically, creating a detailed, organized, interactive and user-friendly patient file. It can be asserted that addressing the proposed system’s limitations and creating a more mature product would be attractive not only to the developers of CAD systems, but also to the clinicians using them.

8. Future Work
The proposed system was successful in allowing a proof-of-concept to be established. The proposed future work frameworks are mainly ways to turn the system from a proof-of-concepts status to a fully working system.

The main changes to be made to the system pertain mainly to the upgrade of the components of the system. Replacing Google Glass with a voice activated software kit in a computer coupled with a camera in a fixed environment would, in one step, take care of the limitations imposed by Google Glass’s immature OS and address the lack of quality in its camera. Also, replacing the detail camera
with a dermoscope would enable the system to incorporate a CAD module in order to aid the clinician. Using this module it would be possible to, in one step, create a labeled state-of-the-art database and use the acquired data to learn and train a system, with feedback from specialists.

Another way to perform the image registration would be to leave the point-of-view image collection and instead equip a room with a set of cameras used in conjunction to perform more accurate image registration. The increase in the quality of the registration process would be at the cost of the portability of the system.

References


