(Excerpt from the PhD Thesis entitled 'Using 3D Facial Motion for Biometric Identification', 11/2009)

PhD Thesis - Chapter 3

Facial Motion in Biometrics: System Design and Specifications

Author: Lanthao Benedikt

Supervisors: Dr. David Marshall, Dr. Paul Rosin

April 5, 2010

Cardiff University - School of Computer Science

Address: Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom

Website: http://www.cs.cf.ac.uk

Thesis Abstract

In the current context of heightened security, expanded research effort is needed for exploring novel biometric modalities, while continuing to improve existing solutions. The purpose of this research is to investigate whether the 3D spatiotemporal dynamics of facial expressions can be used for person identification, and to compare the advantages of this novel approach to classic face recognition using static mugshots. The working hypotheses are formulated as follows:

> ***Hypothesis 1:*** *Human facial movements are stable over time and sufficiently distinctive across individuals to be used as a biometric identifier. However, there exists a hierarchy in the reproducibility and uniqueness of facial actions, such that some facial expressions are more suitable for person recognition compared to others.*
>
> ***Hypothesis 2:*** *There exists a motion signature associated with moving faces, which describes not only the behavioural traits of a person, but also reveals physical idiosyncrasies that becomes visible only when the face is in motion. This identity cue is independent from the physical information conveyed by static faces, such that the temporal dynamics of facial expressions contribute to improve identity perception.*

In this thesis, a systematic study is carried out to quantitatively assess the usefulness of facial motions for person identification, considering both the face verification and the face identification applications. The purpose of the present Chapter is to outline the architecture of an identity recognition system using facial dynamics and the performance evaluation protocol.

1 Architecture of the face recognition system

Face recognition refers to a very broad range of applications including secure access control, crowd surveillance, forensic facial reconstruction and police identification. Many systems are today commercially available, for instance, one can name several 2D systems developed by Cognitec Systems GmbH [1], Neven Vision [2], Viisage [3], and Indentix [4]. More recently, there has been an emergence of 3D face recognition systems e.g. A4Vision [5], Geometrix [6], and Genex Technologies [7]. All these systems belong to the category of static face recognition, and there exists currently no solution that exploits the idiosyncrasies of facial expressions for person identification, which is the purpose of this thesis.

In the context of a *feasibility study*, quantitative evaluations will be carried out in order to ascertain that facial dynamic is a viable biometric. To this end, a prototype of a 'facial dynamic-based' recognition system needs to be built, the architecture of which can be inspired from that of static face recognition systems, with some modifications. There are typically four basic components:

1. **The data acquisition module:** this module comprises the imaging and lighting apparatus necessary for capturing video data of the users. The present application requires an off-the-shelf 3D video camera that needs to be sufficiently fast to accurately capture real-time facial deformations. A survey of commercially available imaging technologies can be found in Appendix A.

   Since all 3D dynamic image capture systems today operate within a limited field of vision[1], they are not suitable for recognition from far distance, e.g. crowd surveillance. For this reason, the present study will focus solely on applications where the users are cooperative e.g. computer login and secure

[1] e.g. the 3dMD Dynamic System requires the user to be within $\approx 1m$ from the camera pods.

access control. The users are typically required to perform some short verbal or nonverbal facial actions, the recording conditions are designed to model as closely as possible real-life situations.

2. **The feature extraction module:** this module comprises the algorithms necessary for extracting and quantifying facial motion characteristics from videos. At enrolment, facial performances of known persons are collected and used to train a 3D spatiotemporal statistical model, as explained in Chapter 5. This model can be subsequently used to estimate/extract facial dynamics from a given video footage. There are two feature extraction scenarios:

   - Ground-truth : very accurate technique for extracting facial motions, but requires manual data annotation, as discussed in Chapter 5. This method is preferred whenever the feature extraction does not need to be performed automatically in real-time. For example, during the enrolment phase, biometric samples of known persons are extracted and stored in a database, the *gallery*. Because this operation is performed *behind-the-scene*, manual landmark placement can be considered.

   - Model fitting: automatic feature extraction, typically by fitting a deformable model to unseen test scans and solving an optimisation problem. As discussed in Chapter 5, developing robust fitting algorithms is still a largely unsolved problem, therefore this method is much less reliable compared to the ground-truth method. It is however indispensable for the deployment of a fully automatic face recognition solution.

3. **The gallery:** this is a database that contains the biometric templates of known persons who are enrolled into the system. Typically, a biometric template is a pair $(I_k, \mathbf{v_k})$ where I_k is a known identity - e.g. name - and $\mathbf{v_k}$ is

the facial motion signature associated with I_k.

The gallery can be either centralised or distributed e.g. a smart card carried by each user. The centralised scenario is more efficient to avoid counterfeit or lost cards, but it has to deal with scalability issues. In large-scale systems of thousands of identities, it can be computationally taxing to match the biometric sample of an unknown user - the *probe* - against all stored biometric templates in the gallery. In order to reduce the computational overhead of pattern matching, it is usually necessary to classify the templates in the centralised gallery. For example, if we know that the user is a man, there is no need to check his probe against women's templates.

4. **The decision making module:** this module comprises the pattern matching algorithms necessary for measuring the resemblance between two biometric samples, together with the implementation of the policy related to the matching process. The basic principles will be discussed in sections 2 and 3, and a survey of suitable algorithms is carried out in Chapter 6. Pattern matching involves comparing a probe of an unknown user to a biometric template in the gallery. The matching process generates a numerical estimation of the similarity. A threshold is usually defined by the system constructor, above which the biometric samples are considered as belonging to the same person. The choice of the threshold is a matter of policy. In a high-security application where the cost of a false acceptance could be high, system policy might prefer very few false acceptances and many more false rejections. In a commercial setting where the cost of a false acceptance could be small and treated as a cost of doing business, system policy might favor false acceptances in order not to falsely reject and thereby inconvenience large numbers of legitimate customers.

1.1 The recognition process: a UML diagram

There exist two scenarios of face recognition: *face verification* and *face identification*, as shown in Figure 1. They share similar feature extraction and pattern matching algorithms, but differ in the decision making process:

- The verification problem is a *one-to-one* comparison where the user presents a probe $\mathbf{v_Q}$ and a claimed identity I_k, the system compares $\mathbf{v_Q}$ to $\mathbf{v_k}$, the biometric template corresponding to person I_k in the gallery. Examples include electronic banking, access to secure buildings and user account login.

- The identification problem is a *one-to-many* comparison in which the probe $\mathbf{v_Q}$ of an unknown person is compared to all biometric templates $\{(I_k, \mathbf{v_k}), k = (1, ..., N)\}$ of known subjects enrolled in the gallery, e.g. in police identification. The system can operate in two modes. Watch list: are you in my database? and basic identification: you are in my database, can I find you?

2 Face Verification

2.1 System architecture

The face verification problem can be formally posed as follows: given a probe $\mathbf{v_Q}$ and a claimed identity I_k, determine if $(I_k, \mathbf{v_Q})$ belongs to class w_1 or class w_2, where w_1 indicates that the claim is true (a genuine user) and w_2 indicates that the claim is false (an impostor). Typically, $\mathbf{v_Q}$ is matched against $\mathbf{v_k}$, the biometric template corresponding to person I_k enroled in the Gallery. Thus

$$(I_k, \mathbf{v_Q}) \in \begin{cases} w_1, & \text{if } S(\mathbf{v_Q}, \mathbf{v_k}) \geq t \\ w_2, & \text{otherwise} \end{cases} \tag{1}$$

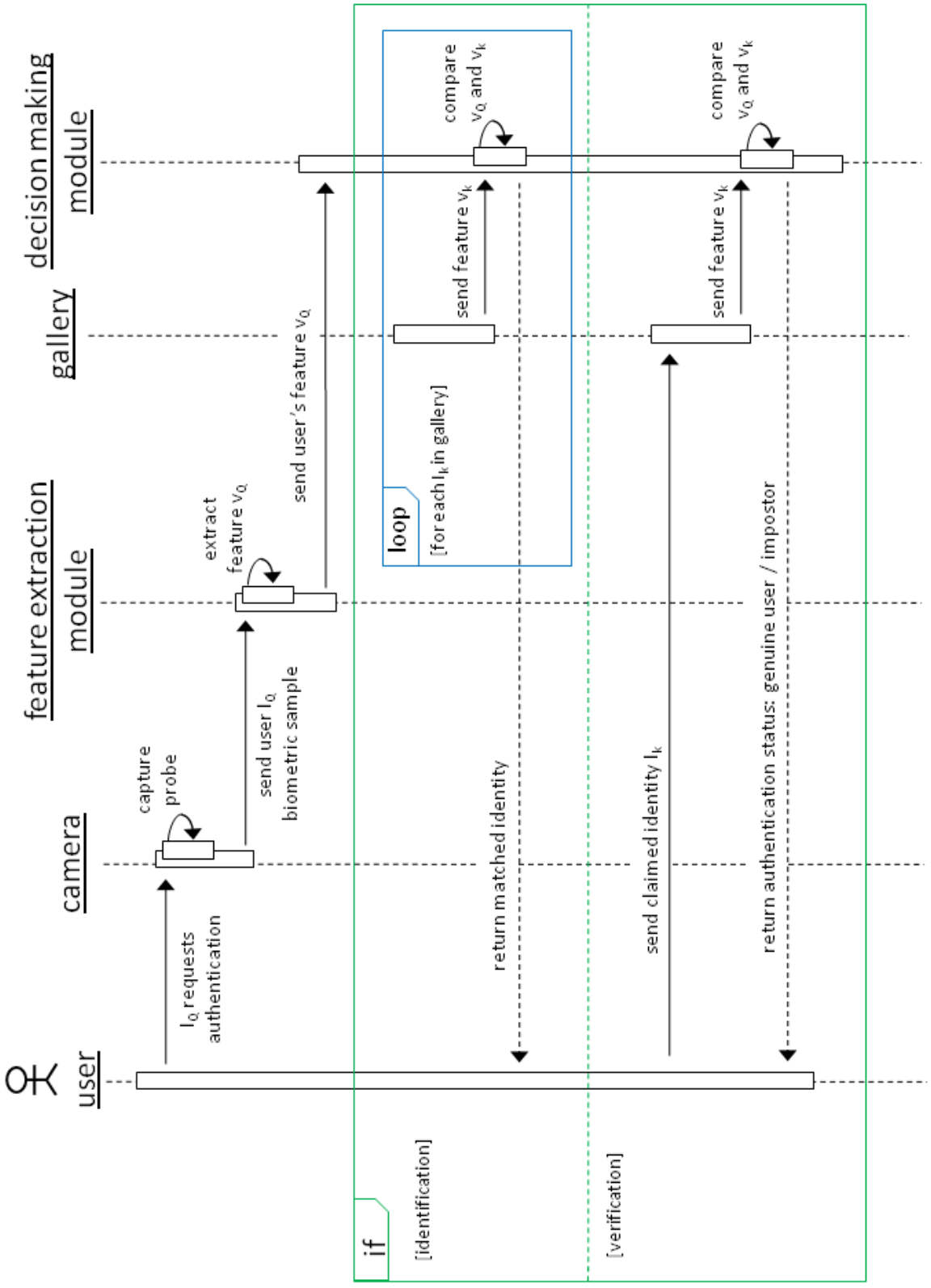

Figure 1: UML (Unified Modeling Language) sequence diagram of the recognition process.

where S is a function that measures the similarity between $\mathbf{v_Q}$ and $\mathbf{v_k}$. t is a threshold chosen by the system constructor according to the required security level. The value $S(\mathbf{v_Q}, \mathbf{v_k})$ - typically a single number - is called the matching score. The higher the score, the more certain the system is that $\mathbf{v_Q}$ and $\mathbf{v_k}$ belong to the same person. Figure 2 depicts the main modules in a face verification system.

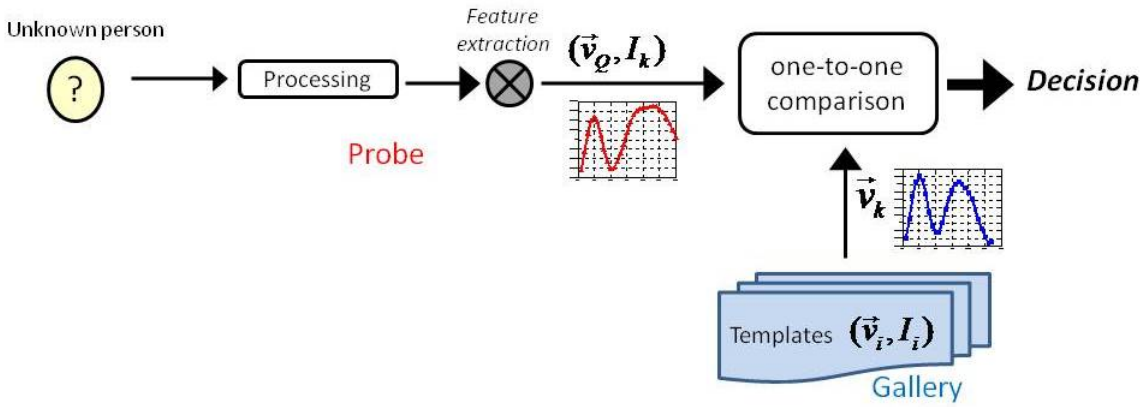

Figure 2: Face verification: given a probe $\mathbf{v_Q}$ of an unknown person and a claimed identity I_k, determine if the person is a genuine user or an impostor. Typically, $\mathbf{v_Q}$ is matched against $\mathbf{v_k}$, the biometric template of I_k.

2.2 Performance evaluation

The performance of a verification system is characterised by two error statistics: the false-reject rate (FRR) when the system mistakes two biometric samples of the same person to be from different persons, and the false-accept rate (FAR) when the system mistakes biometric samples of two different persons to be from the same person. Both FAR and FRR are functions of the threshold t and there is a trade-off between these. The system performance at all the operating points (thresholds t) can be depicted in the form of a Receiver Operating Characteristic (ROC), as

depicted in Figure 3.

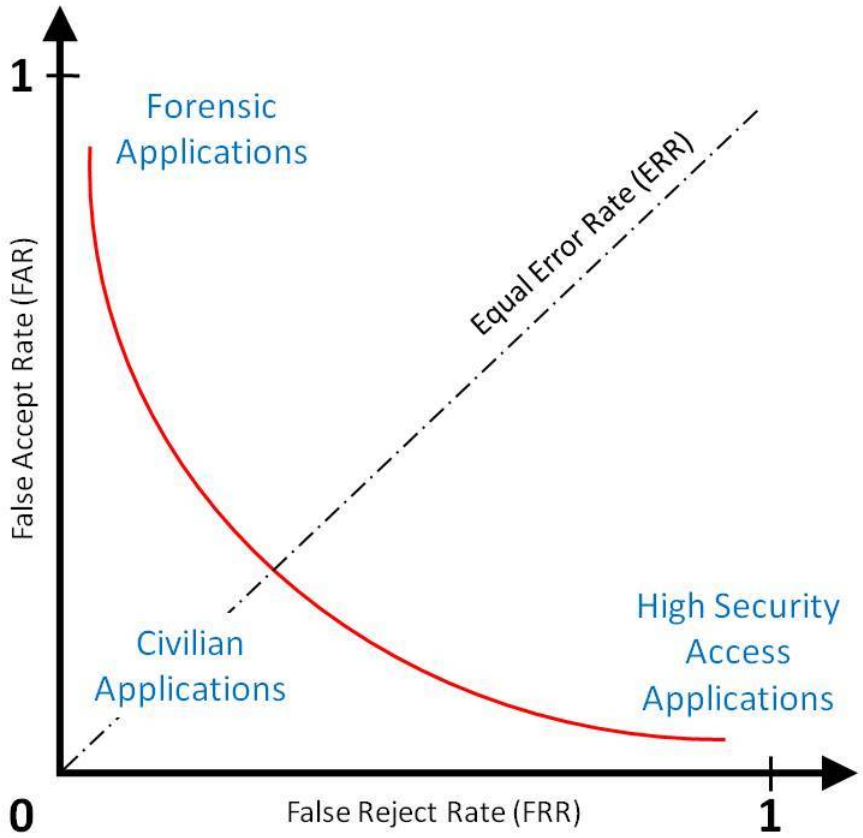

Figure 3: ROC curve of a face verification system. Each operating point on the curve corresponds to a particular value of threshold t. t is chosen according to the required security level of the application.

3 Face Identification

3.1 System architecture

The face identification problem can be formally posed as follows: given a probe $\mathbf{v_Q}$ of an unknown person, determine the identity $I_i, i \in \{1, 2, ..., N, N+1\}$ where

$I_1, I_2, ..., I_N$ are the identities enrolled in the Gallery and I_{N+1} indicates the reject case where no suitable identity can be determined for the user. Thus

$$\mathbf{v_Q} \in \begin{cases} I_i, & \text{if } \max_i S(\mathbf{v_Q}, \mathbf{v_i}) \geq t, i = 1, 2, ..., N \\ I_{N+1}, & \text{otherwise} \end{cases} \tag{2}$$

where $\mathbf{v_i}$ is the biometric template corresponding to identity I_i, S is the function that measures the similarity between $\mathbf{v_Q}$ and $\mathbf{v_i}$, and t is a predefined threshold. Since biometric samples of the same individual taken at different times are almost never identical (due to different imaging conditions or different interactions between the user and the system, for example), it is difficult to determine the perfect match. Thus, instead of reporting a unique identity I_i, the system outputs a set of possible matches which can be determined in two ways: threshold-based or rank-based. Figure 4 depicts the main modules in a face identification system.

- In a threshold-based mode, t is preset by the constructor according to the security level required. If t is too large, the system may fail to identify many users. If t is too small, the system will output a large number of possible matches, many of these are inaccurate.

- The more common approach to design an identification system is the rank-based mode. The system typically determines a set of m best matches which present the highest matching scores. m is usually referred to as the *rank*. The system always returns a fixed number of outputs: the m known identities enrolled in the gallery who resemble the most to the unknown user, sorted in some order.

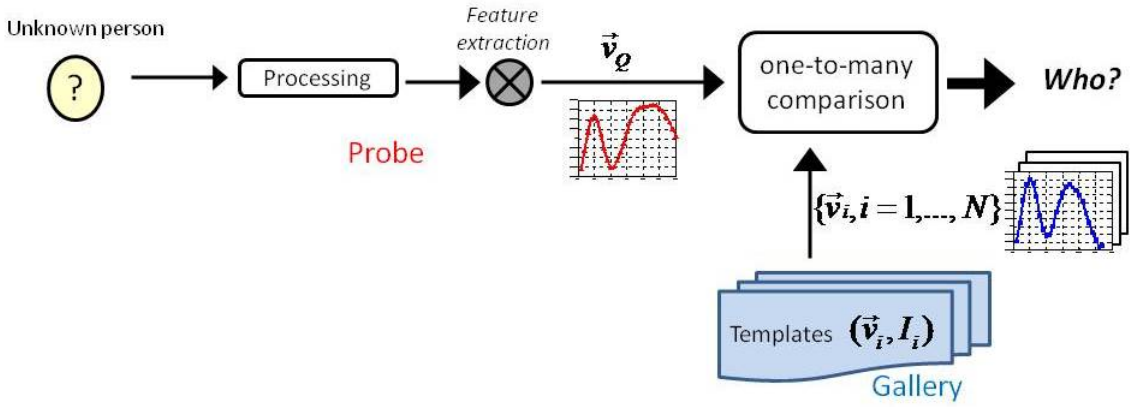

Figure 4: Face identification: the probe $\mathbf{v_Q}$ of an unknown user is matched against all biometric templates $\mathbf{v_i}$ in the Gallery to determine his/her identity. The system can operate in two modes. Watch list: are you in my database? and basic identification: you are in my database, can I find you?

3.2 Performance evaluation

The identification rate of the system is traditionally measured in terms of cumulative match, i.e. the percentages of times the correct identity can be found among the m best matches. The performance of the face identification system is usually plotted as the Cumulative Match Curve (CMC) where the identification rate is plotted against the rank m [8].

Appendix A: Survey of 3D motion capture systems

In order to identify a suitable image capture system and adequate recording conditions, one possible approach is to consider the problem from the system usability point of view. An acceptable solution must typically satisfy the following requirements:

- The recognition must be accurate, even when operating under imperfect real-world conditions, e.g. poor lighting, unpredictable interactions of the user and the system, variability of the physical and emotional conditions of the user, etc. The data acquisition in this study is designed to model as closely as possible such real-life imperfections rather than to acquire high quality video data.

- The solution must be user-friendly, which implies that the data acquisition is contact-less and non-invasive (i.e. no make-up or physical markers). This eliminates the use of marker-based motion capture systems e.g. PhaseSpace [9] and OptiTrack [10], some of which have been employed in related works, where the users are required to wear reflective markers and headband [11, 12].

- Last but not least, the solution must be affordable, in particular if one considers large scale deployment at airport security checkpoints or at ATM machines. This raises a particular question whether low-cost 2D video cameras should be preferred to the far more precise but cost-prohibitive 3D motion capture. In this study, we propose to compare the performances of face recognition using respectively 2D and 3D video data in order to understand whether a trade-off can be found between accuracy and cost.

The development of 3D surface-imaging technology has only occurred recently, which is a much safer solution compared to traditional 3D medical imaging techniques such as CT (computed tomography) that involves X-ray radiations. Furthermore, though still expensive, 3D surface-imaging is more affordable than both CT and fMRI (Functional Magnetic Resonance Imaging). Marker-free systems fall into two categories:

- Laser-based: a laser beam is scanned across the target object band-by-band and a point cloud representing the object's surface is generated. A human face requires thousands of bands to achieve a good accuracy, and the scanning of the entire face may take up to $20s$ from top to bottom. This technique is highly intrusive, causes great discomfort for users, and the band-by-band scanning is not suitable for capturing real-time facial dynamics [13].

- Optical-based: these systems are divided into three sub-categories [14]: passive stereo e.g. the Geometrix system [15], pure structured-light e.g. the Minolta sensor [16], and the most sophisticated technology to date is the stereo photogrammetry [17,18]. Stereo photogrammetry computes the depth information of an object from two or more pictures taken from different viewpoints using triangulation. The position of each camera relatively to the others needs to be known and can be calculated during an initial calibration using a known target object [19]. Stereo photogrammetry can be passive or active:

  - Passive stereo e.g. the Dolphin-Di3D Facial Camera System $Di3D^{TM}$ [18] and the passive stereo system developed by Dimensional Imaging [20] rely on natural patterns such as skin pores, freckles, scars, etc to find common points on photographs. It depends tremendously on the integrity of the pixels, requires high resolution cameras to ensure that there is enough

surface details. Lighting conditions must be carefully controlled.

– In contrast, the active stereo approach e.g. the $3dMDFace^{TM}$ Dynamic System [13] projects unstructured light pattern to give the stereo algorithms as much information as possible to triangulate the geometry. Typically, active stereo is much more resilient to variances in lighting conditions.

References

[1] "Cognitec Systems GmbH," Source: http://www.cognitec-systems.de. Last visited 9^{th} August, 2009 .

[2] "Neven Vision Inc.," Source: http://www.nevenvision.com. Last visited 9^{th} August, 2009 .

[3] "Viisage, Littleton, MA," Source: http://www.viisage.com. Last visited 9^{th} August, 2009 .

[4] "Identix, Minnetonka, MN," Source: http://www.identix.com. Last visited 9^{th} August, 2009 .

[5] "A4Vision Inc.," Source: http://www.a4vision.com. Last visited 9^{th} August, 2009 .

[6] "Geometrix Inc.," Source: http://www.geometrix.com. Last visited 9^{th} August, 2009 .

[7] "Genex Technologies Inc.," Source: http://www.genextec.com. Last visited 9^{th} August, 2009 .

[8] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior, "Guide to Biometrics," Cambridge University Press, New York, USA vol. 22(4) (2004).

[9] "PhaseSpace Motion Capture," http://www.phasespace.com. Last visited 9^{th} August, 2009 .

[10] "OptiTrack," http://www.naturalpoint.com/optitrack. Last visited 9^{th} August, 2009 .

[11] B. Knappmeyer, I. M. Thornton, and H. H. Bulthoff, "The use of facial motion and facial form during the processing of identity," Trends in cognitive sciences **vol. 43,** pp. 1921–1936 (2003).

[12] J. F. Cohn, K. Schmidt, R. Gross, and P. Ekman, "Individual Differences in Facial Expression: Stability over Time, Relation to Self-Reported Emotion, and Ability to Inform Person Identification," in Proc of the International Conference on Multimodal User Interfaces **vol. 116,** pp. 491–498 (2002).

[13] C. Lane and W. H. Jr, "Completing the 3-dimensional picture," American Journal of Orthodontics and Dentofacial Orthopedics **vol. 133(4),** pp. 612–620 (2007).

[14] K. W. Bowyer, K. Chang, and P. J. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," Computer Vision and Image Understanding **vol. 101(1),** pp. 358–361 (2006).

[15] G. Medioni and R. Waupotitsch, "Face recognition and modeling in 3D," in IEEE International Workshop on Analysis and Modeling of Faces and Gestures pp. 232–233 (2003).

[16] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in Multimodal User Authentication Workshop pp. 25–32 (2003).

[17] G. U. 3dMD, Atlanta, "http://www.3dmd.com. Last visited 9^{th} August, 2009," .

[18] "Dolphin Imaging," http://www.dolphinimaging.com. Last visited 9^{th} August, 2009 .

[19] R. Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition pp. 364–374 (1986).

[20] "Dimensional Imaging," http://www.di3d.com. Last visited 9^{th} August, 2009 .