A Method for Describing and Recognizing Fragments of Visual Scenes Based on Analysis of Descriptors Trajectories in the Feature Space

Mikhail V. Petrushan, Mikhail V. Kopeliovich, Aik V. Kurdoglyan, and
Anatoly I. Samarin

Southern Federal University, Rostov-on-Don, Russia,
`drn@bk.ru`

Abstract. We consider the problem of feature description of the object in video image. Classical approaches to the problem are oriented on synthesis and optimization of object descriptors. We propose alternative approach to object description based on the analysis of trajectory of its descriptor in the feature space when the object moves relatively to observer. The approach was evaluated on the facial recognition problem, where it provides high accuracy.

Keywords: Feature Space, Descriptor Trajectory, Motion Analysis

1 Introduction

The problem that we address herein is the problem of distinctive description of objects images. We consider one typical case: visual system observes two objects (not necessarily at the same time) that are placed in its field of view. Images of the objects are described by some n-dimensional descriptor that can be treated as a point in n-dimensional feature space. When motion of these objects occurs relative to visual system, descriptors positions change in feature space. The simple case of 2-dimensional descriptors motion in $\{f1, f2\}$ feature space that is caused by motion of objects in real world is shown in Fig. 1. Displacements of object in real world relative to the visual system generate the trajectory of descriptor's displacements in feature space. For different objects, the common scenario is such that some points of their descriptor's trajectories could be close to each other and even trajectories could have intersections (Fig. 1).

Such trajectories closeness or intersection means that in particular conditions different objects can not be distinguished by comparing their descriptors since they are equal or very close to each other in feature space. An illustration of such phenomena is shown in Fig. 2. When observing cylinder and cone from particular direction of view their images look equal meaning that their descriptors are equal as well.

One possible way to eliminate uncertainty is to perform active observing procedure when visual system moves relative to the object and descriptors begin moving along different trajectories in feature space. Certain regularities and
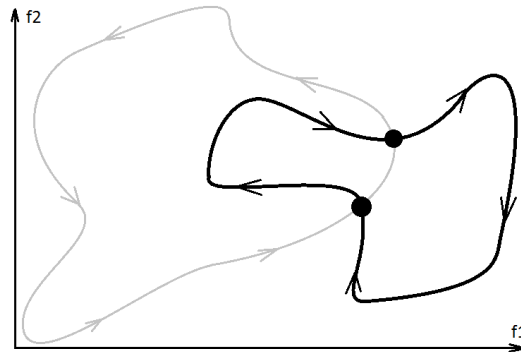
Fig. 1. Possible descriptors trajectories in 2-dimensional feature space $\{f1, f2\}$. First trajectory is drawn in black, second – in gray. Trajectories intersections are marked with spots
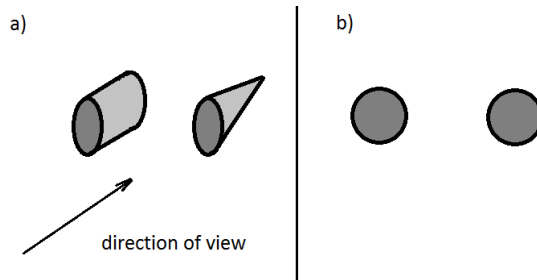

Fig. 2. An example of visual uncertainty. Two different objects look equal (b) from particular direction of view (a)

attributes of descriptor's trajectory can be used as a new high-level descriptor that we propose to use to enhance object recognition in a case of active observing. In order to distinguish between object descriptor and new proposed descriptor that encodes the trajectory of object descriptor in feature space, further we refer to the object descriptor as to the first-order descriptor, and to the trajectory descriptor as to the second-order descriptor. In this paper, we evaluate efficiency of proposed approach on the task of facial images recognition. We consider this work as a proof of concept. We describe a new approach and do not focus on quantitative comparison of the proposed approach and existing analogues.

2 Related Works

Different works on analysis of object motion are mainly aimed at recognition of motion type or action [1,2] and rarely at recognition of object itself. Different image-domain or "optical flow"-domain descriptors are used to encode motion in images [3,4]. Dalal et al. [5] proposed the motion boundary histograms (*MBH*)

descriptor that encodes the relative motion between pixels. It is based on computing derivatives separately for the horizontal and vertical components of the optical flow. Another descriptor used for motion encoding is *HOF* [6] – histograms of quantized optical flow orientations that encodes local distribution of optical flow. Trajectory shape descriptor [1] is also used to encode local motion patterns in images. Motion trajectory is described by a sequence of displacement vectors. The resulting vector is normalized by the sum of displacement vector magnitudes. In contrast to these image-domain descriptors, we propose features-domain descriptor that encodes trajectory of object descriptor's motion in feature space.

An approach similar to our was developed by Favorskaya [7]. She proposed to recognize dynamic patterns by applying descriptors trajectories analysis in the feature space. The problem is solved by constructing a predictive filter for each pattern, which could indicate the future state of the dynamical object, basing on some part of its previous states. The decision depends on comparison of predicted and real descriptor in the $(t + 1)$-th observation. Thus, the dynamical pattern recognition problem is reduced to the filter constructing problem, which predicts trajectory behavior in a multidimensional space. Methods, based on such formulation of the problem of dynamical patterns recognition are called "the methods of learning predicating filters". However, in spite of conceptual closeness between the method of M. N. Favorskaya and our approach, there is principal difference: we don't offer to use trajectory analysis for descriptor prediction (and further comparison of observed and predicted results). Instead of this, we directly describe the trajectory or its characteristics by new higher order descriptor – trajectory descriptor, and during classification we compare these trajectory descriptors, that characterize not the static view of the describing object, but the dynamics of its descriptor in feature space.

3   Methods

We used two first-order descriptors in the study: *Simple Descriptor* (*SD*) and *Gradient Descriptor* (*GD*). *SD* is a vector of intensity values of image pixels in a local vicinity of the describing point. In particular, intensities of all pixels in rows of rectangular area ($R \times R$ pixels) around central pixel of facial area in image are aligned into one n-elements row ($n = R^2$). R was chosen equal to face width. In order to reduce computational complexity and to make trajectory visualization possible we project n-dimensional descriptor to 2-dimensional feature space by Principal Components Analysis (*PCA*).

The next type of first-order descriptor we've experimented with is a *Gradient Descriptor* [8]. It describes local spatial distribution of orientations of over-threshold gradients. In particular, intensity gradients $Grad_x$, $Grad_y$ (in horizontal and vertical directions) in a given point are calculated by applying Sobel operator. Then magnitude and orientation of gradient are calculated:

$$Magn = \sqrt{Grad_x^2 + Grad_y^2};\qquad(1)$$

$$Orient = \arctan \frac{Grad_y}{Grad_x}. \tag{2}$$

Orientation angle is quantized by 16. After that, descriptor's values $D(i,j)$, $i,j = 1..16$ can be computed as a number of gradients in i-th sector (16 sectors total) of descriptor's window (Fig. 3) such that their quantized orientation are equal to j and magnitude is higher than a chosen threshold T. In this work, threshold was chosen as a 10% of maximal gradient magnitude over the image. The center of descriptor's window was placed at the center of the face in image, it's diameter was equal to face width. Facial area could be found by applying Viola-Jones method [9].

Fig. 3. An example of positioning of *GradientDescriptor* input window

Further processing involves projection of descriptor onto 2 principal components in the same way as for SD.

We studied three second-order descriptors of trajectory: *Angles Histogram* (*AH*), *Transitions Histogram* (*TH*), and *Probabilistic Map* (*PM*).

We define *AH* as follows. Relative motion of observation system and the object causes changes in the object's descriptor and in its projection onto principal components. This change of the descriptor projection forms the displacement vector, which is directed at the certain angle (α and β in Fig. 4) to the first principal component (X axis in Fig. 4) – to eigenvector of the descriptors covariance matrix, which corresponds to maximal eigenvalue. Histogram of such angles is the first considered trajectory descriptor. All possible angles from $0°$ to $360°$ are quantized by 100. The i-th element of such histogram represents proportion of the angles in the whole trajectory, which belong to the i-th quantum.

The second considered trajectory descriptor is a *Transitions Histogram*, which is defined as matrix *TH* of dimension 360×360. The matrix element value $TH[\alpha, \beta]$ represents proportion of the angels α followed by angles β in the whole trajectory (see Fig. 4). α, β are the consequent angles, rounded up to integer

values, between displacement vectors in trajectory and first principal component. After processing all adjacent segments of the trajectory, the matrix TH is smoothed by Gaussian filter in window 51 ($sigma$ is equal 8.5). Then all matrix elements are transformed in the follow manner: the minimal element value of TH is subtracted from all its elements, and then elements are divided by difference between the values of maximal and minimal elements of the matrix TH. Such transformation maps the elements values onto the range $[0, 1]$. Resulting matrix is treated as a trajectory descriptor.
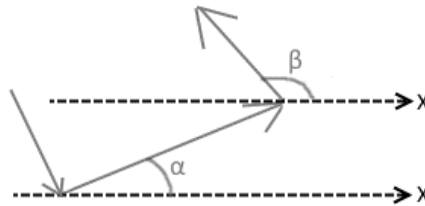

Fig. 4. An example of consequent displacement vectors in descriptor trajectory in feature space. X axis corresponds to the direction of the first principal component

The third variant of the trajectory descriptor is a transition probability map (PM) from certain area of the feature space at a certain angle. It is defined as follows. For particular observing object, area in 2-dimensional feature space, which include all possible projections of object description onto 2 principal components, is divided into regular grid 100×100. Every three adjacent points of descriptor trajectory in feature space constitute an angle, which is quantized by 8 (from $0°$ to $360°$ by step of $45°$). Descriptor's element $PM[i, j, \alpha]$ represents proportion of the quantized transition angles α in the whole trajectory appeared in the cell $[i, j]$ of the regular grid. For each angle α, the part of probabilistic map $PM_{i,j}$ corresponding to α is processed by Gaussian smoothing in window 11 ($sigma$ is equal to 1.8).

All considered second-order descriptors (trajectory descriptors) describe statistical regularities in trajectories of first-order descriptors in feature space during continuous observation of moving object (relatively to observer). The trajectory descriptors proximity is defined as Euclidean distance between them. Classification rule that determines if trajectories descriptors belong to the same class or separate is based on the threshold value of their proximity: if the distance is less than the threshold value, then they are treated as the same class objects, else – to different classes.
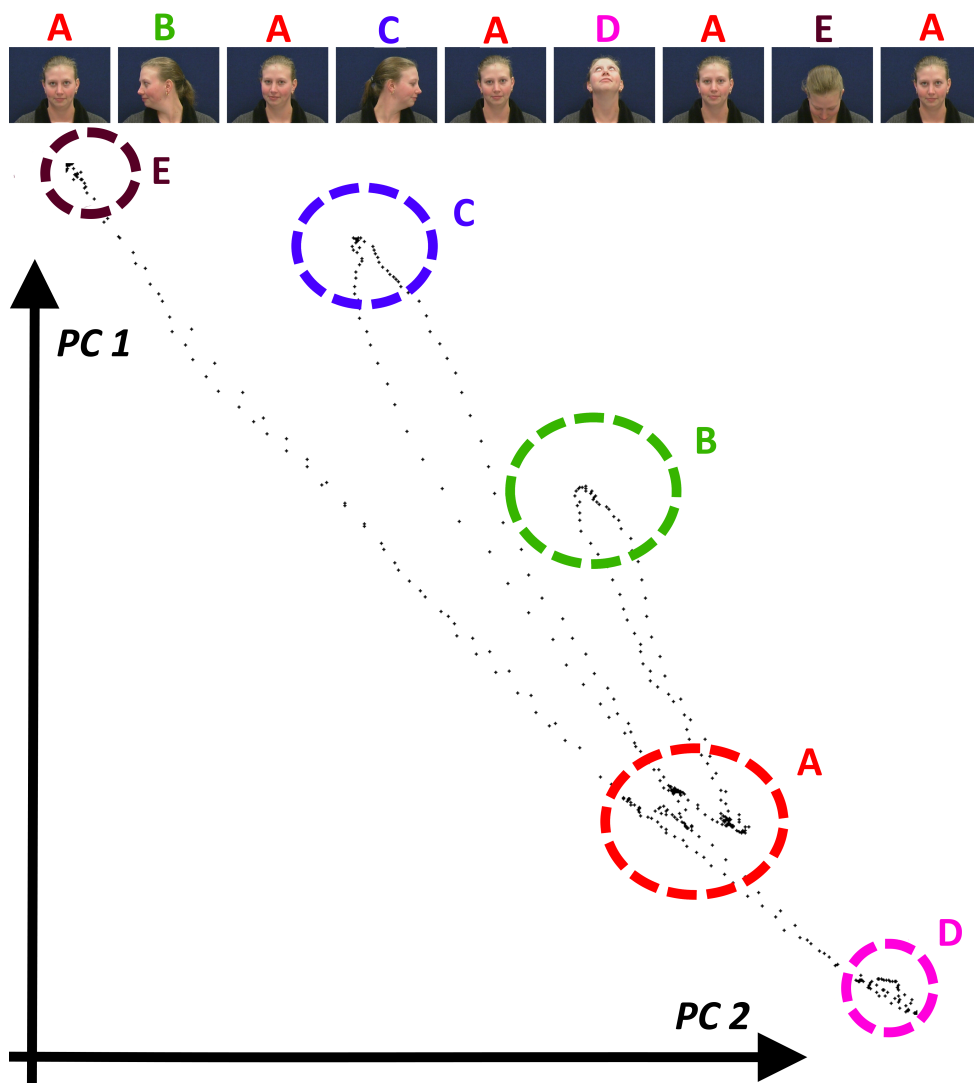
Fig. 5. An example of frames obtained from one video session of VidTIMIT dataset and the descriptor trajectory obtained for the session (in the space of two principal components *PC 1* and *PC 2*). Black points correspond to values of the projections of first-order descriptor. Head positions and corresponding areas in the feature space are marked by letters A-E

4 Experiments and Results

We have evaluated our approach of descriptor trajectory analysis applying it to the facial recognition problem on VidTIMIT dataset [10]. We have used a part of

the dataset that consists of video of 43 volunteers (19 female and 24 male). The video were recorded in 3 sessions, with delays of 6-7 days between sessions. The hairstyle and clothing usually varied between sessions. Additionally, the zoom factor of the camera was randomly perturbed after each recording. Each person performed multidirectional head rotation in each session. The actions sequence consists of the person moving their head to the left, right, back to the frontal position, up, then down and finally return to frontal position (Fig. 5). In each record, face is located within the same square area; therefore, we calculated frame descriptors using the area of 384×384 pixels size that was cropped from the central part of frame.

Firstly, we obtained first-order descriptors for each frame of the dataset video. After that, we calculated *PCA*-basis in the descriptors feature space and projected first-order descriptors onto two principal components with highest variances (Fig. 5). Finally, for each recording session we obtained trajectory descriptors of three types. That is, with given type of trajectory descriptor, there are 43 persons and 3 trajectory descriptors corresponding to each one (because of 3 sessions for a person).

To evaluate the efficiency of trajectory descriptor, we built a binary classifier that estimates Euclidean distance between objects and compares it to a given threshold. For a threshold, we calculated true positive and true negative rate by the following algorithm: for each trajectory descriptor considered, we calculate the reference descriptor for the person by averaging another two descriptors corresponding to the person. If the distance between the considered descriptor and the reference descriptor is below the threshold then the comparison is counted as true positive, otherwise it is counted as false negative. Then we compare the reference descriptor to the each of the trajectory descriptors corresponding to others persons: if the distance is below the threshold, then the comparison is counted as false positive and otherwise as true negative.
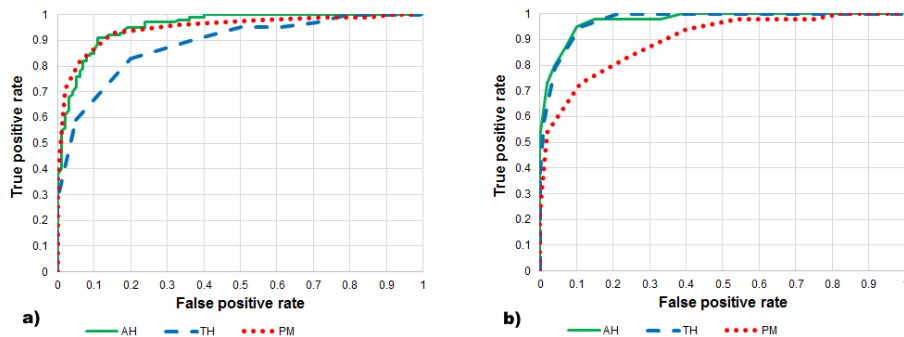

Fig. 6. ROC curves calculated for *Simple Descriptor* (**a**) and *Gradient Descriptor* (**b**)

Varying the classification threshold, we constructed ROC (receiver operating characteristic) curve and calculated the area under the curve. Fig. 6 illustrates ROC curves for each frame descriptor and trajectory descriptor. The values of area under the ROC curves are shown in Table 1.

Table 1. Calculated areas under ROC curve

	AH	TH	PM
Simple Descriptor	0.99	0.88	0.94
Gradient Descriptor	0.97	0.96	0.89

5 Conclusion

The alternative approach to object classification problem was proposed. The approach is based on analyzing trajectory of the object descriptor obtained from video image. Three types of trajectory descriptors were evaluated in application to facial recognition problem. Angles Histogram that represents distribution of displacement vectors angles in feature space was shown to be the most effective among considered trajectory descriptors in terms of maximal area under ROC curve.

References

1. Wang, H., Kläser, A., Schmid, C., Liu, C.L.: Dense Trajectories and Motion Boundary Descriptors for Action Recognition. Int J Comput Vis **103** (2013) 60–79
2. Sanath, N., Ramakrishnan, K.R.: A Cause and Effect Analysis of Motion Trajectories for Modeling Actions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2014) 2633–2640
3. Jeannin, S., Divakaran, A.: MPEG-7 Visual Motion Descriptors. IEEE Transactions On Circuits And Systems For Video Technology **11**(6) (2001)
4. Tran, K., Kakadiaris, I., Shah, S.: Part-based motion descriptor image for human action recognition. Pattern Recognition **45**(7) (2012) 2562–2572
5. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: European Conference on Computer Vision. (2006)
6. Wang, S., Zhu, E., Yin, J., Porikli, F.: Anomaly Detection in Crowded Scenes by SL-HOF Descriptor and Foreground Classification. In: 23rd International Conference on Pattern Recognition (ICPR), Cancún Center, Cancún, México (2016)
7. Favorskaya, M.: The dynamic pattern recognition based on predicating filters. Vestnik SibGAU **1**(1) (2009) 64–68 (in Russian).

8. Petrushan, M., Samarin, A.: The method of contrasting of facial images descriptors for the authorized access system. Information Technologies **3** (2012) 74–78 (in Russian).
9. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition (CVPR) **1** (2001) 511–518
10. Sanderson, C., Lovell, B.: Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference. Lecture Notes in Computer Science **5558** (2009) 199–208