2009 Advanced Video and Signal Based Surveillance

Automatic Initiation of the Periorbital Signal Extraction in Thermal Imagery

Dvijesh Shastri
Dept. of Computer Science
University of Houston,
Houston, TX, USA
dshastri@uh.edu

Ioannis Pavlidis
Dept. of Computer Science
University of Houston,
Houston, TX, USA
ipavlidis@uh.edu

***Abstract*— User intervention in the periorbital thermal signal extraction process breaks down automation. This paper proposes a novel way to minimize user intervention. While previous work demonstrated the importance of accurate computation of the periorbital signal, the present method enables its automatic extraction at a reduced processing time. The proposed algorithm capitalizes on detection of involuntary eye blinking in the thermal imagery. The need for automation has emerged because of repetitive processing of the same subjects, aiming to validate improvements in the periorbital tissue tracking or segmentation algorithms. The proposed approach initiates the tracking and segmentation algorithms on the same spatio-temporal location in repetitive runs of the thermal clip. Thus, it does not only automate the process, but also eliminates the variability introduced by manual intervention. We have tested the algorithm on thermal video clips of 39 subjects who faced stressful interrogation for a mock crime. The results show that the proposed method has reduced total processing time from a week down to a day.**

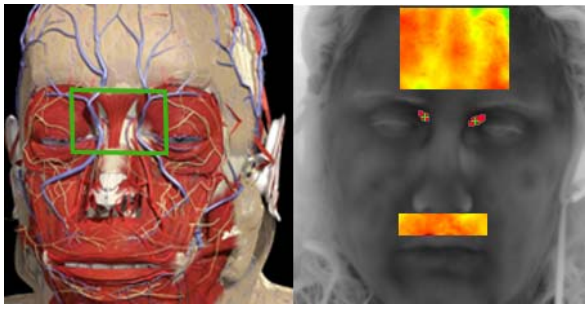
Keywords - thermal imaging, eye blinking detection, image procesing automation

I. INTRODUCTION

Process automation is an essential component of any mass production industry such as the automobile, food, and drug industry. In recent years, process automation is becoming an increasingly popular trend in the video and image processing field, because acquiring and processing massive data sets is crucial to many real-life applications. Such applications include video-based border surveillance, screening at security checkpoints, and the like. Process automation mainly serves two purposes: (a) It eliminates user intervention and thereby reduces human errors. (b) It allows mass production in a short period of time and thereby reduces labor. To achieve these benefits for the periorbital signal extraction process, we propose an automatic initiation approach based on eye blinking detection.

In previous work, we have demonstrated the importance of the periorbital signal in thermal-based lie-detection analysis [1]. The periorbital signal captures the peripheral sympathetic response of the autonomic nervous system during 'flight or fight'. Realizing the importance of accurate extraction of the periorbital signal, we have developed a segmentation algorithm [2] to delineate the periorbital tissue and a tracking algorithm to track the segmented region over time in thermal imagery [3]. The combination of segmentation and tracking algorithms precisely extracts the periorbital signal despite head motion. However, to initiate the signal extraction process, the segmentation algorithm requires the user to provide seed pixels while the tracking algorithm needs the region of interest to be delineated in the first frame. This user intervention introduces significant overhead in a large dataset that is repeatedly processed. Each clip in our dataset takes approximately 45 minutes to process on a standard Dell PC with 2GB RAM, Pentium processor, and Windows XP operating system. Thus, it requires close to 30 hours (45 minutes/subject * 39 subjects) to process the 39 clips of our dataset. Furthermore, every modification in the segmentation and tracking algorithms has to be validated afresh on the dataset. This is quite labor-intensive. The algorithm validation also requires the clips to be initiated on the same spatiotemporal location in order to avoid any variability, which could result in inconsistent measurements. To eliminate user intervention, we propose a method to automate the periorbital signal extraction process.

The periorbital region sits atop of the facial and ophthalmic arterial-venous complexes, which supply with blood the orbicularis oculi muscle (see Figure 1). Thus, the periorbital region is proximal to the eyes, which involuntarily blink at certain intervals. By capitalizing upon the eye properties and face geometry, one can compute the seed pixels and tracker locations.



(a) (b)

Figure 1. Anatomical and thermal images of the face. (a) Facial anatomy [4]. (b) Facial areas of sympathetic importance: (From top to bottom) Supraorbital, Periorbital and Maxillary regions.

In the remainder of the paper, we first discuss the details of our approach in Section 2. In Section 3, we discuss the

978-0-7695-3718-4/09 $25.00 © 2009 IEEE
DOI 10.1109/AVSS.2009.93

experimental results. Finally, we conclude the paper in Section 4.

II. METHODOLOGY

Our approach to process automation is divided into four parts: (1) Segment skin from background pixels by applying a Bayesian classification algorithm. (2) Compute blobs that represent motion in the current frame using a second order change detection method. (3) Find the best eye blinking candidate pair. (4) Use geometric information to compute the location of the seed pixels from this pair.

A. Step 1: Skin Segmentation

The skin segmentation method has at its core a Bayesian algorithm, which we proposed previously [5]. The probability of a pixel being skin (*s*) is calculated by the Bayesian formula:

$$p^{(t)}(s \mid x_t) = \frac{\pi^{(t)}(s) f(x_t \mid s)}{\pi^{(t)}(s) f(x_t \mid s) + \pi^{(t)}(b) f(x_t \mid b)}, \quad (1)$$

where, $\pi^{(t)}(s)$ is the prior skin probability, $f(x_t \mid s)$ is the likelihood of pixel *x* representing skin, $\pi^{(t)}(b)$ is the prior background probability, and $f(x_t \mid b)$ is the likelihood of pixel *x* representing background, at time *t*.

Since the temperature range of the coldest parts of the human face, such as nose and eyebrows, partially overlaps the temperature range of the background, the Bayesian classifier may misclassify cold body parts pixels as background pixels. This creates holes in the segmented skin area. On the other hand, the temperature range of covered skin (e.g., subject's cloth) overlaps with the naked skin temperature range. Therefore, the algorithm may misclassify covered skin as naked skin pixels and generate outliers in the image. These outliers and holes in the skin area confuse the change detection algorithm. Therefore, it is necessary to remove them from the Bayesian output. We use connected component labeling to fill holes and remove outliers from the binary image.

The connected components technique, also known as the blob-coloring technique, is a simple method for region classification [6][7]. It lists all the regions or blobs in an image by assigning a distinct integer number or color to each blob. We keep the largest blob and remove all other blobs. This eliminates all the spurious objects, which are mainly non-skin areas. We generate a complement of the binary image by applying the *NOT* binary operator. The holes in the image now appear as objects and the background of the image is the largest object in the complementary image. We again apply the connected components technique to list the regions in the image and remove all the regions except the largest. Next, we apply the *NOT* binary operator to inverse the complementary binary image. Thus, the output image is free from spurious objects and holes. The image is treated as a binary mask and convolved with the raw thermal image to get optimal skin segmentation (see Figure 2). Localizing the skin region not only reduces false detection of blinking but also improves the time complexity of the eye detection algorithm.

B. Step 2: Change Detection

The simplest way to detect object motion in a video clip is to take image differences between subsequent frames. This first order differentiation identifies global as well as local motion in the imagery. By global motion, we refer to a motion that persists for an extended period of time, such as face motion and body motion. Local motion refers to changes that appear instantaneously, like involuntary eye blinking.

The first-order change detection is computed by taking the absolute difference of the current (I_t) and previous (I_{t-1}) thermal frames:

$$dI_t = |I_t - I_{t-1}|, \quad (2)$$
$$dI1_t = |I_{t-1} - I_{t-2}|.$$

The resultant image dI is then converted into a binary image by selecting an appropriate threshold (*T*) value. The threshold value depends on the noise level in the image. The first-order change is sufficient to detect eye blinking when the face is stationary. However, when the face moves, many

(a) (b) (c)

Figure 2: Skin segmentation: (a) Raw thermal image. (b) Bayesian skin segmentation. (c) Final output after post-processing. The Bayesian skin segmentation segments the skin area, but leaves outliers and holes in the image. Post-processing is necessary for optimal skin segmentation.

blobs appear around the face boundary as well as on the face itself, after the first-order differentiation (see Figure 3a and Figure 3d).

We remove the blobs that represent global motion by taking a second-order differentiation [8]. The second-order differentiation takes three consecutive image frames and computes a first-order differentiation between the first and second frames (dI_t) as well as between the second and third frames ($dI1_t$), as shown in Equation (2). The resultant binary images, dI_t and $dI1_t$, are then used to compute the velocity $\vec{V}(v_i, v_j)$ of the moving object:

$$\vec{V}(v_i, v_j) = \text{argmin} \parallel dI_t - dI1_t(\vec{V}) \parallel, \quad (3)$$

where, ||.|| represents the L1 norm of a vector.

We generate the $dI1_t(\vec{V})$ image by shifting the $dI1_t$ by (v_i, v_j) pixels. The image $dI1_t(\vec{V})$ is dilated via a binary morphological operator to cover not only the changed pixels but also the neighborhood of the changed pixels.

Finally, we compute the ddI_t image by subtracting $dI1_t(\vec{V})$ from dI_t (see Figure 3c and Figure 3f). The isolated pixels in the ddI_t image are removed via a median filter.

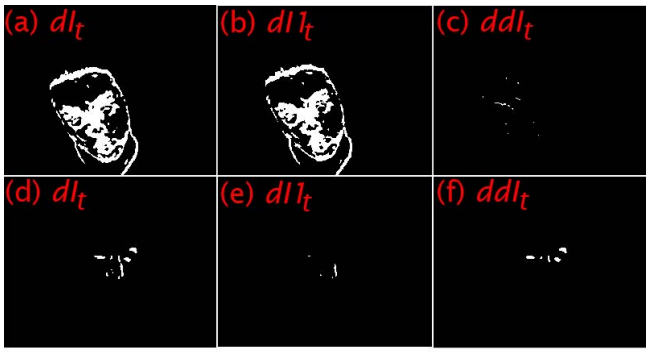


Figure 3: Second order change detection for two different instances. Top instance: (a) Global changes (head motion) at time *t*. (b) Global changes at time *t-1*. (c) Second-order change. It does not contain any useful information. Bottom instance: (d) Local (eye blinking) and global changes. (e) Global changes. (f) Second-order changes (include local changes and some noise).

C. Step 3: Blob Analysis

In this step, the best candidate eye blinking pair is computed. So far in the algorithm, we have analyzed individual blobs. Next, we explore pairs of blobs in order to find the best blinking pair from all possible candidate pairs. The possible number of candidate pairs can be computed using the following equation:

$$Candidate\_Pair, P = \frac{n(n-1)}{2}, \quad (4)$$

where n is the number of blobs in the current frame. For example, if n = 4, the number of possible pairs are 6, $P(b_1, b_2)$: (1,2), (1,3), (1,4), (2,3), (2,4), (3,4).

We compute the Euclidian distance ($d(b_1, b_2)$) between blobs in each pair and the orientation of the line that connects the blob centroids in each pair $\theta_L(b_1, b_2)$. We then apply a distance constraint to discard the pairs that have blobs too close or too far from each other. The distance constraint is set based on anthropometric knowledge [9] and the camera model.

Next, we calculate the orientation of each blob in the following two steps. First, we apply a skeletonization algorithm that transforms an elliptical blob into a line. In the second step, we fit a linear polynomial to find the orientation $\theta_B(b)$ of the line that represents the blob. If both blobs in a pair represent blinking blobs, their orientation ($\theta_B(b_1)$ and $\theta_B(b_2)$) should match the orientation of the line that connects them ($\theta_L(b_1, b_2)$). We use this information to compute the orientation error ($E_O(P)$) for each pair using the following equations:

$$E_O(b_1) = \text{Normalized}(\mid \theta_B(b_1) - \theta_L(b_1, b_2) \mid), \quad (5)$$
$$E_O(b_2) = \text{Normalized}(\mid \theta_B(b_2) - \theta_L(b_1, b_2) \mid),$$
$$E_O(P) = E_O(b_1) + E_O(b_2).$$

If a pair $P(b_1, b_2)$ represents an eye blinking set, it should have blobs of nearly the same size. Based on this assumption, we compute the size ratio error as follows:

$$E_{SR}(P) = \begin{cases} Size_{b_1} / Size_{b_2}, & \text{if } Size_{b_1} > Size_{b_2}, \\ Size_{b_2} / Size_{b_1}, & otherwise. \end{cases} \quad (6)$$

The values are then normalized between 0 and 1. Finally, the total error is computed from Equations (5) and (6):

$$E_T(P) = E_O(P) + E_{SR}(P). \quad (7)$$

The minimum total error $E_T(P)$ corresponds to the best possible eye blinking pair.

D. Step 4: Seed Pixels Computation

Once eye blinking is detected, we use the x- and y-coordinates of the blinking blobs to compute the four corners (top left, bottom left, top right, and bottom right) of the tracking region of interest (ROI), as follows:

$$TopLeft(X) = CG_L(X) - [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))] - [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))], \quad (8)$$

$$TopLeft(Y) = CG_L(Y) + [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))] - [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))],$$

$$BottomLeft(X) = CG_L(X) - [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))] + [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))],$$

$$BottomLeft(Y) = CG_L(Y) + [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))] + [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))],$$

$$TopRight(X) = CG_R(X) + [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))] - [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))],$$

$$TopRight(Y) = CG_R(Y) - [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))] - [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))],$$

$$BottomRight(X) = CG_R(X) + [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))] + [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))],$$

$$BottomRight(Y) = CG_R(Y) - [(d(b_1,b_2)/4)\sin(\theta_L(b_1,b_2))] + [(d(b_1,b_2)/4)\cos(\theta_L(b_1,b_2))],$$

where $CG_L(X)$ and $CG_L(Y)$ are x- and y-coordinates of the center of the left eye blob, respectively, while $CG_R(X)$ and $CG_R(Y)$ are x- and y-coordinates of the center of the right eye blob, respectively. This information is then supplied to the tracking algorithm to initiate the tracker.

Having defined the ROI, we subdivide it to roughly localize the periorbital area. The entire ROI is first divided into two equal size columns: a left column and right column. Each column is then subdivided into four equally sized cells. The top right cell of the left column is labeled as the left periorbital region and the top left cell of the right column is labeled as the right periorbital region. Pixels with the highest temperatures in the left periorbital region and the right periorbital region are then labeled as the left seed pixel and the right seed pixel respectively. The seed information is then used to initiate the periorbital segmentation algorithm.

III. Experimental Results

We have tested the proposed algorithm on thermal clips of 39 subjects. The thermal clips were captured during the interrogation of subjects who were suspects of check stealing in a mock crime scenario [1]. The goal of the proposed approach was to detect at least one eye blinking before the interrogation began. Figure 4 illustrates that the algorithm successfully achieved this goal for all the subjects. On average, it used 300 thermal frames to detect the first blinking in the thermal clips.

There is inter-individual variation in the number of blinks detected (see Figure 4) that is due to variation in the involuntary blinking rate, variation in the number of thermal frames before the beginning of the interview, and the algorithm's error in the presence of excessive head motion. Since false negatives (fail to detect eye blinkings) do not produce adverse outcomes, we have biased the parameters to minimize false positives (false blinking detection). As a result, the algorithm misses blinking in excessive head motion. However, we are not worried about these cases as just one involuntary blinking is required to initiate the process.

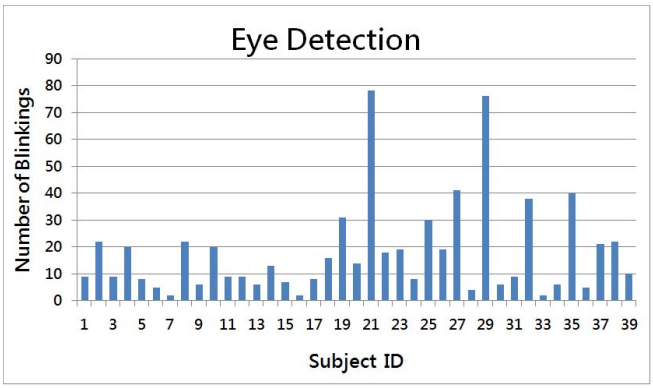


Figure 4: Number of blinkings detected before the start of the interrogation.

There were five instances of false positive detection in the dataset (see Figure 5). The main reason for the false positive detection was that whenever a subject raised his/her eyebrows involuntarily (e.g., in surprise or frustration emotions) or fast enough, the algorithm detected this change as eye blinking. This happened because the eyebrows and eye blinking blobs share similar geometric characteristics in the thermal domain. One possible solution to this problem is to use the thermal signature of the eyebrows in the detection algorithm, which is typically lower than that of eyelids.

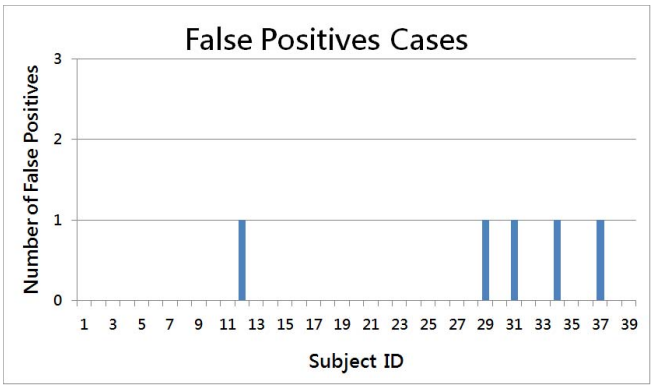


Figure 5: Number of false positive blinks.

The algorithm works optimally at approximately 40 frames per second (fps) video capture rate. Eye blinking detection at a higher video capture rate may require larger inter-frame distance among the three frame samples.

The proposed approach eliminates user intervention in the initiation of the periorbital signal extraction process. This significantly reduces the amount of labor hours. Roughly

speaking, the total labor to initialize 39 clips has been reduced from 4 hours (5mins/clip x 39 clips) to just a few minutes. More importantly, the method has accelerated data processing speed as the process can now be run in a batch mode. The time to complete one run of the signal extraction process for the 39 clips has been reduced from a week (45 min/clip x 39 clips) to a little over a day.

The approach is computationally expensive as the *second-order change detection* computes the motion velocity at every frame and uses median filtering to remove isolated pixels. Therefore, once the signal extraction process is initialized, we stop detecting the blinking.

Figure 6 shows outputs from every algorithmic step that was discussed in Section 2.

IV. Discussion and Conclusions

User intervention was one of the major obstacles in achieving automation in the deception detection process. We propose a novel approach to solve the problem. The periorbital region is localized via detection of the eyes in the thermal imagery.

This approach allows batch-processing of data as user intervention is not required for initiation of the periorbital signal extraction process. Thus, it saves long labor hours, especially when massive amounts of data need to be processed. It also speeds up data processing as it can run day and night without waiting for user input. The other advantage of the method is that it guarantees to select the same spatiotemporal location for repetitive processing of a clip. This is very important when comparison is made among different versions of the tracking and segmentation algorithms.

Ultimately, this approach will prove very convenient when a deception analysis system is deployed in the field, especially, for quick screening at security checkpoints where minimal user intervention is preferred to accelerate the screening process.

By locating the eyes on the face, one can locate other facial areas of sympathetic importance, such as the supraorbital and maxillary areas (see Figure 1). Therefore, this approach can be extended to automate more than one facial thermal signal extraction processes. A natural expansion of our method is to perform blinking quantification and investigate its relationship to deception analysis.

Acknowledgment

This material is based upon work supported in part by the Defense Academy for Credibility Assessment (DACA). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

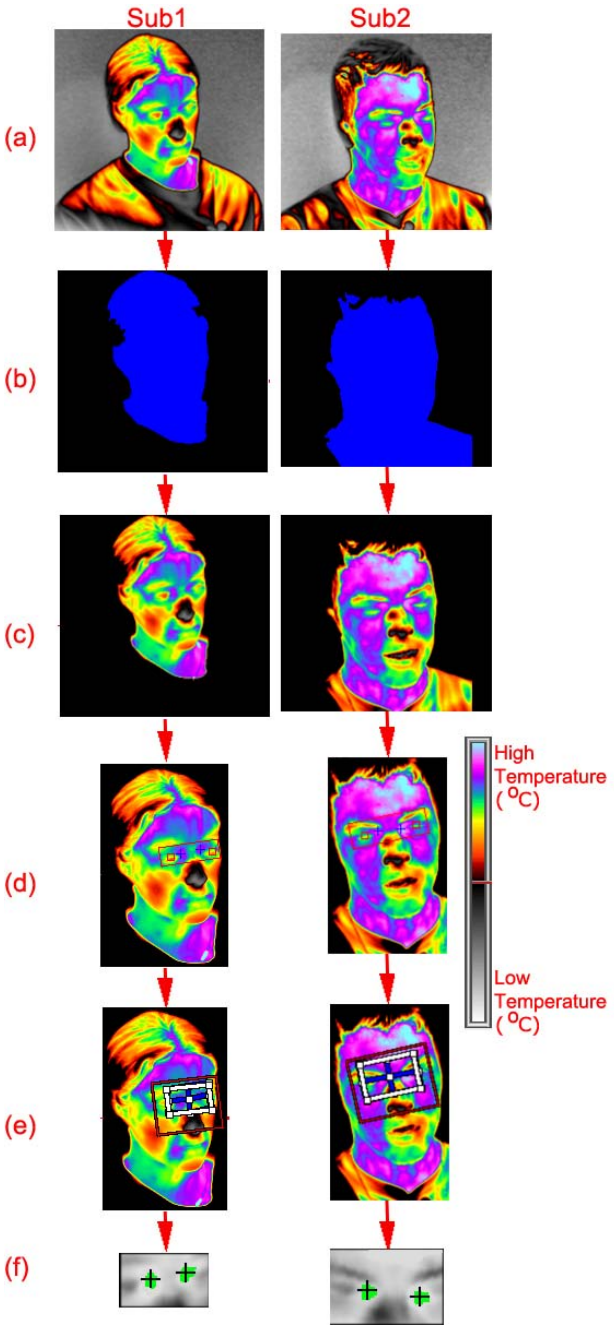


Figure 6: Automatic initialization of the periorbital signal extraction process for two subjects. (a) Original thermal image. (b) Binary mask that represents skin area. (c) Skin segmentation output. (d) Blinking detection, the location of the seed pixels and the tracking region of interest. (e) Tracking initialization. (f) Periorbital tissue segmentation.

REFERENCES

[1] P. Tsiamyrtzis, J. Dowdall, D. Shastri**,** I. Pavlidis, M.G. Frank, and P. Ekman, "Imaging facial physiology for the detection of deceit", *International Journal of Computer Vision,* vol. 71, no. 2, pp. 197-214, October 2006.

[2] D. Shastri, P. Tsiamyrtzis, and I. Pavlidis, "Periorbital thermal signature extraction and applications", in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 102-105, Vancouver, British Columbia, Canada, August 20-24, 2008.

[3] J. Dowdall, I. Pavlidis, P. Tsiamyrtzis. "Coalitional tracking", *Computer Vision and Image Understanding*. vol. 106, no. 2-3, pp. 205-219, May-June 2007.

[4] B. J. Moxham, C. Kirsh, B. Berkovitz, G. Alusi, and T. Cheeseman, "Interactive head & neck (CD-ROM)," Primal Pictures, December 2002.

[5] P. Buddharaju, I. Pavlidis, and P. Tsiamyrtzis, "Pose-invariant physiological face recognition in the thermal infrared spectrum", in *Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition,* pp. 53-60, New York, New York, June 17-22, 2006.

[6] H. Samet, and M. Tamminen, "An Improved approach to connected component labeling of images", *Proceedings of Computer Vision and Pattern Recognition*, pp. 312-318, 1986.

[7] M. Dillencourt, H. Samet, and M. Tamminen, "A general approach to connected component labeling for arbitrary image representations", *Journal of the ACM*, vol. 39, no. 2 pp. 253-280, April 1992.

[8] D. Gorodnichy, "Towards automatic retrieval of blink-based lexicon for persons suffered from brain-stem injury using video cameras", *Conference on Computer Vision and Pattern Recognition Workshop*, vol. 5, pp. 68, 2004.

[9] Leslie G. Farkas, *Anthropometry of the head and face in medicine*, 1981.