We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Digital Signature: A Novel Adaptative Image Segmentation Approach

David Freire-Obregón[1], Modesto Castrillón-Santana[1]
and Oscar Déniz-Suárez[2]
[1]*SIANI, Universidad de Las Palmas de Gran Canaria*
[2]*VISILAB, Universidad de Castilla-La Mancha*
Spain

1. Introduction

Partitioning or segmenting an entire image into distinct recognizable regions is a central challenge in computer vision which has received increasing attention in recent years.

It is remarkable how a simple task to humans is not so simple in computer vision, but it can be explained. To humans, an image is not just a random collection of pixels; it is a meaningful arrangement of regions and objects (Figure 1 shows a variety of images).

Despite the variations of these images, humans have no problem interpreting them. We can agree about the different regions in the images and recognize the different objects. Human visual grouping was studied extensively by the Gestalt psychologists. Thus, Wertheimer pointed out the importance of perceptual grouping and organization in vision and listed several key factors Wertheimer (1938), that lead to human perceptual grouping: similarity,

Fig. 1. Some challenging images for a segmentation algorithm. Our goal is to develop a single grouping procedure which can deal with all these types of images. Image Source: Sample images in MIT/CMU test set for frontal face detection (Sung & Poggio (1999)).

proximity, continuity, symmetry, parallelism, closure and familiarity. In computer vision, these factors have been used as guidelines for many grouping algorithms. Thus, the most studied version of grouping in computer vision is image segmentation. Image segmentation techniques can be classified into two families: (1) region-based, and (2) contour-based approaches.

- Region-based approaches try to find partitions of the image pixels into sets corresponding to coherent image properties such as brightness, color and texture.

- Contour-based approaches usually start with a first stage of edge detection, followed by a linking process that seeks to exploit curvilinear continuity.

On the other hand, in order to distinguish good segmentations from bad segmentations the, already mentioned, classical Gestalt theory has developed various principles of grouping (Palmer (1999); Wertheimer (1938)) such as proximity, similarity and good continuation. As Ren & Malik (2003) pointed out, the principle of good continuation states that a good segmentation should have smooth

1. Intra-region similarity: the elements in a region are similar. This includes similar brightness, similar texture, and low contour energy inside the region;

2. inter-region (dis)similarity: the elements in different regions are dissimilar. This in turn includes dissimilar brightness, dissimilar texture, and high contour energy on region boundaries.

These classical principles of grouping have inspired many previous approaches to segmentation. However, the Gestalt principles distinguish competing segmentations only when everything else is equal. Many of the previous works have made ad-hoc decisions for using and combining these cues. However, even to this day, many of the computational issues of perceptual grouping have remained unresolved.

In this work, we present an image descriptor based on self-similarities which is able to capture the general structure of an image. Computed descriptors are similar for images with the same layout, even if textures and colors are different. Similarly to Shechtman & Irani (2007), images are partitioned into smaller cells which, conveniently compared with a patch located at the image center, yield a vector of values that describes local aspect correspondences. In this Chapter, we demonstrate the effectiveness of our approach on two main topics in computer vision's research: facial expression recognition and hair/skin segmentation.

The outline of the Chapter is as follows. Section 2 reviews the previous segmentation approaches. The theoretical concept of digital signature and its main features will be discussed in Section 3 at a high level. In Section 4, some experimental results for classification and segmentation process in order to illustrate the effectiveness and usefulness of the proposed approach. The conclusions of this work in Section 5 will conclude this Chapter.

2. Related work

Many image segmentation methods have been proposed over the last several decades. As new segmentation methods have been proposed, a variety of evaluation methods have been used to compare new segmentation methods to prior methods. These methods are fundamentally very different, and can be partitioned based on the diversity in segment types:

- Algorithms for extracting uniformly colored regions (Comaniciu & Meer (2002); Shi & Malik (2000)).

- Algorithms for extracting textured regions (Galun et al. (2003); Malik et al. (1999)).

- Algorithms for extracting regions with a distinct empirical color distribution (Kadir & Brady (2003); Li et al. (2004); Rother et al. (2004)).

Thus, most of the methods require using image features that characterize the regions to be segmented. Particularly, texture and color have been independently and extensively used in the area. On the other hand, some algorithms can be also categorized on unsupervised (Shi & Malik (2000)), while others require user interaction (Rother et al. (2004)). Some algorithms employ symmetry cues for image segmentation (Riklin-Raviv et al. (2006)), while others use high-level semantic cues provided by object classes (i.e., class-based segmentation, see Borenstein & Ullman (2002); Leibe & Schiele (2003); Levin & Weiss (2006)).

There are also variants in the segmentation tasks, ranging from segmentation of a single input image, through simultaneous segmentation of a pair of images (Rother et al. (2006)) or multiple images.

Bagon et al. (2008) proposed a single unified approach to define and extract visually meaningful image segments, without any explicit modelling. Their approach defines a 'good image segment' as one which is 'easy to compose' (like a puzzle) using its own parts. It captures a wide range of segment types: uniformly colored segments, through textured segments, and even complex objects.

Shechtman & Irani (2007) proposed a segmentation approach based on 'local self-similarity descriptor'. It captures self-similarity of color, edges, repetitive patterns and complex textures in a single unified way. These self-similarity descriptors are estimated on a dense grid of points in image/video data, at multiple scales. Moreover, our Digital Signature is based in this work.

3. Digital signature

We present an image descriptor based on self-similarities which is able to capture the general structure of an image. Computed descriptors are similar for images with the same layout, even if textures and colors are different, similarly to Shechtman & Irani (2007). Images are partitioned into smaller cells which, conveniently compared with a main patch located in the image, yield a vector of comparison results that describes local aspect correspondences.

The Digital Signature (DS) descriptor is computed from a square image subdivided into $n \times n$ cells, where each cell corresponds to an $m \times m$ pixels image patch. The number of cells and their pixel size have effect on how much an image structure is generalized. A low number of cells will not capture many structural details, while too many small cells will produce a too detailed descriptor. The present approach will consider overlapping cells, which may be required to capture subtle structural details.

Once an image is partitioned, an $m \times m$ main patch located in the image (which does not have to correspond to a cell in the image partition) is compared with all partition cells. In order to achieve greater generalization, image patches are compared computing the Sum of Squared Differences (SSD) between pixel values (or the Sum of Absolute Differences (SAD), which is computationally less expensive). Each cell-center comparison is consecutively stored in a $m \times m$ dimensions DS descriptor vector.

In a more general mathematical sense, a Digital Signature z for a proposed cell p, is a function that makes a cell comparative that fall into each of the disjoint categories (similar to histogram's bins), whereas the graph of a digital signature is merely one way to represent a digital signature descriptor. Thus, if we let *ncells* be the total number of cells, *Wcsize* be the
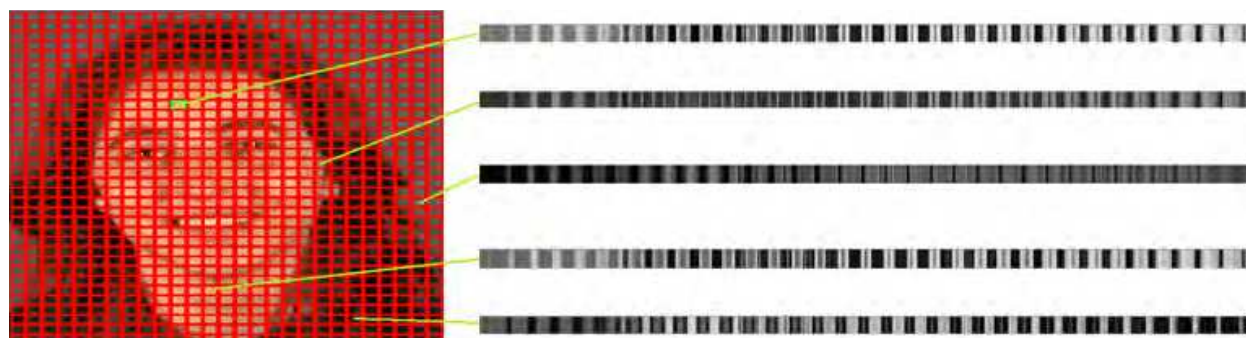
Fig. 2. Digital Signature Descriptor example using a 11x11 partition. The barcode-like vector represents all comparisons between each cell and the other cells. Image source: DaFEx Database (Battocchi & Pianesi (2004)).

width for each cell and *Hcsize* be the height for each cell, the descriptor meets the following conditions for each channel:

$$Digital\_Signature_z = \sum_{p}^{ncells} \mid \int_i^{Wcsize} \int_j^{Hcsize} cell_z(i,j) - \int_u^{Wcsize} \int_v^{Hcsize} cell_p(u,v) \mid \quad (1)$$

Such description overcomes color, contrast and textures. Images are described in terms of their general structure, similarly to Shechtman & Irani (2007). An image showing a white upper half and a black lower half will produce exactly the same descriptor as an image showing a black upper half and a white lower half. Local aspect correspondences are exactly the same: the upper half is different from the lower half. Rotations, however, are not considered.

Digital Signature descriptors are specially useful to describe points defined by a scale salient point detector (like DoG or SURF (Bay & Tuytelaars (2006))). The DS descriptor is shown as a barcode for representation purposes.

Thus, given an input image (i.e. a scale salient point or a known region like detected mouths), a number of cells *n* and their pixel size *m*, DS is computed as follows:

1. The image is divided into different cells inside a template sized $(n \times m) \times (n \times m)$ pixels.

2. The template is partitioned into $n \times n$ cells, each of them sized $m \times m$ pixels.

3. Each cell is compared with any other template cell, and each result is consecutively stored in the $n \times n$ DS descriptor vector. Thus, each cell provides its own Digital Signature.

In order to illustrate the effectiveness and usefulness of the proposed approach we present two different tests. For the first test, similar Digital Signatures from different images are applied to classify mouths into smiling or non-smiling gestures, for the second test, different Digital Signatures from the same image are compared in order to obtain a good image segmentation (See Figure 2).

4. Experimental results

4.1 Facial expression recognition

After extensive research in Psychology, it is now known that emotions play a significant role in human decision making processes (Damasio (1994); Picard (1997)). The ability to show and interpret them is therefore also important for human-machine interaction. Perceptual User Interfaces (PUIs) use multiple input modalities to capitalize on all the communication cues, thus maximizing the bandwidth of communication between a user and a computer.

Examples of PUIs have included: assistants for the disabled, augmented reality, interactive entertainment, virtual environments, intelligent kiosks, etc.

Some facial expressions can be very subtle and difficult to recognize even between humans. Besides, in human-computer interaction the range of expressions displayed is typically reduced. In front of a computer, for example, the subjects rarely display accentuated surprise or anger expressions as he/she could display when interacting with another human subject.

The human smile is a distinct facial configuration that could be recognized by a computer with greater precision and robustness. Besides, it is a significantly useful facial expression, as it allows to sense happiness or enjoyment and even approval (and also the lack of them) (Ekman & Friesen (1982)). As opposed to facial expression recognition, smile detection research has produced less literature. Lip edge features and a perceptron were used in Ito et al. (2005). The lip zone is obviously the most important, since human smiles involve mainly the Zygomatic muscle pair, which raises the mouth ends. Edge features alone, however, may be insufficient. Smile detection was also tackled in the BROAFERENCE system to assess TV or multimedia content (Kowalik et al. (2005)). This system was based on tracking the positions of a number of mouth points and using them as features feeding a neural network classifier. The work Shinohara & Otsu (2004), in turn, used Higher-order Local Autocorrelation, achieving near 98% recognition rates. More recently, the comprehensive work Whitehill et al. (2008) contends that there is a large performance gap between typical tests made in the literature and results obtained in real-life conditions. The authors conclude that the training set may be all-important, specially in terms of variability and size, which should be on the order of thousands of images.

The Sony Cybershot DSC T-200 digital camera has an ingenious "smile shutter" mode. Using proprietary algorithms, the camera automatically detects the smiling face and closes the shutter. To detect the different degrees of smiles by the subject, smile detection sensitivity can be set to high, medium or low. Some reviews argue that: *the technology is not still so much sensitive that it can capture minor facial changes. Your facial expression has to change considerably for the camera to realize that* (*Entertainment.millionface.com: Smile detection technology in camera* (2008)), or *The camera's smile detection - which is one of its more novel features - is reported to be inaccurate and touchy* (*Swik.net: Sonys Cyber-shot T200 gets its first review* (2008)). Whatever the case, detection rates or details of the algorithm are not available, and so it is difficult to compare results with this system. Canon also has a similar smile detection system. This section describes different techniques that are applied to the smile detection problem in video streams and shows the benefits of our new approach.

4.1.1 Representation

In order to show the performance of our new approach, it will be compared with other image descriptors such as Local Binary Patters and Principal Components Analysis. The Local Binary Pattern (LBP) is an image descriptor commonly used for classification and retrieval. Introduced by Ojala et al. (2002) for texture classification, they are characterized by invariance to monotonic changes in illumination and low processing cost.

Given a pixel, the LBP operator thresholds the circular neighborhood within a distance by the pixel gray value, and labels the center pixel considering the result as a binary pattern. The basic version considers the pixel as the center of a 3×3 window and builds the binary pattern based on the eight neighbors of the center pixel, as shown in Figure 3. However, the LBP definition can be easily extended to any radius, R, considering P neighbors Ojala et al. (2002):
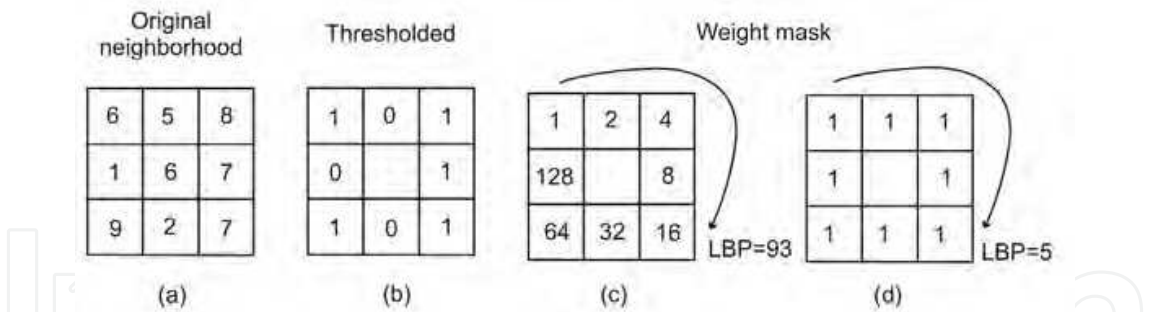
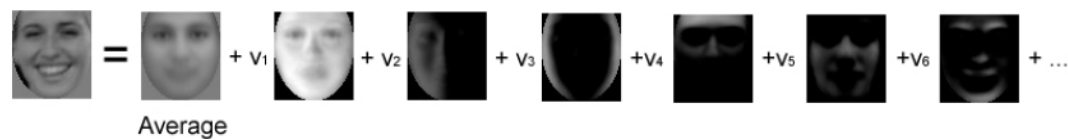Fig. 3. The basic version of the Local Binary Pattern computation (c) and the Simplified LBP codification (d).

Fig. 4. The Principal Components Analysis (PCA) decomposition.

$$LBP_{P,R}(x_c, y_c) = \sum_{k=0}^{P-1} s(g_p - g_c)2^k \; , \; s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (2)$$

Rotation invariance is achieved in the LBP based representation considering the local binary pattern as circular. The experience achieved by Ojala et al. (2002) suggested that just a particular subset of local binary patterns are typically present in most of the pixels contained in real images. They refer to these patterns as uniform. Uniform patterns are characterized by the fact that they contain, at most, two bitwise transitions from 0 to 1 or viceversa. For example, 00000000, 00011100 and 10000011 are uniform patterns. In the experiments carried out by Ojala et al. (2002) with texture images, uniform patterns account for a bit less than 90% of all patterns when using the 3x3 neighborhood.

More recently LBPs have been used to describe facial appearance. Once the LBP image is obtained, most authors apply a histogram based representation approach (Sébastien Marcel & Heusch (2007)). However, as pointed out by some recent works, the histogram based representation loses relative location information (Sébastien Marcel & Heusch (2007); Tao & Veldhuis (2007)), thus LBP can also be used as a preprocessing method. Using LBP as preprocessing method, having the effect of emphasizing edges and noise. To reduce the noise influence, Tao & Veldhuis (2007) proposed recently a modification in Equation 2. Instead of weighting the neighbors differently, their weights are all the same, obtaining the so called Simplified LBPs, see Figure 3-d. Their approach has shown some benefits applied to facial verification, due to the fact that by simplifying the weights, the image becomes more robust to illumination changes, having a maximum of nine different values per pixel. The total number of local patterns are largely reduced so the image has a more constrained value domain.

$$LBP_{P,R}(x_c, y_c) = \sum_{k=0}^{P-1} s(g_p - g_c) \; , \; s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (3)$$

In the experiments described below, both approaches will be investigated, i.e. using the histogram based approach, but also using Uniform LBP and Simplified LBP as a preprocessing step.
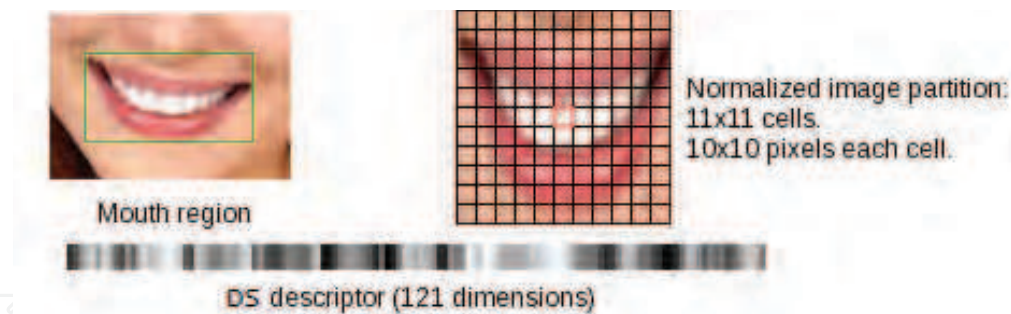
Fig. 5. DS Descriptor example using a 11x11 partition. The barcode-like vector represents all comparisons between each cell and the central patch. In this case, only the center patch will be considered to obtain our descriptor.

Raw face images are highly dimensional. A classical technique applied for face representation to avoid the consequent processing overload problem is Principal Components Analysis (PCA) decomposition (Kirby & Sirovich (1990)). PCA decomposition is a method that reduces data dimensionality, without a significant loss of information, by performing a covariance analysis between factors. As such, it is suitable for highly dimensional data sets, such as face images. A normalized image of the target object, i.e. a face, is projected in the PCA space, see Figure 4. The appearance of the different individuals is then represented in a space of lower dimensionality by means of a number of those resulting coefficients, *vi* (Turk & Pentland (1991)).

We now discuss the central contribution of this section, the incorporation of the Digital Signature descriptor. However, images containing smiling mouths require local brightness to be preserved: teeth are always brighter than surrounding skin and that must be captured by the descriptor. Thus, instead of using SSD, patches are compared using Sum of Differences. Otherwise, a closed mouth would produce the same descriptor as a smiling mouth: lips are surrounded by differently colored skin, exactly as teeth are surrounded by differently colored lips. Figure 5 shows an example with an 11×11 cell partition, each cell sized 10×10 pixels. For this experiment, the DS is applied to classify mouths found by a face detector (Castrillón Santana et al. (2007)) into smiling or non-smiling gestures. Smiling mouths look similar no matter the skin color or the presence of facial hair. This generality can be registered by a self-similarity descriptor like DS.

Thus, for this test, our Digital Signature descriptor is computed as follows:

1. The image is resized to a template sized $(n \times m) \times (n \times m)$ pixels.

2. The template is partitioned into $n \times n$ cells, each of them sized $m \times m$ pixels.

3. A main cell sized $m \times m$ pixels is captured from the template image.

4. The main cell is compared with each template cell, and each result is consecutively stored in the $n \times n$ DS descriptor vector.

4.1.2 Classification and face detection

Two more points are needed to be justified. The first one is the classification method. In order to tell wether two images (smiling or not smiling) have a similar structure, their corresponding DS descriptors can be compared computing SAD between both vectors. However, given that the present work aims at classifying mouth images in two categories, a Support Vector Machine approach (Burges (1998)) is used. The SVM is a set of related supervised learning

methods used for classification and regression. They belong to a family of generalized linear classifiers. A property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers.

The second point is to explain how are faces going to be extracted from videos. Several approaches have recently appeared for real-time face detection (Schneiderman & Kanade (2000); Viola & Jones (2004)), all of them aiming at making the problem less environment dependent. Focusing on live video stream processing, a face detector based on cue combination (Castrillón Santana et al. (2007)) outperforms well known single-cue based detectors such as Viola & Jones (2004). This approach provides a more reliable tool for real time interaction in PUIs. The face detection system used to extract faces from video streams in this work (see Castrillón Santana et al. (2007) for details) integrates, among other cues, different classifiers based on the general object detection framework by Viola and Jones Viola & Jones (2004), skin color, multilevel tracking, etc. In order to further minimize the influence of false alarms, the facial feature detector capabilities were extended, locating not only faces but also eyes, nose and mouth. This reduces the number of false alarms, for it is less probable that multiple detectors, i.e. face and its inner features, are activated simultaneously with a false alarm. Its important to point that each detected face is normalized to 59×65 pixels using the position of the eyes.
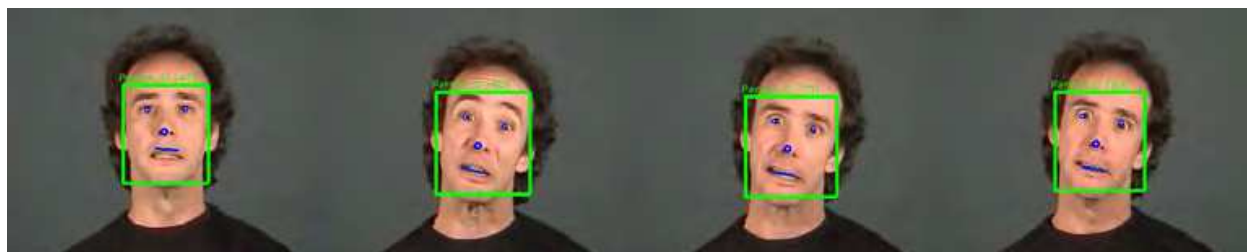

Fig. 6. Facial element detection results for a video stream extracted from the DaFEx Database.

The facial element detection procedure is only applied in those areas which bear evidence of containing a face. This is true for regions in the current frame, where a face has been detected, or in areas with a detected face in the previous frame. Figure 6 shows the possibilities of the face detector.

4.1.3 Performance and evaluation

The DaFEx (Battocchi & Pianesi (2004)) database was used in this experiment. In this database 8 professional actors showed 7 expressions (6 basic facial expressions + 1 neutral) on 3 intensity levels (low, medium, high) twice. The frames of the 48 'happy' videos were extracted of the database sequences (see Figure 7) for test. The total number of images contained in the 'happy' videos of dataset is 12,988, from which, 6,783 are smiling images and 6,205 are no smiling images. These images have been also normalized according to eye positions obtaining 59×65 samples and have been annotated manually by a human.

As briefly mentioned above, the experimental setup considers one possibility as input: the mouth. The input image is a grayscale image, and for representation purposes we have used the following approaches for the tests:

- PCA. A PCA space obtained from the original gray images of 59×65 pixels.

- ULBP Hist. A concatenation of histograms based on the gray image or the resulting ULBP image.

Fig. 7. Low, medium and high intensity happy expressions from DaFEx

DaFEx	Approach	False Negative	False Positive	Error Rate
Complete Database	PCA	11.1%	16.3%	13.5%
	ULBP Hist.	19.5%	23.6%	21.4%
	SLBP Im. Val.	20.2%	21.3%	20.7%
	Digital Signature	10.2%	13.0%	11.5%
LOW Intensity	PCA	13.6%	19.1%	16.2%
	ULBP Hist.	24.9%	24.0%	24.4%
	SLBP Im. Val.	22.1%	23.8%	22.9%
	Digital Signature	12.3%	16.6%	14.3%
MED Intensity	PCA	13.6%	14.4%	13.9%
	ULBP Hist.	19.9%	22.5%	21.1%
	SLBP Im. Val.	20.9%	20.4%	20.6%
	Digital Signature	11.6%	13.7%	12.6%
HIGH Intensity	PCA	6.3%	15.5%	10.6%
	ULBP Hist.	13.7%	24.5%	18.8%
	SLBP Im. Val.	17.6%	19.7%	18.6%
	Digital Signature	6.8%	8.7%	7.7%

Table 1. Error rates achieved using each approach over the DaFEx database, considering the complete database or dividing it into different intensities (see text). Four different approaches were applied: PCA on grayscale images considering 130 coefficients, Normalized Histogram method on the Uniform LBP representation, Normalized Image Values method on the Simplified LBP representation and the Digital Signature on grayscale images.

- SLBP Im. Val. A concatenation of the image values based on the gray images or the resulting SLBP image.

- Digital Signature. DS descriptor obtained from the original gray images of 59×65 pixels.

Similar experimental conditions have been used for every approach considered in this setup. The test sets are built randomly, having an identical number of images for both classes. Results presented correspond to the percentage of wrong classified samples of all test samples.

As it can be appreciated in Table 1, best results in almost every situation are achieved with our new approach. None of the LBP based representations outperforms that approach. However, even if the Uniform LBP approach evidences a larger improvement when normalized histograms are used, the Simplified LBP approach reported better results than Uniform LBP. As already stated in Tao & Veldhuis (2007) this preprocessing provides benefits in the context of facial analysis.

The PCA based representation achieves a better error rate than the LBP's approaches. However, it doesnt keep a good balance between false postive and false negative. PCA
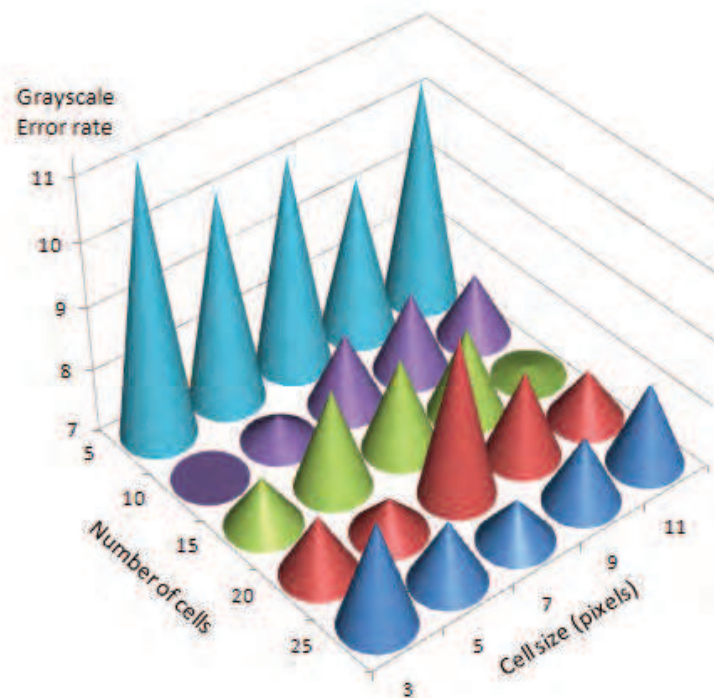
Fig. 8. Results achieved by the Digital Signature approach for the high intensity test. As it can be appreciated, best results are achieved for $n = 10$ and $m = 3$.

deserves an additional observation, not always the increasing of the space dimension for PCA reports better results. This is the main reason for keeping only 130 coefficients for the image descriptor. When the DS descriptor is used, the test achieved the lowest error rate.

For the Digital Signature approach, error-rate behavior is also quite similar to behaviour obtained previously with PCA and Image Value tests rates. Again, the lowest error rates were achieved by the grayscale image test. For DS, it is important to mention that overlap is not considered between cells. Firstly, several tests without overlapping were made in order to find optimun DS parameters (number of cells and cells' size yielding the lowest error rate). For smile detection it was found that 10×10 cells of 3×3 pixels performed best (See Figure 8) thanks to the closeness between the size of the extracted DS main cell (30×30 pixels) and the original size of the mouth capture (20×12 pixels). It is also shown that worst results are achieved for configurations less than 10×10 cells because of the loss of information due to resizing in the Normalization step. Beyond that number of cells and for bigger sizes, behaviour is irregular due to the fact that information extracted is not reliable because of the false information introduced when the mouth is resized to fit the DS cell. When images are upsampled, redundant and useless information is created. Unfortunately, when overlap was introduced, the achieved error rates were higher than without overlapping. Used images were too small for overlapping regions to be significant. Unlikely what happens with the PCA approach in this case there is a good balance between false positive and false negative. Thus, the new approach has got better results than other approaches that have been successfully used before for smile detection (Freire et al. (2009a;b)).
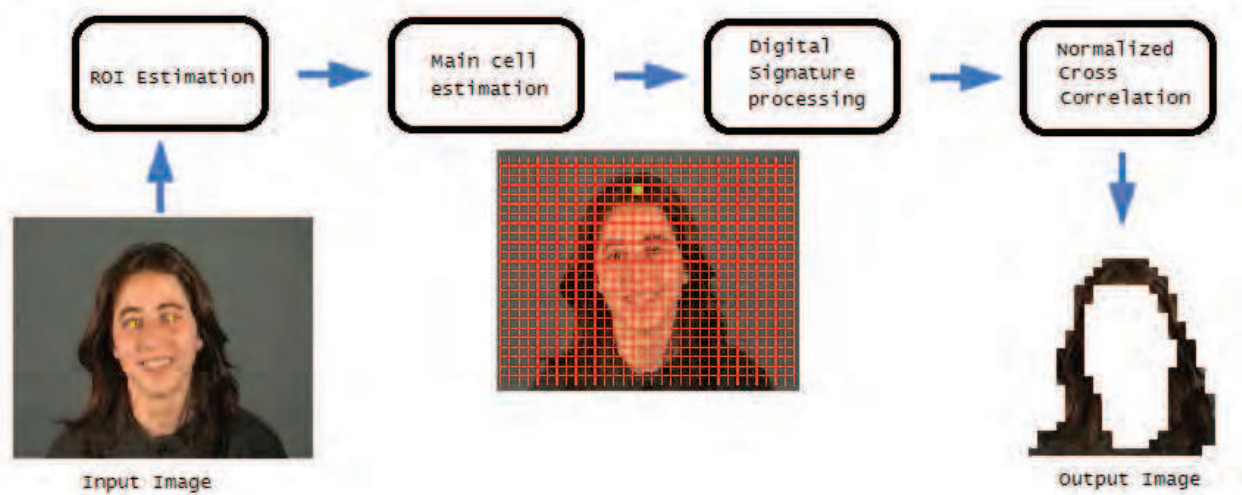
Fig. 9. Block diagram of our segmentation method: (a) input image, (b) Region of interest estimation, (c) main cell estimation (highlighted in green), (d) Digital Signature processing for each cell, (e) Normalized Cross Correlation between each cell's DS and the main cell's DS, and (f) the hair extracted by the proposed method.

4.2 Image segmentation

Face recognition techniques have attracted much attention over the years and many algorithms have been developed. Since it has many potential applications in computer vision and automatic access control system, its research has rapidly expanded by not only engineers but also neuroscientists. Especially, face segmentation is an essential step of face recognition system since most face classification techniques tend to only work with face images. Therefore face segmentation has to correctly extract only face part of given large image. However, apart from facial features, hair style or clothing also reflects important personal traits.

Clothing segmentation is widely used in many computer vision tasks, such as dressed people detection (Ioffe & Forsyth (2001); Ronfard et al. (2002)), identification, image editing, human sketches and portraits for graphics rendering (Chen et al. (2004; 2006)). However, existing segmentation methods suffer from variations in colors and styles, different lighting conditions, cluttering backgrounds and occlusions generated by poses or other objects. Most existing methods use clothing models (Chen et al. (2006); Lee & Cohen (2006); Sprague & Luo (2002)). Typically, clothing models are trained with tagged samples, and then the clothing is extracted from images by comparing to the models. For example, in Sprague & Luo (2002), a dress model and a shirt model were trained, respectively, and then clothing detection was performed on the pre-segmented image by selecting a better match against the trained models. Such methods can only handle a few clothing styles and rely on the pre-segmentation accuracy.

On the other hand, human hair related applications have attracted increasing interest in recent years, since hair plays a significant role in the overall appearance of an individual. To achieve these tasks, hair segmentation is generally the first prerequisite step. However, to our knowledge, in most previous studies, hair is assumed segmented already or manually labeled. Furthermore, besides the above hair-related applications, many computer vision tasks can also benefit from segmented hair. For instance, it provides an important clue for gender classification, since hair styles of male and female are generally different. Hair can also facilitate age estimation since hair distribution and color gradually changes with the increase
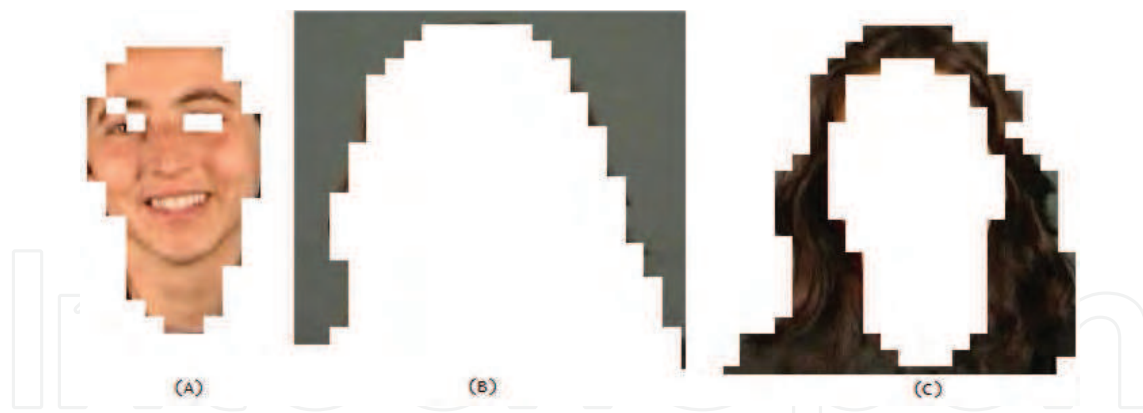
Fig. 10. Segmentation results on this frame image sequence are shown in subimages (a) to (c). Segment in (a) correspond to the person's skin, segment in (c) correspond to the person's hair and (b) correspond to the background. This results are achieved for $n = 35$ and $m = 11$.

of age, especially for old men/women. Hair information can even contribute a lot to face recognition, considering that the hair style of one normal subject does not abruptly change frequently. To sum up, more attention should be paid to automatic hair segmentation. The work of Yacoob and Davis Yacoob & Davis (2006) is the only prior work we can find on hair detection. Their approach constructs a simple color model and uses it to recognize the hair pixel. However, their detection can only work under controlled background environment and very less hair color variation. On the other hand, hair modelling, synthesis, and animation have already become active research topics in computer graphics (Kajiya & Kay (1989); Marschner et al. (2003); Moon & Marschner (2006); Paris et al. (2004); Wei et al. (2005)).

In this section, without any predefined model, we propose a new method to segment parts of an image (e.g. clothing, hair style or skin) by using the Digital Signature technique.

4.2.1 The proposed method
Thus, given an input image, a number of cells n and their pixel size m, the proposed method is computed as follows (See Figure 9):

1. The facial detector described in Section 4.1.3 provides the eyes position. Thus, a region of interest can be estimated.

2. The image is divided into different cells inside a template sized $(n \times m) \times (n \times m)$ pixels.

3. The template is partitioned into $n \times n$ cells, each of them sized $m \times m$ pixels.

4. Taking the eye positions into account, it is possible to estimate a main cell position.

5. Each cell is compared with any other template cell, and each result is consecutively stored in the $n \times n$ DS descriptor vector. Thus, each cell provides its own Digital Signature (See Figure 2).

6. Each cell's DS will be compared to the main cell's DS. This will help to decide whether or not the proposed cell is similar to the main cell.

4.2.2 Results
Figure 10 displays a segmentation results with a simple color frame. The Digital Signature succeeded to generate accurate skin and background segmentation. It fails to generate accurate hair segmentation because the hair and cloth have similar color in the example. It can be also appreciated an spatial aliasing effect due to the fact that the cell's shape is square.

One way to reduce the spatial aliasing is to decrease the cell's size but, this fact could be harmful for the method because it will cause an information loss for each cell.

RGB color images were considered here for test. In each segmentation, the value of the one free parameter, cell size in Equation 1, was kept constant: 11×11, despite the different characteristics of the images. Figure 11 shows the results of applying the segmentation algorithm to two more images under unconstraint lighting environment. The overall result of this study was that the segmentation results were generally stable to perturbations of the cell size as far as the distance between eyes remains similar between images.

Fig. 11. Some components of the partition applying a Digital Signature cell size of 11×11 ($n = 11$). Images source: askmen.com(*Askmen.com* (2011)).

5. Conclusions

In this Chapter, we have introduced a novel descriptor for segmenting an image into a regular grid of cells. We have argued that the regular grid confers a number of useful properties, such as a powerful representation approach for classification purposes. We have also demonstrated that despite this topological constraint, we can achieve segmentation performance comparable with current algorithms.

For the first test, this Chapter described a smile detection using different LBP approaches, as well as PCA image representation, combined with SVM. The potentiality of the Digital Signature based representation for smile verification has been shown. The DS based representation presented in this paper outperforms other approaches with an improvement over a 5%. Uniform LBP does not respond to a statistical spatial patterns locality. This means, that there is no gradual change between adjacent blocks preprocessed with Uniform LBP. Depending on the value of a pixel inside one of the blocks, codification between two adjacent pixels can be, for example, from pattern 2 to pattern 9. Translated to the space domain of SVM, this means that dimensions can be too far away. Simplified LBP keeps the statistical spatial patterns locality. There is a gradual change between adjacent preprocessed pixels. Translated to the SVM's space domain, this gradual change means that similar points are closer in this space. Our future line is focus on the potentiality of the DS descriptor for generic applications such as image retrieval. In this paper we have developed a static smile classifier achieving, in

some cases, a 93% of success rate. Due to this success rate, smile detection in video streams, where temporal coherence is implicit, will be studied in short term, as a cue to get the ability to recognize the dynamics of the smile expression.

For the second test, we have analyzed how the Digital Signature cell's descriptors can be used to segment an input image without any previous training. Once the eyes position is obtained, the template of the main cell can be find inside the image by comparing their digital signatures. The model works very well under certain circumstances:

1. There is a variety of patterns between the main cell and the rest of cells.

2. Under constraint and unconstraint lighting environment.

3. The face/hair/clothes features are relatively small to fill holes (See false negatives for the first dress in Figure 11)

The above insight opens the door for further research into the balance between complexity of the model and the way in which it is used for inference. In order to reduce the aliasing effect other geometrical shape for the cells can be also tested. On the other hand the cell's size depends on image cues such as the image's size or other heuristic method depending on the segmentation process (i.e. for face segmentation, the distance between the eyes could be an important cue for the cell size).

6. Acknowledgments

7. References

Askmen.com (2011).

Bagon, S., Boiman, O. & Irani, M. (2008). What is a good image segment?, *ECCV*.

Battocchi, A. & Pianesi, F. (2004). Dafex: Un database di espressioni facciali dinamiche, *SLI-GSCP Workshop Comunicazione Parlata e Manifestazione delle Emozioni*.

Bay, H. & Tuytelaars, T. (2006). Surf: Speeded up robust features, *Proceedings of the Ninth European Conference on Computer Vision*.

Borenstein, E. & Ullman, S. (2002). Class-specific, top-down segmentation, *ECCV*.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* 2(2): 121–167.

Castrillón Santana, M., Déniz Suárez, O., Hernández Tejera, M. & Guerra Artal, C. (2007). ENCARA2: Real-time detection of multiple faces at different resolutions in video streams, *Journal of Visual Communication and Image Representation* pp. 130–140.

Chen, H., Liu, Z., Rose, C. & Xu, Y. (2004). Example-based composite sketching of human portraits, *Proceedings of Third International Symposium on Non Photorealistic Animation and Rendering (NPAR)*, pp. 95–102.

Chen, H., Liu, Z., Zhu, S. & Xu, Y. (2006). Composite templates for cloth modeling and sketching, *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 943–950.

Comaniciu, D. & Meer, P. (2002). Mean shift: a robust approach toward feature space analysis, *PAMI* .

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*, Picador.

Ekman, P. & Friesen, W. (1982). Felt, false, and miserable smiles, *Journal of Nonverbal Behavior* 6(4): 238–252.

Entertainment.millionface.com: Smile detection technology in camera (2008).

Freire, D., Castrillón Santana, M. & Déniz Suárez, O. (2009a). A novel approach for smile detection combining ulbp and pca, *Proceedings of the EUROCAST*.

Freire, D., Castrillón Santana, M. & Déniz Suárez, O. (2009b). Smile detection using local binary patterns and support vector machines, *Proceedings of the VISIGRAPP*.

Galun, M., Sharon, E., Basri, R. & Brand, A. (2003). Texture segmentation by multiscale aggregation of filter responses and shape elements, *ICCV*.

Ioffe, S. & Forsyth, D. (2001). Probabilistic methods for finding people, *International Journal of Computer Vision* 43(1): 45–68.

Ito, A., Wang, X., Suzuki, M. & Makino, S. (2005). Smile and laughter recognition using speech processing and face recognition from conversation video, *Procs. of the 2005 IEEE Int. Conf. on Cyberworlds (CW'05)*.

Kadir, T. & Brady, M. (2003). Unsupervised non-parametric region segmentation using level sets, *ICCV*.

Kajiya, J. & Kay, T. (1989). Rendering fur with three dimensional textures, *Computer Graphics* 23(3): 271–280.

Kirby, Y. & Sirovich, L. (1990). Application of the Karhunen-Loéve procedure for the characterization of human faces, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 12(1).

Kowalik, U., Aoki, T. & Yasuda, H. (2005). Broaference - a next generation multimedia terminal providing direct feedback on audience's satisfaction level, *INTERACT*, pp. 974–977.

Lee, M. & Cohen, I. (2006). A model-based approach for estimating human 3d poses in static images, *IEEE Trans. Pattern Analysis and Machine Intelligence* 28(6): 905–916.

Leibe, B. & Schiele, B. (2003). Interleaved object categorization and segmentation, *BMVC*.

Levin, A. & Weiss, Y. (2006). Learning to combine bottom-up and top-down segmentation, *ECCV*.

Li, Y., Sun, J., Tang, C. & Shum, H. (2004). Lazy snapping, *ACM TOG*.

Malik, J., Shi, J., Leung, T. & Belongie, S. (1999). Textons, contours and regions: Cue integration in image segmentation, *ICCV*.

Marschner, S., H.W.Jensen, M.Cammarano, Worley, S. & Hanrahan, P. (2003). Light scattering from human hair fibers, *Proceedings of the SIGGRAPH*, pp. 780–791.

Moon, J. T. & Marschner, S. R. (2006). Simulating multiple scattering in hair using a photon mapping approach, *Proceedings of the SIGGRAPH*.

Ojala, T., Pietikäinen, M. & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 24(7): 971–987.

Palmer, S. (1999). Vision science: Photons to phenomenology, *MIT Press*.

Paris, S., Briceno, H. M. & Sillion, F. X. (2004). Capture of hair geometry from multiple images, *Proceedings of the SIGGRAPH*.

Picard, R. W. (1997). *Affective Computing*, MIT Press.

Ren, X. & Malik, J. (2003). Learning a classification model for segmentation, *9th Int. Conf. Computer Vision*, Vol. 1, pp. 10–17.

Riklin-Raviv, T., Kiryati, N. & Sochen, N. (2006). Segmentation by level sets and symmetry, *CVPR*.

Ronfard, R., Schmid, C. & Triggs, B. (2002). Learning to parse pictures of people, *Proceedings of ECCV*, pp. 700–714.

Rother, C., Kolmogorov, V. & Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts, *SIGGRAPH*.

Rother, C., Minka, T., Blake, A. & Kolmogorov, V. (2006). Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs, *CVPR*.

Schneiderman, H. & Kanade, T. (2000). A statistical method for 3d object detection applied to faces and cars, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1746–1759.

Sébastien Marcel, Y. R. & Heusch, G. (2007). On the recent use of local binary patterns for face authentication, *International Journal of Image and Video Preprocessing, Special Issue on Facial Image Processing* .

Shechtman, E. & Irani, M. (2007). Matching local self-similarities across images and videos, *IEEE Conference on Computer Vision and Pattern Recognition 2007 (CVPR'07)*.

Shi, J. & Malik, J. (2000). Normalized cuts and image segmentation, *PAMI* .

Shinohara, Y. & Otsu, N. (2004). Facial expression recognition using Fisher weight maps, *Procs. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*.

Sprague, N. & Luo, J. (2002). Clothed people detection in still images, *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, pp. 585–689.

Sung, K. & Poggio, T. (1999). Example-based learning for view-based human face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(1): 39–51.

Swik.net: Sonys Cyber-shot T200 gets its first review (2008).

Tao, Q. & Veldhuis, R. (2007). Illumination normalization based on simplified local binary patterns for a face verification system, *Proc. of the Biometrics Symposium*, pp. 1–6.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition, *J. Cognitive Neuroscience* 3(1): 71–86.

Viola, P. & Jones, M. J. (2004). Robust real-time face detection, *International Journal of Computer Vision* 57(2): 151–173.

Wei, Y., Ofek, E., Quan, L. & Shum, H. (2005). Modelling hair from multiple views, *Proceedings of the VISSAP*.

Wertheimer, M. (1938). Laws of organization in perceptual forms (partial translation), *A Sourcebook of Gestalt Psychology* pp. 71–88.

Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M. & Movellan, J. (2008). Developing a practical smile detector. Submitted to Transactions on Pattern Analysis and Machine Intelligence. Last accessed: March 2008.

Yacoob, Y. & Davis, L. S. (2006). Detection and analysis of hair, *IEEE Trans. Pattern Analysis and Machine Intelligence* 28(7).
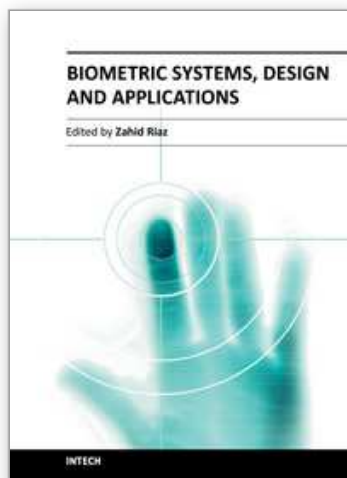
Biometric Systems, Design and Applications

Edited by Mr Zahid Riaz

Biometric authentication has been widely used for access control and security systems over the past few years. The purpose of this book is to provide the readers with life cycle of different biometric authentication systems from their design and development to qualification and final application. The major systems discussed in this book include fingerprint identification, face recognition, iris segmentation and classification, signature verification and other miscellaneous systems which describe management policies of biometrics, reliability measures, pressure based typing and signature verification, bio-chemical systems and behavioral characteristics. In summary, this book provides the students and the researchers with different approaches to develop biometric authentication systems and at the same time includes state-of-the-art approaches in their design and development. The approaches have been thoroughly tested on standard databases and in real world applications.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

David Freire-Obregón, Modesto Castrillón-Santana and Oscar Déniz-Suárez (2011). Digital Signature: A Novel Adaptative Image Segmentation Approach, Biometric Systems, Design and Applications, Mr Zahid Riaz (Ed.), ISBN: 978-953-307-542-6, InTech, Available from: http://www.intechopen.com/books/biometric-systems-design-and-applications/digital-signature-a-novel-adaptative-image-segmentation-approach

INTECH
open science | open minds