

Reliable Distortion Detection in Compressed Fingerprint Videos

Chitra Dorai, Nalini Ratha, and Ruud Bolle

IBM T.J. Watson Research Center

P.O. Box 704, Yorktown Heights

New York 10598, USA

{dorai,ratha,bolle}@watson.ibm.com

November 2, 1999

Abstract

Distortions in fingerprint images arising due to the elasticity of fingerprints and pressure and movement of fingers during image capture lead to great difficulties in establishing a match between two images of a single fingerprint. In this paper, we propose a novel approach to detect and estimate distortion occurring in fingerprint videos. Our approach directly works on MPEG- $\{1,2\}$ encoded fingerprint video bitstreams to estimate interfield flow without decompression and uses flow information to determine elastic distortion.

Key words: Fingerprints, live-scan, CCIR-601 format, fingerprint videos, image sequences, flow, distortion.

1 Introduction

Distortions in fingerprint images arising due to the elasticity of fingerprints and pressure differences during image capture lead to difficulties in establishing a match between two images of a

single fingerprint. Some of the problems of interest in this area include (i) investigation of plastic distortions using flow; (ii) extraction of key frames based on degree of distortion; and (iii) estimation of affine motion model parameters to characterize interfield motion in fingerprint videos.

2 Previous Work

Previous work on distortion uses a composite matched filter method with Fourier transform based correlation.

3 Distortion Detection in Fingerprints

Fingerprint matching is a challenging problem. Fingerprint images are acquired by touching a scanning surface with one or more fingers. During image acquisition, the pressure and movement applied by a finger on the scanning surface causes distortion in the image of the topological features of the finger. Often this distortion can lead to poor matcher performance since high tolerances are required to account for the distortion in images. In this paper, we describe a novel system and method for detecting distortion from a sequence of fingerprint images acquired at a sampling rate of 30 fps.

Our approach employs an optical flow-based analysis of fingerprint frames to detect distortion in a video bitstream. Since MPEG- $\{1,2\}$ -based compression is typically used to compress fingerprint image sequences, our approach directly works with the embedded information in the encoded bitstream such as motion vectors of macroblocks in a frame, without having to uncompress the bitstream into their individual images. This savings can be quite huge when large fingerprint images are sampled at high frame rates.

Our novel approach has the following steps:

- Fingerprints are obtained as a continuous sample of images over a specific time interval, \square and the resultant sequence of grey scale images are encoded into a video bitstream using MPEG-1 or MPEG-2 video compression standard.
- Compute the interfield flow of each macroblock present in a field (frame) which represents

the motion that may be present in any field (frame) with respect to its immediate next field (frame) in the video bitstream without decompression. A *flow* characterization of a video bitstream is a frame-type-independent uniform motion representation amenable for consistent interpretation and computed from the raw motion vectors encoded in the MPEG- $\{1,2\}$ bitstreams.

- Using the flow computed for each frame in a given bitstream, determine frames which exhibit a high level of inter-field motion activity. This is based on a novel measure $Z2NZ$ computed for each frame, which is the ratio of the number of macroblocks of zero flow vectors to the number of macroblocks exhibiting non zero flow vectors in a frame. If $Z2NZ < 1$, implying that macroblocks undergoing nonzero motion exceed those with no motion, the frame is deemed to be a candidate frame for distortion investigation.
- With the subset of candidate distorted frames, our analysis proceeds further by identifying the stationary region in the fingerprint present in each frame using the flow field of the frame. This region is a non-moving region of the finger around which the skin is pressed hard or rotated or twisted or rolled to induce distortion. A connected component analysis is performed on a frame obtaining connected regions containing macroblocks that show zero motion. The largest (in area) connected component is selected as the “core” unmoving region around which further analysis takes place. We compute the bounding box of the pivotal region determined in the frame.
- For each frame, compute motion parameters from the fingerprint region around this central area by imposing an affine motion model on the frame flow and sampling the moving region radially around the bounding box of the pivotal region. The affine motion parameters quantify translation, rotation, and elastic stretching due to fingerprint motion. Six parameters, a_1, \dots, a_6 are estimated in this process, where a_1 and a_4 correspond to translation, a_3 and a_5 correspond to rotation, and a_2 and a_6 correspond to scale. We compute curl [1] in each frame j , $\mathcal{C}_j = -a_3 + a_5$. The curl in each frame quantitatively provides the extent of rotation, or the spin of the finger skin around the pivotal region. We also compute the magnitude of the translation vector, $\mathcal{T}_j = (a_1, a_4)$ of the frame.

- For each frame, we next compute a smoothed curl and translation magnitude values by computing the average values of curl and translation magnitude over its neighborhood. The temporal window that was used for smoothing spanned one-tenth of a second or in other words over 3 frames.
- A simple thresholding classifier is employed to signal distorted and non-distorted frames. This classifier takes into account the temporal extent of the distortion frame and also the range of the curl values in its final determination of distorted frames.
- A grouping process is carried out to collect consecutive non-distorted (“good”) frames to result in undistorted set of frames from which a frame can be chosen as key frame for final matching. This frame is expected to yield better matching scores with tighter tolerance values in the matching system.

3.1 Distortion in Fingerprint Sequences

Distortion in fingerprints can be viewed as that of adding a dynamic component to static fingerprints, thus imposing a behavioral aspect on the physiological aspect of the finger that is normally captured by a fingerprint scanner. The appearance of a fingerprint image is changed over time by various factors: A user may exert force on the finger with respect to the sensing device that typically results in images with thick ridges. The force may vary at different time instants during the interval when the print is being sensed. Or a user may apply torque which results in rotated fingerprint frames. Or it could be rolling a fingerprint which results in frames containing partial views of the finger to full views. One can also envision a fingerprint sequence, in which a user uses a combination of translation, rotation, and force to embody specific patterns of movements across the scanner surface. All these result in sequences, where the fingerprints are continuously elastically deformed according to the force, torque, and roll.

3.2 Flow extraction from MPEG-2 Encoded Bitstreams

The MPEG-2 video standard [2] aims at higher compression rates and contains all the progressive coding features of MPEG-1. In addition, MPEG-2 has a number of techniques for coding interlaced video.

A digital video sequence is a series of one or more images, often referred to as frames. MPEG-2 provides a choice of two picture structures to code a frame [2]. *Field pictures* consist of individual fields (assemblies of alternate lines of a frame) that are each divided into macroblocks (16×16 blocks of pixels) and coded separately [3]. A *frame picture* is one whose interlaced field pair is interleaved together into a frame which is then divided into macroblocks and coded. Thus, a frame can be encoded using two *field pictures* or a single *frame picture*. The pictures can be further coded as *I pictures*, *P pictures*, and *B pictures*. MPEG-2 video encoding utilizes five different motion compensation modes to reduce the temporal redundancy between adjacent pictures in a sequence [3]: (i) frame prediction for *frame pictures*, (ii) field prediction for *frame pictures*, (iii) field prediction for *field pictures*, (iv) 16×8 motion compensation for *field pictures*, and (v) *dual-prime* for P pictures. We refer the reader [2] for a comprehensive discussion of these prediction schemes.

In order to capture the flow characteristics present in fingerprint frames of a MPEG-2 encoded bitstream, we exploit the MVs of the MBs from each frame or field that provide the temporal information in the stream. A major difficulty with MPEG-2 encoded videos is that pictures in the sequences can be of different structures (*frame pictures* or *field pictures*); they can be of different types, i.e., I, P, or B frames or fields; and they can occur in a variety of GOP patterns. This poses difficulties in obtaining MVs directly from any frame in the stream without the knowledge of its frame type and coding specifics. The additional flexibility to predict MVs from different fields or frames (field prediction and frame prediction, to name a few) in MPEG-2 further aggravates the generation of consistent frame-to-frame motion information and comparison of MVs across different frame types.

To summarize, since raw motion vectors (MVs) from MPEG- $\{1,2\}$ video can be from different picture structures (*frame pictures* in progressive scanning mode or *field pictures* in interlaced mode) and from different picture coding types (I, P, or B) which can occur in a variety of frame orderings (*IBBPBBPBBPBBBI* . . . , *IPPPP* . . . , etc.), utilizing them directly in video analysis leads to problems in generation of consistent frame-to-frame motion information and in comparison of

motion across different frame types. We make use of a frame-type-independent motion representation called *flow vectors* [4] for a coherent motion analysis spanning multiple temporally contiguous frames. Our flow representation handles MPEG-2 specific enhancements related to picture structures, picture coding types, and additional motion-compensation schemes.

Our approach involves representing each MV in an MB in a picture regardless of its frame and picture type, with a backward-predicted vector determined with respect to the *next immediate field*. The set of backward-predicted vectors (called the *flow*) computed for each field in the stream then represents the direction of motion of each MB in the field with respect to the next field in the sequence. Observe that even if the picture in the video was encoded as a *frame picture*, our approach computes the flow fields for its top and bottom fields; this aids us in characterizing high motion changes that may be present even between fields of the frame. Further observe that not all MBs will necessarily have their associated flow vectors; but the number of such MBs is rarely large to affect our aggregate analysis of flow vectors during annotation. An input video stream consisting of *frame* and *field pictures* is thus, transformed into a sequence of *flow fields* using our new and efficient flow estimation techniques. As a result of utilizing this frame/picture/motion-type-independent framework, a video stream that has been encoded into different bitstreams with differing IPB patterns can be analyzed easily under a single common framework and compared. Experiments on thousands of frames from SD and HD video bitstreams demonstrate the good performance of our flow estimation process. The experiments also show that the flow estimates generated from our algorithm are consistent and robust, leading to correct annotation of global motion occurring in video streams.

We have developed accurate flow estimation procedures and details of relating to generation of reliable flow vectors from **MPEG-2** encoded bitstreams are found in [4].

3.3 Detection of Candidate Distorted Frames

Using the flow computed for each frame in a given bitstream, we next determine frames which exhibit a high level of inter-field motion activity. This filtering allows us to perform a first level filtering of those frames that likely to contain a high degree of distortion in their finger. This is based on a novel measure $Z2NZ$ computed for each frame, which is the ratio of the number of

macroblocks of zero flow vectors to the number of macroblocks exhibiting non zero flow vectors in a frame. During distortion, while a portion of the finger is held stationary, the rest of the finger is twisted, rolled or pressed hard on the scanning surface and this results in a flow field in which there are a few zero flow macroblocks and a substantial number of macroblocks show some flow. Thus, the measure, $Z2NZ$ provides a quantitative characterization of the total flow present in the frame. If $Z2NZ < 1$, implying that macroblocks undergoing nonzero motion exceed those with no motion, the frame is deemed to be a candidate frame for distortion investigation.

3.4 Stationary Region Detection in Candidate Frames

When a person knowingly or unknowingly attempts to result in distorted images of their fingers during image capture, a portion of a finger is typically held stationary and the rest of the finger around this pivotal unmoving region is pressed, twisted and moved around to introduce distortion in its image. Therefore, determining the unmoving region in a fingerprint frame aids in focusing the motion estimation process around this region.

With the subset of candidate distorted frames, our analysis proceeds further by identifying the stationary “core” region in the fingerprint present in each frame using the flow field of the frame. A connected component analysis is performed on a frame using zero flow as the basic merging criterion. This results in connected regions containing macroblocks that show absolutely no motion. The largest (in area) connected component is selected as the “core” unmoving region around which further analysis takes place.

3.5 Affine Motion Estimation in Candidate Frames

For each frame, compute motion parameters from the fingerprint region around this central area by imposing an affine motion model on the frame flow and sampling the moving region radially around the bounding box of the pivotal region. The affine motion parameters quantify translation, rotation, and elastic stretching due to fingerprint motion. Six parameters, a_1, \dots, a_6 are estimated in this process, where a_1 and a_4 represent horizontal and vertical translation respectively, a_3 and a_5 correspond to rotation, and a_2 and a_6 correspond to scale.

A parameterized model of image motion such as affine motion model is invoked when it can

be safely assumed that spatial variations of pixels in a region can be represented by a low-order polynomial. With fingerprint frames, we are especially interested in studying the rotational and translational effects due to force, twist and roll in finger movements and therefore an affine model is deployed as a first level approximation of our requirement to obtain rotation, translation, and scale parameters. Within small regions, in our case, annular regions around the pivotal region in a frame, an affine model of global motion is invoked to model the flow of a macroblock in the region:

$$u(x, y) = a_1 + a_2x + a_3y \quad (1)$$

$$v(x, y) = a_4 + a_5x + a_6y \quad (2)$$

where (x, y) denote the coordinates of a macroblock in the frame, (u, v) , the flow vector associated with that macroblock, and a_1, \dots, a_6 , the affine transformation parameters. Let \mathcal{U} denote $\begin{bmatrix} u & v \end{bmatrix}^T$, \mathcal{X} denote $\begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix}^T$, and $\vec{\mathbf{a}}$ denote $\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{bmatrix}^T$.

We can now estimate $\vec{\mathbf{a}}$ by minimizing the error between the flow estimated using Equations 1 and 2 and the actual flow determined from the encoded bitstream.

$$S(\vec{\mathbf{a}}) = \sum_x \sum_y [(\hat{u}_{xy} - u_{xy})^2 + (\hat{v}_{xy} - v_{xy})^2] \quad (3)$$

where $\begin{bmatrix} \hat{u} & \hat{v} \end{bmatrix}^T$ is the estimated flow vector. We solve for $\vec{\mathbf{a}}$ using the least squares estimation technique [5]. For different radii, starting from the periphery of the bounding box of the pivotal region and incrementing the radius in unit step, all valid macroblocks in the annular region given by each radius are used in the flow estimation process. For details about setting up the equations and their solutions to compute $\vec{\mathbf{a}}$, we refer the reader to [5].

Once the affine transformation vector is estimated, we compute *curl* [1] in each frame j , $\mathcal{C}_j = -a_3 + a_5$. The curl in each frame quantitatively provides the extent of rotation, or the spin of the finger skin around the pivotal region. The rotation about the viewing direction captured by curl

proves to be useful in determining and describing the distortion in frames.

3.6 Final Selection of Distorted Frames

For each candidate frame, we next compute a smoothed curl and translation magnitude values by computing the average values of curl and translation magnitude over its neighborhood. The temporal window that was used for smoothing spanned one-tenth of a second or in other words over 3 frames. The candidate frames near the beginning and the end of the sequence retain their curl values if an adequate neighborhood cannot be established for smoothing. A sequence of fingerprint frames can be viewed at this juncture as a set of contiguous candidate distorted frames separated by groups of undistorted frames.

In each group of contiguous candidate distorted frames, a simple thresholding classifier is employed to signal whether this interval contains distorted frames. This classifier takes into account the temporal extent of the distortion frame and also the range of the curl values in its final determination of distorted frames. If the temporal extent of a group of frames exceeds a threshold, t_b , then it is established that the group cannot be a mere noisy blip but rather it contains frames that have low $Z2NZ$ ratio and have non zero curl values. Therefore the group is marked to contain distorted frames.

On the other hand, if the temporal length of a group is small (less than t_b), then the classifier investigates more closely to investigate whether this small distortion blip is a true distortion interval. This is carried out using two sequential tests: (i) the first test checks whether the difference in curl values of frames in this group exceeds a certain threshold \mathcal{C}_v ; this indicates that there are abrupt changes in the curl values which is an indicator of distortion. (ii) the second test examines whether the maximum translation magnitude in this group of frames is greater than a certain threshold, \mathcal{T}_v ; this is to ensure that the group of frames does not undergo pure translation only which can be due to a finger being moved from one point on the scanning surface to another without any twist or roll.

Once the group establishes that its curl variation is high and its translation is small, the classifier labels the group as true distortion frame group. If the curl variation is low, it is inferred that there is no strong evidence of distortion and the group is labeled as “undistorted”. Finally, a

merging process is carried out to group consecutive intervals of same distortion flag to form large meaningful intervals.

Each frame is given one of the two labels. The labels of successive frames are compared to extract a grouped contiguous sequence of frames with identical labels. Presence of a few noisy labels (e.g. over one sixth of a second or five frames) within a sequence is tolerated and handled by assigning them the same label as that of the longer enveloping sequence. Key frames can be chosen for final matching from each “undistorted” intervals. This frame is expected to yield better matching scores with tighter tolerance values in the matching system.

4 Highlights of Our Approach

The advantages of this technique are the following: (i) Our technique can work in both compressed and uncompressed image sequences. With compressed bitstreams, we estimate the flow using the raw motion vectors encoded in the streams. With uncompressed fingerprint image sequences optical flow between frames can be estimated using one of many well-established methods from the literature [?, 6]. (ii) With fingerprint scanning systems that output compressed data, computations are performed directly in compressed domain without uncompressing to an image sequence. This leads to less memory requirement and efficient computations. (iii) Robust flow extraction from raw motion vectors of a compressed video bitstream is unique to our approach. (iv) Sequence analysis in fingerprint images has not yet been reported in the literature and our work is the first to report. (v) A novel method for locating non-moving fingerprint region during distortion using the flow is proposed in our approach.

5 Experimental Results

5.1 NIST Special Database 24

NIST Special Database 24 [7] contains MPEG-2 encoded videos of live-scan fingerprints. There are two datasets:

- 100 MPEG-2 video files obtained from 10 fingers of 5 males and 5 females; each video

stream is 10 seconds long (with 30 fps and 720×480 frame resolution). The important point about this dataset is the prints are with plastic distortions.

The distortions were induced by rolling and twisting each finger. Some blurring is also present in the data.

- 100 MPEG-2 video files obtained from 10 fingers of 5 males and 5 females; each video stream is 10 seconds long (with 30 fps and 720×480 frame resolution). This differs from the other set in the fact that the fingerprints are at various rotated angles.

The finger rotation is from left extreme to the right extreme. Sampling angle might not be uniform across images. The expectation was to create five evenly sampled rotated prints between the two extremes.

Observe that the databases is a random selection of individual fingerprints.

5.2 Ground Truth

Twenty streams obtained from two different fingers, left and right index fingers of ten people (both males and females) were visually analyzed and the frame numbers were marked as those containing distortion and those which did not. The ground truth was compared with the results obtained from our system to estimate both false positive and missing intervals. The results were found to be promising.

References

- [1] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *ICCV*, (Boston, MA), pp. 374–381, 1995.
- [2] Joint Technical Committee ISO/IEC JTC 1, "ISO/IEC 13818-2. "information technology—Generic coding of moving pictures and associated audio information:Video," May 1996.
- [3] B. G. Haskell, A. Puri, and A. N. Netravali, *Digital video: An introduction to MPEG-2*. Chapman and Hill, 1997.

- [4] C. Dorai and V. Kobla, “Extracting motion annotations from MPEG-2 compressed video for HDTV content management application,” in *Proc. International Conference on Multimedia Computing and Systems*, vol. 1, (Florence, Italy), pp. 673–678, IEEE, June 1999.
- [5] J. Meng and S.-F. Chang, “Cveps – a compressed video editing and parsing system,” in *Proceedings of the ACM Multimedia 96 Conference*, (Boston, MA), November 1996.
- [6] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, “Hierarchical model-based motion estimation,” in *2nd European Conference on Computer Vision*, pp. 237–252, 1992.
- [7] C. I. Watson, “Nist special database 24 digital video of live-scan fingerprint data.” National Institute of Standards and Technology, July 1998.