Content Based Retrieval of Emotions in Face Images

J Nileema, Priti Chandra, C. Bhagvati, Arun K Pujari, B L Deekshatulu
Department of Computer & Information Sciences
University of Hyderabad, Hyderabad, 500 046, INDIA
akpcs@uohyd.ernet.in

Abstract

Content-based image retrieval is an important area of research involving techniques of image processing, information retrieval, and data mining. Mathematical morphology is a powerful set of tools in image processing and normally used for filtering, shape description etc.. Sequence data mining and clustering are the popular data mining techniques that are successfully applied to varieties of applications. In the present work, we demonstrate that the pattern spectrum, a well known morphological technique, can be used to capture emotions in a face image. We also show, by devising a novel data mining technique for sequence database, that the set of pattern spectra corresponding to face image database can be searched to retrieve faces having emotion similar to that of a query image. We make use of link-based clustering combined with density based clustering to handle the data mining tasks. We demonstrate the efficiency of our system experimentally.

1. Introduction

Mathematical morphology uses concepts from set theory, geometry and topology to analyze geometrical structures in an image. This characteristic of mathematical morphology is the motivation behind using it for feature extraction in content-based image retrieval [18], [19]. The aim of the present study is to demonstrate the applicability of mathematical morphology in capturing the facial expression of human face. We show that by blending the morphological method with data mining technique, it is possible to use a simple but robust way of emotion-based search of human face database. We demonstrate here that for a given face image as a query, it is possible to retrieve a set of all images containing similar emotional expression as that of the query image, from a database of face images. We make use of pattern spectra of mathematical morphology theory to capture these features. Morphological operations such as opening distribution and pattern spectrum act as shape descriptors in analyzing an image. The pattern spectrum can be viewed as a sequence of numbers and hence, we employ the well-known techniques of similarity search in sequence database. However, since the database consists of very large number of short sequences in the present context, we make use of a novel combination of binning and clustering for this purpose. The main aim is to retrieve almost all relevant images as well as to minimize the number of irrelevant images. To achieve this, the system considers all the close clusters (Inter-Cluster search) and then search within that cluster (Intra-Cluster search). Our experimental results corroborate this hypothesis.

The main contribution of the present work is two-fold.

- It demonstrates that pattern spectra can capture emotion features of a face image.
- It also devises a novel method of clustering and searching the sequence database arising out of the set of pattern spectra.

The existing work on facial expression analysis is discussed in section 2. In section 3, the basic notions of mathematical morphology and the pattern spectra are introduced. In section 4, we justify the relevance and importance of pattern spectra in the context of face analysis. Once the pattern spectra of the image database are generated, it is necessary to carry out the search in this database. We discuss some of the major clustering algorithms available in data mining literature in section 5. Section 6 deals with the proposed system of content-based image retrieval of face images based on emotion. We report experimental results in section 7.

2. Facial Expression Analysis

Facial expression is among the most powerful, natural, and immediate means for people to communicate their emotions and intentions [4], [6]. In recent years, significant progress has been made in computer-vision based approaches to discriminating emotions [10], [3], [2], [13], [16]). Ekman and Rosenberg [6], [7] propose that there are six basic emotions, namely *joy, surprise, anger, sadness, fear* and *disgust*. The Facial Action Coding System(FACS) [6] is a system designed for human observers to detect subtle changes in facial appearance. It is a system that linguistically describes all possible visually detectable facial changes in terms of 44 Action Units(AU). So far, several studies on vision-based facial gesture analysis have suggested that FACS AUs could be detected from digitized face images. Black and Yacoob [1] describe a system that recognizes facial expressions in the presence of significant head motions. In an earlier version, they use optical flow at high gradient points on the face to recognize facial expressions. Essa and Pentland [8] propose a method for recognition of facial expressions based on differential patterns of optical flow. Kearney and McKenzie [14] propose a self-adaptive expert system that converts facial data into a set of face actions and then into a set of emotion labels. The system recognizes 36 different face actions but uses hand-measured manually supplied face image data that is difficult to track automatically. Cohn [4] proposes an optical-flow-based method for discriminating between Action Units in the eyebrow, eye and mouth regions. The method can identify eight individual AUs and seven AUs combinations. Zhang [20] compares the use of geometrical features with a multi-scale, multi-orientation Gabor wavelet based representation to identify expressions.

3. Pattern Spectrum

Binary Morphology[17] deals with the binary input images and binary structuring elements. The gray-scale images are converted into binary by thresholding. The set of all white pixels in a white and black image constitutes a complete description of the image.

All the operations of mathematical morphology are usually defined between two sets (say *A* and *B*), where *A* is the set we want to examine and *B* is called a *structuring element*. By performing morphological operations on set *A* with different structuring elements $\{B_i\}$, we get a resulting set series $\{Y_i\}$. Each specific structuring element B_i filters out a kind of information from the original image. By analyzing the resulting set sequence $\{Y_i\}$, we can obtain sufficient heuristic information. The fundamental operations associated with an object are the standard set operations of union, intersection, complement $\{\cup, \cap, ^c\}$ and translation[21]. Given a vector x and a set A, the *translation*, $A+x$ is defined as:

$$A+x=\{\alpha+x \mid \alpha\in A\}$$

The *dilation* of *A* and *B* is defined as the union of all translates of the set *A* by each of the elements of the set *B*.

$$D(A,B)=\bigcup_{\beta\in B}(A+\beta)$$

The *erosion* of *A* and *B* is defined as the intersection of all translates of the set *A* by each of the elements of the set (-*B*) where $-B=\{-\beta \mid \beta\in B\}$,

$$E(A,B)=\bigcap_{\beta\in B}(A-\beta)$$

Opening of an image *A* by a structuring element *B* is defined as:

$$A\circ B=D(E(A,B),B)$$

The opening of an image is just a sequential application of erosion followed by dilation by the same structuring element. *Closing* of an image *A* is defined as:

$$A\bullet B=E(D(A,-B),-B)$$

Let *A* be finite extent discrete binary image and the discrete binary pattern *B* be any finite subset of Z^2. It can be shown that $\forall k\geq 1$,

$$\ldots\subseteq A\circ(k+1)B\subseteq A\circ kB\subseteq A\circ(k-1)B\ldots\subseteq A\ldots$$

Pattern Spectrum of an image A with respect to a structuring element B is defined as $PS_B^A(k)=[Area(A\circ kB)-Area(A\circ(k+1)B)]$ where $Area(A\circ kB)$ indicates the area of the image when opened with B of size k. We shall drop the prefixes A and B to denote the pattern spectrum as PS(k) whenever the context is unmistakable.

Pattern Spectrum is a shape size descriptor of an object content of a digital image. The area obtained by opening *A* by a structuring element *B* of size *r* is a measure of the pattern content of *A*, relative to the pattern *rB*. By varying both size *r* and the shape of *B*, we obtain a shape size spectrum of *A* which is the full pattern spectrum of *A* relative to all the patterns that can fit inside *A*. By keeping *B* fixed, we get a size histogram of *A* relative to *B*. It can be clearly seen that pattern spectrum acts as a shape descriptor when a structuring element of suitable shape and size is applied. In a sense, the pattern spectrum provides useful information about the shape similarity between A and B. At small sizes of B information about high-frequency noise and boundary roughness relative to B is given. Any sharp or significant spikes reveal the existence of long capes or bulky protruding parts. The similarity between A and B is given by the magnitude of spikes at larger scales. The higher the spikes, the greater is the similarity between A and B.

4. Facial Images with Emotions

The purpose of the present system is to use this concept of pattern spectrum to identify the different emotions expressed by different individuals. The retrieval is based on the emotion irrespective of the individuals. The emotions that an individual expresses provide a source where the shape and size of the face change. Consider an individual with normal and laughing expression as shown Figure 1.

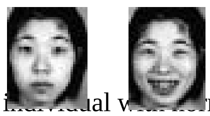

Figure 1. An individual with normal expression and laughing expression

It can be clearly seen that there is a considerable change in the shape of the face and this change can be captured using the pattern spectrum. The pattern spectra of the above two images are in Figure 2.

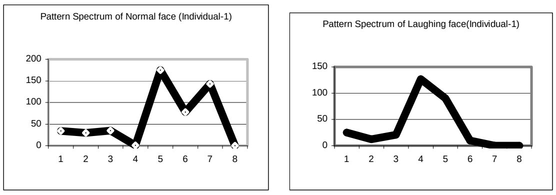

Figure 2. Pattern Spectra of an individual with normal and laughing expressions

Let us consider two different individuals with the same laughing expression (Figure 3). The pattern spectra of the two individuals is shown in Figure 4. It is observed that there is some similarity between the two graphs. This similarity is used to show that the emotions expressed by the two individuals are same. It can be seen that a smiling face usually implies an open mouth. As an open mouth is wider, a smile is revealed by the presence of a significant peak in the pattern spectrum at medium to large scales.

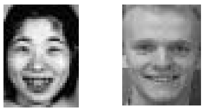
Figure 3. Two different individuals with laughing expression

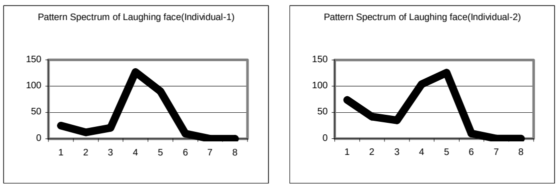

Figure 4. Pattern Spectra of two different individuals with laughing expression

5. Clustering

Clustering is a useful technique for discovery of data distribution and patterns in the underlying data. The goal of clustering is to discover dense and sparse regions in a data set. ROCK[11] and DBSCAN[9] clustering techniques are employed in the present system to carry out clustering of images.

ROCK (RObust Hierarchical Clustering with links)[11], a robust hierarchical clustering algorithm, is an adaptation of an agglomerative hierarchical clustering algorithm. It heuristically optimizes a criterion function defined in terms of the number of "links" between objects. Informally, the number of links between two objects is the number of common neighbors they have in the dataset. Starting with each object in its own cluster, they repeatedly merge the two closest clusters till the required number of clusters remain.

The ROCK algorithm starts with each cluster being a single data point and keeps merging the pair of clusters with the best positive goodness measure. In [11] the algorithm is stated to have the complexity $O(n^2 logn)$. But in [5], Dutta and Bhattacharya observe that the straightforward implementation of the algorithm has actually the complexity $O(n^3)$. They show in [5] that by carefully handling the link-data and using a method of deferred deletion from the local heaps, the complexity can be improved to $O(n^2 logn)$. We use the improved implementation of the algorithm. The following observations are made while experimenting with this algorithm.

If an object has common neighbors with two or more of existing clusters then they are all merged. Any object belongs to one and only one cluster. The tree like construction of clusters leads to the formation of some clusters, which contain a large number of images, as large as 80% of the images of the database and most clusters are left with only a few images. One can see that some clusters have a very large number of objects whereas other clusters contain very few objects. Thus, it is felt that some sort of post processing is necessary when we use ROCK. We propose to combine another clustering algorithm together with ROCK.

DBSCAN[9] uses density-based notion of the clusters to discover the clusters of arbitrary shapes. The key idea is that, for each object of a cluster, the neighborhood of a given radius has to contain at least a minimum number of data objects. We can use this concept on individual clusters obtained from ROCK. Considering one cluster at a time, the objects in the cluster are marked as either core or non-core objects. The non-core objects from the cluster are removed and tested whether some or all of them can be categorized into the succeeding clusters. The process is repeated until all the clusters are checked for accommodating the non-core objects of the clusters excluding it. Finally the objects that cannot be categorized into any cluster are retained as noise objects. These are the objects that do not match with any of the clusters formed. The noise objects are then discarded from the clusters.

6. Emotion Based Face Image Retrieval

Based on the foregoing discussion, we propose a novel emotion based image retrieval system. The system consists of mainly three phases:

- Extraction of Feature Vector Table for images in the database.
- Clustering of images in the database.
- Retrieve images closest to the query image.

The first phase i.e. feature vector extraction is done using morphological operations discussed in section 3. The concept of pattern spectrum is used. Sequences are handled for similarity search of images with various emotions, based on comparison of pattern spectrum. The second phase i.e. clustering the images is done using data mining clustering algorithms. In the third phase, images that are similar in emotion to the query image are retrieved.

6.1 Feature Vector Extraction

A gray scale image set is converted into a binary image set using thresholding. The image set is opened with a particular structuring element starting from a size until the area of the resulting image after opening is zero. Various structuring elements are used such as square, horizontal line, vertical line and principal and non-principal diagonal. Opening distribution of an image is used to calculate pattern spectrum and feature vector. The sequence of numbers in the pattern spectrum is processed to produce a *bitstring*. The sequence of numbers is scanned sequentially for each consecutive pair of numbers. If the value is increasing, then we assign 1; otherwise, 0 is assigned.
Bitstring bs_A for any sequence A is defined as:

$$bs_A(k) = \begin{cases} 1, & \text{if } PS(k+1) > PS(k) \\ 0, & Otherwise \end{cases}$$

The pattern spectrum of each image is essentially a sequence of numbers and a similarity-based search technique of sequence data would apparently help in emotion-based retrieval. Recently, in [15] an efficient technique of similarity based search is proposed in which the sequences are first converted to bitstrings and then indexed by binning the equivalent decimal number. The similarity is determined by the closeness of the BIN number. We observe that this method does not yield good result in the present context. This is mainly because of following reasons. First, the sequences corresponding to pattern spectra are very short and hence there is a major effect in any minor variation in the sequences corresponding to similar expression. Second, since there are variations within a class, it is worthwhile to determine one representative sequence that can represent the class. It is therefore worthwhile to determine a sort of representative sequence, which captures all the common features of all the pattern spectra with respect to an emotion. To accomplish this objective, the bitstrings are first clustered and then all the bitstrings within a cluster are aggregated by OR operation. The resulting bitstring is the representative bitstring of the cluster. We follow the binning technique to determine the BIN number of the decimal equivalent of the representative bitstrings. The similarity-based retrieval is used to determine the closeness of the BIN number of the query image to the BIN number of the representatives and all the images of a cluster are returned.

6.2 Clustering of Images in the Database

The basic principle of clustering images hinges on a concept of distance metric or similarity metric. Instead of comparing the BIN number of the query image with the BIN numbers of all the images in the database, the images in the database are clustered, the representative BIN number of each cluster is found and the query image is compared with only the cluster representatives. For clustering the images, the bitstrings that represent the images are considered as feature vector.

ROCK clustering technique is applied to arrive at an initial set of image clusters in the database. Initially all the objects are considered as separate clusters. A cluster is randomly selected. To this cluster all the images that are in the neighborhood are added. This procedure is repeated for each image in that neighborhood in turn. The above procedure is continued iteratively till there are no images to fall into this cluster. Again a randomly selected image from the remaining images is selected to start a new cluster and the above procedure is repeated. This process is repeated till all the images fall in one or more cluster. This is because of the linear growth of the clusters. In order to overcome the drawback of ROCK algorithm, DBSCAN is applied on the already formed clusters. Instead of considering the entire database for clustering, the adapted algorithm considers the clusters already formed to re-cluster them so that more similar images are in one cluster. Representative BIN numbers of all the clusters are found. *Representative BIN* number of a cluster C is defined as bitwise-OR of BIN numbers of all the images in the cluster. It is calculated by the expression:

$$bs_C(k) = \vee_A bs_A(k), \quad \forall A \in C$$

BIN(C) is the bin number of the decimal equivalent of bs_C

6.3 Query Matching

When a query with a particular emotion is given, the system has to retrieve those images in the database that are more similar in emotion to the query image i.e. the cluster whose features are closer to the query image feature vector, is retrieved.

Given the query image q, the feature vector is found using pattern spectrum. The bitstring bs_q that represents the image is converted into its corresponding BIN number *BIN*(*q*). The concept of handling longer and shorter queries is also considered at this point. *BIN*(q) is compared with *BIN(C)* of all the clusters. The whole cluster is retrieved finally.

7. Results

The image collection is downloaded from the various database sources. More than 200 images consisting of facial expressions under similar conditions are collected from various sources: JAFFE, CMU, AT&T. The images are scaled to a size of 75 × 75 and are converted into bitmap and gif images in gray scale format. The database is divided into 3 image collections: imgset1, imgset2, imgset3. The imgset1, training set, is formed keeping in view that the background and foreground intensities of all the images are similar and the two colors are distinguishable. There are 21 individuals with various expressions. All the three image sets are used as test images in our experiments. For testing the robustness in the clustering phase we consider three expressions: normal, laugh and surprise; other expressions are used for detection of outliers. Images with all the six basic expressions are present in image set 2 and image set 3. Sample images from the other image sets, which can be used for querying are shown in Annexure II. We describe below the working of the proposed procedure by using a test case.

Test Case

This section discusses the process, which the system follows in retrieving an image. First the feature vectors for every image in the database are calculated.

Figure 5. An image in the database

The image in Figure 5 is a gray-scale image of size 60×69. This is thresholded to obtain a binary image and scaled to the required size. The structuring element is a square with initial size 5×5. We have experimented with different types of structuring elements but we observe that the best result is obtained in the present context when the structuring is a square. The opening distribution and the pattern spectrum of this image are extracted as shown in Figure 6.

Size	**Area**	**Pattern Spectrum**
5	1549	70
6	1479	158
7	1321	115
8	1206	40
9	1166	162
10	1004	148
11	856	274
12	582	192
13	390	28
14	362	362
15	0	0

Figure 6. Pattern spectrum of the image in Figure 5

The pattern spectrum of all the images in the collection are extracted and stored in a representative file for a particular structuring element. The feature vector of the image is derived from the pattern spectrum using the binning concept described in section 6. The length of the bit string of various images varies depending on the nature of the image. We normalize the length by 0-padding method. Each new image considered is compared with the maximum length of the bit strings already existing and then it is updated. For image set 1, the number of columns in the feature vector table is 17. Also the bin numbers of all the images are calculated and stored. The bin number of the image in Figure 5 is 297.

Using the feature vector table, the images in the collection are clustered using ROCK and DBSCAN algorithms. For ROCK we consider the image set 1 with threshold as 0.8; Sample Record set as 1 to 51; Number of clusters as 3 (corresponding to the three basic expressions considered). The parameters for DBSCAN are Epsilon= 0.02 and Minpoints = 9.

The pattern spectrum of the query image (Figure 7) is extracted using the square structuring element of initial size 5 × 5. It is observed that the length of the bit string is 18. The result is shown in Figure 8. We observe that out of 9 images retrieved by our algorithm, eight of these are smiling faces and one is though different one it has a face with wide mouth.

Figure 7.

When a query with the emotion other than the three basic emotions analyzed is given, it is observed that the system retrieves images with emotions most similar to that of the query image. If the other image collections are used for clustering and querying, then the input datasets given to the clustering algorithms and the structuring element used have to be chosen such that retrieval is efficient, for which much experimentation is necessary. The present system is able to retrieve similar images based on the emotion expressed by the query image. The retrieval is completely based on the emotions irrespective of the different individuals present. Even if the emotion expressed by the query image is not present in the database of the system, it retrieves those images with most similar emotion to the query.

It is to be noted that the number of clusters is equal to number of distinct gestures that we wish to identify. We consider in our experiment the five basic expressions. Thus it becomes very efficient to find the similarity matching of the query BIN with one of five representative BINs. Moreover, due to clustering and aggregating the cluster to determine the representative most of the errors are smoothened out.

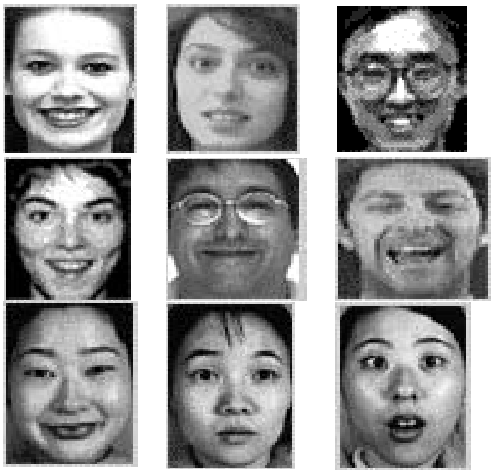

Figure 8. Matching images to the query image in Figure 7

The efficiency of the proposed system can be measured on two parameters. First is the quality of clustering and the second is correct retrieval. We observe that classification of the expressions are not very accurate as we see in the above example too that a face with an expression different from laughing expression is in the cluster corresponding to laughing. It is expected as it is often difficult to identify the correct expression from the face image for even a human expert. However, our classification scheme classifies more or less all the images of an expression to a cluster. The accuracy is nearly 85%. The similarity matching is found to be very efficient. It is nearly 100% correct retrieval by our method. We propose to carry out a rigorous benchmarking and add more expressions to the database.

8. Conclusions

In this paper a novel application of mathematical morphology is demonstrated for content based image retrieval of face images. The major contributions of the present paper are: a new approach to extract the features related to emotional expression in a face image, a modified clustering algorithm that takes into consideration aggregation of feature values of all the face images containing an emotion, and the similarity search in a set of sequence. Our experimental experience corroborate our hypothesis that pattern spectra can indeed capture the distinctive features related to expressions in face images.

Acknowledgement**-**
This work is supported by the research grant of AICTE.

References

[1] M. Black, and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *Int. Journal of Comput. Vision*, 1997.

[2] J. F. Cohn, Gary S. Katz, *Bimodal Expression of Emotion by Face and Voice.*

[3] J. F. Cohn, Karen Schmidt, Ralph Gross, and Paul Ekman, " Individual Differences in Facial Expression: Stability over Time, Relation to Self-Reported Emotion, and Ability to Inform Person Identification**,"** in *IEEE Int. Conf. on Multimodal Interfaces*, 2002.

[4] J.F. Cohn, A.J. Zlochower, J.J. Lien, and T. Kanade, "Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression," in *Third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, April 14-16 1998, pp. 396-401.

[5] M. Dutta, D.K. Bhattacharya, "On the complexity of the ROCK algorithm for clustering of categorical data," In National Seminar on Soft Computing and Data Mining, Tezpur, India, 2001

[6] P. Ekman, E. Rosenberg: *Facial Action Coding System*: Consulting psychologists press, Palo Alto, 1978.

[7] P. Ekman, E. Rosenberg: *What the face reveals*: NY: Oxford University.

[8] I. Essa, and A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions," *IEEE Trans. PAMI,* vol.19 no.7, pp. 757-763, 1997.

[9] M. Ester, H.P. Kreigel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. On Knowledge discovery and Data Mining (KDD-96),* Portland, OR, 1996, pp. 226-231.

[10] B. Fasela, Juergen Luettin B, "Automatic facial expression analysis :A survey," Pattern Recognition 36, pp. 259 –275, 2003.

[11] S. Guha, R. Rastogi, K. Shim, "ROCK: A robust clustering algorithm for categorical attributes," in *Proc. IEEE Int. Conf. On Data Engineering,* Sydney, March 1999.

[12] R.M. Haralick, S.R. Sternberg, X. Zhuang,"Image Analysis using Mathematical Morphology,"in *IEEE Trans.* PAMI, vol 9, July 1987.

[13] A. Kapoor, "Automatic Facial Action Analysis," Submitted to the MSc. Program in Media Arts and Sciences, School of Architecture and Planning MIT, June 2002.

[14] G.D. Kearney, S. McKenzie, *"Machine Interpretation of Emotion," Cognitive Science,* vol. 17 no.4, pp.589-622, 1993.

[15] E. J. Keogh, M. J. Pazzani,"An indexing scheme for fast similarity search in large time series database," in *11th Int Conf. on Scientific and Statistical Database Management*, Ohio, pp. 28-30, July 1999.

[16] M. Pantic, L.J.M. Rothkrantz, *An Expert System for Multiple Emotional Classification of Facial Expressions*

[17] J. Serra, *Image Analysis and Mathematical Morphology,* London Academic Press, 1982.

[18] P. Soille, M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," in *IEEE Tr. on Geoscience and Remote Sensing*, 2002.

[19] P. Soille, *Morphological Image Analysis, Principles and Applications*, Springer, Berlin, 1999.

[20] Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu,"Comparison between Geometry-based and Gabor Wavelets-based Facial Expression Recognition using Multi-layer Perceptron," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, 1998.

[21] http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip-Morpholo.html