Face Alignment under Partial Occlusion in Near Infrared Images

Sifei Liu[1,2], Dong Yi[1], Bin Li[2], Stan Z. Li[1,0]

[1] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
E-mail:{sfliu,dyi,szli}@cbsr.ia.ac.cn

[2] University of Science and Technology of China, Hefei, China
E-mail:binli@ustc.edu.cn

Abstract: Face alignment and recognition in less controlled environment are one of the most essential bottlenecks for practical face recognition system. Recently several researches have focused on partial face recognition problem, but few works have addressed the problem of face alignment under partial occlusion. In this paper, we present a robust face alignment method by combining local feature matching and Probabilistic Hough Transform (PHT) for partial face alignment in near infrared (NIR) images. Given a set of well aligned faces as target, and for face images with occlusions, their correspondences are established by local feature matching. For faces with missing components, many false matches of local features will be built due to lack of holistic information. The PHT approach aims to find correct correspondences and resist the inevitable false ones by taking each parameter candidate generated by correspondences pair as a vote in the 4-D in-plane transform parameter space. We also employ geometric constraints and appearance consistency and combine them with PHT in an probabilistic hough optimization function, so that each vote is weighted by a probabilistic score. Experiments of alignment on both MBGC portal face video and facial images with Glass-face occlusions show that our approach can reliably and accurately deal with missing data of facial components caused by partial occlusion.

Key Words: Face Alignment, partial faces, probabilistic Hough transform (PHT).

1. INTRODUCTION

Facial images with partial occlusions often appears in real applications when environments are not strictly controlled. For NIR face recognition system in video surveillance, faces are often partial, a typical case is portal face video in Multiple Biometrics Grand Challenge (MBGC) [9]. Furthermore, in NIR face recognition system, near infrared light is easily reflected by glasses, and generating highlight around the region of eyes, as shown in Fig. 1. These partial occlusion reduce the performance of face alignment and recognition. Highlight is generated in the region of eyes, which is another case of occlusion in facial components. There are many researches focused on partial face recognition problem, such as methods based on partial similarity measure [15], sparse representation [16], and have obtained some significant achievements. But the key ingredient step alignment is almost neglected. While face images are incomplete, alignment of faces will become very challenge.

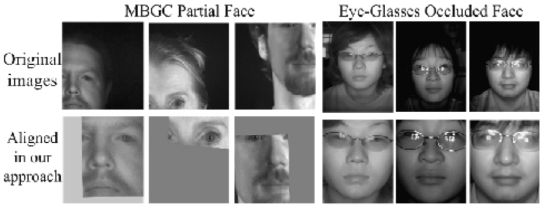


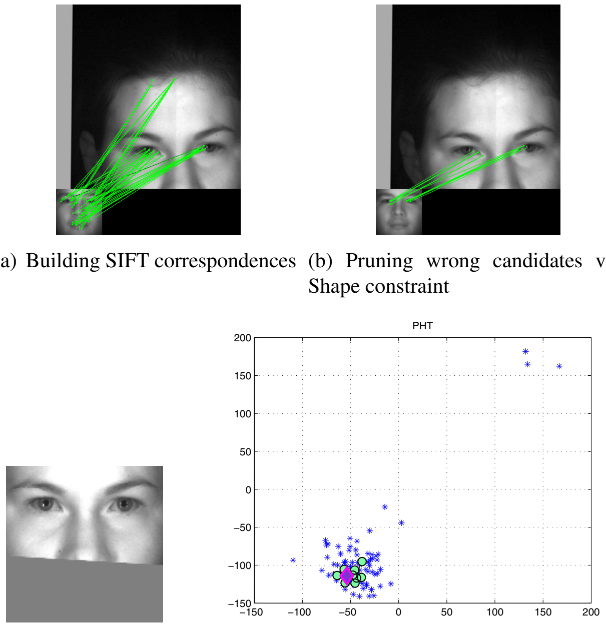
Fig. 1: Challenging samples in face alignment

For accurate automatic face alignment which is a key ingredient to affect the performance of face recognition, various techniques have been proposed in the past two decades. Cootes proposed a series of deformable landmark models based methods, such as ASM [1] and AAM [2], and CLM [3], which obtain good results but need intensive manual labeling and model training. Unsupervised approaches [4] [5] are proposed recently to avoid tedious manual labeling and training, which are more preferable for real-world applications. In practical frontal face recognition when only in-plane affine transform is needed, a simple and common way is to locate specific points on faces, such as eye center, nose tip, and transform it into canonical pose. However, this approach cannot well handle face images with low quality, especially when there are noises or occlusions on the those points to be detected.

However when faces are occluded, approaches discussed above will ineffective or even unable to handle because part of local features or image information are missing. To deal with the partial occlusion problem, Yang [10] proposed a SIFT [6] based feature codebook learning and RANSAC based parameter estimation algorithm. This approach needs a training process of codebook to represent features in subspace, it performs good in Multiple Biometrics Grand Challenge (MBGC) [9] but has a low detection rate of local features on faces.

The Hough transform [11–13] provides another way of parameter estimation via a voting strategy and has been widely used in object detection and pose estimation problems. Taking

[0] Stan Z. Li is the corresponding author.



(a) Building SIFT correspondences (b) Pruning wrong candidates via Shape constraint

(c) Aligned face (d) Hough votes in parameter space

Fig. 2: The overview of our approach: (a)Building correspondences between partial frame and canonical facial image, (b)Pruning wrong candidates by Geometric Constraint designed according to geometric relationship between correspondences, (c)Transformed partial face, and(d)the weighed hough votes' distribution in parameter space, the green dots are valid candidate with maximum votes, and we set the result as the median of these valid ones, as illustrated in purple diamond

the advantage of voting, this approach can effectively find inliers aggregated in space and prune outliers that usually more sparse and scattered in space in many parameter estimation cases, see Fig. 2(d).

In this paper, we base our work on NIR face recognition system, which provides an effective way to solving the illumination problem and performs well when incorporated with effective algorithms. We propose a novel, Hough transform based method to deal with images with missing data of facial components caused by partial occlusion (highlight, out-of-view), incomplete face and is training free. Given a test face images and a target face image set, their correspondences are established by local feature matching. However, it is very difficult to ensure perfect correspondence as part of facial component are missing(eg. when left eye are occluded, local features on right eye are likely to match with left eye in facial images of canonical set). Therefore, PHT is employed in which correspondences cast votes for possible parameters of affine transformation. To obtain globally consistent face alignment from imperfect data, we put it in an probabilistic formulation through combination of hough voting, Geometric Constraints, and prior of matching quality. We call our approach "Local Feature based probabilistic hough transform (LFPHT)". In addition to that, we apply this framework to the whole sequences of each subject rather than single image, to overcome bad matches caused by low quality images within each subject.

Two experiments have carried out on both MBGC-08 NIR partial face portal videos and self-collected NIR Glass-face databases demonstrated the validation of LFPHT. Our approach outperforms the approach in [10] and is training free. Also good result showed in NIR glass-face alignment and recognition, indicating that the LFPHT is a good way to improve NIR face recognition performances in cases of wearing glasses.

2. ALIGNMENT ALGORITHM

This section will describe the proposed method for aligning incomplete NIR faces, our approach aims to find correspondences between incomplete face sequences and a set of well aligned face images, and through this to find the in-plane affine transformation parameters. Because part of faces are occluded, stable keypoints detection(AAM, ASM, eye-detection) is invalid, for this reason we based our method on uncertain correspondence detection technique like SIFT matching.

In the proposed method, alignment between a test image to target face set is done in three steps: (1) detecting face images' feature points respectively by SIFT based methods and building their correspondences; (2) Creating PHT model by incorporating hough voting, correspondences relationship, and prior of matching quality, and (3) solving the alignment transform parameter (scale, offset and rotation) by maximize the probability of the model.

Building local correspondences by SIFT Scale Invariant Feature Transform (SIFT) developed by Lowe [6] finds local feature correspondences between images by combining a scale invariant keypoint detector and gradient histogram based descriptor. Descriptors are constructed through information of locations and orientations, which is robust to small geometric distortions and small errors in the keypoint detection, and proved by Mikolajczyk and Schimid [7] to be the best comparing to existing descriptors.

We extract keypoints and their descriptors based on SIFT, both on a set of canonical face images as gallery, and images with missing face component to be aligned, as probe. Their correspondences are established through nearest neighborhood(NN) matching. However, as relationship between correspondences are ignored in this way, many incorrect matches will be build in the process. Fig 2(a)shows the initial built correspondences by matching descriptors with NN.

Framework of Probabilistic Hough Transform (PHT) Through grouping every pairs in correspondences set collected by SIFT matching and NN classifier, a set of candidate affine parameters, named initial transformation set, can be generated in a 4-D parameter spaces. In this way, each pair of correspondences make a probabilistic decision in the spaces. We formulating hough voting strategy in a probabilistic way and taking the affine parameter estimation as an optimization problem.

Let C represents the set of matched correspondences between test probe and the gallary(canonical pose), (X,Y) are sets of correspondences respectively for probe and the gallery, (x_i, y_i) represents the ith correspondences. $S(\Phi,C)$ is the score of similarity transformation parameters $\Phi = (s, \theta, x, y)$ to be estimated, and is defined as:

$$S(\Phi, C) = p(\Phi, X, Y) = p(\Phi | x_1, y_1; x_2, y_2; ...; x_n, y_n); \quad (1)$$

Each candidate parameters $\varphi_i \in \Phi$ can be generated from a pair of correspondences, for each pair in the parameters set generated by SIFT matching, we set $C_i = (x_i, y_i)$. An extreme version is to ignore all correlation between correspondences and the grouped pairs:

$$S(\Phi, C) = \prod_{i=1}^{N} \prod_{k=1;k\neq i}^{N} P(\Phi, C_i, C_k); \qquad (2)$$

Hence we can modeling parameters distribution by grouping every two correspondences to calculate an affine parameter and cast a vote in the 4-D affine parameter spaces. However, this formulation assumes completely independence between pairs. To solve this problem, and rather than counting the correlations in probabilistic forms, we take a "decorrelation" step by pruning correspondences with incorrect geometric structure.

Pruning correspondences with Shape Constraint The correlation of correspondences are mainly rely on its geometric relationship. For a pair of correspondences (x_i, y_i) and (x_j, y_j), the relative orientations between (x_i, x_j) and (y_i, y_j) are supposed to be consistence, as shown in Fig. 2(b), we name it geometric orientation constraint(GOC). To deal with joint probability in (1), and making the problem tractable, we taking the GOC between the best matched correspondence and the other ones into consideration.

Geometric Constraint is done when initial SIFT correspondences set are generated. For each keypoint K_i in probe image, we compute the distance from its descriptor to each of descriptor in gallary image. The matching quality of local descriptors is measured by ratio: $ri = d1 \setminus d2$, where $d1, d2$ are respectively the minimum and the second minimum distance. In order to pruning wrong correspondences, we use Shape Context [14] described in [10] to record the relative orientations between correspondences with best matching quality(named "main correspondence") and the other ones both in probe and gallery. Correspondences will be eliminated if the difference of relative orientation with main correspondence are larger than threshold of 30°, as moderate errors are permitted.

Face Alignment via Probabilistic Hough Transform By pruning candidates of geometrically conflicting we get filtered set of correspondences $(\tilde{X}, \tilde{Y})$. We can use the same form of (2) as Geometric Constraint has been processed:

$$\begin{aligned} S(\Phi, \tilde{C}) &= \prod_{i=1}^{N} \prod_{k=1;k\neq i}^{N} P(\Phi, \tilde{C_i}, \tilde{C_k}) \\ &= \prod_{i=1}^{N} \prod_{k=1;k\neq i}^{N} P(\Phi | \tilde{C_i}, \tilde{C_k}) P(\tilde{C_i}, \tilde{C_k}); \end{aligned} \qquad (3)$$

where $P(\tilde{C_i}, \tilde{C_k})$ is defined as cooccurrence of this pair:

$$P(\tilde{C_i}, \tilde{C_k}) = \begin{cases} P(\tilde{C_i}) P(\tilde{C_k}) & \tilde{C_i}, \tilde{C_k} \text{ agree with GOC} \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

The first term represents the hough voting and its parameter distribution can be approximated using Parzen-window estimation [11–13]. For the sack of computing complexity we make the 4-D space into discrete grids where each parameters will fall into its closest grid node, as following:

$$p(\varphi_j | \tilde{C_i}, \tilde{C_k}) = \begin{cases} 1 & \text{if } s(\tilde{C_i}, \tilde{C_k}) \rightarrow \varphi_j \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

Where $\varphi_j \in \Phi$ is the jth node of the gird and $(\tilde{C_i}, \tilde{C_k})$ is a pair in the set. $s(\cdot; \cdot)$ is the function that mapping pair to the closest node in discrete space, this mapping can be easily calculated by its coordinates.

The second term, when GOC are met, is the correspondences' prior defined by both distance of descriptors and distance ratio for matching quality, defined as:

$$p(C) = \frac{\exp(-ratio \| \tilde{D}_{xi} - \tilde{D}_{yi} \|)}{Z} \qquad (6)$$

where $\tilde{D}_p$ and $\tilde{D}_g$ are descriptors in probe and gallery images, and Z is a constant to make $p(C)$ a probability distribution. To avoid obvious wrong transformations, the constraint of transformed local feature appearance consistency is utilized. In this way, all valid candidates, as illustrated by green dots in Fig. 2, will be judged jointly by their PHT score and the consistency of patches between transformed partial face images and the canonical sets. The PHT based parameter estimation problem can thus be described as an optimization of energy function $E(\Phi)$ that combine constraint term with weighing coefficient β:

$$E(\Phi) = \frac{1}{S(\Phi, \tilde{C})} + \frac{\beta}{N} \sum_{i=1}^{N} | I(T(X, \Phi)) - V_i(X) \| \qquad (7)$$

where T denotes transform of image, and V is the image of target.

3. JOINT FACE ALIGNMENT

Face alignment with local feature matching between a single probe image and the target set turns out to be unreliable when there are few correspondences. To make it robust, we introduce a joint face alignment strategy in which all images within a sequence, named a joint set, can share their correspondences for each other. For sharing correspondences, joint set must be aligned firstly. Because these images are from the same person and quality, they can easily aligned by SIFT. Through summing up each joint set, joint alignment energy is written as:

$$E(\Phi) = \frac{1}{\sum_s S(\Phi, C_s)} + \frac{\beta}{N} \sum_s \sum_{i=1}^{N} \| I_s(T(X, \Phi)) - V_i(X) \|$$

$$S(\Phi, C_s) = \sum_{i=1}^{N} \sum_{k=1, k \neq i}^{N} \sum_{j=1}^{P} p(\varphi_j | \tilde{C}_i, \tilde{C}_k) p(\tilde{C}_i) p(\tilde{C}_k) \qquad (8)$$

Where s represents sequences. Here the step of Geometric Constraint is performed on joint set, where main correspondence defined by best matching are defined throughout a sequence. This strategy is reliable because main correspondence is more likely to be correct in a joint set-to-target setting.

4. EXPERIMENTS

We evaluated the SIFT-based Probabilistic Hough Transform on two challenging datasets in Fig. 1. In MBGC-08 Near Infrared partial face video sequences, we compared our approach with Yang's work [10] as well as RANSAC, which is a traditional method and is widely used in parameter estimation as well as pruning local correspondences' outliers [8, 10]. In addition to that we compared our approach with eye-detection face alignment method on our self-collected Glass-face face database. Through comparison in face recognition performances in baseline algorithms we proved the validation of our work and furthermore provide another way to deal with alignment and recognition for NIR faces with glasses.

MBGC-08 partial face sequences The MBGC-08 partial face database [9] contains 139 sequences for 139 subjects with 2286 images with resolution 2048 × 2048. The canonical face set are 128 × 128 NIR still images of holistic frontal faces with fixed eye-coordinates. We firstly aligned faces of the same subject as described in section 3: As at least two correspondences are needed for affine parameter which has 4 variables(scaling factor, rotation angle, x and y offset), faces are matched to the one that has maximum number of keypoints within a sequence and subjects with less than 2 valid correspondences will be discarded as cannot be detected. In this experiment, 115 sequences are alignment feasible, and 5 strategies will be compared at this subset: (1) PHT without Joint Alignment(PHT); (2) PHT with Joint Alignment(PHT+Joint); (3) RANSAC without Joint Alignment(RANSAC); (4) RANSAC with Joint Alignment(RANSAC+Joint); and(5) Yang's work which used a SIFT codebook learning strategy. In (5), we use the canonical face set described above as training set. All parameters for SIFT matching as well as GOC are settled to be the same in these methods. We manually aligned the partial set to get the ground-truth parameters for each subject, through back projecting two stable points on images using parameters acquired in these 4 method compared with the ground-truth, and set a least displacement tolerance $R = 6pixel$, we can evaluate the affine transform parameters error through the distance of projected points. Table 1 shows the detection and alignment performance on this dataset.

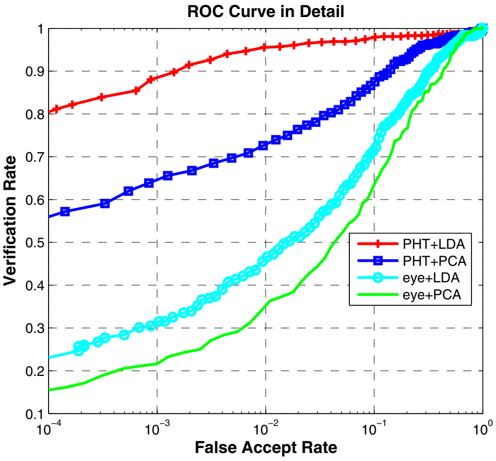


Fig. 3: Glass-face face recognition results

Alignment for Subjects with glasses In NIR faces with glasses, eyes are always occluded by highlight generated via reflection. We consider eyeglasses as an occluded region, and roughly estimate and neglect the region via a face detector during local features matching. We compare our method with the widely used alignment approach based on eyes localization and evaluate its performance by face recognition. Without loss of generality, PCA and LDA are used to do face recognition. Firstly, we prepare a target set including 19 persons, 300 face-without-glasses images to supply alignment criterion. Then, a database of NIR faces-with-glasses of 252 subjects(1934 images) is collected for alignment and recognition. 3 data sets are constructed from it: training set for PCA and LDA- 106 subjects randomly selected from 252(860 images); gallery set- 2 images per subject randomly selected from the other 146 subjects; probe set- another 2 images per subject randomly selected from that 146 subjects. Face alignment is processed separately in probe and gallery set, there is no intersection for both face image and subjects between probe and gallery set.

Fig. 3 illustrates the performance of face recognition in PCA and LDA respectively in the eye-detector approach and our proposed PHT approach. At FAR=0.001, the proposed PHT achieves 64.5% and 88.4% verification rate respectively in PCA and LDA, outperforms the eye-detector method whose verification rates are 21.6% in PCA and 31.5% in LDA.

Table 1: Comparison of different methods on MBGC-08 partial face database, PHT outperforms the others, while Joint alignment strategy can improve the performance

	Missed sequences	Incorrect alignment	Correct rate
PHT	1	18	84.35%
PHT+Joint	1	13	**88.70%**
RANSAC	1	28	75.65%
RANSAC+Joint	1	26	77.39%
Yang et al. ICB'09	4	19	83.48%

5. CONCLUSION

In this work, we proposed a novel approach for partial face alignment in near infrared (NIR) images, which can deal with faces with partial occlusion and is training free. Thereinto, a effective PHT model is constructed to find the correct correspondences, which combining hough voting, Geometric Constraints, matching quality prior, and local appearances consistency. Furthermore, using the information of image sequences, a joint alignment strategy is proposed to enhance the PHT model. Experiments on MBGC-08 partial face database proved the validation and advantage of our method. The second experiment in a Glass-face database provides a novel way deal with glasses highlight, which can efficiently improve the accuracy of face alignment and recognition.

ACKNOWLEDGEMENTS

This work was supported by the Chinese National Hi-Tech(863) Program Projects #2008AA01Z124, and the AuthenMetric R&D Fund.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, Active shape models - their training and application. *Computer Vision and Image Understanding.* vol. 61, no. 1, pp. 38-59, January 1995.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor, Active appearance models. *Lecture Notes in Computer Science.* vol. 1407, 1998.

[3] David Cristinacce and Tim Cootes. Feature Detection and Tracking with Constrained Local Models. *17th British Machine Vision Conf, 2006,* pp. 929-938, 2006.

[4] G. Huang and V. Jain and Erik Learned-Miller. Unsupervised joint alignment of complex images. *Proc. ICCV 2007*, Rio de Janeiro, pp. 1 - 8, Oct. 2007.

[5] Jianke Zhu and Luc Van Gool and Steven C.H. Hoi. Unsupervised Face Alignment by Robust Nonrigid Mapping. *Proc. ICCV 2009*, Kyoto, pp. 1265 -1272, Sep. 2009.

[6] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision.* vol. 60, no. 2, pp. 91-110, 2004.

[7] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[8] Mustafa Özuysal and Michael Calonder and Vincent Lepetit and Pascal Fua. Fast Keypoint Recognition Using Random Ferns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 32, no. 3, pp. 448-461, March 2010.

[9] NIST. Multiple Biometric Grand Challenge (MBGC). http://face.nist.gov/mbgc. 2008.

[10] Jimei Yang and Shengcai Liao and Stan Z. Li. Automatic Partial Face Alignment in NIR Video Sequences. *Proc. ICB 2009.* vol. 5558, pp. 249-258, 2009.

[11] S. Maji and J. Malik. Object detection using a max-margin Hough transform. *Proc. CVPR 2009.* pp. 1038-1045, June 2009.

[12] Juergen Gall and Victor Lempitsky. Class-specific hough forests for object detection. *Proc. CVPR 2009.* pp. 1022-1029, June 2009.

[13] Rozenn Dahyot. Statistical Hough Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* pp. 1502-1509, Aug. 2009.

[14] Gustavo Carneiro. Pruning Local Feature Correspondences using Shape Context. *Proc. ICPR 2004*, vol. 3, pp. 16- 19, Aug. 2004.

[15] Yi, Dong and Liao, Shengcai and Lei, Zhen and Sang, Jitao and Li, Stan. Partial Face Matching between Near Infrared and Visual Images in MBGC Portal Challenge. *Proc. ICB 2009.* pp. 733-742, 2009.

[16] Wright, John and Yang, Allen Y. and Ganesh, Arvind and Sastry, S. Shankar and Ma, Yi. Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 31, no. 2, pp. 201-227, 2009.