

# **INCITS M1/03-0507**

## **Contribution From Identix Incorporated**

### **Consolidation of Unposted Comments On Face Recognition Format for Data Interchange Version N\***

This document contains a consolidation of comments, previously unposted to M1, from various sources, which Identix Incorporated believes to be appropriate to the further development of the Face Recognition Format for Data Interchange standard.

**Source: Identix Incorporated**

**Date: October 21, 2003**

---

\* Changes are incorporated into Version 2.0, M1/03-0494.

## Document M1/03-0494 Changes from Document M1/03-0349

This document describes the bulk of the changes and the editor's instructions that lead to the revised document.

### **General**

Re-organization is described in the upcoming sections in some detail. The major changes to the document structure area. In particular,

- a. Move compression of full frontal images and canonical images to an informative annex, as there is still much research to perform before definitive statements can be made on compression.
- b. Clearly define Normative, vs. Best Practices vs. Informative. Normative, Best Practices, Informative discussions must be labeled with respect to each data type.
- c. For each item to be discussed, only the normative portion will be in the main body of the text, all best practice, and informative sections will be in appendices.

Editor, please add a global view of the document as an image, showing the process, and where the standard applies in the process. Please use the figure supplied by the French group as a starting point.

Place information contained in 'Section 2 Conformance' into 'Section 1 Scope'. Remove 'Section 2 Conformance'. Change all instances of "conformance" to "compliance". Final terms and definitions should be aligned with SWG1 work in this regard.

Editors are instructed to consider the general comment of unifying the data structure types using the supplied table as a guideline. Use the recommendations of the full WG3 to resolve the question of what to include from CBEFF versus the existing header. Continue to use FRDI as the header major, as consistent with document 185. Resolve with other proposals is necessary.

Editor is instructed to show that other images are derivatives of a basic image that satisfies only the most basic parts of the image formats. Change other to basic and change the other to reserved. Informative comments shall be put in appendices, including compression and most of sections 5.

The basic image class allows for arbitrary pose and potential for future classification is there.

Editor is instructed to separate definition of grayscale and color images.

Image types:

Basic  
Grayscale Full Frontal  
Color Full Frontal  
Token

The editor is instructed to check with the Rapporteur of SC29 WG1 on advice for an ISO standard lossless compression/storage algorithm.

## **References**

Editor, E1. page 3, clause 4.1.2 (now on page 25 of M1/03-0494, clause 5.7.2): Correct the publication year and the reference numbers. JPEG baseline standard was published in 1994 as ISO/IEC 10918-1 and ITU-T Rec. T.81.

Change E2. page 4, clause 4.1.3 (now on page 25 of M1/03-0494, clause 5.7.2): Change “ISO 15444” to “ISO/IEC 15444-1.”

## **Header**

Include a well-defined record format header that follows as closely as possible guidance from CBEFF and secondarily other data format documents. The global record header and specific image headers must include the following information placed as appropriate. CBEFF now is an abbrev. for “Common Biometric Exchange Formats Framework”

Add image width and height

Add version number of specification

All header format data types must have a zero entry specifying the value of undetermined (verses other). Include CBEFF header for the global (entire) record format. Use the CBEFF format if possible to imbed multiple face images into the global data structure.

Include a format for the storage of multiple images. State clearly that it is intended that this format allow for the encapsulation of multiple face images. Best-practices: keep only the same type of images in the same record.

Define the data structure(s) as Big-Endian.

If eye positions are not determined then the positions are both set to (0,0). In the features area, a flag should indicate the validity of the eye positions. Editor is instructed to solve problem of indicating when a landmark is not determined.

The primary data types will be labeled in the header. The four types will be basic, color full frontal, grayscale full frontal, and token. State clearly that basic is satisfied by any face image data that meets r is intended for use with video media and other face applications requiring novel imaging. State clearly that other is not intended as a placeholder for images that could in principle, but do not in practice, conform to the other three classes.

Include pose information. This includes pose classification (see the FERET work) and precise pose angle determination, for yaw, pitch, and roll.

The header will include a definition of the allowed Color Spaces, grayscale, RGB888, YUV422.

Addition fields include Hair Color, gender (Australia will provide), Eye color, and a landmark/feature structure, based on the ISO/MPEG4 definition of feature positions must be used in an extensible way.

## ***Photographic Properties of Images***

Allowed Expressions will be encapsulated.

Expression for full and token images.

Best Practices

Mouth closed, neutral expression

Allowed in Normative:

- a) A smile where the inside of the mouth and/or teeth is exposed (jaw open)
- b) Raised eyebrows
- c) Eyes looking away from the camera
- d) Squinting
- e) Frowning

Not allowed:

- a) Closed eyes
- b) Hair covering eyes
- c) Rim of glasses covering part of the eye

Backgrounds will be labeled as plain vs. non-plain as defined currently in the text.

The following features will also be included for exception handling:

1. Left Eye patch
2. Right eye Patch
3. Both Eye Patch
4. Sunglasses (for medical reasons)
5. Major medical condition that would make landmark detection difficult. Includes eyes, nose, face deformity.

Note that non-zero values are not normative.

Editor: Allow for the specification features that would impact facial recognition performance, including details such as tattoos, scars and facial jewelry. (However, general descriptive information that cannot be discerned by objective analysis of the photo is beyond the scope of this document.)

Image Source: The editor is instructed to add the image source table as defined below

Field name:

Image Source

Identifier: IMSRC

Data Type: UINT8

Byte Count: 1

This field would be an enumeration:

0 Unknown.

1 Static photograph from an unknown source.

2 Static photograph from a digital still-image camera.

3 Static photograph from a scanner.

4 Single video frame from an unknown source.

5 Single video frame from an analog video camera.

6 Single video frame from a digital video camera.

When data in the data format header(s) (which is technology independent) is present in the CBEFF patron format, the editor is instructed not to include this information in the specific data format, and to use CBEFF data structures for that data. If data elements of this type are missing, the editor is instructed to request additions to the CBEFF structures. This is in conformance to the general recommendation to use the CBEFF patron format structure for this data format.

The editor is instructed to add informative references to the question of aging and it's effect on performance on face recognition algorithms, particularly the impact on the False Rejection Rate.

Please add a reference to the data types UINT8, UNIT16, UNIT32.

Quality should be discussed.

Include pose angle in the header and use of the other category for pose images.

### ***Digital Properties of Images***

Remove redundancy in square and pixel aspect ratio.

The Editor is instructed to specify a minimum of 90 true/independent pixels from eye to eye verses 120 as best practices. In both cases, the original data source should have a non-resampled resolution accordingly – 90 or 120 minimum. Refer pixels to definition below of inter-ocular distance.

Editor is instructed to change 8 bits on luminance on the face to 7 bits of luminance per pixel on the face. Include the comments that: “For 8 bit dynamic range cameras, there is no single f-stop that will work for all skin tones. Therefore special care must be taken to adjust the gain, dynamic range, and f-stop of the imaging system.

The editor is instructed to make the following color changes: The color and grayscale full frontal image types must be separately defined. For color, color spaces of RGB888 and 4:2:2 are allowed.

Editor, move origin to (0,0). Upper left will be (0,0) and is instructed to verify consistency with other data formats.

Include as informative the comments that: “For 8 bit dynamic range cameras, there is no single f-stop that will work for all skin tones. Therefore special care must be taken to adjust the gain, dynamic range, and f-stop of the imaging system.”

### ***Photographic Properties of Images***

Make clear that much of this section refers to full frontal images.

In sections 5.2.3.1 and 5.2.3.4, the following is to be used: a solid background, as defined, is to be best practice. Specification of background is not required for normative. However editor should note that special consideration should be given to the other face lighting constraints in this case. If a background is uniform, it must be specified as such in the header for each image. 18% Gray discussion is to be moved to an informative paragraph.

Best Practice: A background should be stored unaltered by image processing algorithms that assume proper segmentation of the head/face from the background.

Editor is instructed to make clear that goal of the document is to specify the face image from the photographer’s point of view - no shadows, etc. This means that how this image was obtained is outside of the scope of normative discussion. Therefore, the number of light sources is to be informative, three-point lighting is informative, diffuse lighting is informative. The one exception is the normative statement that lighting from an undiffused direct source is not acceptable.

For medical reasons, the wearing of glasses and patches is allowed. In these cases, the specification of the patch, and/or dark glasses in the header structure is mandatory.

Make clear that flash is allowed as long as some diffusion or other techniques are applied. Lighting Reflection Due to Glasses->(This can typically be achieved by increasing the angle between the lighting, subject and camera to 45° (degrees) or more. Move to informative annex)

In section 5.6, make clear that the goal is to eliminate distortion and leave the methods to an index.

Remove "a minimum of three point-balanced (lights) source".

Editor is instructed to make modifications changing the 1.33 ration to 1.34 in section 5.2.2.3.6

Editor is instructed to remove this application specific reference to data storage and displayed image in section 5.5.2.

## ***Landmarks***

Make clear definitions of left and right eyes. Explicitly support the addition of outer eye corners and inner eye corners in the data header. Furthermore, the center of the eyes will be computed as midpoint of the eye corners. A diagram or photograph showing eye centers when eyes looking straight forward versus turned to side must be included. The inner corner of the eyes should be added to the landmark list. Add lower and upper lip landmarks.



For determination add the following proposed sentence:

“It is possible:

- 1 - to use computer inspection
- 2 - to use human visual inspection
- 3 - to use computer and human visual inspection.”

## ***Compression***

Compression should be consistent with specified ISO standards. We believe that JPEG and JPEG2000 are appropriate. Move compression of images into an informative annex.

The editor should note that images from interlaced video are not allowed for full frontal grayscale or color, and token images. All interlacing must be absent (not simply removed, but absent).

Add the following note to the informative appendix: For an actual application it is important to consider both the actual performance and the degradation with compression.

Recheck that these images conform to the new definition for normative digital image properties.

Make consistent text discussion and graphs.

## ***Expression***

Expression for full and token images.

Best Practices: Mouth closed, neutral expression

Allowed in Normative:

- a) A smile where the inside of the mouth and/or teeth is exposed (jaw open)
- b) Raised eyebrows
- c) Eyes looking away from the camera
- d) Squinting
- e) Frowning

Not allowed:

- a) Closed eyes
- b) Hair covering eyes
- c) Rim of glasses covering part of the eye

## ***Pose***

It should be defined what “frontal” means with respect to the up/down direction.

## ***Centering/Portrait Properties***

Centering of the head will be normative.

The face of the assisting person shall not be visible.

The width:height ratio is to be made a best practice. Should make clear that these refer to the color full frontal.

Replace “vertically and horizontally centered” by “vertically aligned and horizontally centered” in section 6.4.2.2

## ***Video***

Interlaced or interlaced corrected images are not allowed for full-face and token image types.

Interlaced images can be labeled as “other”, however. We allow 256x256 images for storage in a data packet for the purposes of video capture and processing. Editor, discuss video in 4th paragraph of introduction.

## ***Token Images***

Canonical images are to be renamed as Token images.

There is typo when defining inner and outer regions.

Editor, further clarify the fact that eyes are always in the same position.

Creation of token image is to be informative only.

Bilinear interpolation to be discussed in informative section.

Please be more precise about the 60 pixels as they are eye to eye inclusive. The distance from eye to eye is actually 59.

Definition of the padding in the background of the token image is best practice. Best Practice: Padding should be done by repetition of the last “valid” pixel per row for additional rows and per column for additional columns of the transformed original image. If padding in the corners of the images is necessary, linear interpolation between the corresponding valid horizontal and vertical pixels should be done. Normative: any color in the undefined region.

The editors are requested to consider a new data type that extends the definition of token image and allows for arbitrary distance from eye to eye.

### ***Cropped Images***

Editor, we look forward to a new data type that would more properly define a cropped image. In the meanwhile, please remove this.

### ***Compression***

Will move compression to informative annex.

Make more precise the reasons for the cutoffs that lead to these values.

Editor will clarify the changes in compression performance and remove discussion of ROI for JPEG (verses JPEG2000).

Editor is instructed to include settings and other information to allow for reproducibility of results.