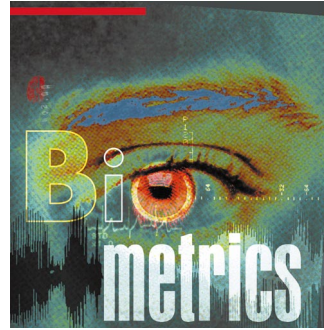


BioID: A Multimodal Biometric Identification System



By combining face, voice, and lip movement recognition, this identification system yields highly accurate results. Because it includes a dynamic feature, it provides more security than systems using only static features.

Robert W. Frischholz
Ulrich Dieckmann
 Dialog
 Communication
 Systems AG

Most systems that control access to financial transactions, computer networks, or secured locations identify authorized persons by recognizing passwords or personal identification numbers. The weakness of these systems is that unauthorized persons can discover others' passwords and numbers quite easily and use them without detection. Biometric identification systems, which use physical features to check a person's identity, ensure much greater security than password and number systems. Biometric features such as the face or a fingerprint can be stored on a microchip in a credit card, for example. If someone steals the card and tries to use it, the impostor's biometric features will not match the features stored in the card, and the system will prevent the transaction.

A single feature, however, sometimes fails to be exact enough for identification. Consider identical twins, for example. Their faces alone may not distinguish them. Another disadvantage of using only one feature is that the chosen feature is not always readable. For example, some five percent of people have fingerprints that cannot be recorded because they are obscured by a cut or a scar or are too fine to show up well in a photograph.

Therefore, our company, Dialog Communication Systems (DCS AG), developed BioID, a multimodal identification system that uses three different features—face, voice, and lip movement—to identify people. With its three modalities, BioID achieves much greater accuracy than single-feature systems. Even if one modality is somehow disturbed—for example, if

a noisy environment drowns out the voice—the other two modalities still lead to an accurate identification. BioID is the first identification system we know of that uses a dynamic feature, lip movement.¹ This feature makes BioID more secure against fraud than systems using only static features such as fingerprints. Commercially available since 1998, the system has demonstrated its reliability in many installations around the world.

SYSTEM FUNCTIONS

Figure 1 diagrams BioID's functions. The system acquires (records), preprocesses, and classifies each biometric feature separately. During the training (enrollment) of the system, biometric templates are generated for each feature. For classification, the system compares these templates with the newly recorded pattern. Then, using a strategy that depends on the level of security required by the application, it combines the classification results into one result by which it recognizes persons.

Data acquisition and preprocessing

The input to the system is a recorded sample of a person speaking. The one-second sample consists of a 25-frame video sequence and an audio signal. From the video sequence, the preprocessing module extracts two optical biometric traits: face and lip movement while speaking a word. To extract those features, the preprocessing module must have exact knowledge of the face's position. Since this recognition system should be able to function in any arbitrary environment with off-the-

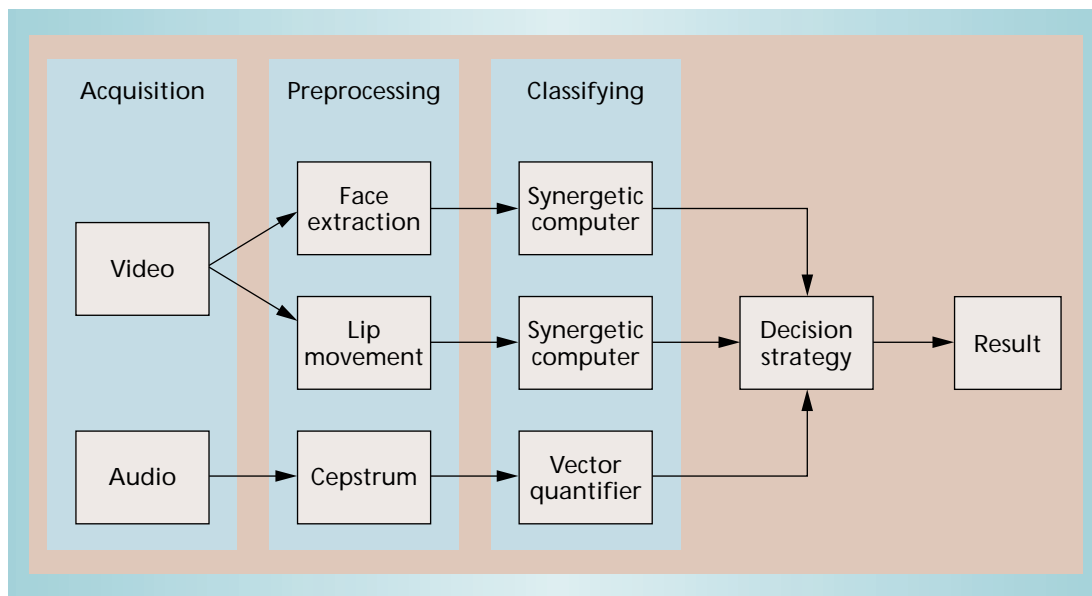


Figure 1. BioID's main functions. From video and audio samplings of a person speaking, the system extracts facial, lip movement, and voice features (a cepstrum is a special form of the frequency spectrum). Synergetic computers and a vector quantifier classify the recorded pattern and combine the results.

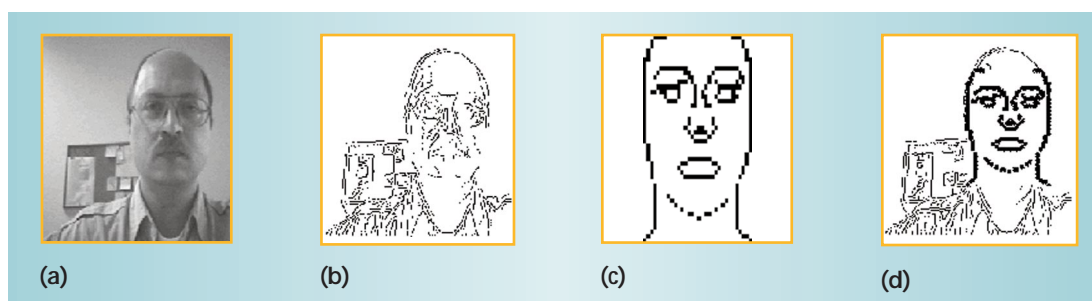


Figure 2. Face location: (a) original image, (b) edge-extracted image, (c) face model, and (d) face model overlaid on the edge-extracted image.

shelf video equipment, the face-finding process is one of the most important steps in feature extraction.

Using Hausdorff distance for face location. To detect the location of a face in an arbitrary image, identification systems often use neural-net-based algorithms,² but these approaches are very time-consuming. Instead, BioID uses a newly developed, model-based algorithm that matches a binary model of a typical human face to a binarized, edge-extracted version of the video image. Figure 2 illustrates this process. The face extractor bases its comparison on the modified Hausdorff distance,³ which determines the model's optimal location, scaling, and rotation. The Hausdorff distance uses two point sets, A and B. To obtain the Hausdorff distance, we calculate the minimum distance from each point of set A to a point of set B and vice versa. The maximum of these minimum distances is the Hausdorff distance. Point set A represents the face model, and point set B is a part of the image. The minimum of the calculated maximum distances determines the part of the image where the face is located.

After detecting the face boundaries, the preprocessing module locates the eyes from the first three

images of the video sequence, under the assumption that a person often closes his eyes when beginning to speak. As with face location, eye location also relies on an image model and the Hausdorff distance. Locating the eye positions allows all further processing to take place.

Facial features. For face recognition, the preprocessing module uses the first image in the video sequence that shows the person with eyes open. Once the eyes are in position, the preprocessing module uses anthropomorphic knowledge to extract a normalized portion of the face. That is, it scales all faces to a uniform size, as shown in Figure 3. This procedure ensures that the appropriate facial features are analyzed—not, for example, the head size, the hairstyle, a tie, or a piece of jewelry. After rotating and scaling the image, the preprocessing module extracts a gray-scale image. Some further preprocessing steps take care of lighting conditions and color variance.

Lip movement. BioID collects lip movements by means of an optical-flow technique⁴ that calculates a vector field representing the local movement of each image part to the next image in the video sequence.

Figure 3. Samples of extracted faces: BioID scales all faces to the same size and crops the images uniformly for easier comparison. This photograph collection shows 12 individuals; note the uniformity that the system achieves.

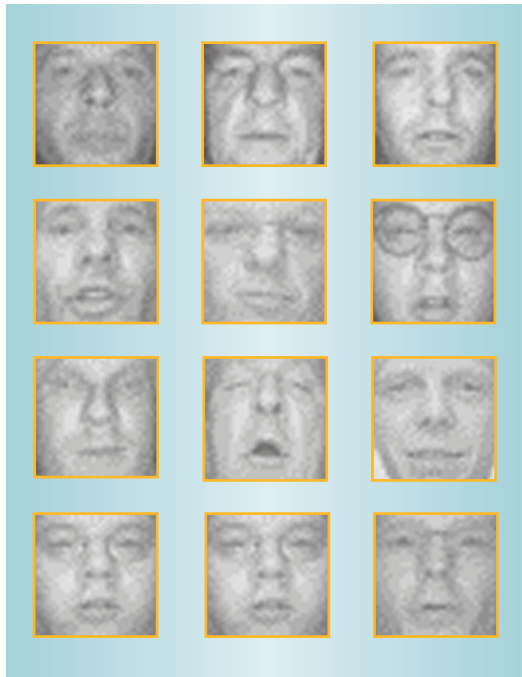
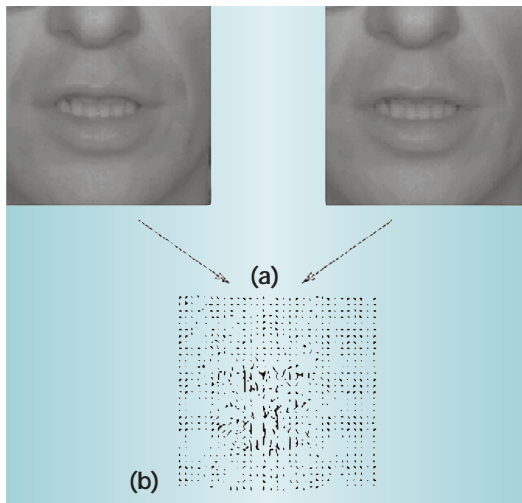


Figure 4. Example of an optical-flow vector field. The lip movement between (a) the two images is defined by (b) the vector field.



For this process, the preprocessing module cuts the mouth area out of the first 17 images of the video sequence. It gathers the lip movements in 16 vector fields, which represent the movement of identifiable points on the lip from frame to frame. Figure 4 shows the optical-flow vector field of two consecutive images. To reduce the amount of data, we reduce the optical flow resolution to a factor of four through averaging. Finally, a 3D fast Fourier transformation of the 16 vector fields takes place. The result is a one-dimensional lip movement feature vector, which the system uses for training and classification of lip movement. Essentially, we are condensing the detailed movement defined by several vector fields to a single vector. Figure 5 presents an overview of the optical preprocessing steps.

Acoustic preprocessing. We record the speech sample using a 22-kHz sampling rate with 16-bit resolution. After channel estimation and normalization, the preprocessing module divides the time signal into sev-

eral smaller, overlapping windows. For each window, it calculates the cepstral coefficients, which form the audio feature vector. The vector quantifier uses this feature vector for classifying audio patterns.

Classification

We use the so-called synergetic computer to classify the optical features, and a vector quantifier to classify the audio feature. (The synergetic computer is a set of algorithms that simulate synergetic phenomena in theoretical physics.^{5,6}) Tests have shown that the synergetic computer performs very well on optical data but not on acoustical data, and the vector quantifier has demonstrated good performance on audio data in previous applications.

Lip movement and face classification. The synergetic computer serves as a learning classifier for optical biometric-pattern recognition. In the training phase, BioID records several characteristic patterns of one person's face and lip movement, and assigns them to a class. Each class represents one person. During the training process, all patterns are orthogonalized and normalized. The resulting vectors, called *adjunct prototypes*, are compressed in each class. This leads to one prototype for each class (person), representing all patterns initially stored in the class without any loss of information. We call this prototype a *biometric template*.

The classification process is fairly easy: We preprocess and multiply a newly recorded pattern with each biometric template. We rank the obtained scalar products, and the highest one (as an absolute value) leads to the resulting class. This strategy is known as winner-takes-all.⁵ Because this principle always leads to a classification—that is, no pattern is rejected—we also take the second highest scalar product into account. If the difference between the highest and the second highest is smaller than a given threshold, we reject the pattern. We judge the classification result as follows: If the two highest scalar products have nearly the same value, the two classes (two people) are indistinguishable, and the classification is “insecure.”

The training process for the optical features of 30 persons with five learning patterns each takes about 15 minutes on an Intel Pentium II. The classification time is very short (several milliseconds) since there are only 30 scalar products to calculate. Figure 6 shows the facial biometric templates of six classes (six people). The figure shows that each template consists of several overlying patterns.

Voice recognition. We use vector quantification to classify the audio sequence. In the system-training phase, the audio preprocessing module analyzes several recordings of a single person's voice. From each voice pattern, it creates a matrix, and the vector quantifier combines these matrices into one matrix. This matrix serves as a prototype (or codebook) that dis-

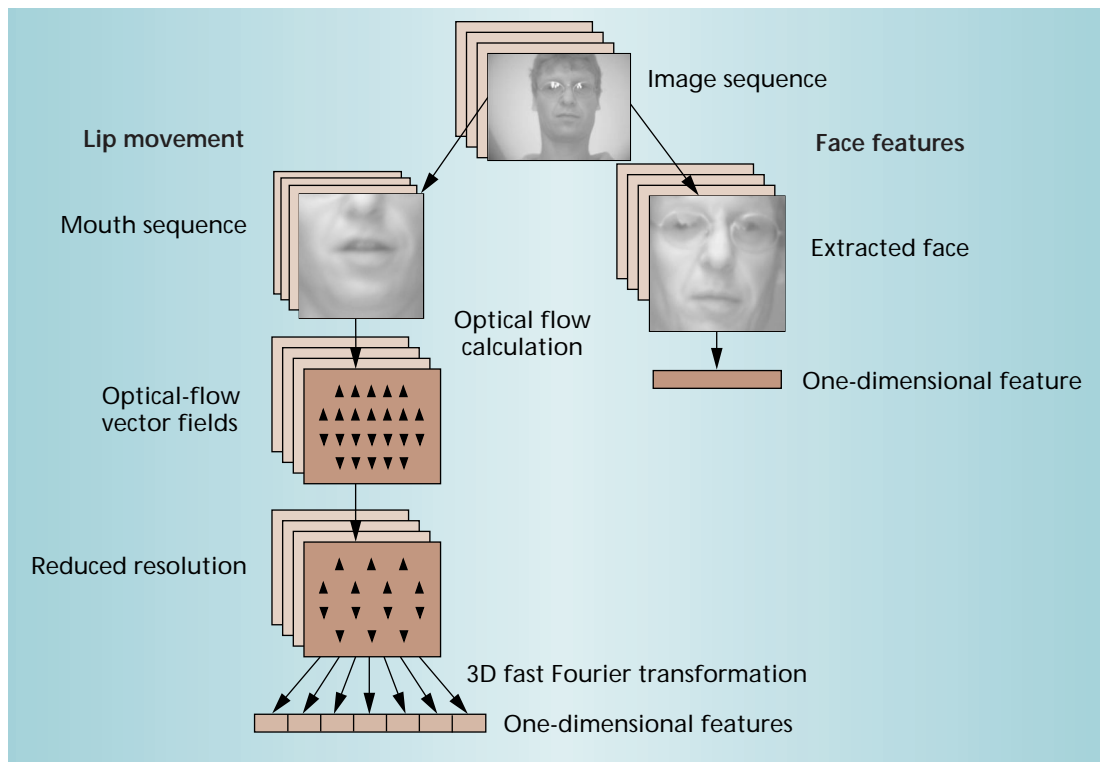


Figure 5. Optical pre-processing of the face and lip movement features.

plays the reference voice pattern. Using this voice pattern, a minimum-distance classifier assigns the current pattern to the class showing the smallest distance.

Sensor fusion

To analyze the classification results, BioID chooses from different strategies to obtain various security levels. Figure 7 shows the available sensor fusion options—that is, the combinations of the three results. For normal operations, the system uses a two-out-of-three strategy, which classifies two of the three biometric features to an enrolled class (person), without falling below threshold values set in advance. The threshold values apply to the relative distances of the best and the second-best scalar product—that is, the two classes that best match—and can be determined by the system administrator.

For a higher security level, the system can demand agreement of all three traits—a three-out-of-three strategy. With this strategy, the probability that the system will accept an unauthorized person decreases, but one must live with the possibility that it will reject an authorized person.

Additional methods make the sum of the classification results of all traits available. These methods allow us to weight individual traits differently. For example, if the system always correctly identifies a person by lip movement, we can give this feature more significance than the others.

BioID is suitable for any application in which people require access to a technical system—computer networks, Internet commerce and banking systems, and ATMs, for example. In addition, this system secures access to rooms and build-

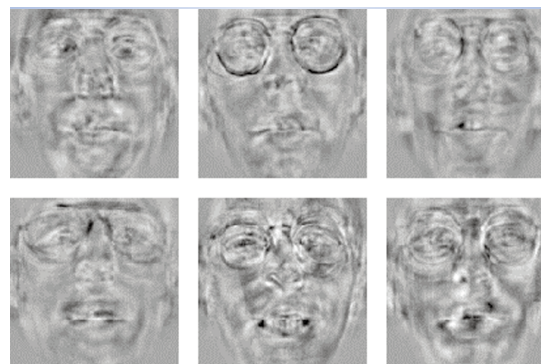


Figure 6. Six examples of synergetic-prototype faces. The white and dark areas show the most distinguishing parts of the face; the gray areas represent the less significant parts.

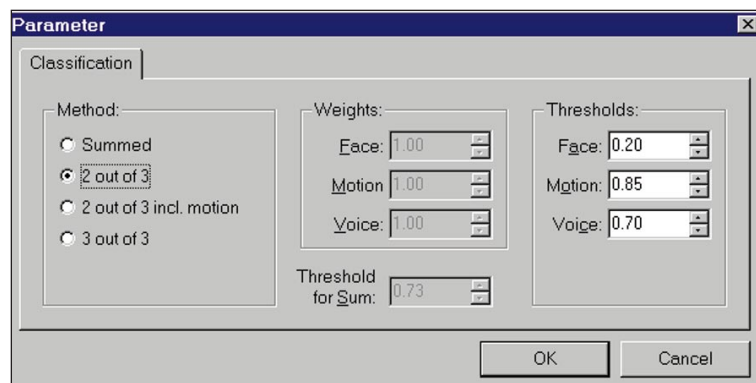
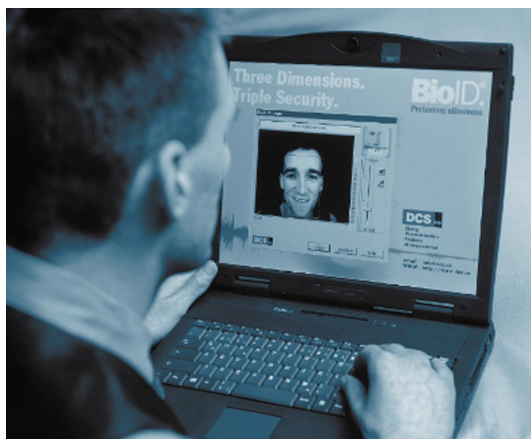


Figure 7. Sensor fusion options. BioID administrators can choose recognition criteria appropriate for the security level desired.

ings. So far, most BioID installations serve physical structures and small office-computer networks.

Depending on the application, BioID authorizes people either through identification or verification. In identification mode, the system identifies a person exclusively through biometric traits. In verification

Figure 8. Interacting with BioID: Seeking access to a computer network, the would-be user poses in front of the PC camera and speaks his name.



mode, a person gives his name or a number, which the system then verifies by means of biometric traits. Figure 8 shows a user interacting with the system.

With its multimodal concept, BioID guarantees a high degree of security from falsification and unauthorized access. It also protects the privacy rights of system users, who must speak their name or a key phrase, and therefore cannot be identified without their knowledge. To guard against the threat of unauthorized use, users can invalidate their stored reference template at any time, simply by speaking a new word and thus creating a new reference template.

In a test involving 150 persons for three months, BioID reduced the false-acceptance rate significantly below 1 percent, depending on the security level. The higher the security level, the higher the false-rejection

rate. Thus, system administrators must find an acceptable false-rejection rate without letting the false-acceptance rate increase too much. The security level depends on the purpose of the biometric system. To guard access to rooms, for example, it may be appropriate to use a very high security level to keep unauthorized people out—even if we sometimes reject an authorized person. ♦


References

1. U. Dieckmann, P. Plankensteiner, and T. Wagner, "SESAM: A Biometric Person Identification System Using Sensor Fusion," *Pattern Recognition Letters*, Vol. 18, No. 9, 1997, pp. 827-833.
2. H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Jan. 1998, pp. 23-38.
3. D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Sept. 1993, pp. 850-863.
4. J. Barron, D. Fleet, and S. Beauchemin, "Performance of Optical Flow Techniques," *Int'l J. Computer Vision*, Feb. 1994, pp. 43-77.
5. R.W. Frischholz, F.G. Boebel, and K.P. Spinnler, "Face Recognition with the Synergetic Computer," *Proc. Int'l Conf. Applied Synergetics and Synergetic Eng.*, Fraunhofer Gesellschaft für Integrierte Schaltungen, Erlangen, Germany, 1994, pp. 107-110.
6. U. Dieckmann, *Kombination verschiedener Merkmale zur biometrischen Personenerkennung* [Combination of Different Features for Biometric Person Recognition], doctoral dissertation, Berichte aus der Informatik [Computer Science Reports], Shaker Verlag, Aachen, Germany, 1999 (in German).

Robert W. Frischholz is senior vice president of the BioID Research and Development Department at Dialog Communication Systems AG, Erlangen, Germany. His research interests are biometric recognition systems and motion analysis. Frischholz has a diploma in computer science and a PhD in electrical engineering from the Fraunhofer Institute of Integrated Circuits in Erlangen. He is a member of the IEEE. Contact him at frz@dcs.de.

Ulrich Dieckmann is the senior researcher in the BioID Research and Development Department at Dialog Communication Systems AG. His research interest is biometric recognition systems, recently focusing on gaze estimation and eye tracking. Dieckmann has a diploma in computer science and a PhD in electrical engineering, both from the Fraunhofer Institute of Integrated Circuits in Erlangen. He is a member of the IEEE. Contact him at ud@dcs.de.

Good news for your in-box.




Sign Up Today for
the IEEE
Computer
Society's
e-News

Be alerted to

- articles and special issues
- conference news
- submission and registration deadlines
- interactive forums

Available for FREE
to members.



computer.org/e-News