M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 307
Appears: International Conference on Computer Vision '95, Cambridge, MA, June 20-23, 1995

Facial Expression Recognition using a Dynamic Model and Motion Energy

Irfan A. Essa and Alex P. Pentland
Perceptual Computing Group, The Media Laboratory,
Massachusetts Institute of Technology
Cambridge, MA 02139, U.S.A.

Abstract

Previous efforts at facial expression recognition have been based on the Facial Action Coding System (FACS), a representation developed in order to allow human psychologists to code expression from static facial "mugshots." In this paper we develop new, more accurate representations for facial expression by building a video database of facial expressions and then probabilistically characterizing the facial muscle activation associated with each expression using a detailed physical model of the skin and muscles. This produces a muscle-based representation of facial motion, which is then used to recognize facial expressions in two different ways. The first method uses the physics-based model directly, by recognizing expressions through comparison of estimated muscle activations. The second method uses the physics-based model to generate spatio-temporal motion-energy templates of the whole face for each different expression. These simple, biologically-plausible motion energy "templates" are then used for recognition. Both methods show substantially greater accuracy at expression recognition than has been previously achieved.

1 Introduction

Faces and facial expressions are an important aspect of human interaction, both in the context of interpersonal communication and man-machine interfaces. In recent years several researchers have attempted automatic recognition and tracking of facial expressions [19, 2, 8, 12, 16, 6]. In this paper we present two new methods for recognizing facial expressions, both of which attempt to improve over previous approaches by using a better model of facial motion and by using facial optical flow more efficiently.

Unlike previous approaches at facial expression recognition, our method does not to rely on a heuristic "dictionary" of facial motion developed for emotion coding by human psychologists (e.g., the *Facial Action Coding System (FACS)* [5]). Instead, we *objectively* quantify facial movement during various facial expressions using computer vision techniques. This provides us with a more accurate model of facial expression, allowing us to more efficiently utilize the optical flow information and to achieve greater recognition accuracy. One interesting aspect of this work is that we demonstrate that extremely simple, biologically-plausible motion energy detectors can be extracted to accurately recognize human expressions.

1.1 Recognizing Facial Motion

To categorize facial motion we need first to determine the expressions from facial movements. Ekman and Friesen [5] have produced a system for describing "all visually distinguishable facial movements", called the *Facial Action Coding System* or *FACS*. It is based on the enumeration of all "action units" (AUs) of a face that cause facial movements. There are 46 AUs in FACS that account for changes in facial expression. The combination of these action units results in a large set of possible facial expressions. For example happiness expression is considered to be a combination of "pulling lip corners (AU12+13) and/or mouth opening (AU25+27) with upper lip raiser (AU10) and bit of furrow deepening (AU11)." However this is only one type of a smile; there are many variations of the above motions, each having a different intensity of actuation.

Recognition of facial expressions can be achieved by categorizing a set of such predetermined facial motions as in FACS, rather than determining the motion of each facial point independently. This is the approach taken by Yacoob and Davis [19, 13], Black and Yacoob [2] and Mase and Pentland [8, 9] for their recognition systems. Yacoob and Davis [19], who extend the work of Mase, detect motion (only in eight directions) in six predefined and hand initialized rectangular regions on a face and then use simplifications of the FACS rules for the six universal expressions for recognition. The motion in these rectangular regions, from the last several frames, is correlated to the FACS rules for recognition. Black and Yacoob [2] extend this method, using local parameterized models of image motion to deal with large-scale head motions. These methods show a 86% overall accuracy (92% if false positives are excluded) in correctly recognizing expressions over their database of 105 expressions. Mase [8] on a smaller set of data (30 test cases) obtained an accuracy of 80%. In many ways these are impressive results, considering the complexity of the FACS model and the difficulty in measuring facial motion within small windowed regions of the face.

In our view the principle difficulty these researchers have encountered is the sheer complexity of describing human facial movement using FACS. Using the FACS representation, there are a very large number of AUs, which combine in extremely complex ways to give rise to expressions. In contradiction to this view, there is now a growing body of psychological research that argues that it is the dynamics of the expression, rather than detailed spatial deformations, that is important in expression recognition [1, 3]. Indeed several famous researchers have claimed that the timing of expressions, something that is completely missing from FACS, is a critical parameter in recognizing emotions [4, 10]. To us this strongly suggests moving away from a static, "dissect-every-change" analysis of expression (which is how the FACS model was developed), towards a whole-face analysis of facial dynamics in motion sequences.

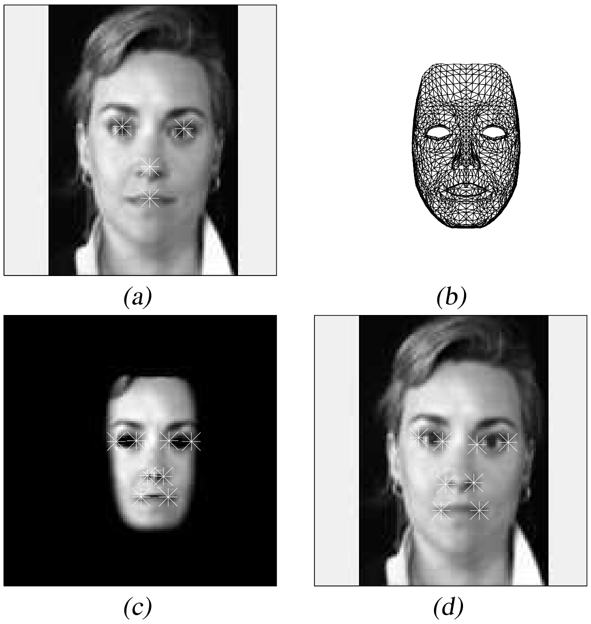


Figure 1: *Initialization on a face image using methods described by Pentland et al. [11, 12], using a canonical model of a face.*

1.2 Two Approaches

We have previously developed a detailed physical model of the human face and its musculature. By combining this physical model with registered optical flow measurements from human faces, we have been able to reliably estimate muscle actuations within the human face (see [7]).

Our *first* method for expression recognition builds on our ability to estimate facial muscle actuations from optical flow data. We have used this method to measure muscle actuations for many people making a variety of different expressions, and found that for each expression we can define a typical pattern of muscle actuation. Thus given a new image sequence of a person making an expression, we can measure the facial optical flow, estimate muscle actuations, and classify the expression by its similarity to these typical patterns of muscle actuation.

Our *second* and more recent approach builds on this methodology by “compiling” our detailed, physical model of facial muscle actuations into a set of simple, biologically-plausible motion energy detectors. For each expression we use the typical patterns of muscle actuation, as determined using our detailed physical analysis, to generate the typical pattern of motion energy associated with each facial expression. This results in a set of simple expression “detectors” each of which look for the particular space-time pattern of motion energy associated with each facial expression.

Both of these methods have the advantage that they make use of our detailed, physically-based model of facial motion to interpret the facial optical flow. Moreover, because we have experimentally characterized the optical flow associated with each facial expressions, we are not bound by the limitations of complex, heuristic representations such as FACS. The result is greater accuracy at expression recognition, and an extremely simple, biologically-plausible, motion-energy mechanism for expression recognition.

Now we discuss these two approaches in more detail.

2 Extracting Facial Parameters

2.1 Initialization

To develop a representation of facial motion for facial expression recognition requires initialization (where is the face?) and registration of all faces in the database to a set of predefined parameters. Initially we started our estimation process by manually translating, rotating and deforming our 3-D facial model to fit each face as has been done in all previous research studies.

To automate this process we are now using the *Modular Eigenspace* methods of Pentland and Moghaddam [11, 12]. This method allows us to extract the positions of the eyes, nose and lips in an image as shown in Figure 1(a). From these feature positions a canonical mesh is generated and then the image is (affine) warped to the mesh and then masked (Figure 1(c)). We can then extract “canonical feature points” on the image that correspond to our mesh (Figure 1(d)), producing a set of registered features on the face image.

2.2 Visually extracted Facial Expressions

After the initial registering of the model to the image the coarse-to-fine flow computation methods presented by Simoncelli [15] and Wang [17] are used to compute the flow within an control-theoretic approach of Essa and Pentland [7]. The model on the face image tracks the motion of the head and the face correctly as long as there is not an excessive amount of rigid motion of the face during an expression.

The method of Essa and Pentland [7] provides us with a detailed physical model and also a way of observing and extracting the “action units” of a face using video sequences as input. The visual observation (sensing) is achieved by using an optimal estimation optical flow method coupled with a geometric and a physical (muscle) model describing the facial structure within a *control-theoretic* framework. This modeling results in a time-varying spatial patterning of facial shape and a parametric representation of the independent muscle action groups, responsible for the observed facial motions. We will use these physically-based muscle control units as a representation of facial motion in our first expression recognition method, while in the second method we will use the estimated and corrected 2-D motion of the face as a spatio-temporal template. Figure 2 shows happiness and surprise expressions, and the extracted happiness and surprise expression on a canonical 3D model, regenerated by actuating the muscles. Figure 2(a) shows the canonical mesh registered on the image while Figure 2(b) shows the muscle attachments within the face model. At present we use the same generic face model for all people. The figure also shows the corrected motion energy for the two expressions. In the next section we discuss the importance of the representations extracted from this kind of analysis and compare them to FACS.

3 Analysis

The goal of this work is to develop a new representation of facial action that more accurately captures the characteristics of facial motion, so that we can employ them in recognition of facial motion. The current state-of-the-art for facial descriptions (either FACS itself or muscle-control versions of FACS) have two major weaknesses:

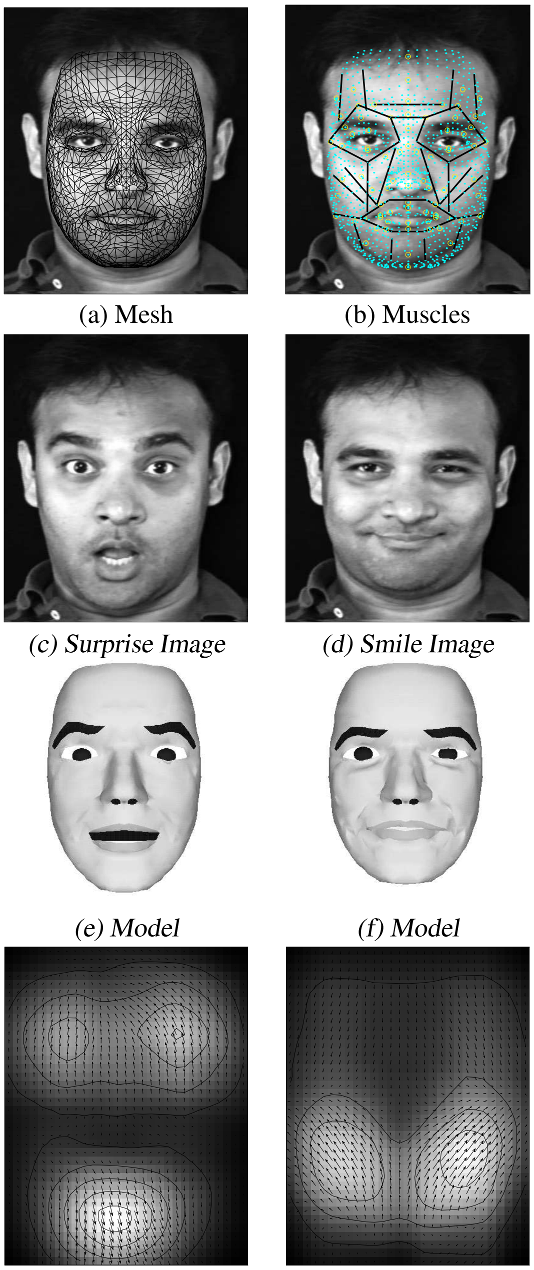


Figure 2: *Determining of expressions from video sequences. (a) Face image with a FEM mesh placed accurately over it and (b) Face image with muscles (black lines), and nodes (dots). (c) and (d) show expressions of smile and surprise, (e) and (f) show a 3D model with surprise and smile expressions resynthesized from extracted muscle actuations, and (g) and (h) show the spatio-temporal motion energy representation of facial motion for these expressions [7].*

- The action units are purely local spatial patterns. Real facial motion is almost never completely localized; Ekman himself has described some of these action units as an "unnatural" type of facial movement. Detecting a unique set of action units for a specific facial expression is not guaranteed.

- There is no time component of the description, or only a heuristic one. From EMG studies it is known that most facial actions occur in three distinct phases: *application*, *release* and *relaxation*. In contrast, current systems typically use simple linear ramps to approximate the actuation profile. Coarticulation effects are also not accounted for in any facial movement, within the FACS framework.

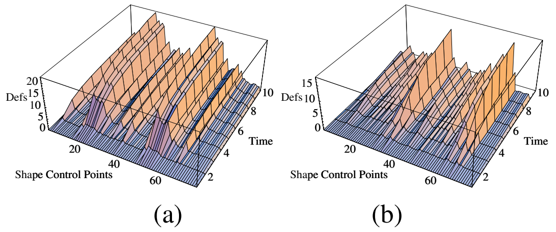


Figure 3: *Observed deformation for the (a) Raising Eyebrow, and (b) Smile expressions. Surface plots show deformation over time for actual video sequence of raising eyebrows and smiling.*

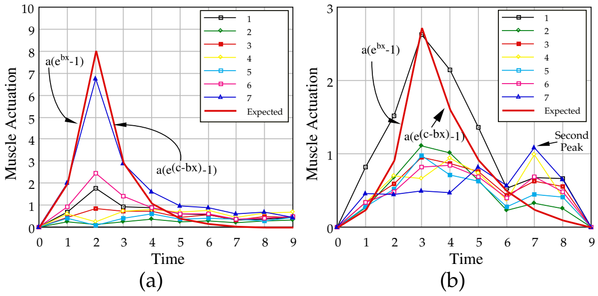


Figure 4: *Actuations over time of the seven main muscle groups for the expressions of (a) raising brow, and (b) smile. These plots shows actuations over time for the seven muscle groups and the expected profile of application, release and relax phases of muscle activation.*

Other limitations of FACS include the inability to describe fine eye and lip motions, and the inability to describe the coarticulation effects found most commonly in speech. Although the muscle-based models used in computer graphics have alleviated some of these problems [18], they are still too simple to accurately describe real facial motion.

Consequently, our earlier method [7] lets us characterize the functional form of the actuation profile, and lets us determine a basis set of "action units" that better describes the spatial properties of real facial motion. In the next few paragraphs, we will illustrate the suitability of these representations using the smile and eyebrow raising expressions.

3.1 Spatial Patterning

Figure 3(a) shows the observed motion of the control points of the face model for the expression of *raising eyebrow* by Paul Ekman. This plot was constructed by mapping the motion onto the face model (a CANDIDE model which is a computer graphics model for implementing FACS motions [14]) measuring the motion of the control points. As can be seen, the observed pattern of deformation

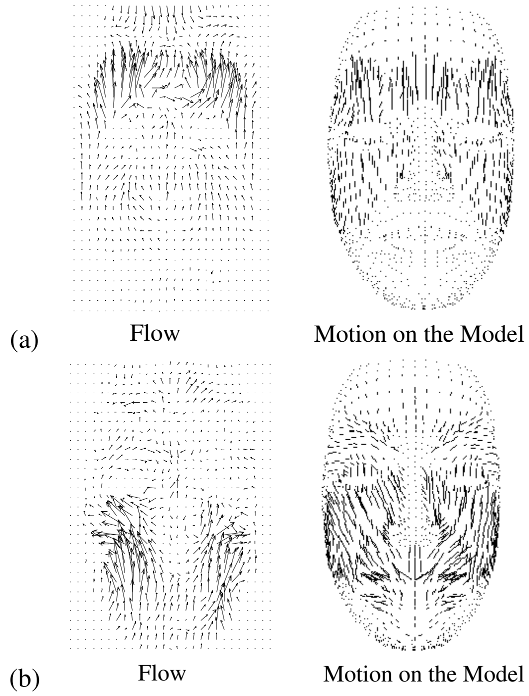


Figure 5: *Left figures show motion fields for the expression of (a) raise eye brow and (b) smile from optical flow computation and the right figures shows the motion field after it has been mapped to a face model.*

is very different than that assumed in the standard implementation of FACS (a linear ramp activating only a select number of localized control points). There is a wide distribution of motion through all the control points, and the temporal patterning of the deformation is far from linear. It appears, very roughly, to have a quick linear rise, then a slower linear rise and then a constant level (*i.e.,* may be approximated as piece-wise linear or as logarithmic).

A similar plot for happiness expression is shown in Figure 3(b). These observed distributed patterns of motion provide a very detailed representation of facial motion which we use in recognition of facial expressions.

3.2 Temporal Patterning

Another important observation about facial motion that is apparent in Figure 3 is that the facial motion is far from linear in time. This observation becomes much more important when facial motion is studied with reference to muscles, which is in fact the effector of facial motion and the underlying parameter for differentiating facial movements. Figure 4 show plots of facial muscle actuations for the observed smile and eyebrow raising expressions. For the purpose of illustration, in this figure, the 36 face muscles were combined into seven local groups on the basis of their proximity to each other and to the regions they effected. As can be seen, even the simplest expressions require multiple muscle actuations.

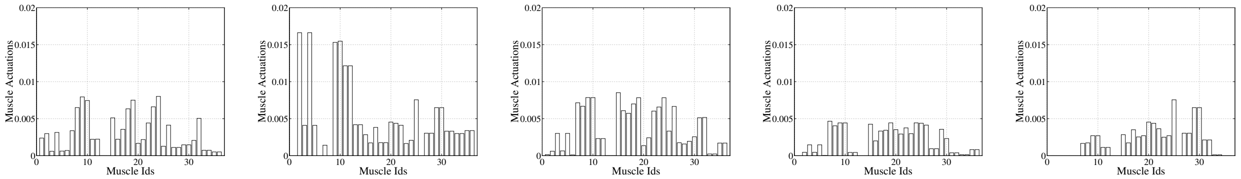
Of particular interest is the temporal patterning of the muscle actuations. We have fit exponential curves to the activation and release portions of the muscle actuation profile to suggest the type of rise and decay seen in EMG studies of muscles. From this data we suggest that the relaxation phase of muscle actuation is mostly due to passive stretching of the muscles by residual stress in the skin.

Figure 6: *Expressions from video sequences for various people in our database. These expressions are captured at 30 frames per second at NTSC resolution and cropped to 450x380.*

Note that Figure 4(b) for the smile expression also shows a second, delayed actuation of muscle group 7, about 3 frames after the peak of muscle group 1. Muscle group 7 includes all the muscles around the eyes and as can be seen in Figure 4(a), is the primary muscle group for the raising eye brow expression. This example illustrates that coarticulation effects can be observed by our system, and that they occur even in quite simple expressions. By using these observed temporal patterns of muscle activation, rather than simple linear ramps, or heuristic approaches of the representing temporal changes (as in [19]), our representation of facial motion is more suitable for recognizing facial motion.

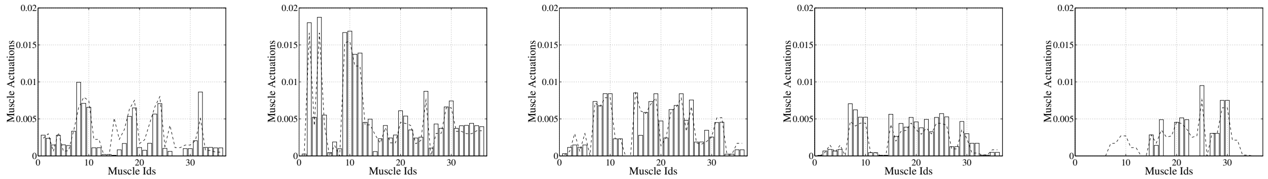
3.3 Corrected Motion Fields using Facial Model

So far we have concentrated on how we can extract the muscle actuations of an observed expression. The control-theoretic approach used to extract the muscle actuations over time can also be used to extract a "corrected" motion field for that expression. Our previously developed method for extracting facial action parameters employs optimal estimation, within an optimal control and feedback framework. It relies on 2-D motion observations from im-



Smile *Surprise* *Anger* *Disgust* *Raise Brow*

Figure 7: *Feature vectors of muscle templates for different expressions.*



GH/Smile [0.9112] *KR/Surprise [0.9962]* *CP/Anger [0.9777]* *SN/Disgust [0.9893]* *SS/R. Brow [0.8898]*

Figure 8: *Peak muscle actuations for several different expressions by different people. The dotted line shows the muscle template used for recognition. The normalized dot product of the feature vector with the muscle template is shown below each figure.*

ages to be mapped onto a physics-based dynamic model, and then the estimates of corrected 2-D motions (based on the optimal dynamic model) are used to correct the observations model. Figure 5 shows the flow for the expressions of raise eyebrow and smile, and also show the flow after it has been applied to the face model as deformation of the skin. By using this methodology to re-project the facial motion estimates back into the image we can remove the complexity of our physics-based model from our representation of facial motion, and instead use only the corrected 2-D observations to describe facial motion (motion energy).

Note that this corrected motion representation is better than could be obtained by measuring optical flow using simple gradient techniques, because it incorporates the constraints from our physical model of the face. Figure 2(g) and (h) shows examples of this representation of facial motion energy. It is this representation of facial motion that we will use for generating spatio-temporal templates for recognition of facial expression.

4 Observing People Making Expressions

One of the main advantages of the methods presented here is the ability to use real imagery to define the representation for each expression. As we discussed in the last section, we do not want to rely on existing models of representing facial expressions as they are not suited to our interests and needs. We would rather generate a representation of an expression by observing people making expressions and then use the extracted profiles for recognition, both in the muscle domain and the corrected motion energy domain.

For this purpose we have developed a video database of people making expressions. Currently these subjects are video-taped while making an expression on demand. These "on demand" expressions have the limitation that the subjects' emotions are not guaranteed to relate to his/her expression. However, at present we are more interested in characterizing facial motion and not human emotion. Categorization of human emotion on the basis of facial expression is a hot topic of research in the psychology literature and our belief is that our methods can be useful in this area. We are at present in collaborating with several psychologists on this problem.

At present we have a database of 20 people making expressions of *smile*, *surprise*, *anger*, *disgust*, *raise brow*, and *sad*. Some of our subjects had problems making the expression of sad, therefore we have decided to exclude that expression from our present study. We are working on expanding our database to cover many other expressions and also expressions with speech. The last frames of some of the expressions in our database are shown in Figure 6. All of these expressions are digitized as sequences at 30 frames per second and stored at the resolution of 450x380. All the results discussed in this paper are based on expressions performed by 8 subjects with a total of 52 expressions. This database is substantially larger than that used by Mase [8] in his pioneering work on recognizing facial expressions. Although our database is smaller than that of Yacoob and Davis [19], we believe that it is sufficiently large to demonstrate that we have achieved improved accuracy at facial expression recognition.

5 Recognition of Facial Expressions

We will now discuss how we use our representations of facial motion for recognition of facial expressions. We will first discuss use of our physically-based motion representation for recognition, followed by our spatio-temporal motion energy model.

5.1 Model-based Recognition

Recognition requires a unique "feature vector" to define each expression and a similarity metric to measure the differences between expressions. Since both temporal and spatial characteristics are both important we require a feature vector that can account for both of these characteristics. We must, however, account for the speed at which the expressions are performed. Since facial expressions occur in three distinct phases: *application*, *release* and *relaxation* (see Figure 4), by dividing the data into these phases and by warping it for all expressions into a fixed time period of ten discrete samples, we can normalize the temporal time

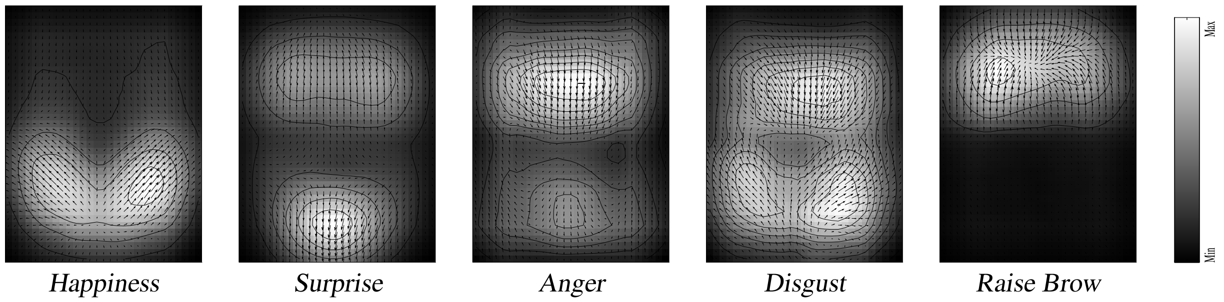


Figure 9: *Spatio-temporal motion-energy templates for the five expressions, averaged using the data from two people.*

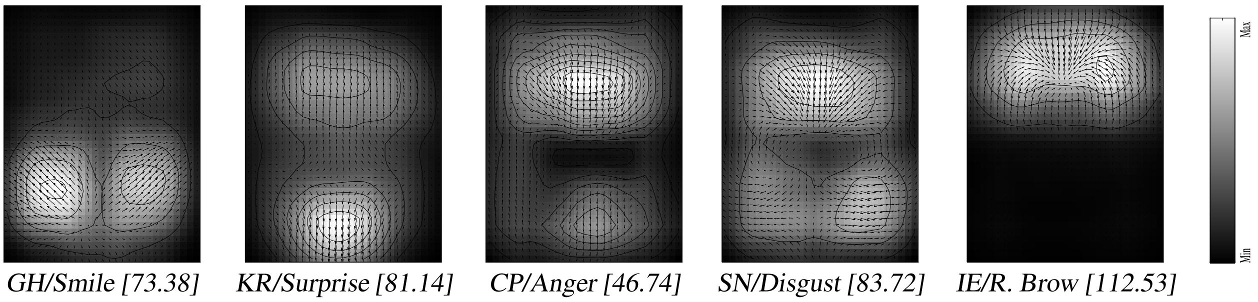


Figure 10: *Motion templates for four expressions from 5 people. Their similarity scores are also shown.*

course of the expression. This normalization allows us to use the muscle actuation profiles to define a unique feature vector for a each facial motion.

We define the peak actuation of each muscle between the application and release phases as the feature vector for each expression. Choosing at random two subjects per expression from our database of facial expressions, we define an (averaged) muscle activation feature vector for each of the expressions *smile*, *surprise*, *anger*, *disgust*, and *raise eyebrow*. These peak muscle actuation features, which we call *muscle templates* are shown in Figure 7. As can be seen, the muscle templates for each expression are unique, indicating that they are good features for recognition. These feature vectors are then used for recognition of facial expression by comparison using a normalized dot product similarity metric.

By computing the muscle activation feature vectors for our independent test data, 8 subjects making about 5 expressions multiple times, we can assess the recognition accuracy of this physical-model based expression recognition method.

Figure 8 shows the peak muscle actuations for 5 people making different expressions. The muscle template used for recognition is also shown for comparison. The dot products of the feature vectors with the corresponding expression's muscle template are shown below each figure.

The largest differences in peak muscle actuations are due to facial asymmetry and intensity of the actuation. The intensity difference is especially apparent in the case of the surprise expression where some people open their mouth less then others. Our analysis does not enforce any symmetry constraints and none of our data, including the muscle templates shown in Figure 7, portray exactly symmetric expressions.

Table 1 shows the results of dot products between peak muscle actuations of the five randomly chosen expressions with each expression's muscle template. It can be seen that for the five instances shown in this table, each expression is correctly identified.

We have used this recognition method with all of our data, consisting of 8 different people making the 52 expressions (including: *smile*, *surprise*, *raise eye brow*, *anger*, and *disgust*) Since some people did not make all expressions, we have 10 samples each for surprise, anger, disgust, and raise eyebrow expressions, and 12 for the smile expression. Table 2 shows the results of the overall recognition results in the form of a *classification matrix*. In our tests there was only one recognition failure, for the expression of anger. Our overall accuracy was 98.0%. Table 5 shows the average and the standard deviation of all the expressions compared to each expression's muscle template with a corresponding plot of mean and standard deviations of similarity scores. This table shows the repeatability and reliability of our recognition method.

5.2 Spatio-temporal Templates for Recognition

The previous section used estimated peak muscle actuations as a feature to detect similarity/dissimilarity of facial expressions. Now we will consider a second, much simpler representation: the (corrected) motion energy of facial motion. Figure 9 shows the pattern of motion generated by averaging two randomly chosen subjects per expression from our facial expression image sequence database. Notice that each of these motion templates is unique and therefore can serve as an sufficient feature for categorization of facial expression. Note that these motion-energy templates are sufficiently smooth that they can be subsampled at one-tenth the raw image resolution, greatly reducing

	Expressions				
Templates	SM	SP	AN	DI	RB
SM	**0.91**	0.36	0.91	0.75	0.17
SP	0.32	**0.99**	0.34	0.28	0.20
AN	0.62	0.28	**0.99**	0.47	0.81
DI	0.32	0.22	0.66	**0.88**	0.43
RB	0.75	0.27	0.84	0.24	**0.98**

Table 1: *Some examples of recognition of facial expressions, using peak muscles actuations. A score of 1.0 indicates complete similarity.*

	Expressions				
Templates	SM	SP	AN	DI	RB
SM	**12**	0	1	0	0
SP	0	**10**	0	0	0
AN	0	0	**9**	0	0
DI	0	0	0	**10**	0
RB	0	0	0	0	**10**
Success (%)	**100**	**100**	**90**	**100**	**100**

Table 2: *Results of Facial Expression Recognition using peak-muscle actuations. The overall recognition rate is 98.0%.*

SM: Smile, SP: Surprise, AN: Anger, DI: Disgust, RB: Raise Brows.

	Expressions				
Templates	SM	SP	AN	DI	RB
SM	**73.4**	255.0	230.3	209.4	294.6
SP	233.1	**81.1**	143.7	141.4	243.4
AN	213.8	187.0	**46.7**	95.2	152.3
DI	154.0	178.3	126.4	**83.7**	227.3
RB	288.9	322.2	147.7	240.1	**46.8**

Table 3: *Example scores for recognition of facial expressions using spatio-temporal templates. Low scores show more similarity to the template.*

	Expressions				
Templates	SM	SP	AN	DI	RB
SM	**12**	0	0	0	0
SP	0	**10**	0	0	0
AN	0	0	**9**	0	0
DI	0	0	1	**10**	0
RB	0	0	0	0	**8**
Success (%)	**100**	**100**	**90**	**100**	**100**

Table 4: *Results of Facial Expression Recognition using spatio-temporal motion energy templates. The overall recognition rate is 98.0%.*

the computational load. We use the Euclidean norm[1] of the difference between the motion energy template and the observed image motion energy as a metric for measuring similarity/dissimilarity. Note that metric works oppositely from the dot-product metric: the lower the value of this metric, more similar the images are.

Using the average of two people making an expression, we generate motion-energy template images for each of the five expressions. Using these templates (shown in Figure 9), we conducted recognition tests for our independent test database of 52 image sequences. Figure 10 shows 5 examples of the motion energy images generated by different people. The similarity scores of these to the corresponding expression templates are shown below each figure.

Table 3 shows the results of this recognition test for five different expressions by different subjects. The scores show that all the expressions were correctly identified. The classification results of this methods over the whole database, displayed as a *confusion/classification matrix*, are shown in Table 4. This table shows that again we have just one incorrect classification of the anger expression. The overall recognition rate with this method is also 98.0%. Conducting this analysis for our independent test database of 52 expressions, we can generate a table which shows the mean and variance of the similarity scores across the whole database. These scores are shown in Table 6. Again it can be seen that the recognitions are quite reliable.

6 Discussion and Conclusions

In this paper we have presented two methods for recognition of facial expressions. Unlike previous efforts at facial expression recognition that have been based on the Facial Action Coding System (FACS), we develop a new, more accurate representations of facial motion and use these new representations for the recognition/identification task. We analyze a video of facial expressions and then probabilistically characterizing the facial muscle activation associated with each expression. This is achieved using a detailed physics-based dynamic model of the skin and muscles coupled with optimal estimates of optical flow in a feedback controlled framework. This analysis produces a muscle-based representation of facial motion, which is then used to recognize facial expressions in two different ways.

The first method uses the physics-based model directly, by recognizing expressions through comparison of estimated muscle activations. This method yields a recognition rate of 98% over our database of 52 sequences.

The second method uses the physics-based model to generate spatio-temporal motion-energy templates of the whole face for each different expression. These simple, biologically-plausible motion energy "templates" are then used for recognition. This method also yields a recognition rate of 98%. The combined accuracy of both these methods on our independent test database is 100%. This level of accuracy at expression recognition is substantially better than has been previously achieved.

We are at the moment working on increasing the size of our database to also include other expressions and speech motions.

[1] We have used other distance/similarity metric with quite similar result, we only report on Euclidean norms here.

Acknowledgments

We would like to thank Baback Moghaddam, Trevor Darrell, Eero Simoncelli, John Y. Wang and Judy Bornstein for their help.

References

[1] J. N. Bassili. Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psyschology*, 4:373–379, 1978.

[2] M. J. Black and Y. Yacoob. Tracking and recognizing facial expressions in image sequences, using local parameterized

	Expressions				
Templates	SM	SP	AN	DI	RB
SM	**0.97 ± 0.03**	0.63 ± 0.04	0.95 ± 0.01	0.86 ± 0.04	0.59 ± 0.16
SP	0.58 ± 0.03	**0.99 ± 0.01**	0.59 ± 0.04	0.57 ± 0.05	0.56 ± 0.09
AN	0.90 ± 0.05	0.55 ± 0.05	**0.97 ± 0.02**	0.91 ± 0.01	0.65 ± 0.14
DI	0.82 ± 0.06	0.57 ± 0.05	0.92 ± 0.03	**0.95 ± 0.03**	0.78 ± 0.10
RB	0.58 ± 0.05	0.57 ± 0.07	0.70 ± 0.05	0.78 ± 0.06	**0.96 ± 0.04**

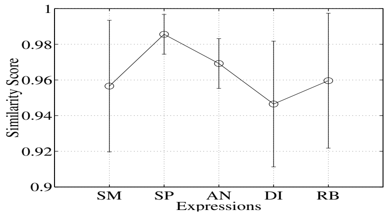


Table 5: *Mean and Standard Deviations of Similarity scores of all expressions in the database. Similarity metric is normalized dot products.*

	Expressions				
Templates	SM	SP	AN	DI	RB
SM	**94.1±34.**	266. ± 52.	234. ± 62.	153. ± 59.	306. ± 15.
SP	230. ± 8.	**123. ±70.**	160. ± 38.	173. ± 14.	233. ± 14.
AN	225. ± 16.	199. ± 76.	**98.±46.**	160. ± 29.	147. ± 15.
DI	149. ± 22.	198. ± 54.	140. ± 43.	**99.±23.**	224. ± 16.
RB	339. ± 32.	321. ± 96.	208. ± 33.	293. ± 26.	**106.8±27.**

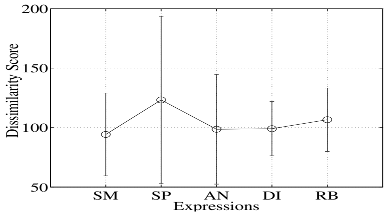


Table 6: *Mean ± Standard Deviation of scores for recognition of facial expressions using the spatio-temporal templates over the whole database. Low Scores show more similarity to the template.*

SM: Smile, SP: Surprise, AN: Anger, DI: Disgust, RB: Raise Brows.

models of image motion. Technical Report CAR-TR-756, Center of Automation Research, University of Maryland, College Park, 1995.

[3] V. Bruce. *Recognising Faces*. Lawrence Erlbaum Associates, 1988.

[4] C. Darwin. *The expression of the emotions in man and animals*. University of Chicago Press, 1965. (Original work published in 1872).

[5] P. Ekman and W. V. Friesen. *Facial Action Coding System*. Consulting Psychologists Press Inc., 577 College Avenue, Palo Alto, California 94306, 1978.

[6] I. Essa, T. Darrell, and A. Pentland. Tracking facial motion. In *Proceedings of the Workshop on Motion of Nonrigid and Articulated Objects*, pages 36–42. IEEE Computer Society, 1994.

[7] I. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 76–83. IEEE Computer Society, 1994.

[8] K. Mase. Recognition of facial expressions for optical flow. *IEICE Transactions, Special Issue on Computer Vision and its Applications*, E 74(10), 1991.

[9] K. Mase and A. Pentland. Lipreading by optical flow. *Systems and Computers*, 22(6):67–76, 1991.

[10] M. Minsky. *The Society of Mind*. A Touchstone Book, Simon and Schuster Inc., 1985.

[11] B. Moghaddam and A. Pentland. Face recognition using view-based and modular eigenspaces. In *Automatic Systems for the Identification and Inspection of Humans*, volume 2277. SPIE, 1994.

[12] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition Conference*, pages 84–91. IEEE Computer Society, 1994.

[13] M. Rosenblum, Y. Yacoob, and L. Davis. Human emotion recognition from motion using a radial basis function network architecture. In *The Workshop on Motion of Nonrigid and Articulated Objects*, pages 43–49. IEEE Computer Society, 1994.

[14] M. Rydfalk. *CANDIDE: A Parameterized Face*. PhD thesis, Linköping University, Department of Electrical Engineering, Oct 1987.

[15] E. P. Simoncelli. *Distributed Representation and Analysis of Visual Motion*. PhD thesis, Massachusetts Institute of Technology, 1993.

[16] D. Terzopoulus and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):569–579, June 1993.

[17] J. Y. A. Wang and E. Adelson. Layered representation for motion analysis. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 1993.

[18] K. Waters and D. Terzopoulos. Modeling and animating faces using scanned data. *The Journal of Visualization and Computer Animation*, 2:123–128, 1991.

[19] Y. Yacoob and L. Davis. Computing spatio-temporal representations of human faces. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 70–75. IEEE Computer Society, 1994.