DOI 10.3233/THC-209004
IOS Press

Multi-target video-based face recognition and gesture recognition based on enhanced detection and multi-trajectory incremental learning

Jirui Lin, Laiyuan Xiao* and Tao Wu
School of Software, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China

Abstract.
BACKGROUND: Video-based face recognition (VFR) is one of the frontier topics in the domain of computer vision, which aims to automatically track and recognize facial regions of interests (ROIs) in video sequences.
OBJECTIVE: In videos with multiple faces, the trajectories of individuals are incredibly complex. This is less studied than videos with a single face per frame.
METHODS: In this paper, we present a multi-trajectory incremental learning (MTIL) algorithm, which categorizes trajectories using a Euclidean distance-based greedy algorithm and estimates the most likely labels for each trajectory by incremental learning to correct their classification and improve the accuracy of recognition. Furthermore, this study proposes an enhanced detection method that combines face detection with a robust tracking-learning-detection (TLD) algorithm to improve the performance of face detection in video. The method can also be extended for medical video recognition applications such as gesture recognition control based medical system.
RESULTS: Experiments on Honda/UCSD and BMP (seq_mb) database demonstrate that our method can improve the face detection and face recognition (single or multiple) performance. The method also performs well on the gesture recognition system.
CONCLUSION: The proposed MTIL algorithm can significantly improve the performance of the VFR system and the gesture recognition system.

Keywords: Multi-target face recognition, enhanced face detection, multi-trajectory, gesture recognition

1. Introduction

Video-based face recognition (VFR) is a comprehensive research field that includes face detection, target tracking, and face recognition, and has been widely studied by researchers. Although less complicated than VFR problem, gesture recognition is also important in several realistic applications such as the medical video recognition system. Generally speaking, face recognition and gesture recognition can be combined into one research topic. VFR can be divided into recognition based on video sequences and image sets, where the former utilizes the dynamic spatiotemporal information from the sequences [1].

*Corresponding author: Laiyuan Xiao, School of Software, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei, 430074, China. Tel.: +86 13807180383; E-mail: xiao.l.y@hust.edu.cn.

Considerable progress has been made by VFR researchers including the probabilities approach [2], adaptive learning [3], hidden Markov model [4] and radon transform [23]. The adaptive multi-classifier system (AMCS) for video-to-video FR in changing surveillance environment has been presented by Pagano [5]. Torre et al. [6] developed a VFR method based on adaptive skew sensitivity. The method improves the accuracy and robustness of the classifier ensembles by selecting training data with varying levels of imbalance and complexity. They also proposed a method for partially supervised learning from facial trajectories [7]. Dewan et al. [8] developed an adaptive appearance model tracker (AAMT) system that attempts to solve the 'single sample per person' (SSPP) problem by creating a track-face-model for each person, which is updated for each frame, and matched to each person's gallery-face-model recorded in the system.

The selection of non-targets is a difficult problem because the human face is a complex non-rigid model that is prone to influences from poses, lighting, expressions, and appearance changes [9]. In videos with a single human face, the FR system only needs to detect or track one face region on each frame. In contrast, in videos with multiple faces, the trajectories of individuals are incredibly complex and appear synchronously. In this study, we tested a multi-valued classifier algorithm based on local binary patterns histogram (LBPH). Results indicate that the proposed multi-trajectory incremental learning algorithm (MTIL) can utilize general multi-valued classifier-based FR algorithms to match multiple human face trajectories in a video to labels. The most probable label for each trajectory can be estimated and updated, which gradually improves the accuracy of recognition results.

Human faces must be detected before they can be recognized. Current approaches to face detection include those based on machine learning [10], average face templates [11] or head-shoulder detectors [12]. One recent popular face detection approach is to base the face detection on the Viola & Jones (V&J) classifier [13]. However, this classifier has been known to have false negatives or positives in tests due to changes in lighting or poses (especially for the right orientation), which might explained why the algorithm's training results are insufficiently accurate [5].

The contributions of this paper are as follows. First, it presents an MTIL algorithm, which can recognize multiple faces that simultaneously appear in a scene. The Euclidean distance-based greedy algorithm is used to categorize the trajectories, and each trajectory is stored using a multi-value classifier into classification forms. Second, the accuracy and reliability of face detection have been enhanced by a creative combination of change detection, V&J face detection, and a robust TLD algorithm.

2. Overview of video-based multi-face tracking and recognition system and change detection

Figure 1 depicts the general framework of the face recognition system based on enhanced detection and MTIL. The system consists of modules of face detection, tracking, face recognition, and trajectory incremental learning. The face detection and tracking systems are connected by the change detection module, which is responsible for determining whether the number of faces have changed and detecting false negatives and positives. The tracking system is ready to constantly make adjustments based on information from the change detectors. The features are extracted using an LBPH operator, which has the advantage of being invariant to rotation and grayscale transformation.

The symbols in Fig. 1 are explained as follows:

CPL $= \{C^i; i = 1, \ldots, n\}$: (coordinate classification table) is a set of individual C^i.
TCC $= \{P^i; i = 1, \ldots, n\}$: (trajectory statistics table) is a set of statistical trajectory P^i.
Tail $= \{T^i; i = 1, \ldots, n\}$: (tail trajectory table) is a set of trajectory T^i.

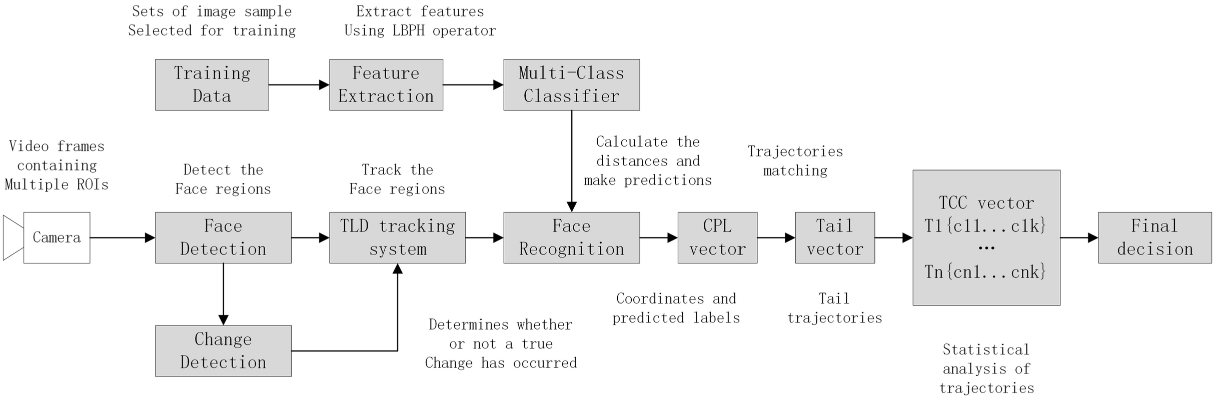

Fig. 1. Architecture of the proposed video-based multi-face recognition system using TLD tracking and trajectory improvement learning.

3. Enhanced face detection based on TLD algorithm

3.1. Choosing the tracking algorithm

The current tracking methods can be categorized as based on regions, dynamic profiles, features, and models [16–21]. A key problem in tracking in the long-term is the variation of the target, such as occlusion, postures, scale, and lighting. It is difficult to ensure the continuity and accuracy of the tracking when the target is obscured or undergoes other local changes from time to time.

The tracking-learning-detection (TLD) algorithm is a long-term tracking algorithm from Kalal et al. [22]. It is extremely robust in handling the occurrence of shape changes, partial occlusions, and other changes to the target using an improved online learning mechanism to continuously update the tracking module's 'significant feature points' and the detection module's target models and relevant parameters. In this study, we used the TLD algorithm to improve our system's performance.

3.2. Face detection and change detection mechanism

The popular Viola & Jones (V&J) classifier is used for face detection in the initial frames of the video. When applied to VFR, the V&J-based face detector can suffer from errors caused by lighting, poses or expressions. For example, the rotation, skewing or intense expressions may all cause the detector to lose track of the face (false negatives, FNs), while inaccuracies in its initial training results may cause it to identify non-face regions as faces from time to time (false positives, FPs). The change detection module can detect real changes of human faces in the scene and provide correction for the tracker on the number and statuses of faces by removing abnormal decrease or increase of human faces caused by FNs and FPs, respectively. Given that the false detections are in a short duration, we can tally the length of time τ in frames where the number of faces shows a sustained change. If τ is between the positive and negative liminal values ($\theta_{dec} < \tau < \theta_{inc} | \theta_{dec} = -3, \theta_{inc} = 3$), the change is considered false and the tracker continues to track the faces; otherwise, it is considered a real change and the tracker updates the number and states of faces accordingly. Here the values of θ_{dec} and θ_{inc} are determined through experiments to effectively eliminate false positives cause the non-face region detected at the same place last for no more than three frames through observation.

4. MTIL-based face recognition

4.1. Multi-trajectory incremental learning (MTIL) algorithm

Recently, good local feature descriptors, such as local binary patterns (LBP) [14] and scale-invariant feature transformation (SIFT) [15], have been widely used in face recognition. We chose the local binary patterns histogram (LBPH) to represent facial features for its moderate computation complexity. Our multi-trajectory incremental learning (MTIL) algorithm tracks multiple face trajectories using Euclidean distance-based greedy algorithm to categorize the trajectories, establish a multi-value classification table for each trajectory, and determine the final result using the majority-voting rule. When a face region is detected, the system marks down its coordinates, checks for the closest trajectories from the tail trajectory table in terms of Euclidean distances, and selects a class label with a majority vote from the trajectory statistics table as the final classification for the region. Thus, it achieves the progressive stabilization of the result.

We have three tables that represent CPL, Tail, and TCC, respectively. CPL is regenerated in each frame of the video, which contains the current information of the faces captured on the screen, whereas the Tail and TCC are created in the first and last frame until the end of the video. Once CPL has been created on a single frame, its information is used to update the information of Tail and TCC. Then, the information of CPL itself will be revised by the aid of Tail and TCC. The final output is the revised information of CPL.

The combined classification process involves the following tasks:

1. For each frame and each individual C^i, input the facial coordinates (x_c^i, y_c^i), the predict recognition label l_c^i, and the corresponding trajectory tr_c^i into the coordinate classification table CPL. CPL is used to create and update the tail trajectory table Tail. If Tail is null, it is created with items from CPL in the same order, with each item T^j in Tail containing the following sub-elements: the tail trajectory number tr_t^i, the tail coordinates (x_t^i, y_t^i), and the classification result l_t^i (here i, j denotes the person number, and c, t denotes the table CPL and Tail). Tail is not null, the Euclidean distance-based greedy algorithm is used to match the coordinates of each individual C^i to the tail trajectory coordinates in Tail. First, traverse through the corresponding coordinates of every element in CPL and Tail to find a matching pair with the least Euclidean distance. Afterwards, the class label l_t^j of T^θ in Tail is updated to the matched class label l_c^i of C^i in CPL. Then, the trajectory tr_c^i of C^i in CPL is updated to the matched trajectory tr_t^θ of T^θ in Tail. Then, execute the next traversal while excluding the matched pairs. The process is repeated until all tail points in Tail have had their matches found (greedy algorithm). In each iteration of the greedy algorithm, a pair with the least Euclidean distance is found while the sum of the all the pairwise distances is minimal. The matched pair is neglected in the next iteration. This strategy ensures the global optimization solution of multiple points pairwise matching, and avoids the results from falling into local optimization (i.e. only ensuring the least distance for some individuals rather than the entire set).
2. Update the trajectory statistical table TCC according to Tail. If TCC is null, it is created with data from Tail, where each trajectory has its initial vote for each class statistics set to 0. If TCC is not null, one vote is added to the class statistics c_k^i that corresponds to the matched classification result l_t^j of the item T^j in Tail.
3. For each individual C^i in CPL, search the trajectory number in TCC and choose the class statistics with maximum votes as the final classification result of this trajectory using the majority-voting rule. Algorithm 1 provides the specific operation process of MTIL.

Algorithm 1: MTIL algorithm

```
Input: coordinate classification table: Cpl $= \{C^i; i = 1, \ldots, n\}$, $C^i(x_c^i, y_c^i, l_c^i, tr_c^i)$
for $C^i(x_c^i, y_c^i, l_c^i, tr_c^i) \in Cpl$ do
    for $T^j(x_t^j, y_t^j, l_t^j, tr_t^j) \in Tail$ do
        $\theta = \text{argmin}_j(||x_t^j - x_c^i||^2 + ||y_t^j - y_c^i||^2)$    /*trajectory matching*/
    update Tail: $l_t^\theta = l_c^i$    /*update the predicted label of the $\theta$ th item $T^\theta$ in Tail*/
    update Cpl: $tr_c^i = tr_t^\theta$    /*update the trajectory number of the i th item $C^i$ in Cpl*/
    eliminate $T^\theta$ from next loop    /*exclude the matched pairs*/
for $P^i(tr_p^i, c_1^i, c_2^i, \ldots, c_m^i) \in Tcc$ do
    for $T^j(x_t^j, y_t^j, l_t^j, tr_t^j) \in Tail$ do
        if $(tr_t^j = tr_p^i)$    /*find the item $j$ with the same trajectory number*/
            update Tcc: $c_k^i + +; (k = l_t^j)$    /*add 1 vote to the corresponding class $k$*/
for $C^i(x_c^i, y_c^i, l_c^i, tr_c^i) \in Cpl$ do
    for $P^j(tr_p^j, c_1^j, \ldots, c_m^j) \in Tcc$ do
        if $(tr_p^j = tr_c^i)$    /*find the item $j$ with the same trajectory number*/
            update Cpl: $l_c^i = \max(c_k^j; k = 1, \ldots, m)$    /*choose the class with maximum votes*/
Output: the revised class label $l_c^i$ of each individual $C^i$ in Cpl
```

4.2. Handling overlapping faces in MTIL

We found that the MTIL method often cannot correctly identify the overlapping face regions because the coordinates of the front face and back face overlapped, and the trajectory was classified as belonging to the back face. To address this problem, we added a balancing rule that states when face regions overlap, only the initial class estimations from the classifier model must be used as the result. Tests indicate that this strategy can significantly reduce these false results.

5. Experimental results

The tests were conducted using the Honda/UCSD Video Database for face tracking and BMP Image Sequences for Elliptical Head Tracking. The Viola & Jones algorithm was used for per-frame face detection and tracking. The robustness of face tracking was achieved through TLD multi-target tracking method combined with a change-detection strategy. The LBPH-trained face classification model used our proposed MTIL algorithm to provide progressive correction to the results of face detection, tracking, and preliminary class estimations of the classifier.

5.1. Video-based face detection based on V&J + TLD algorithm

Figures 2 and 3 depict some results of the face detection using only V&J and V&J + TLD, with Honda/UCSD as test data. False negatives that are caused by lighting, perspectives, expressions or video resolutions are frequent with the V&J system as shown in Fig. 2b; false positives of non-face regions also appear periodically due to inaccuracies of the training as shown in Fig. 2c. Figure 3 shows the experimental results of the V&J + TLD. Comparing Fig. 3 to Fig. 2b and c, it can be seen that the TLD-improved algorithm can detect the majority of regions lost with V&J. False positives have been largely eliminated.

Table 1 compares the detection rates and false positive rates of the two methods on Honda/UCSD database. Table 1 shows that the V&J + TLD method (average detection rate is 89.51%) performed better than the V&J method (average detection rate is 64.97%) in addition to significantly reducing the false

Table 1
Face detection rates (%) and FP rates (%) for VFR on single targets from Honda/UCSD

Method	behzad1	behzad2	chia1	chia2	danny1	danny2	fuji1	harsh1
V&J	71.65 (1.27)	75.96 (1.72)	60.8 (5.26)	59.66 (6.76)	48.72 (1.13)	69.47 (3.31)	54.34 (3.22)	74.69 (1.85)
VJ + TLD	98.7 (0)	98.28 (0)	78.32 (0)	74.88 (0)	65.72 (0)	94.66 (0)	98.39 (0)	98.46 (0)

Method	harsh2	harsh3	harsh4	hector1	hide1	james1	james2	jeff1
V&J	74.16 (4.49)	76.62 (1.82)	58.65 (4.14)	67.27 (2.73)	58.12 (2.35)	59.62 (1.65)	67 (0.67)	62.81 (5.79)
VJ + TLD	97.75 (0)	98.16 (0)	89.85 (0)	90.3 (0)	81.65 (0)	71.7 (0)	95.96 (0)	99.34 (0)

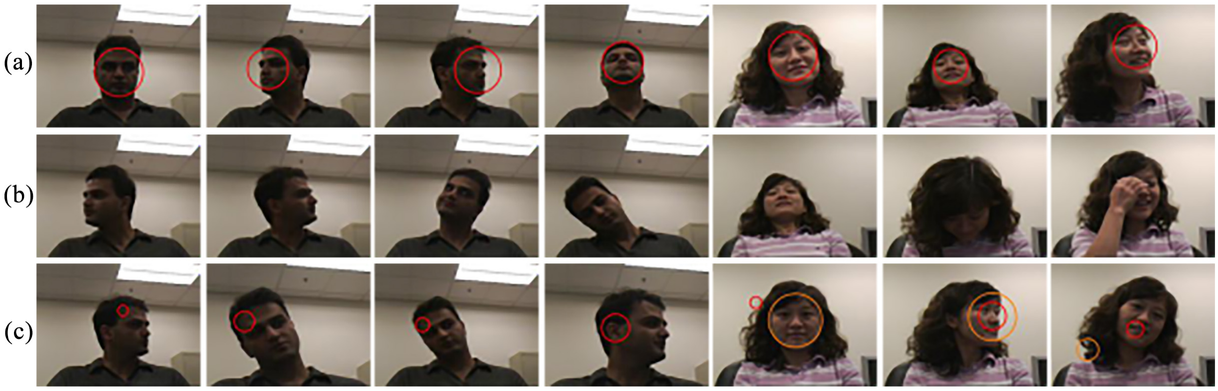

Fig. 2. Face detection with V&J on Honda/UCSD: a. true positives; b. false negatives; c. false positives.

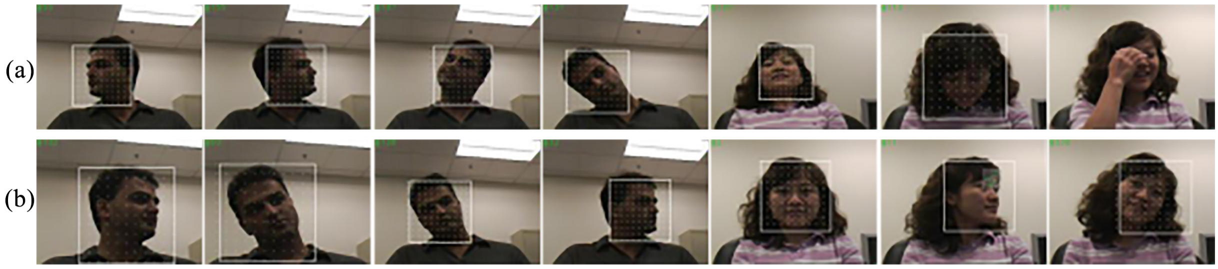

Fig. 3. Face detection with V& J + TLD on Honda/UCSD: a. re-test on false negatives; b. re-test on false positives.

positives. The test shows that TLD combined with a change-detection strategy can greatly improve the accuracy and robustness of VFR.

Another test video is the seq_mb file from the BMP Image Sequences for Elliptical Head Tracking database. This video is characterized by a low video resolution and drastic head movements (360-degree head rotation or horizontal skewing). The former factor may lead to frequent FPs using V&J, e.g. Fig. 4a, while the latter leads to FNs, e.g. Fig. 4c. The use of the V&J + TLD algorithm can effectively alleviate the problem of false positive as shown in Fig. 4b, and partly reduce the false negative as shown in Fig. 4d. The FN errors still exist because the head rotations can cause long periods of failures in face-tracking as seen in the two last images of Fig. 4d.

Table 2
Single-target face detection and tracking using V&J and V&J + TLD algorithms on BMP (seq_mb) (Unit: frame)

Method	TP↑	FP↓	FN↓	Precision↑	Recall↑
V&J	342	398	159	46.22%	68.26%
V&J + TLD	388	0	108	100%	78.23%

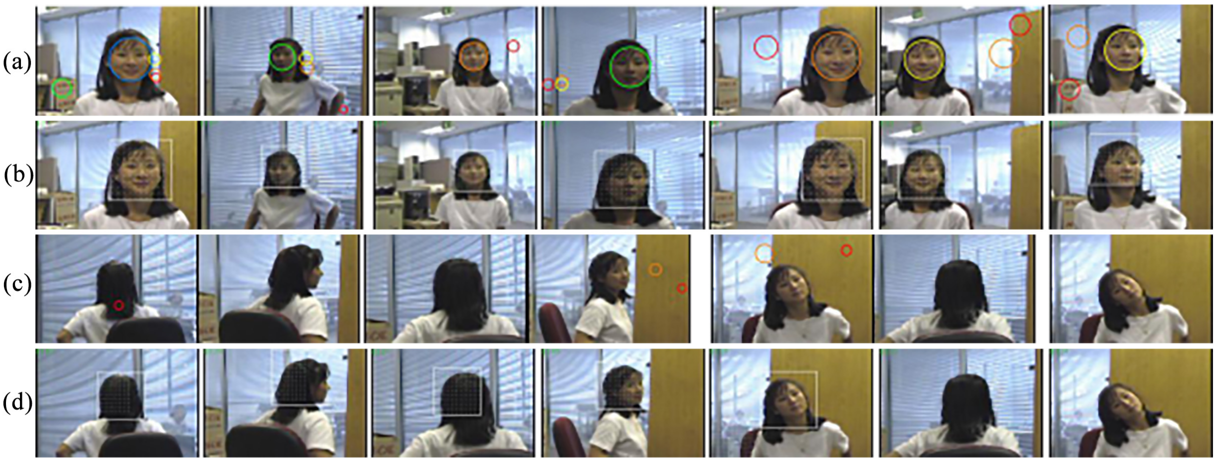

Fig. 4. Face detection with V&J and V&J + TLD on BMP: a and b. false positive tests; c and d. false negative tests.

Table 2 lists the test results of single-target detection and tracking on seq_mp, with the precision and recall rates calculated by the following equations:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}. \quad (1)$$

V&J + TLD has significantly reduced the false positives, while showing some improvement on false negatives. The value of FP has reduced to zero in Tables 1 and 2. As a result, the precision has grown up to 100% according to Eq. (1). As shown in Figs 3 and 4, there are no regions of non-face marked in the frames. TLD ensures the continuous and reliable tracking of the facial region, and the change detection mechanism ensures that a new tracking must be based on a newly appeared face because non-face region detected at the same place can't last for a long time.

5.2. *VFR based on LBPH + MTIL algorithm*

5.2.1. *Single-trajectory video-based face recognition*

The full VFR tests both use V&J + TLD as the detection and tracking algorithm. The single trajectory test uses videos from Honda/UCSD, while the multi-trajectory test uses a video from BMP Head Tracking. V&J + TLD is used for division of face regions.

Table 3 compares the correct recognition rates (frames of correct recognition/total frames) and false recognition rates (false positive frames/frames with detected faces) between LBPH and LBPH + MTIL. The LBPH + MTIL algorithm shows a significant improvement over using only LBPH in accuracy and false positives. In addition, the LBPH + MTIL algorithm has significantly improved the FP error-correcting performances. Some videos (jeff and victor) show higher inaccuracies and lower recognition rates for ROIs, which may be due to the hand-picked training samples being not sufficiently representative causing low accuracies in the initial recognition process.

Table 3
Comparison of correct recognition rates (%) and false recognition rates (%) between LBPH and LBPH + MTIL on single targets from Honda/UCSD

Method	behzad	chia	danny	fuji	harsh	hector	hide	james	jeff
LBPH	52.1 (39)	65.74 (12.22)	53.87 (43.82)	62.09 (37.91)	56.42 (28.85)	32.92 (54.27)	52.62 (28.25)	60.45 (24.65)	33.5 (66.4)
LBPH + MTIL	84.14 (0)	74.68 (0)	95.88 (0)	93.46 (0)	79.62 (0)	72 (0)	73.33 (0)	80.22 (0)	85.17 (14.69)
Method	joey	leekc	louis	miho	ming	rakesh	saito	victor	yokoyama
LBPH	54.96 (40.15)	60.89 (28.05)	64.65 (34.62)	65.6 (31.09)	64.84 (35.16)	55.32 (34.67)	74.56 (18.01)	22.15 (52.11)	65.23 (20.39)
LBPH + MTIL	91.84 (0)	84.64 (0)	98.89 (0)	95.2 (0)	100 (0)	84.68 (0)	90.94 (0)	23.13 (50)	81.94 (0)

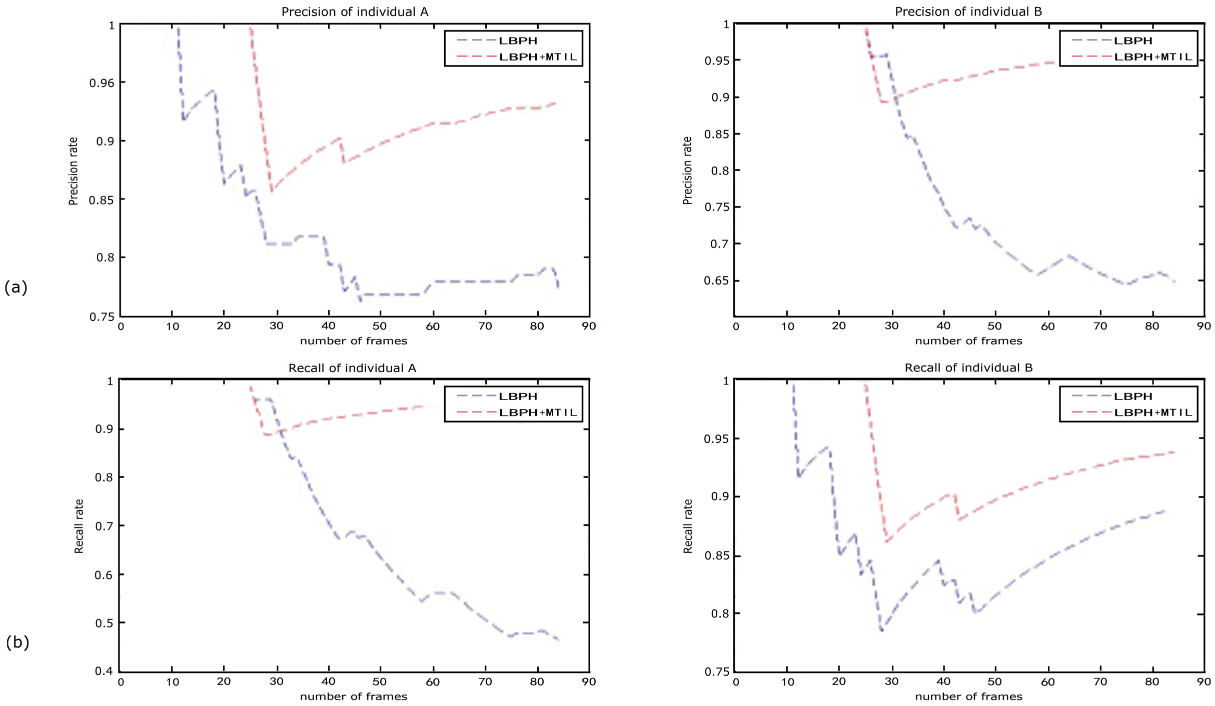

Fig. 5. Variations of precision rates and recall rates for LBPH and LBPH + MTIL. a. Precision rates of individuals A and B; b. recall rates of individuals A and B.

5.2.2. *Multi-trajectory VFR*

The video for multi-trajectory VFR tests is taken from the second half of seq_mb from the BMP Image Sequences. The segment provides the complexity factor for multi-trajectory recognition because it contains two individuals who obscured each other during the video, one of which had first left and then reentered the scene. Figure 5 compares the algorithms' effect on precision rates and recall rates. Table 4 compares the final data.

Experiments show that compared with the original algorithm, this method improves the accuracy of recognition. For LBPH + MTIL, both precision and recall rates show an increasing trend over time, with generally better performance than LBPH.

Table 4
Results of multi-ROI recognition with LBPH and LBPH + MTIL on seq_mb. (Unit: frame)

Individual A:

Method	TP↑	FP↓	FN↓	Precision↑	Recall↑
LBPH	34	10	39	77.27%	46.58%
LBPH + MTIL	69	5	4	93.24%	94.52%

Individual B:

Method	TP↑	FP↓	FN↓	Precision↑	Recall↑
LBPH	72	39	10	64.86%	87.8%
LBPH + MTIL	77	4	5	95.06%	93.9%

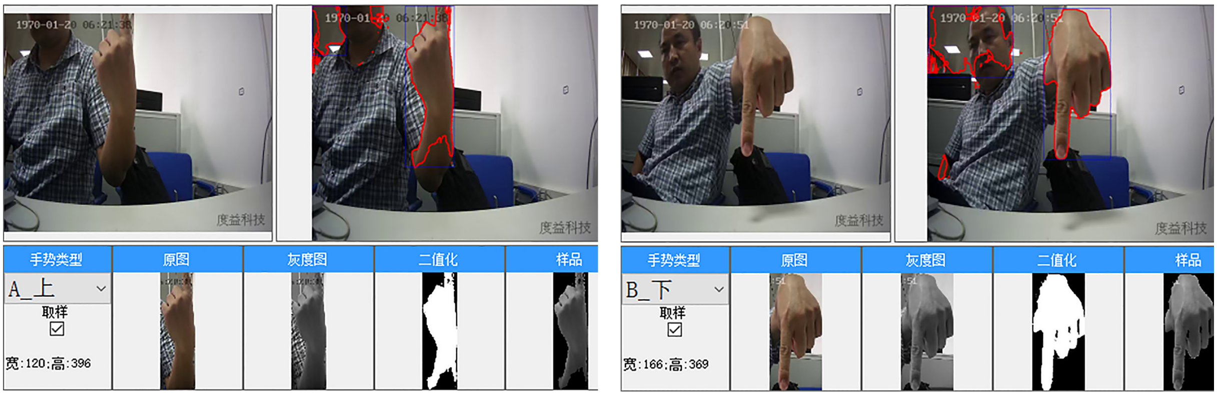

Fig. 6. Hand detection using skin-color algorithm.

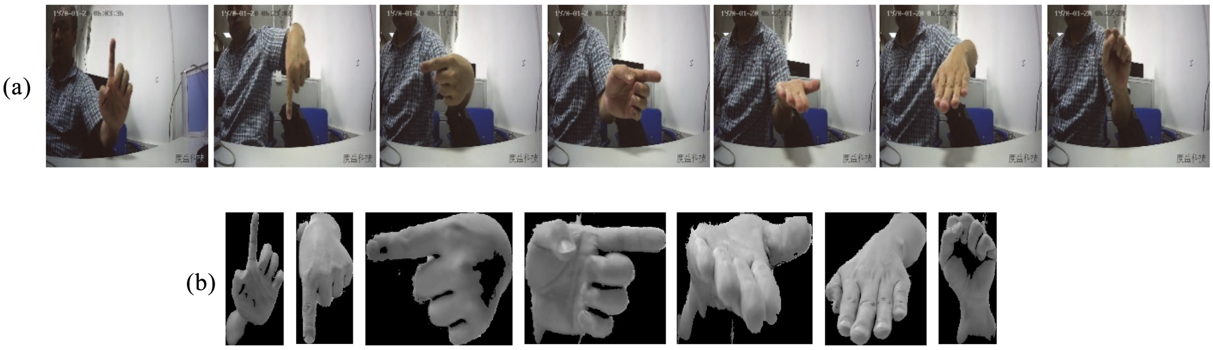

Fig. 7. a. hand gestures used in our experiment; b. gesture samples extracted from the row image.

5.3. Gesture recognition system based on LBPH + MTIL algorithm

The proposed method can also be used for medical system such as gesture recognition based touchless visualization system for medical volume [24]. Instead of the V&J algorithm, we used the skin-color detection algorithm in HSV color space to deal with the hand detection problem as shown in Fig. 6, since V&J face detection algorithm cannot be applied to gesture recognition. We first use skin-color detection to find the proximate area, and then apply binarization to eliminate the redundant part such as the clothing.

As the experiment setting of [24], we adopt 7 gestures to conduct the experiment as shown in Fig. 7a.

Table 5
Gesture recognition result

Finger up	Finger down	Finger left	Finger right	Palm up	Palm down	Grasp
94.3	98	95	97.5	95.6	98.9	100%

They are Finger up, Finger down, Finger left, Finger right, Palm up, Palm down and Grasp. Some samples extracted from the row image are shown in Fig. 7b.

We collected 897 gesture samples and split them into two half parts, that is, training part and test part. The gesture recognition experiment is conducted using LBPH + MTIL Algorithm. The recognition results are shown in Table 5, which illustrate that the proposed method perform well on gesture recognition system.

6. Conclusions

Video-based face recognition is a challenging problem that combines tracking, detection, and recognition. Gesture recognition is similar to face recognition. It can be used on medical recognition based touchless visualization system The Viola & Jones algorithm has been widely used in VFR, but systems based on V&J are known to have false negatives or positives in tests. The accuracy and reliability of face detection can be improved by a combination of TLD and change detection based on video continuity. Tests have shown that our approach can recognize multiple targets from videos, while improving the precision recognition over time. The TLD algorithm combined with a change-detection strategy significantly improved the accuracy and robustness of face detection. Tests on a video from BMP show that the V&J + TLD can increase the accuracy, and can improve without completely eliminating false positives due to the low resolution and drastic head rotations of the video, with a lower increase to the recall rate than the precision rate.

The accuracy of FR tends to progressively regress over time. This enables us to correct the classification results using spatiotemporal information from the video. The establishment and classification of face trajectories are particularly difficult when more than one face appear on the scene. Our proposed multi-trajectory incremental learning algorithm can track and recognize multiple faces in the video using a Euclidean distance-based greedy algorithm to classify the trajectories, storing each trajectory's data in multi-value statistics tables, and basing the final results on the majority-voting rule. Tests on videos from Honda/UCSD show that the LBPH + MTIL algorithm has significantly increased recognition rates compared to LBPH, while significantly decreasing the false recognition rates. Tests with BMP show that it had significantly improved average precision and recall rates. The LBPH + MTIL algorithm's precision and recall rate curves show a trend of increase over time, with better overall performance and final results than LBPH. Because LBPH is a feature extraction method which can be used for more than just face recognition, the proposed LBPH + MTIL method can also be applied to medical video recognition such as gesture recognition control based intelligent medical system, which recognizes the current gesture video images of the operator and then sends a control command.

To conduct the experiment of gesture recognition, skin-detection and binarization are used to detect hand samples, then LBPH + MTIL method is also used for recognition. Results show that the method can perform well on the gesture recognition system.

Conflict of interest

None to report.

References

[1] Huang Z, Wang R, Shan S, and Chen X. Face recognition on large-scale video in the wild with hybrid Euclidean-and-Riemannian metric learning. Pattern Recognition. 2005; 48(10), 3113–3124.

[2] Zhou S, Krueger V, and Chellappa R. Face Recognition from Video: A Condecsation Approach, in: Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, Washington DC. 2002-05, pp. 221–228.

[3] Edwards GT, Talaor CJ, and Cootes TF. Improving Identification Performance by Integrating Evidence from Sequences, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 1999-06, pp. 486–491.

[4] Hadid A, and Pietikainen M. From Still Image to Video-based face Recognition: An Experimental Analysis, in: Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. 2004, pp. 813–818.

[5] Pagano C, Granger E, Sabourin R, Marcialis GL, and Roli F. Adaptive ensembles for face recognition in changing video surveillance environments. Pattern Recognition. 2014; 286(11), 75–101.

[6] De-La-Torre M, Granger E, Sabourin R, and Gorodnichy DO. Adaptive skew-sensitive ensembles for face recognition in video surveillance. Pattern Recognition. 2015; 48(11), 3385–3406.

[7] De-La-Torre M, Miguel E, Radtke PVW, Sabourin R, and Gorodnichy DO. Partially-supervised learning from facial trajectories for face recognition in video surveillance. Information Fusion. 2014; 24(3), 31–53.

[8] Dewan MAA, Granger E, Marcialis GL, Sabourin R, and Roli F. Adaptive appearance model tracking for still-to-video face recognition. Pattern Recognition. 2015; 49(C), 129–151.

[9] Wang G, Zheng F, Shi C, Xue J, Liu C, and He L. Embedding metric learning into set-based face recognition for video surveillance. Neurocomputing. 2015; 151, 1500–1506.

[10] Chouchene M, Sayadi ME, Bahri H, Dubois J, Miteran J, and Atri M. Optimized parallel implementation of face detection based on GPU component. Microprocessors & Microsystems. 2015; 39(6), 393–404.

[11] Phimoltares S, Lursinsap C, and Chamnongthai K. Face detection and facial feature localization without considering the appearance of image context. Image & Vision Computing. 2007; 25(5), 741–753.

[12] Liu Q, Zhang W, Li H, and Ngan KN. Hybrid human detection and recognition in surveillance. Neurocomputing. 2016; 194, 10–23.

[13] Viola P, and Jones MJ. Robust real-time face detection. International Journal of Computer Vision. 2004; 57(2), 137–154.

[14] Ojala T, Pietäikinen M, Member S, and Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 24(7), 971–987.

[15] Moradi M, and Abolmaesumi P. Medical image registration based on distinctive image features from scale-invariant (SIFT) key points. International Congress Series. 2005; 1281, 91–110.

[16] Hu W, Tan T, Wang L, and Maybank S. A survey on visual surveillance of object motion and behavior. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews. 2004; 34(3), 334–352.

[17] Jepson AD, Fleet DJ, and El-Maraghi TF. Robust online appearance models for visual tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003; 25(10), 415–522.

[18] Olson CF. Maximum-likelihood template matching, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island, SC, USA: IEEE, 2007; 52–57.

[19] Avidan S. Support vector tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26(8), 1064–1072.

[20] Comaniciu D, Ramesh V, and Meer P. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003; 25(5), 564–577.

[21] Jun Y, Shunli Z, and Li Z. Object tracking with hierarchical multiview learning. Journal of Electronic Imaging. 2016; 25.

[22] Kalal Z, Mikolajczyk K, and Matas J. Face-TLD: Tracking-Learning-Detection applied to faces, in ICIP, 2010.

[23] Kertész G, Szénási S, and Vámossy Z. Application and properties of the radon transform for object image matching, in: 15th International Symposium on Applied Machine Intelligence and Informatics, Herlany, 2017; pp. 353–358.

[24] Fujii R, and Ryoma T. Touchless A, Visualization System for Medical Volumes Based on Kinect Gesture Recognition, in: International Conference on Innovation in Medicine and Healthcare Springer International Publishing. 2016.