

Lip segmentation based on Lambertian shadings and morphological operators for hyper-spectral images

Alessandro Danielis^{a,*}, Daniela Giorgi^a, Marcus Larsson^b, Tomas Strömberg^b, Sara Colantonio^a, Ovidio Salvetti^a

^a*Institute of Information Science and Technologies (ISTI) of the National Research Council (CNR), via G. Moruzzi 1, 56124 Pisa, Italy.*

^b*Linköping University, SE-581 83 Linköping, Sweden*

Abstract

Lip segmentation is a non-trivial task because the colour difference between the lip and the skin regions maybe not so noticeable sometimes. We propose an automatic lip segmentation technique for hyper-spectral images from an imaging prototype with medical applications. Contrarily to many other existing lip segmentation methods, we do not use colour space transformations to localise the lip area. As input image, we use for the first time a parametric blood concentration map computed by using narrow spectral bands. Our method mainly consists of three phases: (i) for each subject generate a subset of face images enhanced by different simulated Lambertian illuminations, then (ii) perform lip segmentation on each enhanced image by using constrained morphological operations, and finally (iii) extract features from Fourier-based modeled lip boundaries for selecting the lip candidate. Experiments for testing our approach are performed under controlled conditions on volunteers and on a public hyper-spectral dataset. Results show the effectiveness of the algorithm against low spectral range, moustache, and noise.

Keywords: Lip spatial pattern; Segmentation; Blood concentrantion map;

*Corresponding author. Tel. +39 050 621 3138
E-mail address: alessandro.danielis@isti.cnr.it; danielis.alessandro@gmail.com

1. Introduction

Lip contour segmentation has attracted much attention in the last years because is one of the fundamental enabling technologies for many of human-computer interaction applications, such as lip reading [1], audio-visual speech recognition [2], facial expression analysis [3], fatigue syndrome diagnosis [4], and so forth. For a recent review on lip segmentation techniques we refer to [5]. In general, lip segmentation is complicated by similar problems that are encountered with face detection, such as variability among subjects and lighting conditions. However, lip detection tends to be more sensitive to adverse conditions, e.g., lack of a sharp contrast between the lip and face, variability in shape, colour, and relation to surrounding features across individuals. Lip colour is one of the most important features used in lip detection research area [2, 6]. The first step in lip segmentation involves applying a colour transform to enhance the contrast between the lips and surrounding skin. Many approaches make use of pixel intensities in a particular colour space, e.g., HSV (Hue Saturation Value) or CIELAB [7, 8, 9], while several recent techniques use knowledge about the lip's colour to identify and track the lips [10]. However, there is much debate among researchers as to the best transform for this task [11]. Colour differentiation could not always guarantee a high degree of lips discrimination for images acquired with different imaging systems, which produce different colour values, even for the same subject under the same lighting conditions [12].

In this paper, we present our research effort for automatic lip segmentation towards a hyper-spectral imaging (HSI) system with medical applications [13], using spectral bands for human face characterisation and features extraction.

The main contribution of our work lie in the development of a fully au-

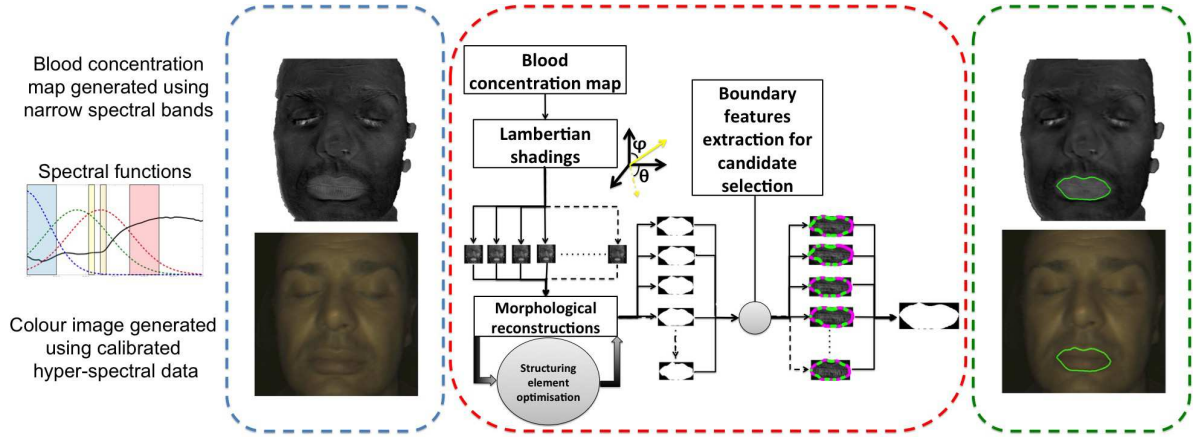


Figure 1: The flowchart of our proposed lip segmentation method.

automatic method able to detect the lip area on blood concentration face maps generated from HSI data by using narrow spectral bands, and robust against low spectral range information, moustache and added noise. We will show the potential of our concentration maps in providing a more discriminating lips' spatial pattern than the one generated by standard colorimetric spaces on hyper-spectral data. The flowchart of our proposed method is shown in Fig. 1. For each concentration map, we generate a subset of Lambertian shaded images as an enhancement step aimed at providing more discriminating lip patterns and making them less-subject specific. Then we perform morphological reconstructions under texture constraints for locating the lip region, and we model reconstructed and shaded lip boundaries to extract the segmented candidate in the subset. The experimental results have been conducted on hyper-spectral face images showing unexpressive lips acquired under controlled conditions by our HSI system on volunteers.

This paper is organised as follows. Section 2 overviews lip segmentation methods mostly related to our method. Section 3 introduces the input data of our algorithm, focusing on the blood concentration map, the technique used for

its generation, and its potential usefulness in lip detection for hyper-spectral images. Section 4 deals with the steps of our algorithm. Section 5 describes the experiments performed for testing our method, the dataset, the metric used for quantitative evaluation and the results obtained for lip detection against added noise. To show the robustness of our method against moustache and noise and the general improvement in performance of our framework, a comparison of the performance achieved with four fully automatic and fast active contour methods is carried out on two different hyper-spectral datasets. Section 6 gives some concluding remarks.

2. Related works

To the best of our knowledge, no lip segmentation method have been developed for hyper-spectral images in the Visible spectrum. Hyper-spectral data have been used in similar problems of tongue segmentation [14] and face recognition using manually selected areas (which include lips) and near infra-red bands [15]. In this Section, we focus on those techniques most related to our segmentation framework, highlighting their dependence on colour space transformations, and their unsuitableness to our problem, which lacks of sharp colour contrast images. Also, we briefly overview techniques using information about lip’s morphology.

Li and Cheung [16] employed colour transformations and morphological filters to segment the desirable lip region approximately based on the memberships. They proposed a colour transformation sequence in 1976 CIELAB and LUX colour spaces to establish a Gaussian model in a modified HSV colour space to calculate the memberships of lip and non-lip region. Using morphological filters to select the lip region candidate based on the two memberships, they finally extracted lip contour via convex hull algorithm.

Nath et al. [17] proposed a two stage lip contour extraction method based on RGB pixel intensity pattern variation analysis and morphological reconstruction. In their scheme, the first mouth region is extracted by using block threshold based binary conversion and some morphological operation. The RGB intensity pattern provided accurate estimate of lower lip region. Further processing was carried out by using morphological opening by reconstruction for detecting the upper lip region, which usually is a critical part to localise. Finally, they performed polynomial curve fitting to obtain the lip contour.

Such colour discriminating techniques will not be appropriate for low colour contrast images: a weak contrast between lip and skin across subjects acquired under controlled conditions makes challenging - especially in more adverse conditions - to find a colour space transform for robust and accurate lip enhancement/localisation.

Techniques that use information about the lip's morphology include active contour models (snakes) [18], shape models [19], and active appearance models [20]. Fast and automatic snake methods [21, 22] allow a contour to deform so as to minimize a given energy functional in order to produce the desired segmentation. An energy functional E is associated with the curve. The energy terms are defined in a way such that the final position of the contour will have a minimum energy. Existing active contour models can be categorized into two major classes: edge-based models, and region-based models. Edge-based use a continuous approximation of the original image function, so that boundary points can be characterized by a differential property (e.g., image gradient) and a contour representation be fitted to the boundary points. Region-based define energy functionals based on region statistics rather than local image gradient. Edge-based segmentation algorithms have better precision on the edges along the object boundary, whereas region-based segmentation algorithms are

robust against initial curve placement and insensitivity to image noise. We will show that both of these methods often converge to the wrong result when the lip edges are surrounded by beard/moustache or when the lip region is noisy. Region-based approaches present several advantages with respect to edge-based ones, including robustness against initial curve placement and insensitivity to noise. However, these techniques are derived on the base of a strong hypothesis, consisting in the fact that regions of interest can be differentiated in terms of their global statistics. In the practical situation of lip segmentation, the region to be segmented may be occluded for instance due to the presence of beard/moustache. In these situations traditional region based algorithms may return erroneous segmentation results. To accurately segment these regions, a new class of active contours energies which utilizes local information [23] instead of global one, appears to be a better solution. The initiative is that, if the homogeneity hypothesis is not satisfied globally, there's a greater probability for it to be satisfied within smaller sub-regions. Again, we will show that such technique is not enough accurate for our problem when compared to our proposed framework.

Shape models and active appearance models are much accurate than the previous snake methods, but require a large amount of storage, which is unattractive from a hardware perspective. Active contour models incorporated with shape prior knowledge [24] have proven to be accurate and fast at the same time, but they are not fully automatic and require detailed mathematical application. Those models are usually composed by terms which minimizes a feature-based energy functional and satisfies the object shape (lips in this context) prior constraint. The predefined lips/mouth shape is obtained from training data. In this sense, they cannot properly be considered fully automatic, since a training step is needed. Machine learning techniques have shown their success, but most

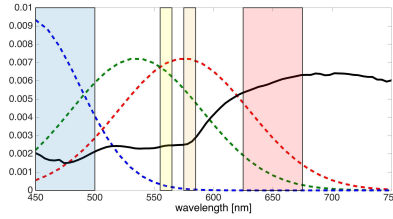
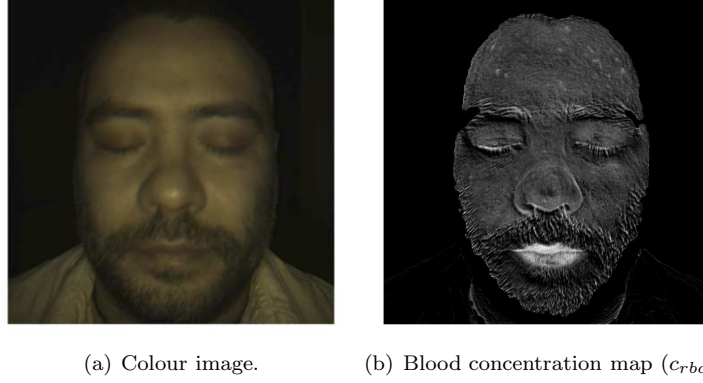
of them require large amount of storage for training [25].

3. Input data

The input data of our algorithm include colour images and blood concentration maps. These images were calculated from HSI data acquired using a monochrome camera equipped with an acousto-optic tunable filter (AOTF), an optical lens system and a white LED ring light. Before any further analysis the HSI data was dark corrected and white normalized using calibration images taken from a *Spectralon*[®] tile. The red, green and blue channels in the colour images were calculated as a weighted spectral average (Gaussian distribution specified by its Full-Width-Half-Maximum) using calibrated HSI data in the wavelength ranges $575 \pm 65\text{nm}$ (red), $535 \pm 65\text{nm}$ (green) and $445 \pm 50\text{nm}$ (blue), respectively. The Blood concentration maps were calculated based on the average spectral intensity in four different wavelength bands: 455-495nm, 558-562nm, 578-582nm and 630-670nm (see Appendix for details). The selected wavelengths are chosen based on existing bandpass filters (Edmund Optics) for estimating the concentration of red blood cells and the haemoglobin oxygen saturation from a simplified multispectral imaging system consisting of monochrome cameras, each with a filter central wavelength as of above. Fig. 2 shows the colour and blood concentration images of a subject in our dataset, the spectral functions used for calculating our RGB images and the spectral bands used for generating the blood concentration maps.

3.1. Blood concentration map

Skin colour is determined by scattering and absorption of light within the skin layers, caused mostly by concentrations of two chromophores, melanin and haemoglobin [26]. Changes in blood supply to sub-cutaneous districts result in change of optical properties of skin due to change in blood volume fraction



(c) Calibrated spectrum (solid black) and RGB weighting factors (dashed) used for calculating colour images, and spectral bands used for generating blood concentration maps.

Figure 2: Colour image and related blood concentration map generated using spectral functions shown in (c) and hyper-spectral data acquired by our HSI system.

and oxygenation. Thus, accurate analysis of optical properties of skin is a way to evaluate elementary haemodynamic parameters. These can relate to heart rate, oxygen saturation, peripheral resistances and endothelial function. Recent studies have demonstrated the capability of only using a subset of spectral bands for estimating the blood saturation and concentration on face images [27].

In this work, the local blood concentration map c_{rbc} is calculated using the four wavelength bands as specified above. These bands were specifically selected to maximise the sensitivity to the amount and saturation of blood while compensate for tissue scattering effects and melanin influence. Based on these four wavelengths, the light absorbed intensity $A(\lambda_i)$ can be iteratively estimated for the wavelengths 560nm and 580nm. Based on these absorbed intensities the

blood concentration maps are calculated as:

$$c_{rbc}l = \frac{A(560nm)}{2\mu_{a,tot}(560nm)} + \frac{A(580nm)}{2\mu_{a,tot}(580nm)} \quad (1)$$

where $\mu_{a,tot}(\lambda_i)$ are the total blood absorption coefficient at wavelength λ_i , respectively and l is the optical pathlength (Appendix). The $c_{rbc}l$ map is non-absolute and depends also on photon pathlength in tissue. The pathlength varies between subjects, but not over short time intervals within a skin site. In such map, the lip spatial pattern appears to be highly enhanced. Fig. 3 shows false-colour surface representations of the lip region obtained from the concentration map shown in Fig. 2b and colour space derived maps computed with standard colorimetric formulas [28] on the HSI colour image shown in Fig. 2a. All maps are represented under the same false coloring scheme. It can be

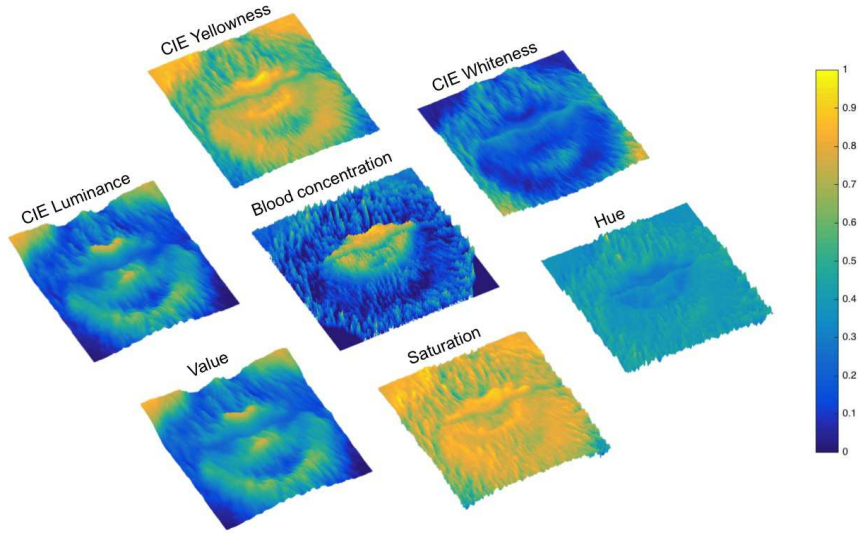


Figure 3: Lip surface (false-colour) pattern comparison: blood concentration (center) and some colorimetric maps computed from the HSI colour image shown in Fig. 2a.

observed that colour spaces do not provide discriminated feature representations

of lips and surrounding skin. On the contrary, concentration map discriminates well the lip spatial pattern, making easier the work of segmentation algorithms. The compactness and the brightness of the lip region suggested us the key idea to view and treat concentration maps as topographic images with specific morphology (height and connectivity).

4. Proposed method

In this Section we describe our segmentation algorithm for unexpressive lips on hyper-spectral images.

4.1. Pre-processing

First, to achieve a good initialisation and lessen the computational time, we pre-processed our colour images by using the Viola-Jones face detector, a method using Haar-like features, efficiently calculated using integral images [29]. Then, we filtered the corresponding and registered blood concentration maps by using very fast average filtering performed on their integral images, in order to make more compact the regions of interest and to overcome drawbacks coming from noisy images. In an integral image, every pixel is the summation of the pixels above and to the left of it. An integral image allows to rapidly calculate summations over image subregions in constant time, regardless of the neighborhood size. The steps of our pre-processing phase are shown in Fig. 4.

Although Viola-Jones face detector gives us useful results, their mouth detector framework is not enough accurate on our low colour contrast HSI images (see Fig. 17), also because not intended to be a precise lip segmentation algorithm.

4.2. Lambertian shadings

Here we report the technique used for enhancing the lip spatial pattern. The base idea is to exploit texture changes generated by different simulated

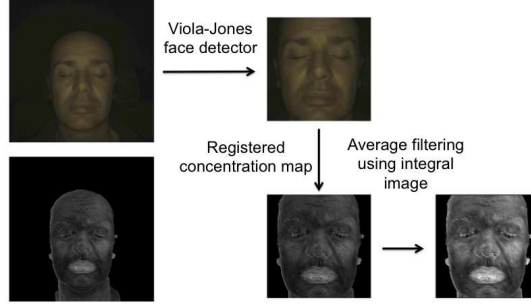


Figure 4: Image pre-processing based on Viola-Jones face detector and average filtering with integral image.

Lambertian illuminations. As a result, we obtain a subset of Lambertian shaded images for each map. Compared to other previous lip detection frameworks, we can view this stage as an alternative method to colour space transformations for lip enhancement.

Viewing the map as a topographic relief, the shading can be obtained by simulating a Lambertian light source, i.e., considering the cosine of the angle between the relief normal and light direction. Mathematically, given the generic normal vector $\mathbf{n} = (n_1, n_2, n_3)$ and the light direction

$$\mathbf{L}(\theta, \phi) = [\cos(\theta)\sin(\phi), \cos(\theta)\cos(\phi), \sin(\theta)] \quad (2)$$

specified by the spherical coordinates θ (elevation) and ϕ (azimuth), the Lambertian shading LS is generated by computing the dot product between surface normal and light as:

$$LS = L_1 \cdot n_1 + L_2 \cdot n_2 + L_3 \cdot n_3 \quad (3)$$

where L_1 , L_2 , and L_3 are the vector components of $\mathbf{L}(\theta, \phi)$. Different shaded relief images can be computed by specifying elevation and azimuth coordinates at which the surface is illuminated and a median gradient scale, used for gener-

ating a reasonable shading result. Given the input greyscale image $I(x, y)$, this is obtained by weighting the surface normals as:

$$\begin{aligned} n_1 &= -\frac{g_x \cdot gs}{\sqrt{1+g_x+g_y}} \\ n_2 &= -\frac{g_y \cdot gs}{\sqrt{1+g_x+g_y}} \\ n_3 &= 1 \end{aligned} \tag{4}$$

with

$$\begin{aligned} g_x &= \frac{\partial I(x, y)}{\partial x} \\ g_y &= \frac{\partial I(x, y)}{\partial y} \end{aligned} \tag{5}$$

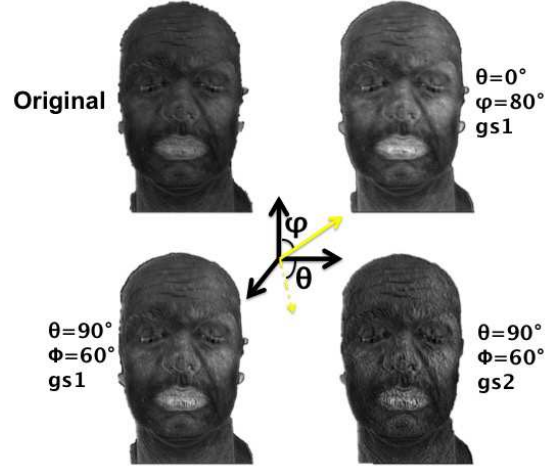
and gs the gradient scale computed as:

$$gs = \frac{sf}{\text{median}(g_x, g_y)} \tag{6}$$

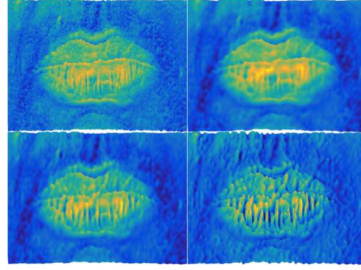
with sf a scale factor ranging in $(0, 1)$. The shaded image $J(x, y)$ is finally computed via matrix multiplication as:

$$J(x, y) = I(x, y) \cdot LS(x, y) \tag{7}$$

To illustrate the effect produced by Lambertian shadings on concentration maps, we report three different shadings performed with two gradient scale factors gs_1 and gs_2 (with $sf = 0.25$, $gs_2 = 3 \cdot gs_1$) and two directions specified by two pairs of (θ, φ) values (Fig. 5). Note that it is not possible to a priori know which shading could be the best enhancement for lip segmentation. A number of shadings, i.e., a number of Lambertian directions, which have the potential of enhancing the lip pattern at different scales, have to be performed in order to obtain a reasonable subset of shaded images, providing a powerful input to



(a) Original concentration map and three Lambertian shadings.



(b) Corresponding surface (false-colour) lip regions.

Figure 5: Lambertian shadings illustration.

the next processing step and making easier the segmentation work.

To motivate and assess the use of Lambertian shadings as a contrast enhancement step, we will compare segmentations obtained from non-shaded concentration maps with our full method, i.e., with shaded and post-processed images (see Section 5.3). Also, we will compare the local enhancement provided by the shadings with one generated by another enhancement operation.

4.3. Constrained morphological reconstructions

At this stage, we morphologically reconstructed Lambertian shaded images in order to detect bright and compact regions, which satisfy specific texture constraints. Morphological reconstruction is one of the major parts of morphological image processing, and has already demonstrated its usefulness for image segmentation tasks [30]. Greyscale reconstruction by geodesic erosions/dilations is a transformation involving two images, the marker and the mask, and a structuring element [31]. The marker image is the starting point of the transformation, while the mask constrains the transformation. The structuring element controls the operation and can be binary or greyscale. The high points, or peaks, in the marker image specify where processing begins. The peaks spread out, or dilate, while being forced to fit within the mask image. The spreading processing continues until the image values stop changing.

Formally, considering two functions f and g , with $f \geq g$, and defined on the same domain D_I , the greyscale reconstruction $R_f(g)$ of mask I from marker g is computed by iterating greyscale geodesic erosions of g above f until stability is reached:

$$R_f^\varepsilon(g) = \bigwedge_{n \geq 1} \varepsilon_f^n(g) \quad (8)$$

where \bigwedge stands for the pointwise minimum and $\varepsilon_f^n(g)$ is the greyscale geodesic erosion of size $n \geq 0$ given by:

$$\varepsilon_f^n(g) = \varepsilon_f^1 \circ \varepsilon_f^1 \circ \dots \circ \varepsilon_f^1(g) \quad (9)$$

iterated n -times, with ε_f^1 the elementary greyscale erosion given by:

$$\varepsilon_f^1 = (g \ominus SE) \vee f \quad (10)$$

where \vee stands for the pointwise maximum and $g \ominus SE$ is the erosion of g by the

flat structuring element SE . The reconstruction by geodesic dilations of mask I from marker g has a dual definition with respect the former one [30], and is obtained by iterating greyscale geodesic dilations of g under I until stability has reached:

$$R_f^\delta(g) = \bigvee_{n \geq 1} \delta_f^n(g) \quad (11)$$

where \bigvee stands for the pointwise maximum and $\delta_f^n(g)$ is the greyscale geodesic dilation of size $n \geq 0$ given by:

$$\delta_f^n(g) = \delta_f^1 \circ \delta_f^1 \circ \dots \circ \delta_f^1(g) \quad (12)$$

iterated n -times, with δ_f^1 the elementary greyscale dilation given by:

$$\delta_f^1 = (g \oplus SE) \wedge f \quad (13)$$

where \wedge stands for the pointwise maximum and $g \oplus SE$ is the dilation of g by the flat structuring element SE .

To discriminate the lip region from other skin areas, we characterised the lip spatial pattern with specific texture properties, accounting for the fact that in the proposed concentration map they appear as compact and bright areas, peaks in topographical terms. Starting with a defined flat ball-shaped structuring element, we performed a sequence of reconstruction by geodesic erosions followed by reconstruction by geodesic dilations, in order to efficiently extract regional maxima areas from our greyscale images. The shape of the structuring element ensured that the scale of erosion/dilation is equal in all directions. This sequence of reconstructions is optimised for localizing bright regions, which satisfy texture constraints. Once the structuring element dimension has been initialised for performing morphological reconstructions, the optimisation is achieved by reducing the structuring element dimension until the constraints were satisfied

(Fig. 7). The implementation of the dimension optimisation is reported in Algorithm 1. Approximating the detected regions of interest with ellipses, we computed the following region descriptors:

- orientation (**or**) (see Fig. 6).
- connectivity (**c**), the number of holes in the region.
- solidity (**sol**), a measure that specifies the proportion of the pixels in the convex hull that are also in the region. Calculated as $\text{Area}/\text{convexArea}$.
- eccentricity (**ecc**) of the ellipse, i.e., the ratio of the distance between the foci of the ellipse and its major axis length.

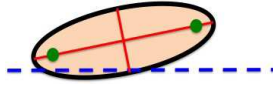


Figure 6: Ellipse approximating a generic detected lip region of interest. The solid red lines represent the axes, the small green circles are the foci, and the orientation **or** is the angle between the horizontal dotted line and the major axis.

In this work, we assumed and characterised the unexpressive lip regions as areas having low (**or**), (**c**)=1 (i.e., closed lips with no holes), and high ($\text{sol} \times \text{ecc}$) values.

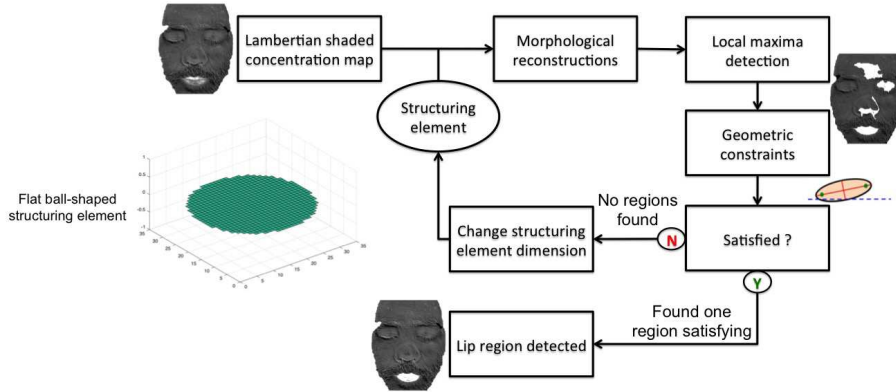


Figure 7: Illustration of the constrained morphological reconstruction algorithm.

Algorithm 1: Structuring element optimisation.

Input: Lambertian shaded concentration image I_{LS} ; flat ball-shaped structuring element SE

Output: Reconstructed image I_o

```
1  $dim$  = initialiseDimension;  
2 while not satisfied do  
3    $dim$  = dim-1;  
4    $I_{RE}$  = reconstructByErosions( $I_{LS}$ ,  $SE(dim)$ );  
5    $I_{RED}$  = reconstructByDilations( $I_{RE}$ ,  $SE(dim)$ );  
6    $M$  = regionalMaxima( $I_{RED}$ );  
7    $P$  = regionProperties( $M$ );  
8   satisfied = checkProperties( $I_{RED}$ ,  $P$ );  
9    $I_o$  =  $I_{RED}$ ;  
10 return  $I_o$ 
```

4.4. Lip boundary extraction

In this post-processing phase, we modeled and characterised lip boundaries in order to automatically extract the boundary candidate from each subset of reconstructed and shaded images.

In the past years, some researchers have adopted different models to keep the geometric shape of lip contours. In computerised shape analysis, it is desirable to classify objects using robust features, which are independent of scaling, translation, and rotation, and possibly not affected by roughness coming from noise. To lessen the effect produced by noisy boundaries mostly caused by moustache and added noise, we extracted features known as Fourier descriptors based shape factors (FDF) and fractional concavity/convexity (FCC, FCV), which have already demonstrated their usefulness for other image processing applications [32, 33]. The Fourier descriptor based shape factor (FDF) is a measure related to the presence of roughness or high-frequency components in a contour. In [34], the normalized magnitudes of the FDFs for various types of shapes have demonstrated to be different enough to aid in classification. Per-

soon and Fu [35] showed that the FD's may be used to characterize skeletons of objects, with applications in character recognition and machine-part recognition. Fourier descriptors (FD) are not only useful for the creation of shape factors, but also for smoothing outlines or reducing an arbitrary outline to a specified number of points. In this study, we used FD to both derive the shape factor FDF and to derive a polygonal model to represent the contour, in order to identify and analyse lip concavities and convexities in a localised manner.

We treated the lip boundary as a union of piecewise continuous and locally-salient concave/convex parts. The segments are used to compute the angles subtended by the individual concave/convex parts at their vertices.

4.4.1. Polygonal modeling

Due to undesirable roughness coming from segmentation process, a smoothing technique, which anticipates boundary approximation, can be useful sometimes. In this work, we generated lips' smoothed contours by computing the forward elliptic Fourier (fEF) transform [36] for each contour. This transform creates a "shape spectrum" of a closed (x,y) outline. Then, we applied the reverse elliptic Fourier (rEF) transform to take a specific number of harmonics from the spectrum and reconstruct the smoothed (x,y) outline version. Given the number of harmonics N_h obtained from the shape spectrum computed with the fEF, the number of boundary pixels N_p , the DC components (i.e., first harmonics) A_0 and C_0 in the x and y projections, the Fourier approximations of a closed contour in the x and y projections are:

$$\begin{aligned} x_n(p) &= A_0 + \sum_{n=1}^{N_h} a_n \cdot \cos\left(\frac{2n\pi t}{N_p}\right) + b_n \cdot \sin\left(\frac{2n\pi t}{N_p}\right) \\ y_n(p) &= C_0 + \sum_{n=1}^{N_h} c_n \cdot \cos\left(\frac{2n\pi t}{N_p}\right) + d_n \cdot \sin\left(\frac{2n\pi t}{N_p}\right) \end{aligned} \tag{14}$$

where p is a generic point of the reconstructed smoothed boundary. The achieved contours are then interpolated with periodic cubic bivariate spline, i.e., with splines that match first and second derivatives at end points. Given the vector coordinates (x^{FD}, y^{FD}) defining the FD-derived boundaries, and the total number N of x^{FD} pixels, the interpolation nodes were defined as a vector containing $N/2$ evenly spaced points with spacing $\frac{N-1}{N/2-1}$. The final result is a smoothed polygonal approximation of lip boundaries (Fig. 8).

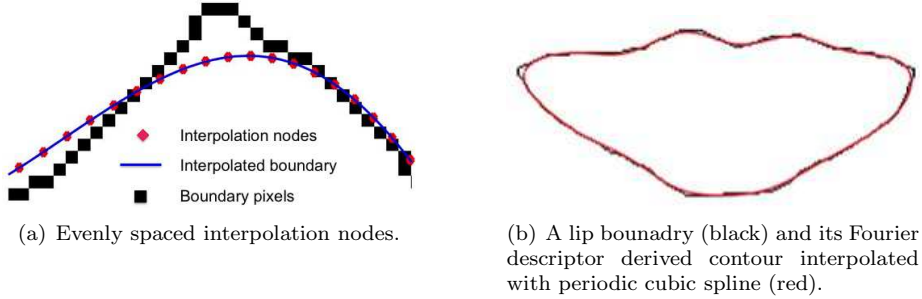


Figure 8: Spline interpolation of a FD-derived boundary with evenly spaced interpolation nodes.

Concavity/convexity modeling. In this study, we labeled the individual segments between inflection points as concave/convex parts by using the polygonal modeling of above. Given the FD-derived smoothed version $f(x,y)$ of a lip contour, and the periodic cubic spline interpolated form $g(x,y)$ of $f(x,y)$, the spline approximation $z(x,y)$ used for the labeling procedure is defined as

$$z(x,y) = g(x,y) + g'(x,y) \quad (15)$$

where $g'(x,y)$ is the first order derivative of $g(x,y)$. Portions of the functions $f(x,y)$ and $z(x,y)$ computed for a segmented lip contour are shown in Fig. 9. We define as concave parts all the $z(x,y)$ segments inside the polygonal region or on

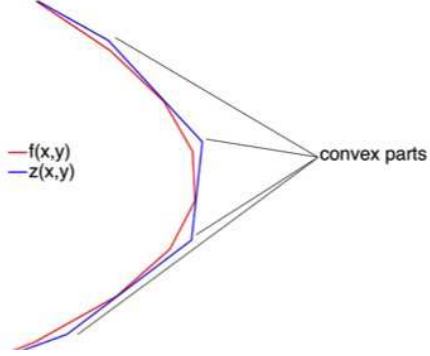


Figure 9: Illustration of the labeling procedure used to discriminate convex from concave parts. Segments (blue) out of the polygonal boundary region (red) are convex parts.

the polygonal boundary represented by $f(x,y)$, as convex parts otherwise. Fig. 10 shows the contour model of a lip region overlaid with magenta and green portions, which represent the convex and concave parts, respectively.

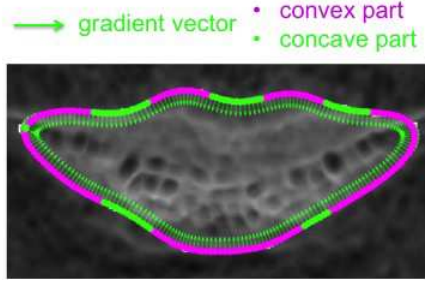


Figure 10: Concavity/convexity lip boundary modeling.

4.4.2. Fractional concavity and fractional convexity

Let S_i , $i = 1, 2, 3, \dots, N$, be the lengths of N segments obtained from the polygonal model of a lip boundary. The total length of the boundary T_l is computed as the cumulative length of the N segments:

$$T_l = \sum_{i=1}^N S_i \quad (16)$$

Given CC_i , $i = 1, 2, 3, \dots, P$, the lengths of P concave segments, and CV_i , $i = 1, 2, 3, \dots, Q$, the lengths of Q convex segments, the cumulative concave length (CC_l) and convex length (CV_l) are computed as:

$$\begin{aligned} CC_l &= \sum_{i=1}^P CC_i \\ CV_l &= \sum_{i=1}^Q CV_i \end{aligned} \tag{17}$$

The fractional concavity (FCC) and convexity (FCV) values are computed by normalising the respective cumulative lengths by the total boundary length, so that the sum of the concavity and convexity fractions for a given boundary equals unity:

$$\begin{aligned} FCC &= CC_l / T_l \\ FCV &= CV_l / T_l \end{aligned} \tag{18}$$

The fractional parameters are normalised with respect to the total length of the boundary, thus making them independent of the size of the boundary. The FCC and FCV parameters are complementary to each other; hence, we used only FCV as a feature in our lip pattern classification experiments.

4.4.3. Fourier descriptor based shape factor

In the computation of FD's, the (x,y) coordinates of each point in the periodic array of N boundary points of an object are represented as complex values, given by $z_i = x_i + jy_i$, with $i=0,1,\dots,N-1$. Note that the boundary points refer to boundary pixels of the original contour of the object. The FD's are defined as:

$$\begin{aligned} Z(k) &= \frac{1}{N} \sum_{i=0}^{N-1} z_i \cdot \exp[-j\frac{2\pi}{N}nk] \\ n &= 0, 1, \dots, N-1 \end{aligned} \tag{19}$$

Using the procedure proposed by Shen et al. [37] to compute a single shape factor, the FD's are normalized as follows: $Z(0)$ is set equal to 0 in order to make the descriptors independent of position, and each coefficient is divided by the magnitude $Z(1)$ in order to normalize for size. After these steps the magnitudes of the FD's are independent of positions, size, orientation and starting point of the contour; also, the orientation and starting point affect only the phase of FD's. The normalized Fourier descriptors (NFD's) are defined as:

$$NFD(k) = \begin{cases} 0, & k=0 \\ Z(k)/Z(1), & k=1,2,\dots,N/2 \\ Z(k+N)/Z(1), & k=-1,2,\dots,-N/2+1 \end{cases} \quad (20)$$

Contours with sharp excursions have more high-frequency components than smooth contour. However, applying a weighting factor that increases with frequency leads to unbounded values that are also sensitive to noise. A shape factor FDF based upon the NFD's was defined by [37] as:

$$FDF = 1 - \frac{\sum_{k=-N/2+1}^{N/2} ||NFD(k)||/||k||}{\sum_{k=-N/2}^{N/2} ||NFD(k)||} \quad (21)$$

The advantage of this measure is that it is limited to the range (0,1), and is not sensitive to noise. This measure is invariant to translation, rotation, starting point, and contour size, and becomes larger as the contour shape becomes more complex and rough.

4.4.4. Implementation

In this work, we used the shape factors described above as indeces of lip boundary regularity. Due to their characteristic shape, Fourier based modeled unexpressive lip boundaries have major portions of convex segments. On the

other hand, noisy segmented boundaries have both concave and convex segments as well as prominent spicules. We characterised this property by computing FCV values for each segmented lip boundary. On the basis of the differences in shape between unexpressive lips and noisy lips, we observed that the former have low FDF and high FCV values. Fig. 11 shows FDF and FCV values of manual and automatic segmentations from two shaded images obtained from one concentration map shaded with different angles (azimuth and elevation) of illumination. Note that the images here reported are two shaded maps of the subset composed by Lambertian shaded images obtained from the same input concentration map (see Fig. 1). The manual segmentation on the right of the 1st row is marked with many spicula in order to enhance the difference with the smooth manual one in the same row, and highlight the numerical evidence for the statement of above. The shading simulated with $\phi = 180$ and $\theta = 60$ permits to obtain a more accurate segmentation (see 2nd row), which can be selected by using the FCV and FDF shape factors.

We implemented simple steps for extracting the candidate lip boundary (Algorithm 2). For each Lambertian shaded and reconstructed subset of images,

Algorithm 2: Boundary candidate extraction.

Input: Modeled segmented boundaries \mathbf{B} ; feature vectors \mathbf{FDF} and \mathbf{FCV}

Output: Candidate boundary \mathbf{C}

```

1  $\mathbf{MSA} = \text{MostSegmentedArea}(\mathbf{B});$ 
2  $\mathbf{mFDF} = \text{evaluateMeanOnMSA}(\mathbf{FDF}, \mathbf{MSA});$ 
3  $\mathbf{mFCV} = \text{evaluateMeanOnMSA}(\mathbf{FCV}, \mathbf{MSA});$ 
4  $\mathbf{ind}_1 = \text{find}(\mathbf{FDF} < \mathbf{mFDF});$ 
5  $\mathbf{ind}_2 = \text{find}(\mathbf{FCV} > \mathbf{mFCV});$ 
6  $\mathbf{C} = \text{merge}(\mathbf{B}(\mathbf{ind}_1), \mathbf{B}(\mathbf{ind}_2));$ 
7 return  $\mathbf{C}$ 

```

we computed the mean of FCV (\mathbf{mFCV}) and FDF (\mathbf{mFDF}) values on the area

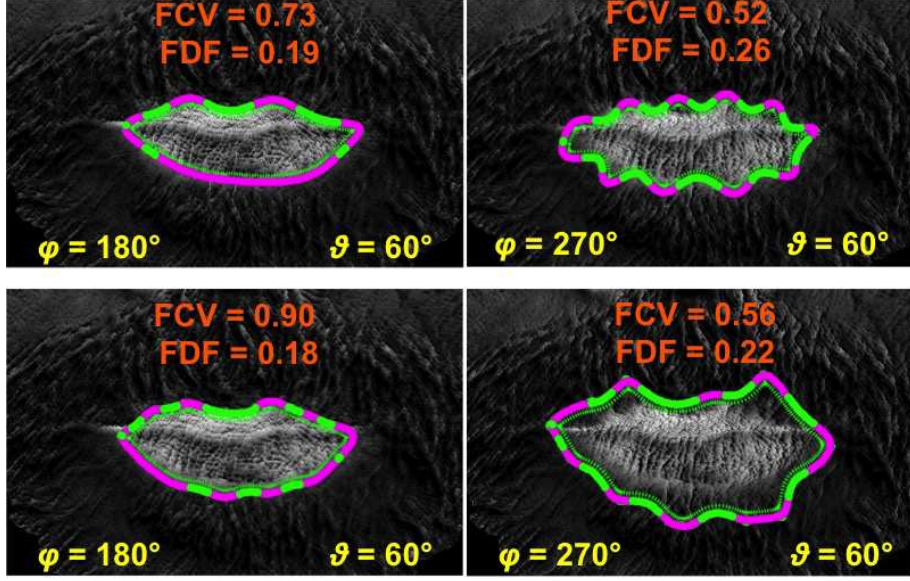


Figure 11: Lips contour manually segmented on two differently shaded images obtained from the same concentration map (1st row), and automatically segmented by our method on the same images (2nd row). The shadings are obtained with the azimuth (ϕ) and elevation (θ) angles here reported.

containing the highest number of segmented boundaries. This avoids extraction of wrong contours coming from noisy regions with similar texture properties of lip area. Then, based on the assumption that regular boundaries have higher FCV and low FDF values, we selected only those with FCV values lower than $mFCV$ and higher than $mFCF$, and, finally, we merged them. The binary images $B(ind_1)$ and $B(ind_2)$ are matrices containing the segmented candidates at pixel locations indicated by the index vectors ind_1 and ind_2 respectively, and zeros elsewhere. The final segmentation C is obtained by simple logical *or* (\vee) operation as:

$$C = B(ind_1) \vee B(ind_2) \quad (22)$$

This is equivalent to perform a simple summation of the matrices $(B(ind_1) + B(ind_2))$.

5. Experimental results and dataset

In this Section, we show the effectiveness of our method for unexpressive lip segmentation against moustache, lack of sharp contrast between lip and face, and three different added random noise distributions.

The implementation of the above-stated method was performed with the following parameters. We used a 5×5 averaging filter to reduce the noise effect. We initialised the ball-shaped structuring element radius to 15 pixels. To generate reasonable shading lip patterns, we computed 12 Lambertian shadings for each image in our dataset by setting: $\theta = [60^\circ \ 70^\circ \ 80^\circ]$, $\phi = [0^\circ \ 90^\circ \ 180^\circ \ 270^\circ]$, and gs is computed with the scale factor $sf = 0.25$. We used 30 harmonics for computing smoothed interpolated forms of lip contours. The experiments ran on an Intel Core i7 2.5 GHz machine with *Matlab*[®] *R2015b* image processing toolkit, and the averaging time consuming in seconds of the proposed algorithm is on average 0.2 s per image subset, 2.4 s (0.2×12) per final segmented image.

The algorithm was tested on experimental data consisting of 1392 face images of sizes 1000×1000 , generated from 29 subjects acquired by the *Wize Mirror* HSI system located at a fixed distance from the subjects. The volunteers, that took part in this experiment, were asked to look into the camera under controlled indoor illumination, and extra requests were made to ensure relatively perfect frontal images, such as a normal pose and a neutral expression.

The dataset was generated as follows. For each subject, we generated 2 images using shading gradient scales $gs_1 = 2 \cdot gs$, and $gs_2 = 3 \cdot gs$, plus a set of 9 images by adding Gaussian, Speckle and Salt & Pepper noise models [31] combined with three different shadings calculated with gs , gs_1 , and gs_2 . We iterated each random noise distribution 5 times, thus obtaining a total of 1 (original) + 2 (gs_1 , gs_2) + 3 (noises) \times 3 (original, gs_1 , gs_2) \times 5 (random noise distributions) = 48 images per subject. Fig. 15 shows a subset of 12 images

obtained with 1 iteration of this process, i.e., with 1 Gaussian, Speckle and Salt & Pepper random noise distributions in combination with three gradient scale shading factors (gs , $gs1$, $gs2$). The numerical parameters used to define the three noise models, are:

- Gaussian: mean = 0; variance = 0.05;
- Speckle: mean = 0; variance = 0.05;
- Salt and Pepper: noise density = 0.05.

All parameters are normalized, i.e., they correspond to operations with images with intensities ranging from 0 to 1.

5.1. Experiment 1

To evaluate the performance of the algorithm, we consider the segmentation problem as a binary classification and use the metric defined in [38], for measuring the overlap between the segmented lip region S and the ground truth GT as:

$$OL = \frac{2(S \cap GT)}{S + GT} \times 100\% \quad (23)$$

The ground truth of each image was obtained by manual segmentation. In order to reduce possible drawbacks coming from manual segmentation, the ground truth was segmented three times independently by the same person, and the experiment was evaluated using the three corresponding ground truths. Fig. 12 shows segmentation results on concentration and colour images of subjects that have authorized for image publication. Results show that lips' spatial pattern of concentration connected with a single subject corresponds well across individuals and against different random noise distributions (see Table 1 and Fig. 15), which are often hard to eliminate. Average rates do not change significantly with the ground truth used for evaluation. To quantify the performance

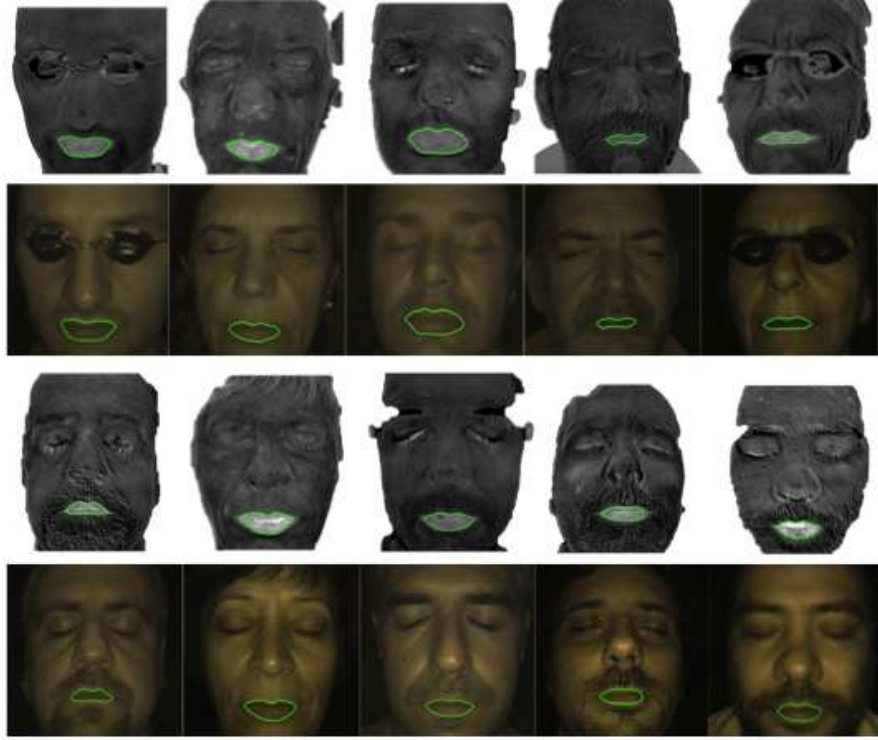


Figure 12: Lips contour segmented on original concentration maps (1st and 3rd rows) by our method, and superimposed on corresponding colour images (2nd and 4th rows).

achieved with respect to the use of different gradient scale factors and against different random noise distributions, we report the histograms and scatter plots of Fig. 13. We constructed a matrix M composed by 3 columns M_1 , M_2 , M_3 each containing $3 \text{ (noises)} \times 5 \text{ (random distributions)} \times 29 \text{ (subjects)} = 435$ OL rates achieved using the gradient scale factor gs , $gs1$, and $gs2$, respectively, and the ground truth $GT1$. Columns M_1 , M_2 , and M_3 have an average rate of about 91%, 90%, and 88% respectively (last row in Table 2). This demonstrates that the Lambertian shading performed with gs for this dataset works better than the other gradient scale factors, and that, overall, Lambertian shading provides meaningful enhancements for the segmentation of the lip spatial

Table 1: Average *OL* rates defined using 3 manual segmentations (*GT1*, *GT2*, *GT3*) for subsets of images degraded with different random noise distributions (*N1*, *N2*, *N3*, *N4*, *N5*) and with three gradient scales.

Segmentation performance			
Noise	GT1	GT2	GT3
<i>N1</i>	90%	88%	89%
<i>N2</i>	90%	89%	90%
<i>N3</i>	90%	89%	89%
<i>N4</i>	89%	90%	90%
<i>N5</i>	89%	90%	90%
All	90%	89%	90%

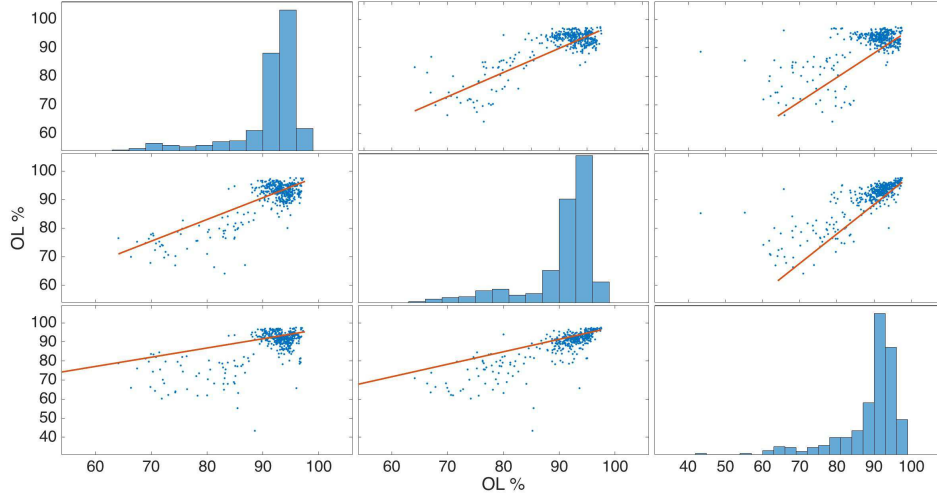


Figure 13: Histograms and scatter plots of the matrix M columns M_1 , M_2 , and M_3 each containing 3 (noises) \times 5 (random distributions) \times 29 (subjects) = 435 *OL* rates achieved using the gradient scale factor gs , $gs1$, and $gs2$, respectively, and the ground truth *GT1*. The subplot in the i th row, j th column of the matrix is a scatter plot of the i th column of M against the j th column of M . Along the diagonal are histogram plots of each column of M .

pattern. Histograms show that the worst segmentation rates achieved with gs and $gs1$ are on average not less than 70%, while the ones obtained by using $gs3$ are around 60%. Figure shows some worst segmentation rates obtained on lips

Table 2: Average OL rates defined using ground truth $GT1$ and obtained using 3 gradient scales ($gs, gs1, gs2$) for subsets of images degraded with different random noise distributions ($N1, N2, N3, N4, N5$).

Segmentation performance per gradient scale factor			
Noise	gs	gs1	gs2
$N1$	91%	90%	88%
$N2$	91%	90%	89%
$N3$	91%	90%	89%
$N4$	91%	89%	88%
$N5$	91%	90%	88%
All	91%	90%	88%

degraded with Gaussian and Speckle noise, and shaded with $gs3$ factor. It can

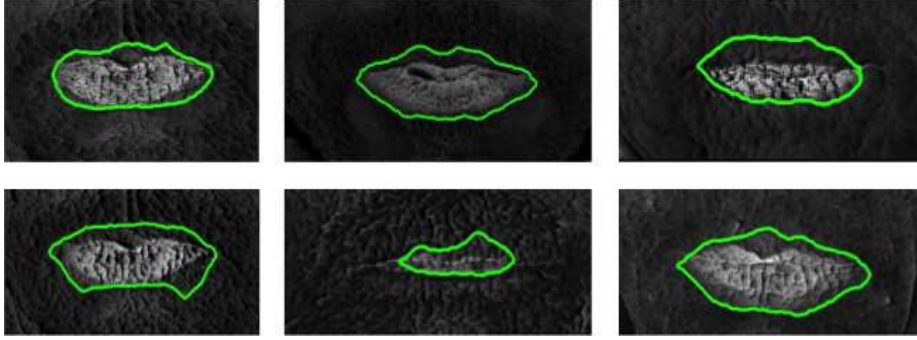


Figure 14: Lips contour segmented with rates equal or less than 70% on concentration maps degraded with Speckle (1st row) and Gaussian (2nd row) noise, and $gs3$ shading factor.

be observed that the lip area is correctly localized but not properly segmented by our method due to the high deterioration mainly generated by the $gs3$ shading factor. The OL rates distribution almost approximate a linear function in all scatter plot combinations, meaning that random noise distributions do not affect significantly the performance of our method.

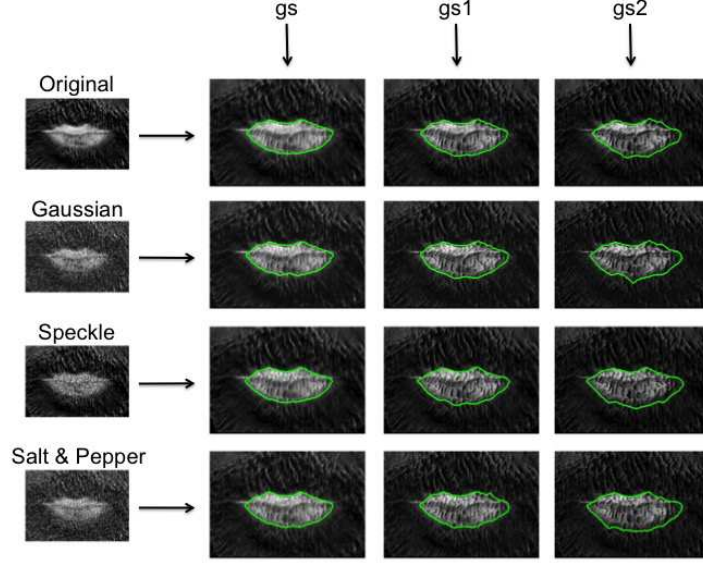


Figure 15: Lips contour segmented on one original concentration map (first row of the figure) and on its corresponding degraded images, by our method. Below the 1st row, the lip images in the i th row of the figure have been degraded with the same random noise distribution (Gaussian, Speckle or Salt & Pepper), while images in the j th column have been shaded with the same gradient scale shading factor (gs , $gs1$ or $gs2$).

5.2. Experiment 2

In this Section, we compare the performance of our method with respect to four techniques based on active contour methods on our dataset and on the Stanford public indoor hyper-spectral dataset established by Skauli and Farrell [39]. The public database contains 43 facial images, which have been recorded with HySpex line-scan imaging spectrometer covering the spectral ranges 0.4 to 1 micrometers. In particular, each hyper-spectral face image includes a total of 148 spectral bands spanning the visible and NIR range from 415 to 950 nm in steps of 4 nm. The image dimensions differ from one subject to another and range from a minimum of 516×351 pixels to a maximum of 1560×1016 pixels. Fig. 16 shows one RGB image rendered under D65 illuminant, and the corresponding blood concentration map generated with the technique described

in Section 3.1 using hyper-spectral data from the Stanford dataset.

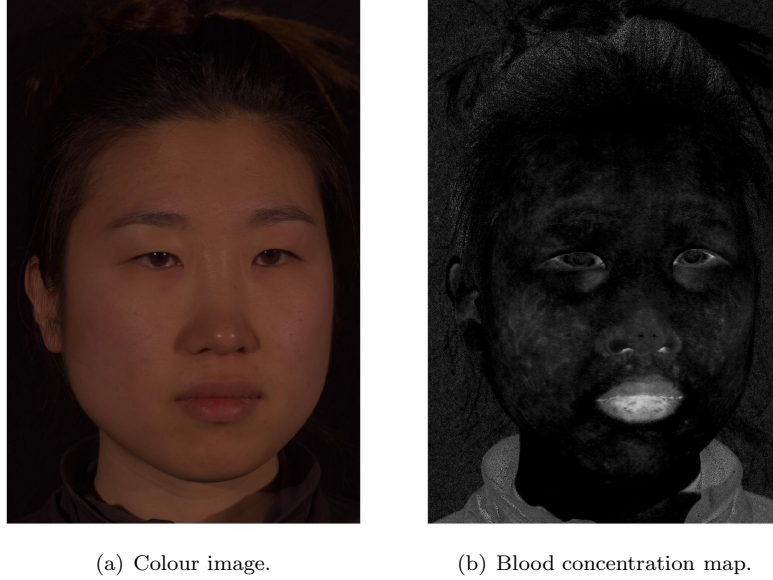


Figure 16: RGB colour image and related blood concentration map generated using hyper-spectral data from the public Stanford dataset.

For all the proposed methods, as it is commonly done with active contour models [18], a mask close to the boundary of interest has to be initialized. For this purpose, we used the Viola-Jones mouth detector (VJMD) framework to automatically detect the mouth region on our colour images, providing an input mask that can be fast processed by snakes. As can be noted, the VJMD is not much accurate in detecting the mouth on such colour images (see Fig. 17).

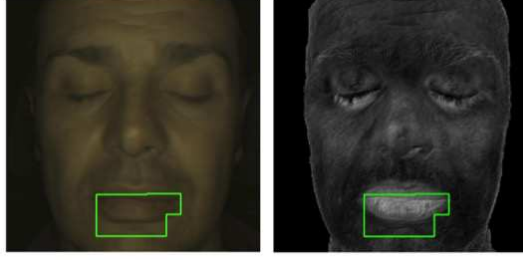


Figure 17: Viola-Jones mouth detector performance on one colour image of our dataset (left), and mouth area superimposition on its corresponding concentration map (right).

Then, the snake evolves the detected contour towards the outline of the object of interest.

The first active contour method we propose is an edge-based (EB) model [21], which uses image gradients in order to identify object boundaries. One benefit of this type of snake is the fact that no global constraints are placed on the image.

The second method here reported is based on the Chan and Vese active contour (CVAC) segmentation algorithm [22], which is a region-based model.

The last method we test on our dataset is the local region-based framework for guiding active contours developed by Lankton and Tannenbaum (LT) [23]. This technique can improve the segmentations provided by the CVAC on heterogeneous objects, i.e., objects that cannot be easily distinguished in terms of global statistics. The LT algorithm can be used to localize any region-based energy. Here we test two different energies localized by the LT framework: the Chan and Vese energy (LTCV) and the Yezzi energy (LTY) [40].

For a quantitative comparison, we report *OL* rates achieved with EB, CVAC, LTCV, LTY and our method on original and degraded maps generated from our dataset (see Table 3) and from the Stanford dataset (see Table 4). The total number of images used in this test from our dataset is (1 original map + 3

noise models \times 5 random distributions) \times 29 subjects = 466, from the Stanford dataset is (1 original map + 3 noise models \times 5 random distributions) \times 43 subjects = 688. Segmentations obtained with active contour methods from the Stanford dataset are on average more accurate than those from our private dataset, mainly because of the reduced presence of faces with beard/moustache. Some contours segmented by CVAC and LTAC are very accurate on concentration maps from the public dataset (see Fig. 18), confirming the effectiveness of our parametric maps. However, as can be observed from the cases reported in Fig. 19, the Chan and Vese model converges to wrong results in presence of noise and/or moustache. As expected, contrarily to our method, snakes are very sensible to noise and moustache.

Table 3: Average *OL* rates defined using ground truth *GT1* and obtained using the gradient scale (*gs*) for original concentration maps and maps degraded with Gaussian, Speckle and Salt & Pepper noise models, with edge based method, Chan and Vese model, localized energy models implemented by Lankton and Tannenbaum, and our proposed method.

Performance comparison on our dataset				
Method	Original maps	Noise		
		Gaussian	Speckle	Salt & Pepper
Edge based	60%	60%	60%	60%
Chan and Vese	69%	65%	59%	60%
Localized Chan and Vese	75%	72%	72%	65%
Localized Yezzi	68%	65%	65%	61%
Our	91%	90%	90%	90%

It is important to note that all the active contour methods here tested need of an initial mask before segmentation. Thus, in presence of noise or low contrast between the lip region and the surrounding area, the segmentation would be more sensitive to the position of the initialized mask. Fig. 20 illustrates some segmentations obtained with LTY method from masks initialized by VJMD detector. As can be noted, segmentations from highly enhanced lip areas are

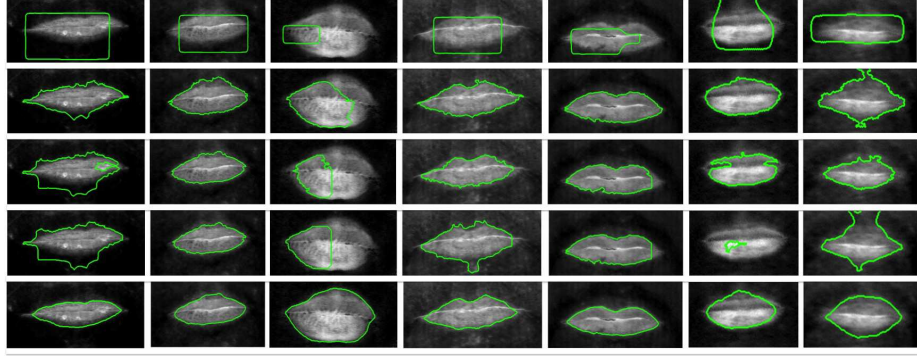
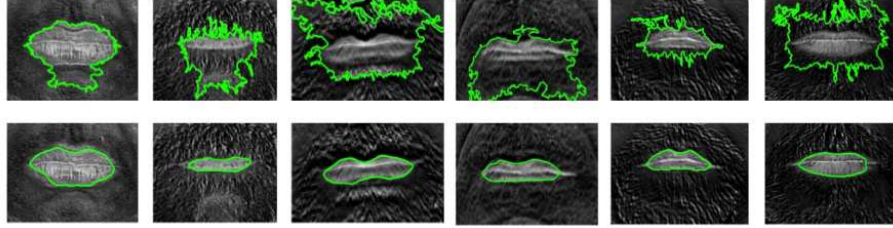
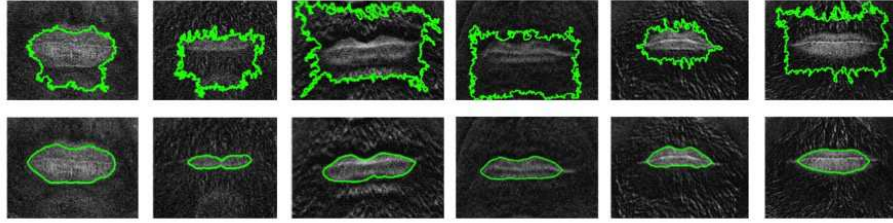


Figure 18: Lips contour segmented by edge-based active contour method (1st row), Chan and Vese model (2nd row), localized Chan and Vese model (3rd row), localized Yezzi method (4th row), and our proposed method (5th row) on concentration maps calculated from seven facial images of the Stanford database.



(a) Lips contour segmented by Chan and Vese model (1st row) and our proposed method (2nd row) on original concentration maps.



(b) Lips contour segmented by Chan and Vese model (1st row) and our proposed method (2nd row) on concentration maps with Gaussian noise.

Figure 19: Segmentation performance comparison on some images from our dataset.

more accurate than others from lower contrast lip patterns and which start from masks initialized within moustache. In cases with very highly enhanced lip areas, segmentations obtained by LT algorithm are still quite accurate, but not as those from our proposed algorithm.

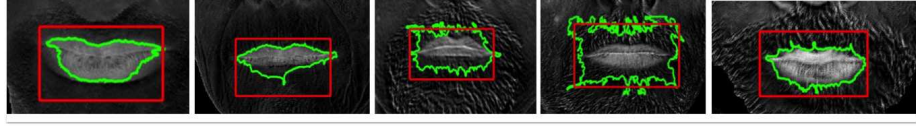


Figure 20: Lips contour segmented by localized Yezzi method (green) and corresponding masks (red) initialized by Viola-Jones mouth detector.

Fig. 21 shows segmentations obtained with our method on some original concentration maps calculated from the Stanford dataset with our algorithm. Contours are also superimposed on the corresponding RGB images calculated by using the standard illuminant D65, the CIE 1964 colour matching functions and all the available wavelenghts within the visible spectrum (from 414 nm to 702 nm with step of 4 nm).

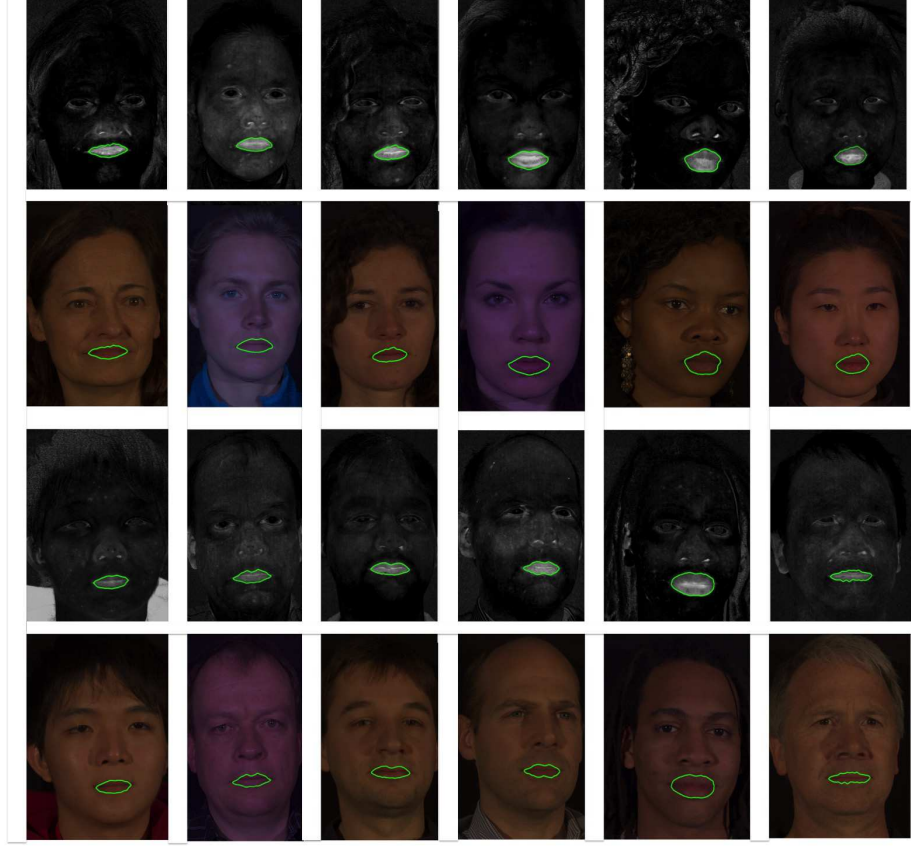


Figure 21: Lips contour segmented on original concentration maps (1st and 3rd rows) by our method on some images from the Stanford dataset, and superimposed on corresponding colour images (2nd and 4th rows).

For both the dataset, we set 200 iterations to perform the active contour methods, and the averaging time consuming in seconds of these four proposed segmentation techniques is on average 1 second for the Edge method, 4 seconds for the Chan and Vese algorithm, and 5 seconds for the two localized active models.

Using much more complex active contour models may be advantageous in terms of achieved accuracy, but useless from a hardware perspective that requires low power consumption and fast calculations. Also, although more com-

Table 4: Average *OL* rates defined using manual ground truth and obtained using the gradient scale (*gs*) for original concentration maps and maps degraded with Gaussian, Speckle and Salt & Pepper noise models from the Stanford database, with edge based method, Chan and Vese model, localized energy models implemented by Lankton and Tannenbaum, and our proposed method.

Performance comparison on the Stanford public dataset				
Method	Original maps	Noise		
		Gaussian	Speckle	Salt & Pepper
Edge based	57%	57%	57%	57%
Chan and Vese	74%	65%	59%	60%
Localized Chan and Vese	77%	73%	72%	67%
Localized Yezzi	59%	56%	55%	53%
Our	89%	88%	88%	88%

plex active contours can be more accurate for lip segmentation, a training step is needed for obtaining accurate segmentations.

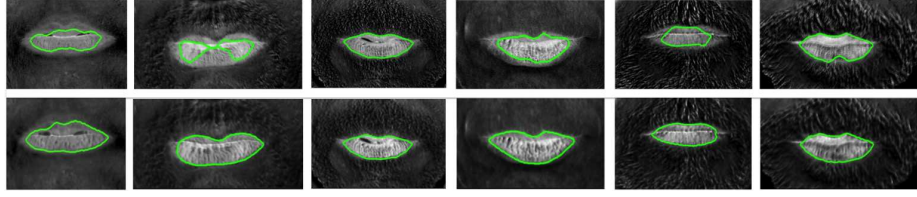
5.3. Experiment 3

To show the effectiveness of Lambertian shadings as an enhancement step, we compare the segmentation results from our full algorithm with those obtained from raw concentration maps, i.e., maps not enhanced by the shading. Note that in this case the lip contours were segmented without the boundary extraction step described in Section 4.4, but by only using the constrained morphological reconstructions. Fig. 22 reports some segmentation results obtained with and without Lambertian shadings on original concentration maps, and on maps with Gaussian noise. The contrast enhancement provided by Lambertian shadings improves the segmentation accuracy on concentration maps (see Table 5).

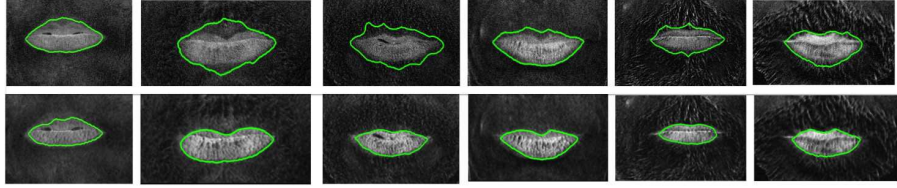
Fig. 23 shows that segmentations from our full algorithm are much more robust against added noise (Gaussian in this case) if compared to segmentations from a non-shaded concentration map. We can distinguish five different segmentations from the same non-shaded map with five different random Gaussian noise

Table 5: Average *OL* rates defined using manual ground truth and obtained on shaded and non-shaded concentration maps from our database.

Performance comparison on the Stanford public dataset				
Method	Original maps	Noise		
		Gaussian	Speckle	Salt & Pepper
Without shading	83%	81%	76%	80%
Our full method	91%	90%	90%	90%



(a) Lips contour segmented on raw (non-shaded) concentration maps (1st row) and our full method (2nd row).



(b) Lips contour segmented on raw (non-shaded) concentration maps with Gaussian noise (1st row) and our full method (2nd row) on maps with Gaussian noise.

Figure 22: Segmentation performance improvement with Lambertian shadings.

distributions, while we obtain almost the same contour on the corresponding Lambertian shaded map. This shows that Lambertian shadings provide better input images for lip segmentation algorithms on blood concentration maps.

6. Conclusion

We proposed an automatic lip detection technique for hyper-spectral images without using colour space transformations, shape or active contour models, and machine learning. We showed that our three-steps algorithm gives promising results in detecting the lip region by using a greyscale face blood concentration

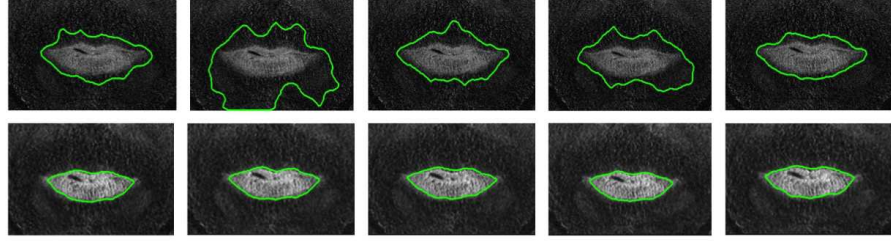


Figure 23: Segmentation robustness versus added noise. Lip contour segmented on one raw (non-shaded) concentration map (1st row) and our full method (2nd row) on the same shaded map. Five different Gaussian random noise distributions (with same constant mean and variance values) were added to the same concentration map.

map computed with few spectral bands. Our experimental results showed a large improvement in performance compared to four competitors on two different datasets. Blood concentration maps have shown its potential in providing a meaningful discriminated lip spatial pattern. The algorithm exploits image/surface variations caused by changes in simulated Lambertian illuminations. Segmentations obtained with Lambertian shadings are much more accurate and robust against added noise than those from raw concentration maps. This enhancement technique is used to generate a subset of differently shaded images, which will be inputs for further processing steps aimed at selecting the best lip candidate obtained from one shaded image. We believe that this novel application of Lambertian shadings as an enhancement step could be carried out or improved by other segmentation techniques. The proposed boundary descriptors well characterise the lip regions acquired under controlled conditions of normal pose and neutral expression. The experimental controlled conditions permitted to evaluate the effectiveness of the concentration map for the segmentation of unexpressive lips on hyper-spectral images against moustache, low spectral information, and added noise. Overall, our method represents a promising approach for lip segmentation on hyper-spectral images with medical applications.

Conflict of interest

None declared.

Acknowledgements

The authors are grateful to the volunteers for photo database capturing, to the bio-optics group at Norwegian University of Science and Technology (NTNU) for the acquisition of the images, and to Dr. Giuseppe Coppini, researcher at the Institute of Clinical Physiology (National Research Council of Italy - CNR, Pisa) for his collaboration. Special thanks goes to Dr. Arianna Mencattini from the University of Rome “Tor Vergata”, for her suggestions. The work described in this paper was performed in the framework of the FP7 Specific Targeted REsearch Project SEMEOTICONS (SEMEiotic Oriented Technology for Individual’s Cardiometa**bol**ic risk self-assessme**N**t and Self-monitoring) partially funded by the European Commission under Grant Agreement 611516.

The authors would like to express their gratefulness to the referees and editor for their useful comments and suggestions that improved this paper.

Appendix

The blood concentration maps were calculated using a modified Beer-Lambert law algorithm based on four wavelength bands. The algorithm was designed to have a maximal sensitivity to variations in blood concentration and oxygenation while compensating for scattering effects and melanin influence. This was achieved by selecting two closely located spectral bands, around 560nm and 580nm, where a maximal difference between the saturated and de-saturated haemoglobin absorption spectra is displayed. These two wavelength bands were also located to a region where blood absorption has a strong influence on the

backscattered light intensity, maximising their sensitivity to blood concentration. To compensate for scattering effects and blood amount variations, two additional wavelength bands were selected around 475nm and 650nm; spectral regions where the blood absorption has a less dominant effect on the diffuse reflectance.

The estimation of blood amount and blood saturation is based on an iterative Beer-Lambert law algorithm that is tailored for the four selected wavelength bands. According to the Beer-Lambert law, the detected intensity can be described by:

$$I(\lambda) = I_0(\lambda)e^{-c_{rbc}l\mu_{a,tot}(\lambda)} \quad (24)$$

with $\mu_{a,tot}(\lambda)$ the total absorption coefficient defined as

$$\mu_{a,tot}(\lambda) = sa_1(\lambda) + (1 - s)a_0(\lambda) \quad (25)$$

$a_1(\lambda) = \mu_{a,HbO_2}(\lambda)$ is the absorption coefficient for saturated blood, $a_0(\lambda) = \mu_{a,Hb}(\lambda)$ is the absorption coefficient for de-saturated blood, c_{rbc} is the tissue fraction blood, s is the blood saturation and l is the photon pathlength in tissue. The absorbed intensity can thus be described by:

$$\begin{aligned} A(\lambda) = \ln(I_0(\lambda) - \ln(I(\lambda))) = \\ c_{rbc}l(sa_1(\lambda) + (1 - s)a_0(\lambda)) \end{aligned} \quad (26)$$

where $I_0(\lambda)$ is the maximal detectable intensity related to a zero-absorption tissue. The magnitude of $\ln(I_0(\lambda))$ for the 560nm and 580nm bands cannot be measured directly. However, an initial estimation can be done using linear regression of the absorbances $\ln(I(475nm))$ and $\ln(I(650nm))$. This estimation is based on the assumption that $\ln(I_0) \approx \ln(I)$ for the 475nm and 650nm band (i.e., $A(475nm) \approx 0$ and $A(650nm) \approx 0$, due to a low absorption coefficient

of blood), and that the spectral appearance of $\ln(I_0(\lambda))$ is fairly linear in the range 475nm to 650nm (Fig. 24a). To further compensate for a non-zero blood absorption in the 475nm and 650nm bands, $A(475\text{nm})$ and $A(650\text{nm})$ are iteratively updated from the previous estimated value of $A(560\text{nm})$ by assuming that the pathlength doesn't change with wavelength. This results in

$$A_n(\lambda_i) = A_{n-1}(560\text{nm}) \frac{\mu_{a,tot}(\lambda_i)}{\mu_{a,tot}(560\text{nm})} \quad (27)$$

where λ_i is either 475nm or 650nm, and $A_n(\lambda_i)$ is the estimated absorbance difference in iteration n. The absorbance difference $A_{n-1}(560\text{nm})$ is calculated using linear regression as described above, with an initial assumption: $A_0(\lambda_i)=0$. Stable estimations of $A(\lambda_i)$ are typically found in less than five iterations (Fig. 24b). The blood concentration dependency and the pathlength dependency can be cancelled out, assuming that the latter is equal for all wavelengths, by calculating the ratio:

$$R = \frac{A(580\text{nm}) - A(560\text{nm})}{A(580\text{nm}) + A(560\text{nm})} = \frac{s(a_1(\lambda_2) - a_1(\lambda_1) - a_0(\lambda_2) - a_0(\lambda_1)) + a_0(\lambda_2) - a_0(\lambda_1)}{s(a_1(\lambda_2) + a_1(\lambda_2) - a_0(\lambda_2) - a_0(\lambda_1)) + a_0(\lambda_2) + a_0(\lambda_1)} = \frac{sk_1 + k_2}{sk_3 + k_4} \quad (28)$$

using the estimated absorbance difference found after a few iterations. Solving for the saturation s gives:

$$\hat{s} = s = \frac{k_2 - k_4 R}{k_3 R - k_1} \quad (29)$$

where k_i can be calculated from tabulated data on blood absorption and \hat{s} denotes the estimated saturation. Based on the calculated saturation, the A_i estimations and tabulated data for the blood absorption coefficients, the relative

tissue fraction of blood (i.e., blood concentration C) can be calculated as:

$$C = c_{rbc}l = \frac{A(560nm)}{2\mu_{a,tot}(560nm)} + \frac{A(580nm)}{2\mu_{a,tot}(580nm)} \quad (30)$$

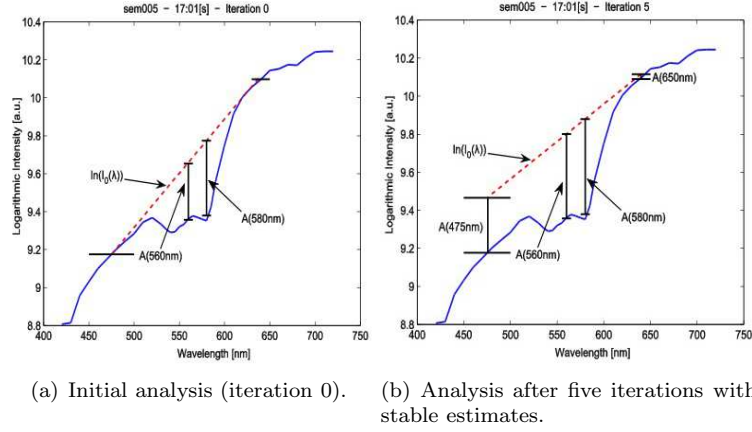


Figure 24: Example of an analyzed HSI spectrum.

Bibliography

- [1] S. Stillittano, V. Girondel, A. Caplier, Lip contour segmentation and tracking compliant with lip-reading application constraints, *Machine Vision and Applications* 24 (1) (2012) 1–18.
- [2] X. Zhang, R. M. Mersereau, Lip feature extraction towards an automatic speechreading system, in: *Image Processing, 2000. Proceedings. 2000 International Conference on*, Vol. 3, 2000, pp. 226–229 vol.3.
- [3] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, K. Scherer, Meta-analysis of the first facial expression recognition challenge, *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 42 (4) (2012) 966–979.

- [4] Y. Chen, W. Liu, L. Zhang, M. Yan, Y. Zeng, Hybrid facial image feature extraction and recognition for non-invasive chronic fatigue syndrome diagnosis, *Comput. Biol. Med.* 64 (C) (2015) 30–39.
- [5] M. Li, On study of lip segmentation in color space, Ph.D. thesis, Hong Kong Baptist University (2014).
- [6] A. W.-C. Liew, S. H. Leung, W. H. Lau, Lip contour extraction from color images using a deformable model, *Pattern Recognition* 35 (12) (2002) 2949–2962.
- [7] X.-g. Shen, W. Wu, An algorithm of lips secondary positioning and feature extraction based on ycbcr color space, in: *International Conference on Advances in Mechanical Engineering and Industrial Informatics*, Atlantis Press, 2015.
- [8] R. Rohani, S. Alizadeh, F. Sobhanmanesh, R. Boostani, Lip segmentation in color images, in: *Innovations in Information Technology, 2008. IIT 2008. International Conference on*, 2008, pp. 747–750.
- [9] S.-L. Wang, W.-H. Lau, A. W.-C. Liew, S.-H. Leung, Robust lip region segmentation for lip images with complex background, *Pattern Recognition* 40 (12) (2007) 3481 – 3491.
- [10] D. Nguyen, D. Halupka, P. Aarabi, A. Sheikholeslami, Real-time face detection and lip feature extraction using field-programmable gate arrays, *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 36 (4) (2006) 902–912.
- [11] A. D. Gritzman, D. M. Rubin, A. Pantanowitz, Comparison of colour transforms used in lip segmentation algorithms, *Signal, Image and Video Processing* 9 (4) (2014) 947–957.

- [12] S. Lucey, S. Sridharan, V. Chandran, Adaptive mouth segmentation using chromatic features, *Pattern Recognition Letters* 23 (11) (2002) 1293 – 1302.
- [13] S. Colantonio, G. Coppini, D. Germanese, D. Giorgi, M. Magrini, P. Marraccini, M. Martinelli, M. A. Morales, M. A. Pascali, G. Raccichini, M. Righi, O. Salvetti, A smart mirror to promote a healthy lifestyle, *Biosystems Engineering* 138 (2015) 33 – 43.
- [14] Z. Liu, J. qi Yan, D. Zhang, Q.-L. Li, Automated tongue segmentation in hyperspectral images for medicine, *Appl. Opt.* 46 (34) (2007) 8328–8334.
- [15] Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyperspectral images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (12) (2003) 1552–1560.
- [16] M. Li, Y.-m. Cheung, Artificial Neural Networks – ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15-18, 2010, Proceedings, Part I, 2010, Ch. Automatic Segmentation of Color Lip Images Based on Morphological Filter, pp. 384–387.
- [17] R. Nath, F. S. Rahman, S. Nath, S. Basak, S. I. Audin, S. A. Fattah, Lip contour extraction scheme using morphological reconstruction based segmentation, in: *Electrical Engineering and Information Communication Technology (ICEEICT)*, 2014 International Conference on, 2014, pp. 1–4.
- [18] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, *International Journal of Computer Vision* 1 (4) (1988) 321–331.
- [19] T. Cootes, C. Taylor, D. Cooper, J. Graham, Active shape models-their training and application, *Computer Vision and Image Understanding* 61 (1) (1995) 38 – 59.

- [20] I. Matthews, T. Cootes, J. Bangham, S. Cox, R. Harvey, Extraction of visual features for lipreading, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24 (2) (2002) 198–213.
- [21] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, *International Journal of Computer Vision* 22 (1) (1997) 61–79.
- [22] T. Chan, L. Vese, Active contours without edges, *IEEE Trans Image Process* 10 (10) (2001) 266–277.
- [23] S. Lankton, A. Tannenbaum, Localizing region-based active contours, *Image Processing, IEEE Transactions on* 17 (11) (2008) 1–11.
- [24] T. H. N. Le, M. Savvides, A novel shape constrained feature-based active contour model for lips/mouth segmentation in the wild, *Pattern Recognition* 54 (2016) 23 – 33.
- [25] Y.-m. Cheung, M. Li, Q. Peng, C. P. Chen, A cooperative learning-based clustering approach to lip segmentation without knowing segment number.
- [26] J. Jimenez, T. Scully, N. Barbosa, C. Donner, X. Alvarez, T. Vieira, P. Matts, V. Orvalho, D. Gutierrez, T. Weyrich, A practical appearance model for dynamic facial color, *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 29 (6) (2010) 141:1–141:10.
- [27] M. Ewerlöf, E. G. Salerud, T. Strömberg, M. Larsson, Estimating skin blood saturation by selecting a subset of hyperspectral imaging data, in: *Proc. SPIE*, Vol. 9328, 2015, pp. 93280J–93280J–10.
- [28] R. S. Berns (Ed.), *Billmeyer and Saltzman’s Principles of Color Technology*, Wiley, 2000.
- [29] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: *Computer Vision and Pattern Recognition, 2001. CVPR 2001.*

Proceedings of the 2001 IEEE Computer Society Conference on, Vol. 1, 2001, pp. I-511–I-518 vol.1.

- [30] L. Vincent, Morphological grayscale reconstruction in image analysis: applications and efficient algorithms, *Image Processing, IEEE Transactions on* 2 (2) (1993) 176–201.
- [31] R. C. Gonzalez, R. E. Woods, *Digital Image Processing* (3rd Edition), Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [32] E. Persoon, K. S. Fu, Shape discrimination using fourier descriptors, *IEEE Transactions on Systems, Man, and Cybernetics* 7 (3) (1977) 170–179.
- [33] R. M. Rangayyan, T. M. Nguyen, Fractal analysis of contours of breast masses in mammograms, *Journal of Digital Imaging* 20 (3) (2006) 223–237.
- [34] L. Shen, R. M. Rangayyan, J. L. Desautels, Application of shape analysis to mammographic calcifications, *Medical Imaging, IEEE Transactions on* 13 (2) (1994) 263–274.
- [35] E. Persoon, K.-S. Fu, Shape discrimination using fourier descriptors, *Systems, Man and Cybernetics, IEEE Transactions on* 7 (3) (1977) 170–179.
- [36] F. P. Kuhl, C. R. Giardina, Elliptic fourier features of a closed contour, *Computer graphics and image processing* 18 (3) (1982) 236–258.
- [37] L. Shen, R. M. Rangayyan, J. E. L. Desautels, Application of shape analysis to mammographic calcifications, *IEEE Transactions on Medical Imaging* 13 (2) (1994) 263–274.
- [38] A. W. C. Liew, S. H. Leung, W. H. Lau, Segmentation of color lip images by spatial fuzzy clustering, *IEEE Transactions on Fuzzy Systems* 11 (4) (2003) 542–549.

- [39] T. Skauli, J. Farrell, A collection of hyperspectral images for imaging systems research, in: IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2013, pp. 86600C–86600C.
- [40] A. Yezzi, A. Tsai, A. Willsky, A fully global approach to image segmentation via coupled curve evolution equations, *Journal of Visual Communication and Image Representation* 13 (1) (2002) 195–216.