CBIR APPROACH TO FACE RECOGNITION

Dimo Dimov, Nadezhda Zlateva, and Alexander Marinov

Institute of Information Technologies at Bulgarian Academy of Sciences (IIT-BAS),
Sofia 1113, Acad G. Bonchev str., block 2 & 29-A, tel. +(359 2) 870 6493, e-mail: dtdim@iinf.bas

Abstract: Face recognition is interpreted as a CBIR (Content Based Image Retrieval) problem. An arbitrary input image of a given face is treated as a sample for search within a database (DB), containing a large enough set of images (i.e. projections from a sufficient number of view points) for each human face of interest. We assume that the faces for recognition are static images, which have been appropriately extracted from an input video sequence. In addition, we have at our disposal a CBIR method for image DB access that is simultaneously fast enough and noise-tolerant. The paper describes both the methodology used for building up the DB of image samples and the experimental study for the noise-resistance of the available CBIR method. The latter is used to acknowledge the applicability of the proposed approach.

Key words: CBIR, face recognition, 3D object recognition, image databases, image/video analysis.

INTRODUCTION

The Face Recognition (FR) area is intensively being explored and developed in the last years. The increasing demands toward the biometric security systems for example [1], definitely count on yet another modality – the face (the person's physiognomy), in addition to the classical modalities – personal code, signature, voice, finger print, iris-image, and other. Many of the FR problems still remain open, for example: (1) detection/identification of a face in a common scene, (2) normalization (by geometric size and illumination) of the faces isolated from the scene, (3) face standardization (i.e. the isolation of temporary/non-significant attributes like hair, moustaches, beard, glasses), (4) storage of the face-samples (patterns, standards) in a DB, as well as the capacity of this DB (i.e. the number of client-physiognomies for recognition), and last but not least – (5) the collision raised by the representation of a 3D object (face) with its 2D projections. A detailed state-of-art survey for the majority of these problems is provided in [2]. Extra information may also be obtained by the specialized surveys, for example [3] that interprets FR mainly in the aspect of emotion recognition. To the set of established FR methods we can definitely add the "Eigenfaces" [4] that allude to problem (2) by Principal component analysis (PCA), and also the "Fisherfaces" [5], where the illumination impact is reduced to 3D-subspace of the whole pixel-feature space and is followed by Fisher's interpretation of Linear discriminant analysis vs. the PCA. Other approaches, such as neural networks, fuzzy sets, colour histograms, template matching and other, have yet a very limited application [2, 3]. A characteristic drawback of the well-known methods is their inability to meet the challenge of a combination of the above mentioned FR problems.

The paper proposes an approach to a combination of the following FR problems: the normalization (2), the IDB size (4) and the 3D-2D-collisions (5). The detection problem (1) is not considered; the standardization problem (3) may be considered in the meanwhile.

BACKGROUND

The proposed approach for FR is vision-based, and uses (minimum) one video camera for capturing the human face in its dynamics. For the sake of simplicity, we consider the recognition of a single moving object, i.e. we assume that the recognized face is „the biggest spot" in the video frames, or at least in the majority of frames incoming from the camera.

We will interpret the task of recognizing a face within the camera frame as a task for the direct comparison of an input example with samples from a given image database (IDB), i.e. as a *CBIR problem* [6, 7, 8].

Essence of the proposed approach: A given face in front of the video camera is considered as a dynamic 3D object, represented by a series of 2D projections, i.e. static images from the camera. If an appropriate part of these images, or similar to them, is already stored in an IDB with representative face images, then we can search into this IDB for the image sample that is the closest (most similar) to the source/input image. Moreover, we can search for a series of image samples, sorted (in descending order) of their similarity to the input image. Of course, a "dictionary" with a large amount of samples will be needed, whose size we will try to estimate experimentally hereinafter.

In the same time, the comparison time of the input image with all samples from the dictionary has to be quick enough to assure operation in real-time. I.e. the realization of the above idea is possible, if we have a noise-tolerant and simultaneously fast enough CBIR method for accessing a large DB with face images. Such CBIR methods are available with the system EFIRS (Effective and Fast Image Retrieval System) developed by IIT-BAS, [9, 8, 7]. Their noise-resistance covers cases of eventual linear transformations in the input (translation, rotation, and scaling) and regular noise, as well as rougher (to a certain level) artefact-noise.

Actually, the primary goal of this research was to clarify if the available (developed by us) CBIR approach is appropriate for the recognition of dynamic objects in a video-clip, in different recognition applications [7]. A preliminary result in this aspect has been already committed for the case of sign language alphabet recognition [10]. In this paper we are considering the case of FR.

Expected problems and allowed limitations: In analogy to [10], the following two problems can be formulated for the FR approach, too:

(1) isolating the object of interest (a human face in dynamics) from the input image (static 2D scene) and/or from the time sequence of similar scenes (video-clip);

(2) accumulating a representative enough IDB, i.e. a dictionary with images of *representative face projections.*

The first problem is characteristic for most of the approaches in the area of computer vision, image processing and recognition, therefore, at this stage, we will consider it a'priory resolved. We will concentrate here on the 2nd problem – gathering the representative samples in the IDB and performing an experiment of evidence for the chosen concept.

Except for being enough in number, the samples from the experimental IDB have to adhere to the general limitation of the available CBIR methods [8, 9] – the images need to be relatively "clean", i.e. to contain the whole object of interest (the human face), in color or gray scale, over an uniform (white) background and, if possible, be devoid of noise-artifacts from the natural surrounding.

All these limitations, related in fact to the well known problems of face segmentation [2, 3], are resolved here in their light form, according to the experiment's specifics. For that purpose, we are using a simplified scene – a motionless human head in front of a dark blue curtain, i.e. with opposite color to the actor's face skin, and facing the camera central position.

We also managed relatively easy the expected difficulties with the scene's surrounding illumination. Thus, we were thoroughly concentrated on the uniformity of the manual scanning with the camera, row by row, as explained below.

APPROACH

As a possible alternative to the needed experimental IDB, we could use an IDB partially available on the Internet, [11]. This approach is not the most suitable for our goals, as revealed in the following analysis.

Three approaches to data gathering: The methodology of gathering data in IDB (containing a lot of projections of different faces) is based on the recording of a short video-clip that traces out the object (a face) through uniform scanning (by position and time) in a spatial sector wide enough and in front of the object. Three possible approaches have been considered:

(1) Static object and moving camera, which scans uniformly the needed spatial sector around/in front of the object.

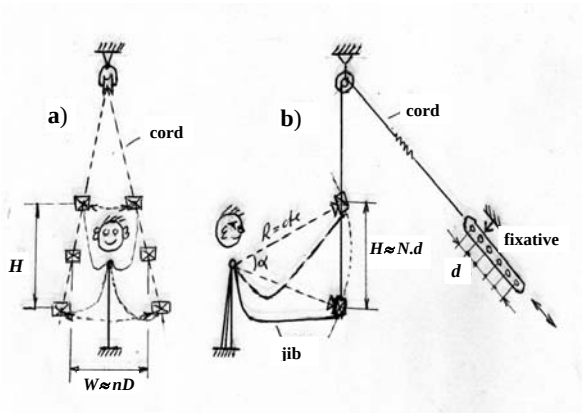


Fig.1. Kinematics' schema of the construction for taking "primary" video-clips, using the "static object – moving camera" approach.

(2) Static camera and moving object, which makes uniform motion in order to expose itself to the camera from all needed points of view.

(3) Static object and static camera.

All the mentioned approaches are valid at the stage of regular exploitation of a given FR system. But at the stage of accumulating different images (views of physiognomies) in an IDB, we will once again adopt approach (1), in view of the similar advantages already pointed out in [10].

The third approach, in comparison to the first two, is not directly oriented towards the creation of a video-clip. The latter needs to be computer generated from a few static photos, taken from different positions within the needed spatial sector around/in front of the object. This approach is comfortable from the actor's point of view, but it is unacceptable for our goals. The needed development efforts on generating the film or the final frame sequence (imitating close enough positions of exposure) are unduly expensive for our goals. The still images produced by this approach have to be considered only as a possible input to a FR system in operation.

The second approach is fairly simple from a photographer's point of view. Yet, it requires high precision and a certain level of "acting" skills on the part of the "object-actor" when performing the uniform movements in front of the camera. If the actor is not trained enough for the role in question, there will be the need of additional efforts on "normalizing" the film. In other words, the approach is unacceptable for our goals, even if it is quite popular from the practice of gathering similar data, [11].

The first approach turns out to be the most acceptable one even though it requires an auxiliary construction for recording a video-clip with a conventional camera; Fig.1 illustrates an idea for this construction. Here, the responsibility for providing the needed uniformity of motion within the video-clip falls on the operator – researcher. This is acceptable and natural as we are speaking of a single (possibly a few multiple) session of unvaried scanning procedures. This process can be automated in the case that the current approach proves to be experimentally efficient enough.

Thus, at this stage, we are using the construction illustrated on Fig. 1 and 2 and propose the following method on providing the experimental data for our IDB:

(s1) Fix the "actor's face" in the necessary pose: in front of the camera and approximately towards its central position.

(s2) Scan the needed spatial sector in front of the actor's face: row by row, top-down, moving in "zigzag" uniformly along

Fig.2. The construction in action.

the rows and having the camera always on while scanning. In vertical transitions (from row to row) use an assistant to provide the "synchronous" release of the vertical restriction (a cord) through equally spaced distances *d*, vertically among the arcs of radius *R*, *R*=const (Fig.1). At the same time the camera operator can simply cover the lens by hand for better segmentation of significant row frames at the video processing step later. It is also desirable that the time for transition (from row to row) does not go beyond 1÷2 seconds, for operativeness.

(s3) If there is next actor face to capture, go to (s1).

Video capture: *The experimental materials* are video-clips, whose unique frames could serve as samples (close-ups) of the human face for recognition. We call these "*primary (video) clips*" and let them meet the following set of requirements:

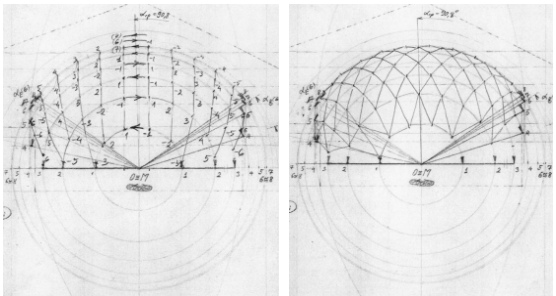
- there is a separate video-clip for the face of each person of interest;
- the video capture is carried out row by row, having each row separated from its neighboring ones by a sequence of empty (black) frames;
- it is advisable that the "manual" scanning with the camera along each row be carried out at a nearly regular speed;
- the average scanning speed along the different rows is tolerated to vary within some not very large bounds.

*Primary video-clips processing***.** Stages:

♦ Separate the motion frames from the entire video-clip.
♦ Derive a *representative set of frames* from the video sequence by, for example, obtaining those frames that fall within a uniform grid ("square" or "triangular" one) over the spherical sector of scanning. The linear parameter *D* of this grid (Fig.1 and 3) will represent the differences between the consecutive representative frames, which contain the object samples from the corresponding view points. *D* is measured in degrees, but also in cm (for the concrete construction, Fig.1), or even in number of frames (by a regular speed of scan).

We assume that the noise-resistance of the used CBIR method could also be measured by means of *D* and *d*, cf. Fig.1. Thus, in a uniform grid with $D > D_0$, where D_0 is an *admissible lower boundary*, the CBIR method would start to err when recognizing by similarity to the samples of the IDB. This D_0 boundary can serve as a measure for the noise-resistance of the CBIR method in use. It will also determine the optimal number of grid nodes ($\approx \pi (N/2)^2 Dd / D_0^2$) for storing the representative samples within the IDB, Fig.3. We will skip the details of defining the geometrical model of the experiment. Below we provide some of the more important parameters:

♠ the radius *R* of the scanned spatial sector, *R* = 51 cm;



(**a**) (**b**)

Fig.3. Spherical sector in front of the camera, scanned row by row. Two main variants for the uniform grid of representative frames: a) square and b) three-angular one.

♠ the average angles of the spatial sector of scanning – horizontally ≈ 80° and vertically ≈ 115° (in angle degrees);
♠ the average speed of scanning, i.e. of the manually moved camera that captures along the scanning rows (horizontal arcs) with ≈ 15÷20 degrees per second, by a camera capture speed of 15÷16 frames per second;
♠ the distance *d* between consecutive scan rows over the spherical sector, fixed to *d* = 10 cm (as cord gaps given by a fixative); at the chosen *R* it refers to *d* ≈ 9.6° (as chord angle);
♠ the number *N* of the scanning rows (arcs over the spherical sector) is chosen to *N* = 8, see Fig.1 and 3;

Thus, for both uniform grids shown at Fig.3 we have the following possible values of *D*:

$D = kd$, for square-type-grid, (Fig.3a)

and respectively

$D = kd\,2/\sqrt{3}$, for triangle-type-grid, (Fig.3b)

where k is the scan row number per grid cell, $k \in \{1,2,...(N-1)\}$.

Significant frames extraction: The used methodology for estimating the identification numbers of the representative frames relies considerably on that during the scanning phase, while jumping from a row end to next row begin, the camera lens is manually hidden, i.e. the corresponding frames from the video-clip are almost black. This simplifies (see Fig.4, 4a, 4b) the separation of the significant frames from the "black" ones via the following 5 steps:

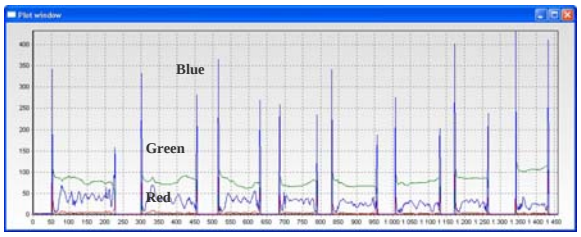


Fig.4. The primary frames for a given face (A2), frames are numbered horizontally. The Green graphic shows the average intensity per frame, the Blue one shows the maximal intensity, while Red one – the average intensity of the differences between two consecutive frames. (R) specifies the zones with significant frames as well as the black zones in between. The significant frame zones can be made more precise through (G) and (B).

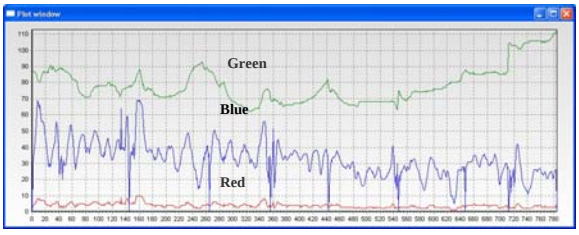


Fig.4a. The significant frame zones only, the black frames zones indicating the (top-down) row changes are removed. The graphics have the same sense as in the above Fig.4.

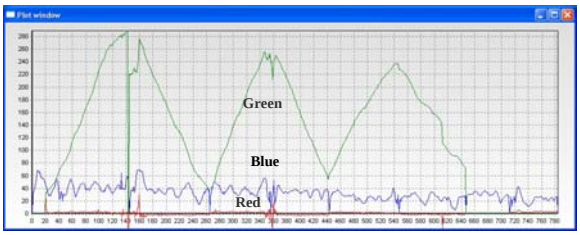


Fig.4b. The significant frames with evaluated the row scan position progress (in Green) in dependence of the evaluated scanning speed (in Red) along the row scan direction. The Blue graphics remain their sense of maximal intensity per frame, like in Fig. 4 and 4a.

(a1) Transform the frames from RGB to a Gray scale;

(a2) Create an image of the differences between each two consecutive (Gray) frames from the video sequence;

(a3) Calculate the values of the average intensity of the original frames, the maximal and average intensity of the differences, (Fig. 4);

(a4) Find a statistical estimation of the corresponding thresholds for the transition from the sequences of insignificant (black) to the sequences of significant (motion) frames, (Fig. 4a);

(a5) Define a regular net of representative frames for the given face using ad-hoc estimations for the row scan position progress and respective scanning speed (Fig.4b).

Face segmentation: Some of the more popular approaches in segmenting the face are: usage of a controlled or known background with a previously acquired background image (i.e. applying background subtraction); using segmentation by motion; or using color segmentation with predefined or generated skin color models [2,3]. In addition, to alleviate the segmentation and recognition process, multiple camera configurations are used [1]. These contribute with additional information such as image depth, 3D shape and motion and thus are helpful in detecting the eye and/or mouth position, etc., [1,2,3].

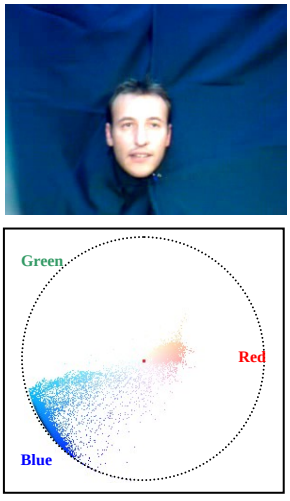


Fig.5. HS-histogram of a given frame.

The problem of segmenting the face from each of the motion frames is simplified by the chosen experimental environment – dark blue background, in contrast to the human skin color. Thus, after a RGB to HSV (Hue-Saturation-Value) color scale transformation of the frames, the segmentation can be carried on mainly in the 2D HS-histogram schema, Fig.5, where two main color sectors are outlined – blue for the background and beige for the face. Hence, after a transformation to a (cyclic) histogram by H only, the segmentation task can be reduced to finding both the optimal thresholds of separation [12]: the blue-beige one and the beige-blue one. That is, at this stage we don't need to use information about the motion. Of course, some other peculiarities are considered in parallel, such as the fact that in some frames the face region contains gleams (too bright pixels) where H is undefined. Here, we also check the illumination (V) of the corresponding pixels, define a threshold, and apply binarization by V, cf. also [12].

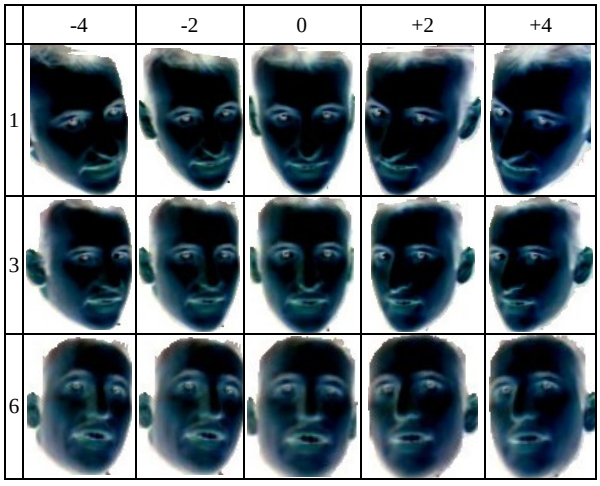


Fig.6. Several representative frames extracted from the set of significant frames for the given face video-clip.

In brief, the chosen segmentation algorithm for a given frame is as follows:

(b1) Compute the HSV space and evaluate the H-sector of the blue background.

(b2) Determine eventual regions surrounding the basic blue background, and repaint them with blue. This is to exclude parts from the environment (floor, ceiling, walls, etc), which might accidentally land in the frame.

(b3) Compute the HSV space again.

(b4) Compute an optimal threshold on the S-histogram. Use low S-values to exclude eventual parts of the experimental construction that might fall into the frame.

(b5) Evaluate the H-sector of the "beige" region of interest (face) and search for the biggest contour (of maximal area) therein.

(b6) Aggregate the other "beige" regions to the biggest one.

(b7) Finally, for the necessities of EFIRS, negate the images (making their colors opposite) and repaint the (blue) background with white (Fig.6).

IDB EXPERIMENTS AND RESULTS

The experimental analysis of the proposed approach has been carried out by the usage of EFIRS [9, 8, 7]. EFIRS is a C/C++ written Windows-XP application operating on an IBM compatible PC. For the experiment objectives, an extra test, functionally similar to the conventional SLT (Simple Locate Test) of EFIRS [9], has been written. The existing IDB structure of EFIRS was used for the generation of experimental IDB of test samples. The chosen CBIR access method was PFWT, as described in [9]. The primary video-clips have been acquired through a construction as in Fig. 1.

Essence of the experiment: For each possible value of the examined parameter D (the basic size of the grid), do:

♣ Generate a separate IDB for EFIRS by loading it with all representative frames for all faces (persons) of interest. It is recommended that the representative frames for each face be chosen regularly positioned (at distances $D \approx d$) over the square grid within the experimental sector of visibility.

♦ For each square on the grid associated with a given primary video-clip, define the closest frame to the center of this square. These central frames are obtained from the set of motion frames along each row of the clip, and are uniformly the most distant ones (at distances $\approx D/2$ or $\approx \sqrt{5}D/2$) from their corresponding 6 neighboring frames (corners of two vertically contacting squares), which have already been stored into the IDB. These central frames are used to provide "the heaviest case" of input precedents for the CBIR search within the IDB.

♥ Carry on a SLT for a CBIR search within the IDB over all square centers, i.e. over "the heaviest cases" of input precedents. Summarize the results for the successful and unsuccessful retrievals from IDB. Practically, for more dependable results, test with all significant image frames extracted from the primary video-clips.

Experiments: At the current stage we have loaded the IDB with the representative frames of 22 faces, belonging to 16 persons, some of them been filmed twice or thrice to capture

different emotional status expressed, e.g., (A1,A2,A3), (Ao,Au), (N1,N2,N3) and (V1,V2), cf. Table 2.

The significant images acquired for these faces are qualitative enough, and enough in their number – 8177 images altogether, on average 378 per face – so that we can carry a preliminary experiment with the EFIRS system expecting the obtained statistical results to be sufficiently reliable. 1251 of the images have been defined as representative ones in the IDB, with an average of 57 representatives per face. The representative frames for each face have been arranged in a uniform square grid, cf. Fig.3a and Fig.6.

The generated IDB applies to the case of $D = d$, where D is the basic size of the grid, and d – the distance between the scanning rows of the video-clip.

Table 1 contains the results for a given face, e.g. "A2". Only 6 rows from all the 8 rows are successfully scanned for this example and the averaged error rate has been evaluated to ~ 2.6%, which is considered a moderately good result among all the face experiments.

Table 2 contains the generalized results for each face experiment, where the general averaged error rate has been evaluated to ~ 6.6%.

Thus, at this stage we can determine the boundary value D_0, i.e. the noise-resistance measure of the CBIR approach in the given FR application, as:

$$D_0 = 1d, \textit{ at FR rate} > 93\%..$$

As for the averaged search time per input face it is ~ 0.2 s.

Table 1. Test results for a given face (A2).

Row No.	Frames ess./repr.	Errors (1) || (2)	Errors (1): Letter err.	Errors (2): row diff.s > ±1	Exact matches	Warns (1) row diff.s = ±1	Warns (2) row diff.s = 0	by w(1) || w(2) avrg of position diff.s	by w(1) || w(2) avrg of abs. position diff.s
1	123 / 11	3	3	0	11	37	72	-1.9	5.8
2	106 / 11	3	1	2	12	11	80	-5.2	7.4
3	91 / 11	8	2	6	11	17	55	13.3	15.9
4	83 / 10	0	0	0	9	20	54	3.9	6.3
5	104 / 10	1	1	0	9	28	66	10.7	13.4
6	65 / 7	0	0	0	8	3	54	0.1	5.8
7	-	-	-	-	-	-	-	-	-
Total	572 / 60	15	7	8	60	116	381	-	-
Avrg.	71 / 10	1.9	0.9	1.0	7.5	14.5	47.6	2.9	9.0
%	-	**2.6**	1.2	1.4	10.5	20.3	66.6	-	-

Table 2. Test results for all faces available into experimental IDB.

Faces IDs	Rows scanned	Significant frames	Represent. frames	% Errors (1) || (2)	% Errors (1): Letter errs.	% Errors (2): row diff.s > ±1	% Exact matches	% Warns (1) row diff.s = ±1	% Warns (2) row diff.s = 0
A1	6	538	63	3.2	1.7	1.5	11.3	21.0	64.5
A2	6	572	60	2.6	1.2	1.4	10.5	20.3	66.6
A3	5	488	53	11.1	9.0	2.0	10.7	15.2	63.1
Ao	5	312	53	9.9	9.9	0.0	16.3	11.5	62.2
Au	5	328	54	8.8	8.8	0.0	15.9	10.1	65.2
DD	5	260	54	16.2	14.2	1.9	19.6	0.4	63.8
GA	4	220	41	0.0	0.0	0.0	18.6	5.9	75.5
GG	3	214	33	4.2	4.2	0.0	15.0	10.3	70.6
HT	6	454	61	14.3	13.2	1.1	12.8	14.5	58.4
IH	5	282	48	0.7	0.7	0.0	16.7	7.1	75.5
LB	6	392	60	0.8	0.0	0.8	14.8	18.9	65.6
LI	6	388	62	3.1	1.5	1.5	16.0	11.6	69.3
LK	3	215	42	17.7	16.7	0.9	16.7	11.2	54.4
MP	6	391	61	3.6	3.6	0.0	15.3	6.4	74.7
MV	7	463	71	4.8	4.8	0.0	14.9	13.2	67.2
N1	6	342	62	4.4	2.3	2.0	16.4	10.5	68.7
N2	8	481	76	3.5	3.1	0.4	14.8	9.6	72.1
N3	7	478	74	6.1	3.1	2.9	13.6	22.6	57.7
PK	6	399	(63)	9.0	5.5	3.5	14.5	14.8	61.7
SK	5	312	53	12.5	11.9	0.6	16.0	5.8	65.7
V1	5	304	53	6.6	4.3	2.3	17.1	15.8	60.5
V2	5	344	54	7.8	5.8	2.0	14.5	16.0	61.6
avrg %	-	-	-	**6.6**	5.3	1.2	14.6	13.4	65.5

Notes to Table 1 & 2:

- "Exact matches" corresponds to the number of samples (for given face and scan row) loaded into the IDB.
- "Errors (1)" count the errors of type 1 ("erroneously recognized face"). These are expected to be highest along the boundary of the scanned spatial sector, and lowest – in its central area.
- "Errors (2)" count the errors of type 2 ("greater than ±1 deviation between both the rows, the input and the found ones"). Their expected behavior is similar to that described for Errors (1).
- "Warns (1)" and "Warns (2)" register the expected situations of "deviations in ±1 from found to current row". The averaged values "Avrg_of_position_diff(erences)" and "Avrg_of_abs(olute)_position _diff" concern both these recognition situations. It is expected that "Avrg_of_position_diff" would be close to zero (if the model is void of geometric inaccuracies), while "Avrg_of_abs_position_diff" – close to the average distance *D/2* (in number of significant frames) between each two consecutive representative frames.

Discussion of results: We can consider as positive the result of $D_0 = d$. It signifies that the size of the needed IDB would be the largest, i.e. ~ 80 = 8*10 as a number of representative images for each face in our case. The latter is not important in terms of the needed resources of (external) memory, since for the proposed CBIR technology, which covers a size of hundreds of thousands of objects, it is enough to only store a key with size of ~ 0.5 KB per object [8].

From the viewpoint of efficiency, the search time remains a logarithmic one [7,8,9,10]. Thus, within an IDB of ~ 80 representative samples for each face (person or physiognomy of interest), the proposed approach would require a $\log_2(80) \approx 6 \div 7$ of extra accesses, in addition to the basic access number of ~ $\log_2 F$, F the number of faces represented in the IDB.

Even a result such as $D_0 < d$ should not be considered hopeless. It signifies that D_0 cannot be found at the chosen accuracy d (i.e. the distance between the rows of scan) of the IDB experiment. The value of D_0 would then be a fractional number $D_0 = \alpha d$, $\alpha \in (0,1)$.

Actually, by definition the result is $D_0 > d$, i.e. the chosen CBIR technology does not err when searching of an exact match, while the percentage of errors from occasional linear transformations on the input (translation, rotation, and scaling) is < 1.5%. The former is fundamental for the used CBIR technology, while the latter has been experimentally proven over a database of a large size ≈ 60 000 of trademark/hallmark images [9,8].

Thus, even at an IDB with $D = d$, we can determine D_0 more accurately as a fraction of d, at least horizontally, on the basis of the average distance between two consecutive motion frames from a primary video-clip. Naturally, this distance is primary dependent on the camera capture speed and speed of the (manual) scanning by rows.

By a comparative analysis of the achieved results towards those of [10], an earlier application of the same approach to sign language recognition problems, one can note that with similar image preprocessing methods in use the results in [10] look better. Nevertheless, the current results are more promising because of the larger experiments over more complicated objects (human faces vs. palm signs).

CONCLUSION AND FUTURE WORK DISCUSSION

The paper proves once again our [7] idea of "surrounding filming" of 3D objects to represent them as a minimal set of representative 2D projections (images) in an IDB and to use this representation to meet the 3D recognition problem as a CBIR one, in behalf of our recent progress in CBIR efficiency [8, 9]. Instead of palm signs [10], we have considered here human faces in their static and/or dynamic.

In the near future we are going to arrange the experiments for $D_0 = 2d$, and even $D_0 = 4d$, and evaluate the respective FR rates. This will give us a basis to define the number of necessary representative frames per face into IDB as a function of a predefined FR rate value.

The currently achieved FR rate of ~ 93% is considered not bad in view of the very early stage of the experimental refinements. The following refinements in the frame image preprocessing are envisaged in the near future:

- a geometrical (shape) normalization of faces, using the distances among the eyes and mouth;
- an intensity (lightening) normalization, considering faces as Lambertian surfaces by analogy of Fisherfaces [5];
- a standardization of faces, removing the extra facial attributes like hair, moustaches, glasses, etc., [2].

In this way we expect to achieve higher FR rate of up to 99% and more, and simultaneously to reduce the number of necessary representative frames per face – up to 5÷9 for relatively small number of faces in the IDB. For large (and very large) IDB of faces it is expected that the number of ~ 32÷64 representative frames per face will be optimal (preferable) for a similar FR rate of ~ 99% (or higher).

ACKNOWLEDGMENTS

This work was partially supported by the following grants: Grant # BG-GR-17/2005 and Grant # VU-MI-204/2006 of the National Science Fund at Bulgarian Ministry of Education & Science, and Grant # 010088/2007 of BAS.

REFERENCES

1. Jain, A.K., A. Ross, and S. Prabhakar, An Introduction to Biometric Recognition, *IEEE Trans. on Circuits and Systems for Video Tech.*, Vol. 14, No. 1, 2004, pp 4-19.
2. Zhao, W., R. Chellappa, A. Rosenfeld, and P.J. Phillips, Face Recognition: A Literature Survey, *ACM Computing Surveys*, 2003, pp. 399-458.
3. Fasel, B., and J. Luettin, Automatic Facial Expression Analysis: A Survey, IDIAP Res. Rep. 99-19, Martigny, Switzerland, Nov. 2002, *J. Pattern Rec.*, vol.36, no. 1, 2003, pp.259-275.
4. Turk, M., and A. Pentland, Eigenfaces for Recognition, *J. Cognitive Neuroscience*, vol.3, no.1, 1991, pp.71-86.
5. Belhumeur, P.N., J.P. Hespanha, and D.J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. on PAMI*, vol. 19, no. 7, 1997, pp.711-720.
6. Smeulders, A.W.M., M. Worring, S. Santini, A. Gupta, and R. Jain, CBIR at the End of the Early Years, *IEEE Trans. on PAMI*, vol. 22, no. 12, 2000, pp. 1349-1380.
7. CBIR/CBOR application(s) based on a FANTIR technology**,** Ref: R&D_BG_14255, Brokerage Event on Information Technologies during 20th INFOSYSTEM, 30.09 - 01.10. 2006, Thessaloniki, Greece, (http://www.innovationrelay.net/bemt/catalog.cfm?status=2&whattodisp=profiledetails&eventid=1405&pr_id=14255&CFID=24699&CFTOKEN=75182066)
8. Dimov, D., Rapid and Reliable Content Based Image Retrieval. In: Proc. of NATO ASI, Multisensor Data and Information Processing for Rapid and Robust Situation and Threat Assessment, 16-27 May 2005, Albena-Bulgaria, IOS Press, 2007, pp. 384-395.
9. Dimov D., A Polar-Fourier-Wavelet Transform for Effective CBIR, In: Proc. of the ADMKD'07, 02.10.07, Varna (in 11-th Int. Conf. ADBIS'07, 29.09-03.10.07, Varna), 2007, pp.107-118.
10. Dimov D., A. Marinov and N. Zlateva, CBIR Approach to the Recognition of a Sign Language Alphabet, In: Proc. of CompSysTech'2007, June 14-15, Rousse, 2007, pp.V.2.1-9, (http://ecet.ecs.ru.acad.bg/cst07/Docs/cp/sV/V.2.pdf)
11. Extended Multi-Modal Face Database, (http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb)
12. Laskov, L., and D. Dimov, Color Image Segmentation for Neume Note Recognition, In: Proc. of the Int.Conf. A&I'07, 03-06.10.07, Sofia, 2007, pp. III.37-41.